DEVOPS-AN OPEN ACCESS JOURNAL
Open Journal

Review

A Comparative Analysis on Artificial Neural Network-Based Two-Stage Clustering

Cheng-Ching Chang; Ssu-Han Chen*

[1]*Department of Library and Information Science, National Taiwan University, Taipei, Taiwan*
[2]*Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei City, Taiwan*

Corresponding author
Ssu-Han Chen
Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei City, Taiwan; E-mail: ssuhanchen@mail.mcut.edu.tw

Article information
Received: October 30th, 2022; **Revised:** January 5th, 2023; **Accepted:** January 28th, 2023; **Published:** February 9th, 2023

Cite this article
Chang C-C, Chen S-H. A comparative analysis on artificial neural network-based two-stage clustering. 2023; 2(1).
doi: https://doi.org/10.70705/ppp.doaj.2023.v02.i01.pp10-15

ABSTRACT

The artificial neural network (ANN), which is capable of noise removal and data complexity reduction, has been regarded as one of outstanding intermediaries in the two-stage clustering procedures. Various ANN-based two-stage clustering procedures have been individually proposed; however, the performance among those methods has not been examined yet. In this study, a preliminary comparative analysis is conducted in four benchmark data-sets and a real-world market data-set, which are used to simulate various conditions for evaluation purposes. The experiment results suggest that high-accuracy self-organizing feature map can potentially improve the effectiveness of decision-making.

Keywords
Two-stage clustering; k-means algorithm; Self-organizing feature map; Adaptive resonance theory.

INTRODUCTION

Life is full of data. Every day, people process massive information that may be stored or represented as data for further analysis and management. One of the essential tools in coping with these data is to assign them into two or more unknown groups, rendering certain similar properties to members in the same group. A general inquiry faced by researchers in conducting many applications is how to discover meaningful structures in observed input vectors. That is, how to develop valid taxonomies (Xu & Wunsch, 2005). Cluster analysis encompasses a variety of unsupervised algorithms. Take the artificial neural network (ANN) for example, it is capable of grouping them into easily interpreted distinct clus- ters if the data-set was well separated (i.e. input vectors with topologically mutually exclusive). Ideal condition, however, is not always the case. Usually, the clustering results of the ANN do not exhibit distinct clusters, but ill-defined boundaries, resulting in undesirable mumbling or ambiguity about the cluster structures and membership. The follow-up in-deep analysis may primarily depend on analyz- ers' intuition or subjective opinion (Abidi & Ong, 2000). An acceptable way to solve the mentioned problems is to draw upon a synergy between the ANN and a partition-based technique such as k-means algorithm. In the first stage, to yield prototypes without any information about clustering numbers, the unsupervised neural model was used to map the input vector. However, it is a universal truth that the number of generated prototypes of ANN is always too huge for quantitative or qualita- tive interpretation. In the second stage, to facilitate cluster identification and interpretation, further grouping of similar prototypes together can delimit cluster numbers and prototype boundaries.

The two-stage clustering procedure has been found excelling in setting boundaries, compared to direct clustering, in many previous studies. Among them, the hybrid of self-organizing feature map (SOFM) and k-means, or SOFM + k-means, is the most well-known procedures. Vesanto and Alhoniemi (2000) and Abidi and Ong (2000) are the forerunners of clustering prototypes obtained by pre-treating input vectors with SOFM. Hereafter, numerous studies applied similar idea on distinct applications. Sugiyama and Kotani (2002) analyzed and categorized the gene expression data. Canetta, Cheikhrouhou, and Glardon (2005) concerned problems on group technology and purchased components in industrial company. Godin, Huguet, and Gaertner (2005) discriminated acoustic emission signals to monitor the chronology of the damaging process. Chaimeun and Srivihok (2005) clustered handicraft customers to understand the customer background and behaviors for the purposes of enhancing the marketing strategy and planning. Khedairia and Khadir (2008) identified the meteorological day types of the Annaba region. Wu, Xia, Chen, and Cui (2011) researched on classifying moving objects into pedestrians, bicycles, and vehicles in traffic video. Chen, Pan, and Jiang (2013) applied rolling SOFM + k-means clustering for fault diagnosis. Hayfron-Acquah and Gyimah (2014) matched fingercodes using SOFM and then classified fingerprints with the help of k-means. In addition, adaptive resonance theory (ART) neural network was recently excavated to be another intermediate step for clustering analysis. Ding, Shi, Shi, and Jiang (2008)

combined ART2 and k-means to detect anomaly (named ART2 + k-means). Park, Suresh, and Jeong (2008) integrated fuzzy ART with k-means to develop a sequence-based clustering for Web usage mining (named Fuzzy ART + k-means). Hung, Chen, Yang, and Deng (2013) used ART2 + k-means to mine the elder self-care cluster patterns. The prior arts agreed with the fact that the two-stage clustering procedure considerably removes potential outliers and noise, as well as reduces data complexity. Since a prototype is the local average of a set of nearby input vectors, the clustering is less sensitive to outliers and noise. These advantages implicitly prompt the two-stage clustering procedures to yield a satisfied quality and efficiency of clustering assignment.

The above two-stage clustering procedures are proposed separately, and performance compari- son among them has not yet been examined. Consequently, we conducted a comparative study on these methods. These comparisons are primarily based on the analysis of four benchmark data-sets: iris, wine, image segmentation, and handwritten digit. Comparing clustering results of using ANN as an intermediate with those obtained from direct method, we can, therefore, demonstrate how the use of ANN improves clustering quality. In order to further evaluate these two-stage clustering pro- cedures, the comparison is extended to a real-world market segmental application.

The rest of this paper is organized as follows. In Section 2, related methodologies are surveyed. In Section 3, the clustering procedures are described and the experimental results are presented. The concluding remarks are discussed in Section 4.

2. Literature review

In view of the ability to compress information, the ANN is an ideal tool for analyzing large data-sets (Juntunen, Liukkonen, Lehtola, & Hiltunen, 2014). In this section, the unsupervised ANN techniques under consideration are sketched. For simplification, we describe the various ANN methods using pseudo-algorithms. Please refer to the original works if interested.

2.1. Introduction of SOFM network

The SOFM network, proposed by Kohonen (1990), converts a higher dimensional input space into a lower dimensional map space and preserves the topological properties of the input vectors by means of a neighborhood concept. SOFM network consists of neurons that associated with weight vectors in input space and a position in the map space. The arrangement of neurons, so called as SOFM lattice, is presented in a triangular, rectangular, or hexagonal grid. The steps of the SOFM algorithm can be summarized as follows:

Step 1: Initialization. Initialize learning rate (α) and synaptic weights (w).

Step 2: Sampling. Sample a training input vector (x) from input space.

Step 3: Matching. Calculate $y=||x-w||$ and find the best matching unit (BMU) with wJ closest to x.

Step 4: Updating. Apply the weight update equation

$$\Delta wi=hJ\times(x-wi) \text{ ----------(1)}$$

where $hic=\alpha\times\exp(||ri-rc||2/2\times\sigma^2)$ is a neighborhood function around the BMU. The weight vectors of the BMU and it neighborhood in the SOFM lattice are updated together toward this input vector.

Step 5: Iteration. Keep returning to Step 2 until the map stops to change.

2.2. Introduction of ART2 network

The ART2 network, proposed by Carpenter and Grossberg (1987), solves the stability and plasticity dilemma in clustering problem. It swiftly updates its model to the new input vector without specify- ing the number of clusters. The main advantages of ART2 are: learning and adapting to a non-stable environment rapidly, stability and plasticity, unsupervised learning of preferences behavior, as well as deciding the number of groups automatically. The steps of the ART2 algorithm can be summa- rized as follows:

Step 1: Initialization. Initialize the nameless parameters (a, b, c, d, and e), learning rate (α), vigilance value (ϱ), synaptic weights of bottom-to-top (bj) and top-to-bottom (tj), and neurons at F1 layer (w, s, u, v, p, q).

Step 2: Activation of short-term memory (STM) in F1 layer. Randomly chosen input vector activates six neurons in F1 layer by calculating the following equations and proceeds along the bot- tom-to-top weight vector (biJ) to F2 layer.

$$w=x+a\times u \text{ ----------(2)}$$
$$s=w/(e+w) \text{ ----------(3)}$$
$$u=v/(e+v) \text{ ----------(4)}$$
$$v=f(s)+b\times f(q) \text{ ----------(5)}$$
$$p=u+d\times tJ \text{ ----------(6)}$$
$$q=p/(e+p) \text{ ----------(7)}$$

where $f(x) = x$ if $x\geq 0$, otherwise $f(x)=0$.

Step 3: Activation in F2 layer. The p is sent to F2 layer for calculating $y=\Sigma(p\times bj)$ and finds the maximum activation unit J in the F2 layer.

Step 4: Pattern Matching. Update p and u and then calculate match value (r)

$$r=(u+c\times p)/(e+u+c\times p) \text{ ----------(8)}$$

Step 5: Reset or resonant. If $||r|| + e < \varrho$, then $yJ = -1$ and inhibit the unit J (go back to Step 3). Otherwise, it represents that STM has gone into resonant state, such that long-term memory (LTM) start to learn. If all of the neurons in F2 layer are inhibited, a brand-new neuron is generated and is initialized in terms of the novelty.

Step 6: Learning process of LTM. The slow-learning rule updates the synaptic weights of unit J is:

$$bJ=a\times d\times u+1+a\times d\times(1-d)\times bJ \text{ ----------(9)}$$
$$tJ=a\times d\times u+1+a\times d\times(1-d)\times tJ \text{ ----------(10)}$$

Step 7: Iteration. Keep returning to Step 2 until the number of epochs is reached.

2.3. Introduction of fuzzy ART network

The fuzzy ART network, proposed by Carpenter, Grossberg, and

Rosen (1991), possesses similar advantages of ART2. It benefits the incorporation of fuzzy set theory and ART principle, thus enhancing generalizability. Another distinct characteristic of fuzzy ART is the complement coding, a mean of normalization process that rescales and doubles the dimension of input vector, which helps avoid the problem of category proliferation but still preserves most information of input vectors. The steps of the Fuzzy ART algorithm can be summarized as follows:

Step 1: Initialization. Set parameters of learning rate (α), constant (β) and vigilance value (ϱ) and ini- tialize synaptic weights (w).

Step 2: Complement coding. Input vector (x) is augmented by complement coding (x□) in F1 layer.

Step 3: Activation in F2 layer. Calculate the net to each template: yj = ||x□ □ w|| / (β + ||w||), where □ is the fuzzy min operator. The largest net of unit J is the winner.

Step 4: Pattern Matching. Calculate match value (r)

r = □x□ □ wJ□ /□x□ ----------(11)

Step 5: Reset or resonant. If ||r|| < ϱ, then yJ=−1 and inhibit the unit J (go back to Step 3). Otherwise, network starts to learn. If all neurons are inhibited, a new neuron is generated.

Step 6: Learning process of weight. The learning rule updates the synaptic weights of unit J is:

w = (1 − □)× w + □ × (x□ □ w)

Step 7: Iteration. Keep returning to Step 2 until the number of epochs is reached.

3. Experiments and results (12)

The flowchart of implemented work is shown in Figure 1. Benchmark and real-world data-sets are used for evaluating both the direct clustering method and the ANN-based two-stage clustering procedures. The clustering results are then evaluated by several cluster validations if there is no class label information in advance; otherwise, not only validations, but also accuracy rate compari- sons will be conducted. Using a self-programming toolkit, the analysis was conducted under the "MATLAB R2007a" environment.

3.1. Clustering analysis

3.1.1. Direct clustering method

Before applying the routine of k-means algorithm on a given data-set, some tricks should be noted. First, the unit and range of the variables may discrepant. It is meaningless if the segment results are obviously dominated by certain variables. Since normalization equalizes the scale for computing distance, the data prepared for feeding for k-means algorithm must be normalized. A normalization equation suggested by Weiss and Indurkhya (1998) is given as follows:

z=0.9-0.1×x−min⁄max−min ----------(13)

Second, the clustering results of k-means algorithm depend on what initial centers were selected. Accordingly, we duplicated each experiment 100 times, each with a new set of initialization, and selected the best results to reduce the effect of local minimum in initialization.

3.1.2. ANN-based two-stage clustering procedures

First of all, the input vectors are preprocessed by means of Equation 13. The parameters and synaptic weights of each network are basically initialized in terms of the recommendation setting of original authors (Carpenter & Grossberg, 1987; Carpenter et al., 1991; Kohonen, 1990), except for the number of topological nodes of SOFM and the vigilance value of ART2 or fuzzy ART. In this study, the number of topological nodes is set to M = 5 × N1/2, where N is the number of input vectors (Vesanto & Alhoniemi, 2000) and the vigilance value is assigned a high value, e.g. ϱ = 0.99. In addition, we re-feed the random arranged input vectors into the learned network for 100 times. Such a design may enable neural networks to learn much sufficiently. After the learning process of each neural network, each input vector is allocated to a most similar synaptic weight/template. For each fired synaptic weight/template, the average of its input vectors is calculated to generate a set of proto- type vectors. Finally, the prototype vectors are conveyed into k-means algorithm for second-stage clustering. Figures 2 and 3 are the flow charts of the two-stage clustering procedures.

3.1.3. Clustering validation

Several clustering validation indexes are used to determine well-defined partitions, which exhibit both strong external isolation and tight internal cohesion. Since no consistent conclusions were drawn to a close in literatures, with respect to clustering validation indexes, Davies–Bouldin Index (DBI) (Davies & Bouldin, 1979), Calinski–Harabasz Index (CHI) (Calinski & Harabasz, 1974), Ray–Turi Index (RTI) (Ray & Turi, 1999) and Dunn Index (DI) (Dunn, 1974), are all included. These validation indexes are adopted since are practically proved to be excellent indexes for evaluating clustering results (Bezdek & Pal, 1998; Bandyopadhyay & Maulik, 2001; Chen & Dai, 2004; Charrad, Lechevallier, Ahmed, & Saporta, 2010; Vesanto & Alhoniemi, 2000). Among those indexes, DBI is a function of the ratio of sum of within-cluster scatter to between-cluster separation (Davies & Bouldin, 1979). The ideal DBI presents minimal ratio of within-cluster scatter and between-cluster separation; therefore, minimizing within-cluster scatter and maximizing between-cluster separation are desired. CHI is a function of the ratio of sum of squares among the clusters to sum of squares within the clusters. A better clustering result is indicated by a higher CH value. RTI is a function of the ratio of the intra- cluster distance to minimal of inter-cluster distance. It took only the minimum of inter-cluster distance because they need the smallest of this distance to be maximized, and the other larger ones will be bigger than this value automatically (Ray & Turi, 1999). The clustering result which gives a minimum RTI will tell us what the ideal number of clusters is since minimizing inter-cluster distance and maximizing inter-cluster one are presented. DI is a function which takes the minimal ratio of inter-cluster distance to maximal intra-cluster distance. The main goal of DI is to maximize inter- cluster distances and minimize intra-cluster distances. Therefore, the number of clusters that maximizes DI is taken as the ideal clustering result.

3.2. Comparison analysis

3.2.1. Comparison bases on benchmark data-sets

The simulating experiments use four benchmark data-sets obtained from distinct domains and sources. First, the iris data-set contains three classes and each class consists of 50 instances, where each class refers to a type of iris plant and each instance is represented by four features (Fisher, 1936). Second, the wine data-set describes three wines derived from different cultivars in terms of 13 chemical constituents (Forina, Lanteri, Armanino, & Lauter, 1991). Third, the image segmentation data-set records seven kinds of outdoor images, which are quantized by 19 continuous geometric attributes. Only 210 samples of train set are adopted from the UCI machine-learning repository (http://archive.ics.uci.edu/ml/datasets.html). Finally, the optical recognition of handwritten digits' data-set includes digits of 0–9 that are scanned as 32 × 32 bitmaps and then characterized by num- ber of pixels of 4 × 4 non-overlapping blocks. Only 1797 samples of test set are adopted from the UCI machine-learning repository. Table 1 summarizes the main characteristics of these data-sets. The sizes of these data-sets range from hundreds to thousands, and the dimensions, from 4 to 64. In this experiment, the clustering result is evaluated by comparing the cluster label (CLi) of each input vec- tor with its true label (Li) provided by the data-set. The accuracy is defined as follows:

$$\text{accuracy} = \Sigma\delta\, Li\,, \text{map}\, CLi\, /n \text{ ----------(14)}$$

where n is number of input vector. The response of the delta function, δ(x, y), equals to one, if x = y; equals to zero, otherwise. The best mapping function, map(x), matches true label and obtained clus- ter label in which the best mapping can be found automatically by Kuhn–Munkres algorithm.

During clustering procedure, the number of clusters of k-means algorithm is equal to the number of classes of each data-set as shown in the fourth row of Table 1. It should be noted that all of these data-sets are labeled, whereas the labels are omitted during clustering and only used for evaluation. As we can observe in Figure 4, the clustering accuracy of data-sets, iris and wine, is quite satisfied. The majority of these two data-sets reaches 90%. While the other two data-sets only express relatively low accuracy, around 80%. Among them, the accuracies of k-means algorithm are worst; follow by ART2 + k-means. The accuracies of SOFM + k-means and fuzzy ART + k-means are nearly equal in the first two data-sets, whereas the accuracies of SOFM + k-means surpass in the last two data-sets.

Another part of this sub-section is to compare the various methods using four benchmark databases, without imposing the number of clusters in advance. The purpose is to prove the fact that the number of clusters is one the most difficult hurdles in clustering analysis. In this experiment, the setting of clustering number of each clustering algorithm segments benchmark data-sets which are initialized from two, and are then added regularly until an empty cluster signal responds. At the same time, piecewise DBI, CHI, IIVI, and DI against number of clusters are recorded. Generally, well- separated clusters are expected to decrease or increase monotonically as increases until the ideal number of clusters is achieved (Bezdek & Pal, 1998). However, finding the minimum or maximum values is less reliable than finding a knee or a sharp change of slope in the plot. In this study, we determine the number of clusters by finding a knee in DBI and CHI, and by finding a sharp change of slope in RTI and DI. Take the data-set of wine as an example, the corresponding plots of different cluster validations against number of clusters, ranging from 2 to 14, are given in Figure 5(a–d). Following the above rule of thumb, the recommended number of clusters is marked by hollow symbol. The results are shown in Table 2 where the correct number is marked in bold italics. According to the obtained results, we proved the fact that specifying correct number of clusters is difficult in clustering analysis. None of indexes under different methods finds out the ideal number of clusters, three, for iris and wine data-sets. As for image segmentation and handwritten digits data-sets, parts of indexes under different methods get the work done properly, especially the SOFM + k-means procedure.

3.2.2. Comparison bases on a real-world marketing data-set

Recency, frequency, and monetary (RFM) analysis, introduced by Hughes (1994), is a marketing technique used to quantitate the activity of customers by examining customers' recent purchased, purchasing frequency, and each transaction amount. With customers' purchasing behavior data, RFM variables can be derived from historical transaction records. After mapping the each customer's records into one of the several clusters, each cluster, then, represents a market segment, to which distinct marketing strategy shall apply, helping companies offer each consumer the right promotions and save either propaganda time or miscellaneous costs.

Our experimental data are adopted from a online tea shop in Taiwan, whose members are about 150,000, and 90% of members are over aged 40. Whenever a member logs in the web, the system records the date, time, and goods selected by customers. The online shop has 4,867 members as effective customers. Each of them has their own three-dimensional RFM values. The unit and range of RFM variables are discrepant; therefore, certain specified variables may dominate clustering results. In order to eliminate scale effects, the data prepared as input of cluster algorithms are normalized by Equation 13. The clustering procedure is similar to what we had done in Section 3.2.1 except the clustering performance evaluation method. In this database, there are no predefined classes, and we do not have a clue about how many clusters exist in the data-set. Accordingly, clustering validation indexes are employed to determine well-defined partitions. DBI, CHI, RTI, and DI are used for evaluation purpose, as in Section 3.2.1.

In this experiment, k-means algorithm directly segmented the customer with setting of number of clusters from three to ten, which yielded the corresponding values of validation indexes. The upper limit of cluster number was set mainly because routine always responds to empty cluster once it was larger than 10. On the other hand, during the first stage of two-stage clustering algo- rithms, 12 of 361 SOFM synaptic weights, five of 210 ART2 templates, and one of 950 fuzzy ART templates were eliminated because there are no customers involved. Then k-means algorithm was applied on the prototypes of individual network with similar setting of the direct method which yielded the corresponding DBI, CHI, RTI, and DI.

The plots of different cluster validations against number of clusters for marketing data-set are given in Figure 6(a–d). We observed that the broken line of different validations of k-means algorithm is uniformly higher or lower than those of two- stage clustering procedures, implying that the two-stage clustering procedures generate better defined clustering results than that of the direct method. Among the two-stage clustering proce- dures, the ART2 + k-means is the worst, whereas SOM + k-means is the best, supporting our observa- tion in Section 3.2.1.

Each position of customer cluster is based on comparing its average RFM values with the total average (Ha & Park, 1998). If the average of a variable in a cluster is greater than the total average, an upward arrow is assigned to that variable; otherwise, a downward arrow is assigned. According to this rule, RFM status as well as strategic positions could be determined. We determine the number of clusters of the better method, SOFM + k-means, by the rule of thumb described in Section 3.2.1, suggesting that the ideal number of clusters is seven with respect to the marketing data-set. The
result is summarized in Table 3. Clusters 2, 3, and 7, which have R↓F↑M↑, shall be considered as loyal
customers who patronize and make a large purchase recently and frequently, especially Cluster 2
whose transaction amount is so huge that this cluster shall be further regarded as a golden customer cluster. Customers in Cluster 1, with R↓F↓M↓, may be new customers who recently visit the website. Cluster 5 possessing R↑F↑M↓ is treated as the potential cluster and top-priority target segment
because each customer in this cluster may be promoted as a loyal customer. Vulnerable customers, with the status of R↑F↓M↓, are presented in Clusters 4 and 6 who are attracted by this website and may have churned.

4. Conclusions

Segments which exhibit both strong external isolation and tight internal cohesion are well defined. The motivation for two-stage clustering architectures was ignited by researched papers that surmount single-stage drawbacks. Four benchmark data-sets and a real-world marketing data-set are used for evaluation purposes. The benchmark data-sets are used for evaluating the clustering accuracy as well as for choosing number of clusters using different indexes, whereas the real-world data-set is used to compare the cluster validations under a series of number of clusters for different methods. The experimental results are unambiguous and agree with previous researches for three points:

(1) The two-stage clustering procedures surpass the direct clustering algorithm because of the intermediary of two-stage clustering procedure, i.e. ANN, which is capable of handling a wide range of messy data. The promises of noise removal and data complexity reduction make the homogeneous clusters easier to get.

(2) Comparing the discovery quality under various conditions, the SOFM + k-means has the lead, then fuzzy ART + k-means, and lastly ART2 + k-means. The SOFM + k-means can potentially improve the quality of decisions that require the cluster analysis, such as market segmenta- tion, credit analysis, or quality grading, to maintain the competitive advantages.

(3) The identification of the number of clusters is one the most tricky problems in clustering analy- sis. None of the two-stage clustering procedures in this study can perfectly assist users to choose proper clustering number for a wide-ranged data-set.

A sensitivity analysis facilitates researchers and practitioners to use and calibrate a method. On the premise of the selection of multiple indexes as the clustering validations, however, doing sensitivity analysis has become an almost impossible mission since multiple parameters are difficult to hold a consistent decision under different validation indexes simultaneously. Among the mentioned two- stage clustering procedures, the SOFM + k-means has four parameters: arrangement type of neurons, number of epochs, learning rate, and smooth factor of neighbor function; the ART2 + k-means has eight parameters: number of epochs, five nameless parameters, learning rate, and vigilance value; and the fuzzy ART + k-means has four parameters: number of epochs, learning rate, constant, and vigilance value. The key idea of this study is to alleviate the issue of setting parameters, and we had tried our best to find out the default or suggested values that were widely examined by practical studies. Nevertheless, we still believe this study shall be examined by further studies.

Funding

This research is partially supported by the Ministry of Science and Technology, Taiwan [grant number MOST 103-2221-E-131-028].

REFERENCES

Abidi, S. S. R., & Ong, J. (2000). A data mining strategy for inductive data clustering: A synergy between self- organising neural networks and k-means clustering techniques. Proceedings of IEEE TENCON, 2, 568–573.

Bandyopadhyay, S., & Maulik, U. (2001). Nonparametric genetic clustering: Comparison of validity indices. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 31, 120–125.

Bezdek, J. C., & Pal, N. R. (1998). Some new indexes of cluster validity. IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), 28, 301–315.

Calinski, R. B., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3, 1–27.

Canetta, L., Cheikhrouhou, N., & Glardon, R. (2005). Applying two-stage SOM-based clustering approaches to industrial data analysis. Production Planning & Control, 16, 774–784.

Carpenter, G. A., & Grossberg, S. (1987). ART 2: Self- organization of stable category recognition codes for analog input patterns. Applied Optics, 26, 4919–4930.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks, 4, 759–771.

Chaimeun, O., & Srivihok, A. (2005). Clustering of Thai handcraft customers using combined SOM and k-means algorithm. International Conference on of Databases and Applications, 35–39.

Charrad, M., Lechevallier, Y., Ahmed, M. B., & Saporta, G.

(2010). On the number of clusters in block clustering algorithms. Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference, 392–397.

Chen, G., & Dai, Y. (2004). A new distance measurement for clustering time-course gene expression data. In 26th Annual International Conference of the IEEE EMBS

(pp. 2929–2932). San Francisco, CA.

Chen, A., Pan, Y., & Jiang, L. (2013). Improving k-means clustering method in fault diagnosis based on SOM network. Journal of Networks, 8, 680–687.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1, 224–227.

Ding, Y. X., Shi, Y., Shi, Y., & Jiang, J. Q. (2008). A hybrid clustering algorithm based on ART2 and its application in anomaly detection. International Conference on Wavelet Analysis and Pattern Recognition, 282–286.

Dunn, J. C. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.

Journal of Cybernetics, 3, 32–57.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, Part II, 179–188.

Forina, M., Lanteri, S., Armanino, C., & Lauter, S. (1991).

PARVUS—An extendible package for data exploration, classification and correlation. Genoa: Institute of Pharmaceutical and Food Analysis and Technologies.

Godin, N., Huguet, S., & Gaertner, R. (2005). Integration of the Kohonen's self-organising map and k-means

algorithm for the segmentation of the AE data collected during tensile tests on cross-ply composites. NDT&E International, 38, 299–309.

Ha, S. H., & Park, S. C. (1998). Application of data mining tools to hotel data mart on the intranet for database

marketing. Expert Systems with Applications, 15(1), 1–31.

Hayfron-Acquah, J. B., & Gyimah, M. S. (2014). Classification and recognition of fingerprints using self organizing maps (SOM). International Journal of Computer Science Issues, 11, 153–159.

Hughes, A. M. (1994). Strategic database marketing. Chicago, IL: Probus Publishing.

Hung, Y. S., Chen, K. L. B., Yang, C. T., & Deng, G. F. (2013).

Web usage mining for analysing elder self-care behavior patterns. Expert Systems with Applications, 40, 775–783.

Juntunen, P., Liukkonen, M., Lehtola, M., & Hiltunen, Y. (2014).

Characterization of alum floc in water treatment by image analysis and modeling. Cogent Engineering, 1, Article ID: 944767. Retrieved from http://dx.doi. org/10.1080/23311916.2014.944767

Khedairia, S., & Khadir, M. T. (2008). Self-organizing map and k-means for meteorological day type identification for the region of Annaba-Algeria. 7th Computer Information Systems and Industrial Management Applications, 91–96.

Kohonen, T. (1990). Self-organizing map. Proceedings of the IEEE, 78, 1464–1480.

Park, S., Suresh, N. C., & Jeong, B. K. (2008). Sequence-based clustering for web usage mining: A new experimental framework and ANN-enhanced k-means algorithm. Data & Knowledge Engineering, 65, 512–543.

Ray, S., & Turi, R. H. (1999). Determination of number of clusters in k-means clustering and application in colour image segmentation. Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques, 137–143.

Sugiyama, A., & Kotani, M. (2002). Analysis of gene expression data by using self-organizing maps and k-means clustering. Proceedings of International Joint Conference on Neural Networks, 1342–1345.

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self- organizing map. IEEE Transactions on Neural Networks, 11, 586–600.

Weiss, S. M., & Indurkhya, N. (1998). Predictive data mining: A practical guide. San Francisco, CA: Morgan Kaufmann.

Wu, J., Xia, J., Chen, J. M., & Cui, Z. M. (2011). Moving object classification method based on SOM and k-means.

Journal of Computers, 6, 1654–1661.

Xu, R., & Wunsch, II, D. (2005). Survey of clustering algorithms. IEEE Transactions on Neural Networks, 16, 645–678.