ICIC Express Letters
Volume **12**, Number **5**, May **2018**

ICIC International ©2018 ISSN 1881-803X
pp. 449–456

SUPPRESSION OF ISOLATED CLUSTERS OCCURRENCE IN SELF-ORGANIZING MAP CONSIDERING THE DISTANCE ON DATA SPACE

Keiji Kamei and Shinji Kawakami

Department of Production Systems
Graduate School of Engineering
Nishinippon Institute of Technology
1-11 Aratsu Kanda, Miyakogun, Fukuoka 800-0394, Japan
kamei@nishitech.ac.jp

Received November 2017; accepted February 2018

Abstract. *This research introduces distance on data space between a training sample and a reference vector into training algorithm of Self-Organizing Map (SOM). The basic SOM only applies distance on data space to determination for the winner neuron on feature space, and that distance is not used for training. By contrast to the basic SOM, approach in this research is that distance on data space is used for training of SOM. The modification of training algorithm of SOM in this research affects neurons which are not close to the winner neuron on feature space, and the winner neuron pulls more neurons into the same cluster. For this, proposal in this research suppresses occurrence of isolated clusters which belong to the same class on feature space. The experimental results show that this approach has succeeded in suppressing occurrence of the isolated clusters. Moreover, the success rate of classification for evaluation data increases by 6% with proposal of this research.*

Keywords: Self-organizing map, Clustering, Suppression of isolated cluster occurrence, Distance on data space

1. **Introduction.** Recently, researches on artificial intelligence based on multilayered artificial neural networks have been studied actively because “Deep Neural Network” (DNN) was succeeded in solving many difficult problems that were not able to be solved by former artificial intelligence. “Deep Convolutional Neural Network” (DCNN) is one of DNN, and DCNN is said that it mimics visual cortex [1,2]. More specifically, DCNN simulates the behaviors of locality of receptive fields and has simple or complex cells on visual cortex. From that structure of network, DCNN achieved good results in the fields of image recognition, which are object classification problems especially.

There is “Self-Organizing Map” (SOM) [3,4] as an approach to simulation of visual cortex different from DCNN. By contrast to DCNN, SOM simulates only primary visual cortex (V1), and SOM does not have simple or complex cells. The different feature of SOM comparing with DCNN is that SOM constructs clusters like the receptive field in V1 based on competitive training described below.

The structure of SOM has 2 layers: one of them is “input layer” and the other one is “competitive layer”. There are some neurons in competitive layer. Neurons in that layer have reference vectors, and that vectors are initialized randomly. A training (input) vector is given to competitive layer from input layer, and then the winner neuron, which is also called the best matching unit, is determined. Reference vectors in each neuron are updated on the basis of the “winner take all”. On the process of updating reference vectors in each neuron, SOM applies the neighborhood function to updating reference vectors, and that function defines influence degree of distance in competitive layer between the winner neuron and a neuron. The neighborhood function is a monotonically decreasing function;

therefore, influence degree of the winner neuron for each neuron decreases as training proceeds. From those characteristics of SOM structure and training algorithm, reference vectors in neurons adjacent to the winner neuron take similar value when training will have been finished. Consequently, neurons that have similar reference vectors gather within local area on feature space; on the other hand, neurons that have big different reference vectors are never adjacent to each other. Input vectors are constituted at high dimension and the competitive layer is constructed at low dimension in general; therefore, SOM has the ability of mapping high dimensional "data space" to low dimensional "feature space". For those reasons, it can be said that SOM has the function of classifier and dimension compressor.

One of problems of SOM is occurrence of isolated clusters that belong to the same class. The causes for that problem are random initialization of reference vectors and mapping of high dimensional data space to low dimensional feature space. In the beginning stage of training, a neuron which is distant from the other winner neurons belonging to the same class may fire for another input vector belonging to the same class due to random initialization. Moreover, it can be said to be difficult to map high to low dimensional space with completely preserving relations among input data. Abe et al. [5] have proposed to solve the problem for initialization of reference vectors based on principal component analysis for training data. The initial values of reference vectors in this proposal are fixed, so there is no diversity of SOM map. For this, it is possible that there are other more suited SOM map. Besides, the aim of their proposal is effective training, and that does not suppress occurrence of isolated clusters. Tokutaka et al. [6] proposed the spherical SOM that was able to preserve relations among input data compared with the basic SOM. It can be said the proposal is such a particular type of SOM.

The aim of this research is to improve the algorithm of SOM training in order to suppress occurrence of isolated clusters without particular preprocessing of training data and initialization of reference vectors. Wakuya et al. [7] proposed to improve SOM algorithm for controlling training speed for a certain training sample. Their proposal is that training coefficient for a certain training sample is larger than that for other training samples, so that firing neurons to a certain training sample pull more neurons into the class of a certain training sample. According to the results of their research, clusters which belong to the class for a certain training sample held large area on feature space. Besides, occurrence of isolated cluster suppressed because the winner neuron for a certain training sample pulled more neurons into the same cluster. By contrast to their proposal, this research does not target only a certain training sample; hence, the coefficient for pulling into a cluster of the winner neuron is determined dynamically on the basis of distance between a training sample and a reference vector on data space. To the best of our knowledge, there are no proposals which apply distance on data space to training of SOM in order to suppress isolated cluster occurrence.

From our proposal, reference vector of a neuron which is close to the winner neuron on feature space and has big difference between a reference vector and a training sample on the data space is enormously updated. Neurons that locate on intermediate positions between the winner neurons belonging to different classes on feature space are affected by the several winner neurons; therefore, those neurons may provide cause of occurrence of isolated cluster in the basic SOM. In contrast to the basic SOM, it is expected that those neurons may not provide cause of occurrence of isolated clusters in our proposal because it applies not only distance on feature space but also distance on data space to training algorithm.

The experimental results show that our proposal has succeeded in suppressing occurrence of isolated cluster. Moreover, our proposal achieved a 96% for the success rate of classification for evaluation data, and that rate increases by 6% compared with the basic SOM.

Section 2 presents algorithm of the basic SOM and its modification in this research, and Section 3 shows the results of experiments. Section 4 concludes and summarizes this paper.

2. **Self-Organizing Map (SOM).** This section presents algorithm of the basic SOM and the method of improvement of introducing distance on data space.

2.1. **The basic SOM.** Self-Organizing Map (SOM) proposed by Kohonen [2,3] is one of clustering algorithms with training from a training dataset, and that is suitable for constructing a classifier. SOM is more superior to the other clustering method in terms of introducing the idea of neighborhood on feature space, so that SOM is able to map a training dataset to feature space with preserving topology among training samples on data space.

SOM is 2-layered structure, i.e., the input layer and the competitive layer. The input layer contains one neuron. That neuron has an input vector, $\mathbf{x}$, and each element of an input vector corresponds to each element of a training sample. The competitive layer contains some competitive neurons. All of competitive neurons are connected to the input layer neuron. A competitive neuron has a reference (weight) vector, $\mathbf{w}$, and each element of a reference vector corresponds to each element of input vector. Dimension of the competitive layer is normally 2 dimensional space. Figure 1 illustrates the overview of SOM structure.

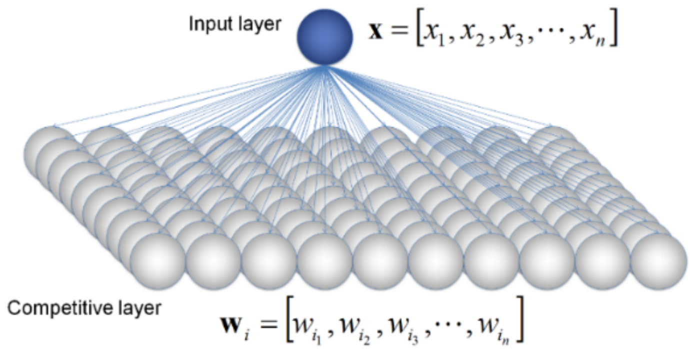


Figure 1. The structure of SOM. The suffix i corresponds to the neuron number on the competitive layer, and its n corresponds to the number of elements.

The competitive layer is often called the "feature space" because clusters are formed by competitive training based on the features among training samples. So, reference vectors in each neuron are updated on the basis of the "winner take all". As mentioned above, SOM preserves topology among training data. This is because neurons which are neighboring to the neuron with the highest fitness are also updated on the basis of distance on feature space. Consequently, neurons that have similar reference vectors gather within local area on feature space; on the other hand, neurons that have big different reference vectors are never adjacent to each other. Here, the neuron with highest fitness is called the "winner neuron" in SOM. The algorithm of SOM updates reference vector of the winner neuron to minimize distance norm between a training sample and its reference vector; then, reference vectors of neurons in neighboring of the winner neuron are also updated simultaneously. There are 2 kinds of SOM, i.e., "sequential SOM" and "batch SOM". This research applies batch SOM to training and clustering. The summary of training procedure for batch SOM is below.

1) Initializing reference vectors in each neuron on competitive layer.

2) Determination for the winner neuron based on following formula

$$k^*(i) = \arg\min_k ||\mathbf{w}_k - \mathbf{x}_i|| \tag{1}$$

3) Calculation for updating coefficient of reference vectors as,

$$\alpha_{k,i} = \frac{h\left(\left\|y_k - y_{k^*(i)}\right\|; \sigma(t)\right)}{\sum_{i'=1}^{N} h\left(\left\|y_k - y_{k^*(i')}\right\|; \sigma(t)\right)} \tag{2}$$

4) Updating reference vectors in each neuron from following equation

$$\mathbf{w}_k(t) = (1 - \varepsilon)\mathbf{w}_k(t-1) + \varepsilon \sum_{i=1}^{N} \alpha_{k,i}\mathbf{x}_i \tag{3}$$

5) Repeat 2 to 4.

Here, k and k^* are the locations of a neuron and the winner neuron on the competitive layer, respectively. $\alpha_{k,i}$ is dynamic coefficient of updating for a neuron k for the i-th training sample. N is the number of training samples. $h(\cdot; \sigma(t))$ is the neighborhood function. ε is the coefficient of training. $\sigma(t)$ is the function for determination of neighborhood size depending on training time step, t.

The neighborhood function is defined as monotonically decreasing according to distance norm on feature space, so that this research applies Gaussian function to that function. The function $\sigma(t)$ is defined as monotonically decreasing according to increasing of training time step, t; therefore, this research applies Equation (4).

$$\sigma(t) = \sigma_f + (\sigma_s - \sigma_f)\exp\left(-\frac{t}{\tau}\right) \tag{4}$$

where σ_s and σ_f are the initial radius and the final radius of neighborhood, respectively. τ is the coefficient to determine the speed of shrinking of neighborhood radius.

2.2. **Improvement of SOM based on introducing the distance on data space.** From the training procedure in above section, the basic SOM only applies distance norm on data space to determination of the winner neuron, and reference vector in a neuron is updated on the basis of distance norm on feature space between the winner neuron and that neuron from Equations (2) and (3). For this, neurons that locate on intermediate positions between the winner neurons belonging to different classes on feature space are affected by the several winner neurons; therefore, those neurons may provide cause of occurrence of isolated cluster in the basic SOM. Additionally, it becomes a cause for necessity of many training steps if reference vectors in neurons in the vicinity of the winner neuron are so different from reference vector of the winner neuron. To overcome those problems, this research introduces distance on data space to the training rule of SOM in addition to basic SOM rule.

The modification for training rule in SOM is as the following. This research regards distance between a training sample and a reference vector on data space as an "error", so that an error, $E_{k,i}$, is defined by distance on data space between a training sample, $\mathbf{x}_i$, and reference vectors in each neuron, $\mathbf{w}_k$; then a reference vector is updated depending on the error and on distance between the winner neuron and a neuron on feature space. Consequently, even if reference vector of the k-th neuron is so different from the i-th training sample and that neuron is not adjacent to the winner neuron, reference vector of that neuron is drastically updated, and vise-versa.

When updating process for reference vector uses distance norm on data space directly, it is not able to stably train under long distance norm because updating values change drastically in that case. Hence, hyperbolic tangent function, tanh, as Equation (6) is introduced in order to solve this problem in this research. The function, tanh, is a nonlinear monotonically increasing function and is saturated to 1 in a certain degree of domain. The aim of our proposal is to consider distance on data space for training of SOM; hence, we introduce the error term to Equation (4) such as Equation (5). The procedure for training of SOM by introducing distance of data space is as below.

1) Initializing reference vectors in each neuron on competitive layer.
2) Determination for the winner neuron as the same manner as the basic SOM.
3) Calculation for updating coefficient of reference vectors as,

$$\beta_{k,i} = \frac{h\left(\left\|y_k - y_{k^*(i)}\right\|; \sigma(t)\right) + E_{k,i}/t}{\sum_{i'=1}^{N} h\left(\left\|y_k - y_{k^*(i')}\right\|; \sigma(t)\right) + \left(\sum_{j=1}^{M} E_{j,i}\right) \Big/ t} \tag{5}$$

$$E_{k,i} = \tanh\left(\frac{\|\mathbf{w}_k - \mathbf{x}_i\|}{\max_k \|\mathbf{w}_k - \mathbf{x}_i\|}\right) \tag{6}$$

4) Updating reference vectors in each neuron from following equation

$$\mathbf{w}_k(t) = (1-\varepsilon)\mathbf{w}_k(t-1) + \varepsilon \sum_{i=1}^{N} \beta_{k,i}\mathbf{x}_i \tag{7}$$

5) Repeat 2 to 4.

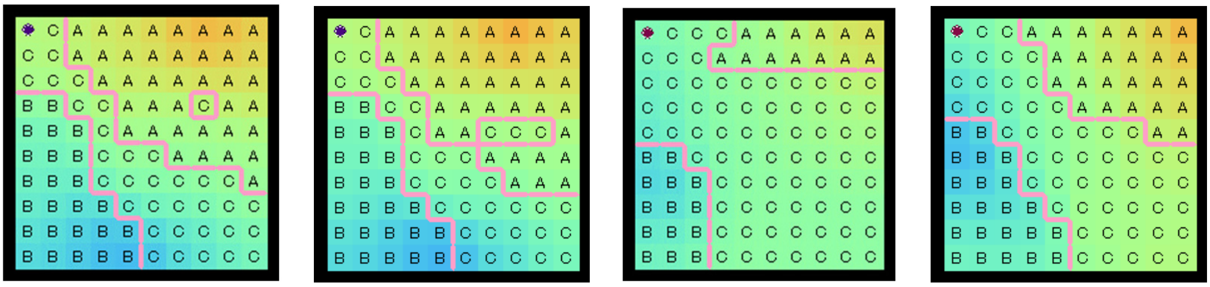
Here, M is the number of neurons on competitive layer.

The argument in Equation (6) takes 0 to 1 because a distance on data space between the i-th training sample and the k-th competitive neuron is divided by the maximum distance, so that the maximum error is about 0.76. The error of the k-th competitive neuron for the i-th training sample is applied to updating coefficient, $\beta_{k,i}$, based on the ratio to the sum of errors in each neuron. An error and the sum of errors are divided by the number of training time steps, t, in order to gradually decrease influence of errors. For this, neurons that locate on intermediate positions between the winner neurons belonging to different classes on feature space are pulled into the certain class, so that it is expected that those neurons may not provide cause of occurrence of isolated clusters in our proposal. The setting for τ is discussed later.

3. **Experiments.** This section presents the experimental conditions and its results. In addition, it discusses the influence of parameters.

3.1. **Experimental conditions.** The training data in this research is "Iris Datasets" provided by UCI Machine learning repository [8]. That training dataset is known for evaluation data of Artificial Neural Networks, Machine Learning, etc. That dataset contains 150 training samples. Training samples are classified into 3 (class A: Iris-setosa, class B: Iris-versicolor and class C: Iris-virginica) classes. The number of attributes is 4, so data space is 4 dimensional space. The dataset is separated into 2 datasets, i.e., 50 training and 50 evaluation samples in this research. In each dataset, the number of samples for the class A is 18, for the class B is 14 and for the class C is 18. Those samples are randomly chosen from each class. SOM is trained by training samples, then class estimation capability for unknown data is evaluated by evaluation samples. In this research, the competitive layer is 2 dimensional space, so feature space is the same size as that space. The competitive layer is constructed by 10×10, total 100 neurons. The parameters for initial neighborhood size, σ_s, and that final, σ_f, are 10 and 1, respectively. The training coefficient, ε, is 1. The number of training steps is 1000 times. The initial values of reference vectors are the same for all of experiments.

The number of neurons is larger than the number of training samples. The reason for the size of competitive layer is intentionally to have the neurons which never become the winner; therefore, it is able to clearly show the effectiveness of our proposal. The determination of other parameters are the same way as general SOM.



(a) $t = 40$ (b) $t = 1000$ (c) $t = 40$ (d) $t = 1000$

FIGURE 2. The results of comparison between the basic SOM and improved SOM. Figures (a) and (c) are after 40 times, and (b) and (d) are after 1000 times of training, respectively.

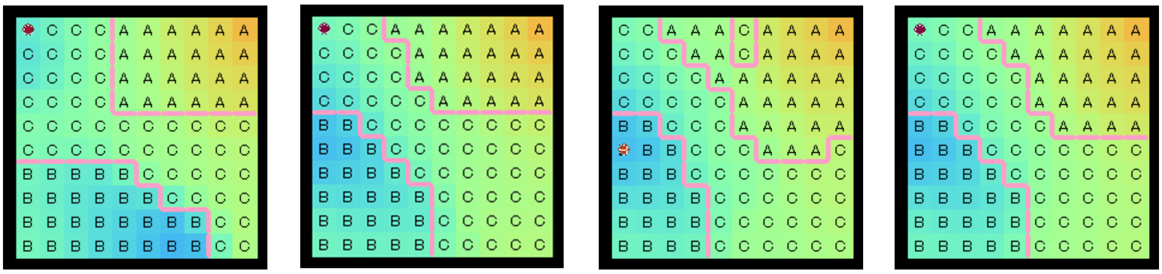
3.2. **Experimental results.** Firstly, the results of comparison between the basic SOM and the improved SOM are shown. Figures 2(a) and 2(c) show the results after 40 times of training in the basic SOM and improved SOM, respectively. Figures 2(b) and 2(d) show the results after 1000 times of training in the basic SOM and improved SOM, respectively. The shrinking parameter, τ, is 10 in this experiment.

In the case of the basic SOM, Figure 2(a) shows that an isolated cluster for the class C occurs in a cluster of the class A. By contrast to the basic SOM, Figure 2(c) shows that there is no isolated cluster in the case of the improved SOM. From Figure 2(a), it can be said that an isolated cluster in 2(b) is formed from the isolated cluster in 2(a). From Figure 2(d), there are no isolated clusters in improved SOM when training is finished. The success rate of classification for evaluation samples of the basic SOM is 90%, i.e., estimations of belonging to classes for 5 unknown samples are not successful. On the other hand, that rate of the improved SOM is 96%, i.e., failures of estimation of belonging to classes for unknown data are only 2 samples. Those results show that SOM which is introduced distance on data space achieved success for suppressing of occurrence isolated cluster and for increase of success rate of classification.

Secondly, it verifies influences of the parameter value of shrinking of neighborhood size, τ, below. From Equation (4), neighborhood size shrinks slowly when that parameter is a large value, and that size shrinks rapidly when that parameter is a small value. For this, neurons that locate on intermediate positions between the winner neurons belonging to different classes on feature space are affected by both of distance on feature space and that on data space for long training time steps if that parameter is a large value. The parameter values of shrinking of neighborhood size are 5, 15, 20 and 30 in this experiment. The results of verification experiments for influences of the parameter values of shrinking of neighborhood size are shown as Figure 3.

From Figure 3, it achieves success for suppressing of occurrence isolated clusters when shrinking parameter is 5, 15 and 30. By contrast to those parameters, an isolated cluster occurs when shrinking parameter is 20. The success rate of classification for evaluation data is 94% when shrinking parameter is 5, 20 and 30. On the other hand, that rate is 96% when shrinking parameter is 15. The cause of those results is that stable training is difficult in neurons which are affected by several winner neurons for long training time steps; hence, it can be said that the determination for the parameter value of shrinking of neighborhood size is key in this research.

4. **Conclusions.** There is "Self-Organizing Map" (SOM) as an approach to simulation of visual cortex different from Deep Convolutional Neural Network (DCNN). The feature of SOM which is different from DCNN is that SOM constructs clusters like the receptive field in V1 based on the competitive training. One of the problems of SOM is occurrence



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

Figure 3. The results of verification experiments for influences of the parameter values of shrinking of neighborhood size. Figures (a) to (d) correspond to $\tau = 5$, $\tau = 15$, $\tau = 20$ and $\tau = 30$, respectively.

of isolated clusters belonging to the same class. The causes for that problem are random initialization of reference vectors and mapping of high dimensional data space to low dimensional feature space. In the beginning stage of training, a neuron which is distant from the other winner neurons may fire for another input vectors belonging to the same class due to random initialization. Moreover, it is difficult to map high to low dimensional space with completely preserving relations among training samples. To overcome this problem, this research introduces distance on data space to the training rule of SOM in addition to the basic SOM.

Firstly, it compares the basic SOM with the improved SOM under the same parameter settings in the experiments. From the results of comparison, occurrence of isolated clusters is suppressed with introducing distance on data space, and the success rate of classification for evaluation data increases by 6%. Second experiments are verification of influences for mapping of the parameter values of shrinking of neighborhood size. The parameter values of shrinking of neighborhood size are 5, 15, 20 and 30 in experiments. From the results of second experiments, it is validated occurrence of isolated clusters depending on setting of shrinking parameter; therefore, it can be said that the determination for the parameter value of shrinking of neighborhood size is key in the approach of this research.

From all of experimental results, introducing of the nonlinear monotonically decreasing function to distance on data space in Equation (5) or applying another function to Equation (6) is further study. Moreover, we intend to apply this proposal to real world problems for evaluation of effectiveness in our future study such as object category recognition from images.

Acknowledgment. This research was supported by Grant-in-Aid for Special Research Equipment from Nishinippon Institute of Technology, 2008.

REFERENCES

[1] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, vol.25, pp.1097-1105, 2012.
[2] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *Proc. of the 3rd International Conference on Learning Representations (ICLR 2015)*, 2015.
[3] T. Kohonen, Essentials of the self-organizing map, *Neural Networks*, vol.37, pp.52-65, 2013.
[4] M. M. Van Hulle, *Faithful Representations and Topographic Maps*, Wiley-Interscience, 2001.
[5] T. Abe, H. Sugawara, S. Kanaya, M. Kinouchi, Y. Matsuura, H. Tokutaka and T. Ikemura, Large-scale Self-Organizing Map (SOM) constructed with the Earth Simulator unveils sequence characteristics of a wide range of eukaryotic genomes, *Proc. of Workshop 2005 on Self-Organizing Map*, pp.187-194, 2005.
[6] H. Tokutaka, M. Ohkita, M. Oyabu and K. Fujimura, The classification of the chain-link and other data by spherical SOM, *Proc. of the 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligence Systems*, pp.458-461, 2012.

[7] H. Wakuya, Y. Horinouchi, H. Itoh, H. Fukumoto and T. Furukawa, An analysis of multidimensional data containing emphasized items by self-organizing map and its application to sightseeing information analysis, *Soft Computing*, vol.21, no.12, pp.3345-3352, 2017.
[8] UCI Machine Learning Repository, *Iris Data Set*, https://archive.ics.uci.edu/ml/datasets/Iris.