A New Clustering Validity Index for Cluster Analysis Based on a Two-Level SOM

Shu-Ling SHIEH, Student Member and I-En LIAO, Nonmember

SUMMARY  Self-Organizing Map (SOM) is a powerful tool for the exploratory of clustering methods. Clustering is the most important task in unsupervised learning and clustering validity is a major issue in cluster analysis. In this paper, a new clustering validity index is proposed to generate the clustering result of a two-level SOM. This is performed by using the separation rate of inter-cluster, the relative density of inter-cluster, and the cohesion rate of intra-cluster. The clustering validity index is proposed to find the optimal numbers of clusters and determine which two neighboring clusters can be merged in a hierarchical clustering of a two-level SOM. Experiments show that, the proposed algorithm is able to cluster data more accurately than the classical clustering algorithms which is based on a two-level SOM and is better able to find an optimal number of clusters by maximizing the clustering validity index.

key words: self-organizing map, clustering, clustering validity index

1. Introduction

Self-organizing map (SOM), originally suggested by Kohonen [10], is a very popular unsupervised neural-network model for the analysis of high-dimensional patterns in data mining applications. It has been used to map high-dimensional input space with the ability to reduce it to two dimensions. Each input data point can be assigned to a cluster according to its nearest output neuron. The neurons in SOM will reveal, in some special topological structure, the output space according to the feature of the input vectors. Such types of topological structures would also denote the characteristic of the input samples. SOM has been employed in a wide range of applications in various different domains, including speech recognition, image data compression, robot control, pattern recognition, and medical diagnosis [13], [14], [17], [18], [20], [22], [24], [26], [27].

The clustering technology is often used to discover the patterns and beneficial relations or the hidden information in a data set. Clustering algorithms partition a data set into several disjointed groups. In this method, points in the same group are similar to each other according to particular similarity metrics. Clustering validity is one of the most essential issues in clustering techniques. This is due to the specific clustering algorithm that is featured in the different input parameter combinations. The method of acquiring the best performance clustering result is a popular topic in SOM’s. Many studies have reviewed the clustering, visualization and data mining capabilities of SOM [2], [10], [11], [19], [23]. Characteristically, there are three types of methods used to evaluate clustering validity: 1) external criteria, 2) internal criteria, and 3) relative criteria [21]. Dragut and Nichitiu [4] have proposed two principles of evaluating clustering results. The two principles are the separation rate of the inter-cluster and the compactness of the intra-cluster. A good clustering validity index is able to help users in choosing the appropriate two-level SOM algorithms.

The traditional SOM requires a value for the number of clusters designed for clustering the input data. The concept of two-level SOM is that the first level is to train data by the SOM and the second level is to cluster data based on the neurons of the SOM. In general two-level SOM, neurons over the expected number of clusters that will be generated should be prepared for the first-level SOM. The second-level clustering algorithm is executed by clustering of output neurons after the training performed by the first-level SOM.

In this paper, a new clustering validity method is proposed that can judge the clustering result of any two-level SOM. This will be achieved by using: 1) the separation rate of inter-cluster, 2) the relative density of inter-cluster, and 3) the cohesion rate of intra-cluster. As a result, by generating evaluating results, we are able to achieve the optimal number of clusters and the highest accuracy from different clustering algorithms. The agglomerative hierarchical clustering is adapted to merge groups recursively from the second-level SOM output until a desired number of clusters are reached.

The remaining sections of this paper are organized as follows. Section 2 briefly presents Kohonen’s SOM algorithm and related work on clustering methods. Section 3 gives a detailed description of the proposed algorithm on synthetic and real data sets. Section 4 provides an overview of the experimental results. These demonstrate that the proposed algorithm is able to cluster the input data and find the optimal number of clusters. The experimental results show that the clustering accuracy of the proposed algorithm is better than that of other clustering algorithms on a two-level SOM. Finally, Sect. 5 gives the conclusions.
2. Related Work

SOM algorithm is applicable to large data sets. The goal of SOM is to transform patterns of high dimensionality into a low-dimensional topological map. The training algorithm proposed by Kohonen for forming a feature map is stated as follows [8], [19].

**Step 1) Initialization:** Choose random values for the initial weights \( w_i \).

**Step 2) Winner Finding:** Find the winning neuron \( c \) at time \( t \), using the minimum Euclidean distance criterion:

\[
    c = \arg \min \| x - w_i \|, \quad i = 1, 2, \ldots, M,
\]

where \( x = [x_1, \ldots, x_m] \in \mathbb{R}^m \) represents an input vector at time \( t \), \( M \) is the total number of neurons, and \( \| \cdot \| \) indicates the Euclidean norm.

**Step 3) Weights Updating:** Adjust the weights of the winner and the neighbor neurons, using the following rule:

\[
    w_i(t + 1) = w_i(t) + \eta(t) h_{ci}(t)(x(t) - w_i(t)),
\]

\[
    h_{ci}(t) = \exp \left( -\frac{\| r_c - r_i \|^2}{2\sigma^2(t)} \right),
\]

where \( x_i(t) \) represents an input data at time \( t \), \( h_{ci}(t) \) is the topological neighborhood function of the winner neuron \( c \) at time \( t \), \( \eta(t) \) is a positive constant called “learning-rate factor”, \( r_c \in \mathbb{R}^2 \) are the location vectors of nodes and \( r_i \in \mathbb{R}^2 \) are the location vectors of node \( i \). \( \sigma(t) \) defines the width of the kernel. Both \( \eta(t) \) and \( \sigma(t) \) will decrease with time. It should be emphasized that the success of the map formation is critically dependent on the values of the main parameters \( h_{ci}(t) \) and \( \eta(t) \), initial values of weight vectors and the pre-specified number of iterations.

In the case of a discrete data set and fixed neighborhood kernel, the sum of the squared-error of SOM can be defined as follows:

\[
    SSE = \sum_{j=1}^{n} \sum_{i=1}^{M} h_{ci} \| x_j - w_i \|^2,
\]

where \( n \) is the number of training samples and \( M \) is the number of map units. A neighborhood kernel \( h_{ci} \) is centered at unit \( c \), which is the BMU of input vector \( x_i \) and evaluated for unit \( i \).

Many researchers proposed a two-level SOM to get a result with hierarchical structures without knowing the cluster number of the input data. Lampinen and Oja [12] use a method consisting of a two-level SOM. The first-level SOM is used as a training procedure, while the second-level SOM is used for clustering. Murtagh [16] applied an agglomerative contiguity constrained clustering method to merge the neurons trained in the first-level SOM. The merging criterion is to apply the shortest distance between the neurons.

Table 1 Four types of definitions of inter-cluster distance.

| Definition          | Inter-cluster distance |
|---------------------|------------------------|
| Single-linkage      | \( d_{\text{single}}(C_i, C_j) = \min_{p \in C_i, q \in C_j} \| p - q \| \) |
| Complete-linkage    | \( d_{\text{complete}}(C_i, C_j) = \max_{p \in C_i, q \in C_j} \| p - q \| \) |
| Centroid-linkage    | \( d_{\text{centroid}}(C_i, C_j) = \| u_i - u_j \| \) |
| Average-linkage     | \( d_{\text{average}}(C_i, C_j) = \frac{1}{n_{C_i}} \sum_{p \in C_i} \sum_{q \in C_j} \| p - q \| \) |
| Ward’s-linkage      | \( d_{\text{ward}}(C_i, C_j) = n_i \| u_i - u_j \| + n_j \| u_j - u_i \| \) |

Kiang [9] proposed the first step as recruiting more neurons than required for executing SOM algorithm and then a refining space topological method is applied by means of a conscience mechanism [3]. After getting the result of the first-level SOM, a merging criterion of minimum variance is applied to the succeeding steps. Wu et al. [25] proposed a clustering validity index based on the inter-cluster, intra-cluster density and inter-cluster distances in the clustering assessments. The assessment is able to find an optimal partition of the input data.

Clustering technology is an essential technique in data mining. In recent studies, clustering technology has broadly been classified into the following categories [6], [7]: 1) Partition clustering algorithms, 2) Hierarchical clustering algorithms, 3) Density-based clustering algorithms, 4) Grid-based clustering algorithms and 5) A mixed clustering category which includes artificial neural network such as SOM and evolutionary computation. This research adopts a two-level SOM network, where this SOM could be satisfied with convergent conditions in the short period of time when it deals with high-dimensional databases.

In the classical hierarchical clustering, a pair of clusters to be merged will have the minimum inter-cluster distance. The inputs to a linkage algorithm can either be distances or similarities between pairs of clusters. The widely used measures of inter-cluster distance in single-linkage, complete-linkage, centroid-linkage, average-linkage and Ward’s-linkage are shown in Table 1, where \( u_i \) is the mean for cluster \( C_i \), \( n_i \) is the number of points in cluster \( C_i \) and \( u \) is the mean after merging clusters. All of this inter-cluster distance generates the similar results, if clusters are well separated and compacted. In contrast, when clusters are closed to one another or if their shapes are not hyper-spherical, the clustering result would have a greater difference. However, the single-linkage clustering algorithm is able to explore any shapes of clusters [4], but the algorithm is easily influenced by outliers and noise [15]. Therefore, the proposed clustering algorithm uses multi-indices to represent a cluster, which uses more information in addition to inter-cluster distance and overcomes the problems of outlier data and non-spherical shapes of clusters.

3. A New SOM-Based Clustering Validity Index

The clustering validity indices of SOM are extremely important for automatically determining the optimal number of clusters. In this paper, a new SOM-based clustering validity index is proposed. In the following subsection, the
clustering validity index that combines the separation rate of inter-cluster, the relative density of inter-cluster, and the cohesion rate of intra-cluster is presented.

3.1 The Separation Rate of Inter-Cluster

To cause clusters to have the lowest similarity, data points that belong to different clusters should be put as far away as possible. There are five popular methods to evaluate inter-cluster distance, as shown in Table 1. The separation rate proposed in this paper utilizes the single-linkage inter-cluster distances, inter-cluster density and the distance between cluster centers. The separation rate \( S_{ij} \) is defined as follows:

\[
S_{ij} = \left( \frac{\| p_i - p_j \|}{\| u_i - u_j \|} \right) \times \frac{1}{1 + \text{density}(u_{ij})}, \quad (5)
\]

where \( p_i \) and \( p_j \) are the closest pair points between the \( i \)th and \( j \)th clusters, \( u_i \) is the center of the \( i \)th cluster and \( u_j \) is the center of the \( j \)th cluster, \( \| u_i - u_j \| \) represents the Euclidean distance between \( u_i \) and \( u_j \), and \( u_{ij} \) is the middle point between the pair points \( p_i \) and \( p_j \) as shown in Fig. 1.

Inter-cluster density is defined as the density in the area between two clusters. The inter-cluster density between cluster \( i \) and cluster \( j \) is represented by \( \text{density}(u_{ij}) \), which is defined as follows:

\[
\text{density}(u_{ij}) = \frac{\sum_{x \in C_i \cup C_j} f(x, u_{ij})}{n_i + n_j}, \quad (7)
\]

where \( n_i \) and \( n_j \) represent the number of data points in the \( i \)th and \( j \)th cluster, respectively; \( x \) is the input vector belonging to the \( i \)th or \( j \)th cluster. The density function represents that the percentage of points in the cluster \( i \) and the cluster \( j \) that belong to the neighborhood of \( u_{ij} \). The neighborhood of \( u_{ij} \) is defined as the circle with \( u_{ij} \) as the center and \((\sigma_i + \sigma_j)/2\) as the radius, where \( \sigma_i \) and \( \sigma_j \) are the standard deviation of cluster \( i \) and cluster \( j \), respectively. It is obvious that a data point belongs to the neighborhood of \( u_{ij} \) if its distance from \( u_{ij} \), defined as \( d(x, u_{ij}) \), is smaller than the average standard deviation of two clusters. The function \( f(x, u_{ij}) \) is defined as:

\[
f(x, u_{ij}) = \begin{cases} 
1 & \text{if } d(x, u_{ij}) \leq (\sigma_i + \sigma_j)/2 \\
0 & \text{otherwise}
\end{cases} \quad (8)
\]

\[
\sigma_j = \sqrt{\frac{\sum_{x \in C_i} \| x - u_i \|^2}{n_i - 1}}, \quad (9)
\]

where the number \( \sigma_i \) is the standard deviation of the cluster \( i \) and the number \( \sigma_j \) is the standard deviation of the cluster \( j \).

The algorithms in [25] use inter-cluster and intra-cluster density, along with inter-cluster distances for clustering. However, this algorithm does not take into account the distance between the \( i \)th cluster center and \( j \)th cluster center. For instance, both the nearest distances of the two adjacent clusters are 3 in Fig. 2 (a) and Fig. 2 (b). The algorithm in [25] calculates the separation rate of Fig. 2 (a) and Fig. 2 (b) as the same. However, the radius of data points in Fig. 2 (a) is smaller than that in Fig. 2 (b). In the proposed algorithm, the distance \( \| u_i - u_j \| \) is taken into consideration in calculating the separation rate.

The separation rate between clusters evaluates the separation level of clusters. It includes the inter-cluster distance, inter-cluster density and the distance between cluster centers. The objective that will be achieved is that the separation rate should be as high as possible, while inter-cluster distance is significantly high, inter-cluster density is low and the distance between cluster centers is high.
3.2 The Relative Density of Inter-Cluster

A good clustering validity criteria must consider whether an inter-cluster separation rate is obvious, along with whether or not intra-cluster data has compactness, but it must also determine a very important factor: the relative density of an inter-cluster. According to recent research, many clustering algorithms put great emphasis on the concept of clustering density, such as density-based DBSCAN clustering algorithm [5]. These methods describe that when the density in the neighborhood of points exceeds some threshold value, it then would become a new cluster. However, if someone only emphasizes the discriminate and cohesion of clustering and does not take the numbers of data in intra-cluster into consideration, it would not effectively represent the separation rate and evaluating result of the cohesion rate. Therefore in this study, it is generalized that the inter-cluster relative density is also the critical factor which should be taken into consideration.

In this paper, the relative density is understood as the ratio between the density of a cluster and that of the remaining clusters. The relative density $R_{ri}$ is defined as follows:

$$R_{ri} = \frac{D_i}{D_r},$$

where $R_{ri}$ is the proportion of the density of cluster $i$ compared with the average density of the remaining clusters, $D_i$ represents the density of the cluster $i$ and $D_r$ stands for the average density of the remaining clusters excluding the cluster $i$. The density of a cluster is the result of dividing the number of data points by the data space of these points. However, since the density would generate extreme values when data points are spread in a high dimensional area, then this density indicator would become invalid. Therefore, this method uses a substitute to calculate the density. The value of intra-cluster density is calculated as the number of data points in the cluster divided by the standard deviation of the cluster. The standard deviation of a cluster implies the sparse ratio of the data points in a cluster. The higher standard deviation represents that the cluster is insubstantial.

$$D_i = \frac{N_i}{\sigma_i},$$

$$D_r = \frac{n_i}{\sigma_i},$$

In the formula above, the number $N_i$ is the data points in the clusters excluding cluster $i$, the number $n_i$ is the number of data points number of the cluster $i$; the standard deviation $\sigma_i$, represents the compactness of the data space excluding cluster $i$ and is defined as follows:

$$\sigma_i = \sum_{j \neq i} \sigma_j / k - 1.$$  \(13\)

When $R_{ri}$ is bigger, it means that the density of the cluster $i$ is higher and the clustering accuracy would be much better, when compared with the average density of the remaining clusters.

3.3 The Cohesion Rate of Intra-Cluster

The cohesion rate evaluates the centralized degree of the data points in the cluster. This research proposes a cohesion rate of the intra-cluster to find the rate of the data points which fall into the average distance of data points in a cluster to its cluster center.

In general, we can assume that the data set are from an approximately normally distributed population. The notation of the cohesion rate is defined as follows: the symbol $\mu$ represents the mean of the distances between every point of the cluster to the cluster center and the standard deviation is $\sigma$. It is derived that 99.74% of data points would fall in $\mu \pm 3\sigma$, 95.44% of data points would fall in $\mu \pm 2\sigma$ and 68.26% of data points would fall in $\mu \pm \sigma$. Using this property, the intra-cluster cohesion rate, $Hi$, of cluster $i$ is defined as the ratio between the number of data points in $\mu \pm \sigma$ and that in $\mu \pm 3\sigma$. This could also be seen as the outlier data points that fall out of $\mu \pm 3\sigma$.

$$Hi = \frac{n_{\mu \pm \sigma}}{n_{\mu \pm 3\sigma}},$$

$$0 \leq Hi \leq 1,$$  \(14\)

$$n_{\mu \pm \sigma} = \sum_{x \in C_i} f(x, u_i),$$  \(15\)

$$f(x, u_i) = \begin{cases} 1 & \|x - u_i\| \leq \sigma_i \\ 0 & \text{otherwise} \end{cases},$$  \(16\)

where $n_{\mu \pm \sigma}$ is the number of data points which fall in the cluster $i$ of range $\mu \pm \sigma$ and $n_{\mu \pm 3\sigma}$ is the number of data points which fall in the cluster $i$ of range $\mu \pm 3\sigma$. When more data points fall into the cluster within the range of $\mu \pm \sigma$, the cohesion rate comes closer to 1, which means the cohesion rate of data points in the cluster would be significantly higher.

3.4 Clustering of SOM Using Clustering Validity Index

In this paper, a new clustering validity is proposed to have three important clustering validity subindices. The generalized clustering validity index contains: 1) the separation rate of inter-cluster, 2) the relative density of inter-cluster and 3) the cohesion rate of intra-cluster. These are used to determine the optimal clustering numbers of the second-level SOM. The resulting clustering will have best accuracy.

The overall clustering validity index, which is called $SOM_{\text{clustering\_validity}}$, is defined by

$$SOM_{\text{clustering\_validity}} = \frac{1}{C_k^2} \sum_{i=1}^{k} \sum_{j=1, j \neq i}^{k} S_{ij} \times (R_{ri} \times H_i) \times (R_{rj} \times H_j),$$  \(17\)

where $C_k^2$ is the number of combinations for choosing two clusters out of $k$ total clusters, and $S_{ij}$ represents the separation rate between cluster $i$ and $j$. Moreover, $R_{ri}$ and $R_{rj}$ represent relative density of cluster $i$ and cluster $j$, respectively, and $H_i$ and $H_j$ represent cohesion rate of cluster $i$ and cluster $j$, respectively. The value of $SOM_{\text{clustering\_validity}}$
reaches the maximum at the optimal number of clusters of a two-level SOM. Therefore, a better clustering accuracy can be achieved.

In the clustering process, the agglomerative hierarchical clustering uses a two-level SOM. The $SOM_{clustering\_validity}$ for all pairs of neighboring clusters fits into the merging criterion. The merging process is described as follows:

- Compute the $SOM_{clustering\_validity}$ for the input data belonging to each neighboring pair of clusters.
- Find the pair of clusters with a minimal value of the $SOM_{clustering\_validity}$, which indicates that the two clusters must be merged into one cluster.
- The merging process is utilized until only two clusters exist or there cannot be any further merging.
- After the merging process, find the optimal number of clusters according to the $SOM_{clustering\_validity}$ for all the input data.

4. Experimental Results

In these experiments, the program is written in Java using Borland JBuilder 9 Enterprise Edition. The platform used is an Intel Pentium 4 3.2 GHz with 512 MB DRAM and 80 GB hard disk running Windows 2000 Server Pack3.

All experiments were performed on the datasets from the University of California at Irvine - Machine Learning Repository [1] to illustrate the effectiveness of the proposed clustering algorithm. Two data sets were used to demonstrate the effectiveness of the proposed algorithm in the experiments. To achieve a better clustering result and to avoid the negative effects produced by noises and outliers, all data sets were pre-processed using data cleaning normalization schemes.

4.1 Wine Data Set

Wine data set [1] has 178 13D data with three known classes. The numbers of data samples in the three classes are 59, 71 and 48. The proposed algorithm is used with a map size of $4 \times 4$ to cluster the data. The $SOM_{clustering\_validity}$ is calculated as a function of the number of clusters at the second-level SOM. Figure 3 shows that the clustering validity index has maximum value of three for the proposed algorithm and k-means on the Wine data set, which matches the actual number of clusters. The clustering accuracies by the proposed algorithms, the CDbw [25], the extended SOM [9], the k-means ($k = 16$), Single-linkage, Complete-linkage, Centroid-linkage, Average-linkage, and Ward’s-linkage are shown in Table 2. The CDbw algorithm includes many preprocessing steps as a prerequisite before clustering of SOM with these steps raising the clustering accuracy. In order to make a fair comparison, this experiment uses the same preprocessing steps to avoid the noises and outliers. The results achieved by the CDbw, the extended SOM and the single-linkage clustering of the SOM are not as good as the result achieved using the proposed clustering of $SOM_{clustering\_validity}$ approach. The proposed clustering of $SOM_{clustering\_validity}$ achieved the best clustering result with a clustering accuracy of 85.4%.

4.2 2D Artificial Dataset

2-D artificial dataset (http://cilab.csie.ncu.edu.tw/course/cluster/FCM’s Data.zip) is a computer simulation data set. The data set contains 579 artificial data points with 2 columns each. The data consists of three shallow elongated parallel clusters in the 2D plane. The proposed algorithm is used with a map size of $5 \times 5$ to cluster the data as shown in Fig. 4.

The $SOM_{clustering\_validity}$ is a function of the number of clusters as plotted in Fig. 5. This indicates that the number of clusters is three, which is equal to the number of classes. For 2D artificial dataset, the proposed clustering of $SOM_{clustering\_validity}$ achieved the best clustering result with the highest clustering accuracy of 98.6% as shown in Table 3. Though the five direct agglomerative hierarchical clustering algorithms, excluding the single-linkage clustering, have the same clustering results with an accuracy of 98.6%, the five direct agglomerative hierarchical clustering algorithms do not find an optimal number of clusters. The proposed clustering algorithm determines the optimal number of clusters by maximizing the validity index and achieved the best clustering result with a clustering accuracy of 98.6%.
Fig. 4 The two-level SOM approach of clustering on 2D artificial data set.

Fig. 5 The clustering validity as a function of the number of clusters for 2D artificial data set.

Table 3 Accuracy comparison of different clustering algorithms for 2D artificial data set.

| Algorithm                  | Accuracy (%) |
|----------------------------|--------------|
| Proposed SOM\(^{\text{clustering validity}}\) | 98.6%        |
| CDbw [25]                  | 78.6%        |
| Extended SOM\[^{9}\]       | 39.7%        |
| k-means \(k=25\)          | 98.6%        |
| Single-linkage             | 69.6%        |
| Complete-linkage           | 98.6%        |
| Centroid-linkage           | 98.6%        |
| Average-linkage            | 98.6%        |
| Ward’s-linkage             | 98.6%        |

5. Conclusions

In this paper, a new clustering validity index for the two-level SOM algorithm is proposed to automatically determine the optimal number of clusters. This clustering validity index includes the separation rate of inter-cluster, the relative density of inter-cluster, and the cohesion rate of intra-cluster. A clustering validity index, SOM\(^{\text{clustering validity}}\), has been defined for evaluating the results of clustering algorithms. The optimal number of clusters is obtained by the maximum value of SOM\(^{\text{clustering validity}}\). The performance of the proposed clustering validity index was tested using a 2D artificial dataset and a real dataset. Its results were compared to those of several other validity indices. The experimental results indicate that the proposed validity index achieves the best clustering result with a clustering accuracy and is reliable on two-level SOM.

References

[1] C.L. Blake and C.J. Merz, UCI repository of machine learning databases, (http://www.ics.uci.edu/~mlearn/MLRepository.html), Department of Information and Computer Science, University of California at Irvine, CA, 1998.

[2] J. Chakma and K. Umemura, “Factor controlled hierarchical SOM visualization for large set of data,” IEICE Trans. Inf. & Syst., vol.E86-D, no.9, pp.1796–1803, Sept. 2003.

[3] D. Desieno, “Adding a conscience to competitive learning,” Proc. IEEE International Conference on Neural Networks, pp.117–124, 1988.

[4] A. Dragut and C.M. Nichitiu, “A monotonic on-line linear algorithm for hierarchical agglomerative classification,” Information Technology and Management, vol.5, no.1-2, pp.111–141, 2004.

[5] M. Ester, H.P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” KDD, pp.226–231, 1996.

[6] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2000.

[7] A.K. Jain, “Data clustering: A review,” ACM Comput. Surv., vol.31, no.3, pp.264–323, Sept. 1999.

[8] J.A. Kangas, T.K. Kohonen, and J.T. Laaksonen, “Variants of self-organizing maps,” IEEE Trans. Neural Netw., vol.1, no.1, pp.93–99, 1990.

[9] M.Y. Kiang, “Extending the Kohonen self-organizing map networks for clustering analysis,” Computational Statistics & Data Analysis, vol.38, no.2, pp.161–180, 2001.

[10] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas, “Engineering applications of the self-organizing map,” Proc. IEEE, pp.1358–1384, 1996.

[11] M.A. Kraaijveld, J. Mao, and A.K. Jain, “A nonlinear projection method based on Kohonen’s topology preserving maps,” IEEE Trans. Neural Netw., vol.6, no.3, pp.548–559, 1995.

[12] J. Lampinen and E. Oja, “Clustering of hierarchical self-organizing maps,” Journal of Mathematical Imaging and Vision, vol.2, pp.261–272, 1992.

[13] L. Leinonen, T. Hiltunen, K. Torkkola, and J. Kangas, “Self-organized acoustic feature map in detection of fricative-vowel coarticulation,” J. Acoust. Soc. Am., vol.93, no.6, pp.3468–3474, 1993.

[14] C.N. Manikopoulos, “Finite state vector quantization with neural network classification of states,” IEEE Proc. Radar and Signal Processing, vol.140, no.3, pp.153–161, 1993.

[15] U. Maulik and S. Bandyopadhyay, “Performance evaluation of some clustering algorithms and validity indices,” IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.12, pp.1650–1654, 2002.

[16] F. Murtagh, “Interpreting the Kohonen self-organizing feature map using contiguity-constrained clustering,” Pattern Recognit. Lett., vol.16, pp.399–408, 1995.

[17] H.B. Perex and F.G. Nocetti, “Fault classification based upon self organizing feature maps and dynamic principal component analysis for inertial sensor drift,” International Journal of Innovative Computing, Information and Control, vol.3, no.2, pp.257–276, 2007.
[18] H. Ressom, D. Wang, and P. Natarajan, “Clustering gene expression data using adaptive double self-organizing map,” Physiol. Genomics, vol.14, pp.35–46, 2003.

[19] S.L. Shieh, I.E. Liao, K.F. Hwang, and H.Y. Chen, “An efficient initialization scheme for SOM algorithm based on reference point and filters,” IEICE Trans. Inf. & Syst., vol.E92-D, no.3, pp.422–432, March 2009.

[20] M.C. Su and H.T. Chang, “A new model of self-organizing neural networks and its application in data projection,” IEEE Trans. Neural Netw., vol.12, no.1, pp.153–158, 2001.

[21] S. Theodoridis and K. Koutroubas, Pattern Recognition, Academic Press, New York, 1999.

[22] L. Vercauteren, G. Sieben, M. Praet, G. Otte, R. Vingerhoeds, L. Boulart, L. Calliauw, and H. Roels, “The classification of brain tumours by a topological map,” Proc. International Neural Networks Conference, pp.287–391, Paris, 1990.

[23] J. Vesanto and E. Alhoniemi, “Clustering of the self-organizing map,” IEEE Trans. Neural Netw., vol.11, no.3, pp.586–600, 2000.

[24] J.A. Walter and K.J. Schulten, “Implementation of self-organizing neural networks for visuo-motor control of an industrial robot,” IEEE Trans. Neural Netw., vol.4, no.1, pp.86–95, 1993.

[25] S. Wu and T.W.S. Chow, “Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density,” Pattern Recognit., vol.37, pp.175–188, 2004.

[26] L. Yun and K. Uchimura, “Using self-organizing map for road network extraction from Ikonos imagery,” International Journal of Innovative Computing, Information and Control, vol.3, no.3, pp.641–656, 2007.

[27] X. Zhang, N. Zhang, J. Lu, and T. Yahagi, “Independent component analysis for image recovery using SOM-based noise detection,” IEICE Trans. Fundamentals, vol.E90-A, no.6, pp.1125–1132, June 2007.

Shu-Ling Shieh received the M.S. degree in information management from National Yunlin University of Science & Technology, Taiwan, in 1996. Presently she is pursuing Ph.D. at the Department of Computer Science of National Chung Hsing University, Taiwan. She is currently a lecturer in the Department of Information Management of the Ling-Tung University, Taiwan. Her research interests are in data mining, neural network, self-organizing map, and information visualization.

I-En Liao received the B.S. degree in Applied Mathematics from National Cheng-Chi University, Taiwan, in 1978, and both the M.S. degree in Mathematics and the Ph.D. degree in Computer and Information Science from the Ohio State University in 1983 and 1990, respectively. He is currently a professor in the Department of Computer Science of National Chung Hsing University, Taiwan. His research interests are in data mining, XML database, wireless networks, and bioinformatics. He is a member of the ACM and the IEEE Computer Society.