Performa (2016) Vol. 15, No.1: 51-58

A Comparison Study: Clustering using Self-Organizing Map and K-means Algorithm

Annisa Uswatun Khasanah[*1)]
1) Department of Industrial Engineering Universitas Islam Indonesia

Abstract

Nowadays clustering is applied in many different scopes of study. There are many methods that have been proposed, but the most widely used is K-means algorithm. Neural network has been also usedin clustering case, and the most popular neural network method for clustering is Self-Organizing Map (SOM). Both methods recently become the most popular and powerful one. Many scholarstry to employ and compare the performance of both mehods. Many papers have been proposed to reveal which one is outperform the other. However, until now there is no exact solution. Different scholar gives different conclusion. In this study, SOM and K-means are compared using three popular data set. Percent misclassified and output visualization graphs (separately and simultaneously with PCA) are presented to verify the comparison result.

Keywords*: clustering, Self-Organizing Map, K-means.*

1. Introduction

In recent years, clustering analysis has been widely applied in many applications, such as social science, biology, medicine, signal processing, engineering, pattern recognition, and computer science(Das et al., 2008; Kwedlo, 2011). Clustering analysis is the process of identifying natural groupings or clusters within multidimensional data, based on some similarities, like Euclidean distance (Jain et al., 1999). Its main purpose is to group samples with the same statistical characteristics together into same cluster in order to achieve higher similarities within same cluster, also there are more significant differences between different clusters (Han and Kamber, 2006). There are many methods that usually used to deal with clustering problems. However, Self-Organizing Map (SOM) and K-means algorithm are the most popular one. SOM, or sometime also called as Kohonen neural network, is powerful data mining tool (Vesanto and Alhoniemi, 2000; Vesanto et al., 1999) and K-means is the most widely used clustering method even though it was proposed over 50 years ago (Jain, 2010). There have been many scholars applied these methods in many different scope of study and there have been a number of papers comparing SOM and K-means algorithm. However, so far there are no definite results, different authors point to different conclusions (Bação et al., 2005). Some authors suggest that SOM performs similar to or slightly better than K-means algorithm (Bação et al., 2005; Flexer, 1999; Waller et al., 1998), while other author conclude the opposite (Balakrishnan et al., 1994).

Waller et al. (1998) try to compare both methodsin social sciences scope. In their study, they compare the classification of the 1-dimensional SOM with two partitioning (one of them is K-means algorithm) and three hierarchical cluster methods. Their study includes 2400 data set with known cluster structure. The program of SOM model was written in S-plus. For this study, two measures of clusters recovery that they used are Cohen's kappa and percent misclassified. Cohen's kapaa is a measurement scale that commonly used in psychological field, which was proposed by Cohen in 1960. From this study, they found that SOM models often produce

[*] *Correspondance* : annisauswatun.kh@gmail.com

cluster solutions that are similar to, or slightly better than those produced by K-means algorithm. In the other paper, Flexer (1999) tried to demonstrate that SOM could be used for the clustering or visualization separately, for simultaneously clustering and visualization, and even for clustering via visualization. For this application, SOM is compared to the other methods (one of them is K-means algorithm) using the data mining tools CLEMENTINE and WEBSOM. He used several data sets, including Iris data set, in this study. He concluded that if the degree of neighborhood set to zero at the end of learning, the result of SOM will equivalent to K-means algorithm. This statement accordance to Waller et al. (1998) which say that SOM and K-means clustering perform equally well in term of data points misclassified when one-dimensional SOM are being used with zero neighborhood and both being better than the other hierarchical cluster methods.

Bação et al. (2005) also try to compare SOM and K-means. The data used in this test is composed of four basic data sets, two synthetics (DS1 and DS2, which were created and consist of 400 and 750 observations respectively) and well known two real world data set (Iris and Sonar data set, which consists of 150 and 208 observations respectively). In order to assess the performance of the two methods, a set of three measurements was used, quadratic error, mean classification error, and structural measurement. Quadratic error is the sum of the squared distances of each point to the centorid of its cluster. This error is divided by the total dispersion of each cluster to obtain a relative measurement. Mean classification error is the number of observations attributed to a cluster where they do not belong. While, structural measurement is used in order to understand if the structural coherence of the groups is preserved by the clustering method. From 100 cycle times for each algorithm, the results show that the clustering obtained with each method is practically the same, but on average SOM outperforms K-means and has far less variation in its results. They also conclude that SOM is less prone to local optima than K-means. On the other hand, K-means gradient orientation forces a premature convergence, which, depending on the initialization, may frequently yield local optimum solutions.

All of the study that have been discussed, conclude that SOM is equivalent to or slightly better than K-means, but Balakrishnan et al. (1994) presents the opposite results. Balakrishnan et al. (1994) compare SOM and K-means algorithm in psychological case. Their results indicates that K-means procedure had fewer points misclassified while the classification accuracy of neural network worsened as the number of cluster the number of clusters in the data increased from two to five. This result is different with what has been stated by Waller et al. (1998). Waller et al. (1998) pointed out that this difference happened due to the difference in programming software. Balakrishnan et al. (1994) used a commercially available product whereas Waller et al. (1998) programmed their own code and this affected the learning rate and neighborhood-size decrement function. In this study, another comparison between SOM and K-means is presented. Percent misclassified and output visualization graphs (separately and simultaneously with PCA) are used to validate the comparison results.

2. Research Method

2.1 Self-Organizing Map (SOM) / Kohonen neural network

SOM or Kohonen neural network was firstly proposed by Teuvo Kohonen in 1972. This method is a kind of unsupervised learning method. Since it belongs to unsupervised learning neural network, target vectors are not needed, since its purpose is to divide the input vectors into clusters of similar vectors. This network can learn to detect regularities and correlations in their inputs (Beale et al., 2011). As illustrated by Figure 1, SOM architecture has two layers, input

and output layer. The input layer consists of *n* input data that will be grouped into *m* clusters. Each unit on the input layer and output layer is called as node/neuron. Input layer and the output layer are connected with weights and these weights will be updated during the training. The distance between input and output node will be measured first using Euclidean distance measurement. The node with the smallest distance will be the winner node. The weights affect the distance between input and output, and these weights will be updated in each iteration in order to minimize the distance. The weights will be randomly generated in the initial step. SOM algorithm can be described as follows,

1. set up network parameters.
   There are several parameters that should be defined in the beginning of the process, such as learning rate *(α)*, shrinking learning rate *(α_rate)*, the topology shape and size,
2. set up connecting weight vector matrix, *w*, randomly,
3. present input vector, *x*, to the network,
4. select the winning weight vector (output node) to find the winner node using Euclidean distance :
   $$d_j = \sqrt{\sum (x - w_i)^2}, \quad (1)$$
   the winning weight vector is denoted as w_c, as shown as follows:
   $$\|x - w_c\| = \min \|x - w_i\|, \quad (2)$$
5. update the weight through equation (3):
   $$w_i(t+1) = w_i(t) + \alpha(t)[x(t) - w_i(t)], \quad (3)$$
6. repeat steps 3-5 until all training samples have been input,
7. shrink learning rate through equations (4):
   $$\alpha(t+1) = \alpha\_rate \times \alpha(t), \quad (4)$$
8. repeat steps 3-7 until the termination criterion is satisfied.

The training will be stopped when the termination criterion is satisfied. The termination criterion is the difference between $w_i(t+1)$ and $w_i(t)$, if there is no significant difference between $w_i(t+1)$ and $w_i(t)$, it means the iteration has been converged and it can be stopped.

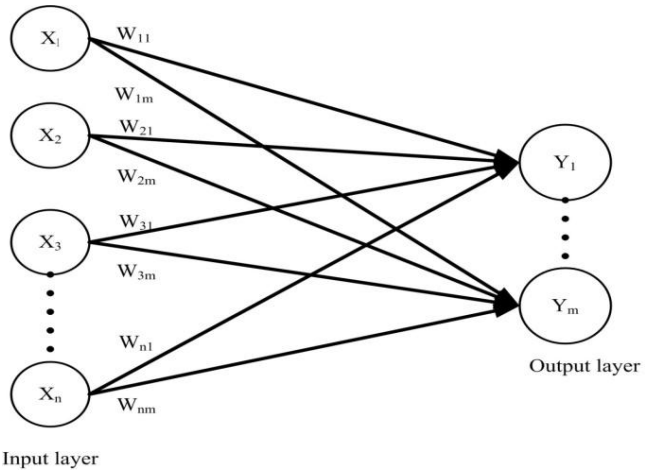


Figure 1. Architecture of one dimensional SOM or Kohonen neural network (Nugraha, 2010)

2.2 K-means algorithm

K-means algorithm is one of the most popular and widely used clustering technique which was firstly proposed by MacQueen in 1967 (Cai et al., 2010; Kwedlo, 2011). K-means algorithm is also well known for its efficiency in clustering large data set (Huang, 1998). K-means algorithm is easy to be implemented and computationally efficient (Das et al., 2008; Kwedlo, 2011; Tan et al., 2006). The K-means algorithm starts with *K* cluster centroids, which are initially randomly selected or derived from some a priori information. Each point in data set is then assigned to the closest centroid, and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is then updated based on the point assigned to each cluster. This process is repeated until no point change clusters, or equivalently, until the centroids remain the same (Tan et al., 2006). K-means algorithm can be further discussed as follows:

1.) determine the number of cluster, *K*,
2.) generate *K* cluster centroids randomly,
3.) calculate the Euclidean distance,
4.) assign each data point to the cluster that the distance between cluster centroid and the data point is the smallest,
5.) recalculate the new centroids,
6.) repeat until the centroids do not change.

Beside the K-means algorithm is easy to implement, the time complexity is only *O(n)* (*n* being the number of data points), which makes it suitable for large data sets. However, K-means algorithm also suffers from several disadvantages, such as, the user has to specify in advance the number of clusters, the performance of the algorithm is data dependent, and this algorithm uses a greedy algorithm approach and is heavily dependent on the initial condition (the centroids of initial cluster). This often leads K-means to converge to suboptimal solution (Das et al., 2008; Kwedlo, 2011).

3. Result and Discussion

In this study, three data sets from UCI Machine Learning Repository Repository (http://archive.ics.uci.edu/ml/) i.e. Iris, Sonar and Wine, are used to verify the clustering results. Table 1 shows the description for each data set. To keep the consistency standards of measurement, this primary component data will be normalized. This can minimize the clustering bias caused by the different units. The normalization transformation procedure is to calculate the distance between two extreme values from the primary data, then minus the minimal value of data to each primary data, then divide it by the difference of extreme values. The algorithms are implemented in C++ Language using Code Block 13.12 on a PC with Intel® Pentium® Dual CPU T3200 @2.00GHz Processor – 1 GB RAM. The three datasets are trained using SOM and K-means algorithm with the same parameters (initialα=1, α_rate = 0,8, number of cycle = 100). The initial weights for SOM and initial centroids for K-means are randomly generated. Then, to validate the comparison results of both algorithms, percent misclassified and output visualization graphs (separately and simultaneously through PCA) are used and will be further discuss as follows.

Table 1. Data sets descriptions

Data set	Number of sample data	Number of attributes	Number of clusters
Iris	150	4	3
Sonar	208	60	2
Wine	178	13	3

3.1 Percent Misclassified

To assess the performance of both algorithms in those data sets, percent misclassified (Bação et al., 2005; Balakrishnan et al., 1994; Waller et al., 1998) is used as the measurement. The clustering results from SOM and K-means algorithm from the result of C++ are compared with the real cluster from the data, the error are calculated and then, the results are presented in Table 2.

Table 2. Percent misclassified for each data set

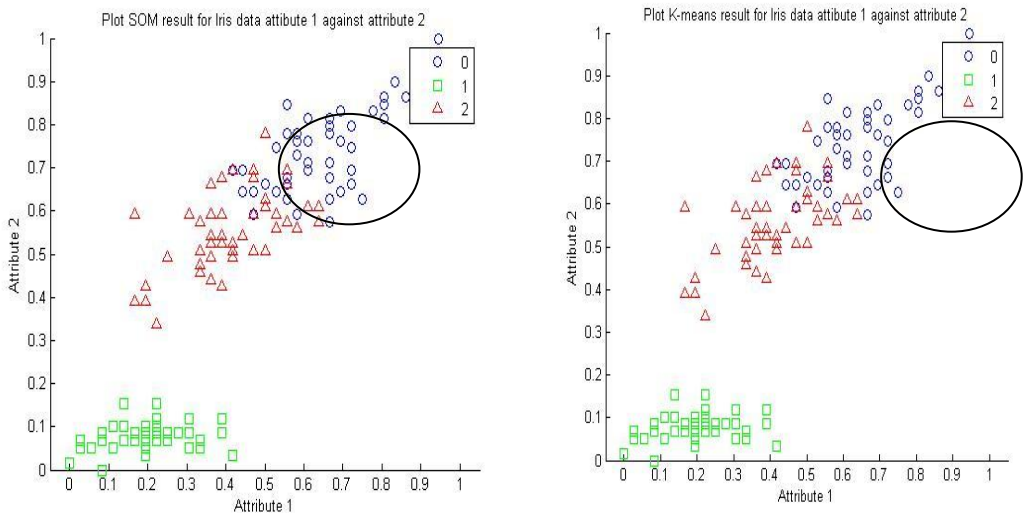
Data set	Percent Misclassified (%)	
	SOM	K-means
Iris	17,30	**12,00**
Sonar	45,19	**43,75**
Wine	13,48	**5,00**

Note: the best results are printed in bold

From overall results in Table 2, it can be indicated that K-means is better than SOM due to the smaller value of percent misclassified. These results are the same with what have been mentioned by Balakrishnan et al. (1994) but different from Waller et al. (1998), Flexer (1999), and Bação et al. (2005). For Sonar data, K-means is just slightly better than SOM, but forIris and Wine data, there are much difference between K-means and SOM. While Bação et al. (2005), who also use Iris and Sonar data sets, showed that for both data sets SOM has smaller value of percent misclassified. As pointed out by Waller et al. (1998) that this difference happened due to the difference in programming software.

3.2 Comparison through Graphics

Another analysis to compare the SOM and K-means clustering results is also conducted through visualization graphic as proposed by Flexer (1999). Theses graphics are presented in two-dimensional space for each attribute against another attribute. In this study not all graphics for each features are presented, only graphics for attribute 1 against attribute 2 for each data set are shown as illustrated by Figure 1. The graphics show the spread of data and the corresponding clusters. Different color indicates different cluster.



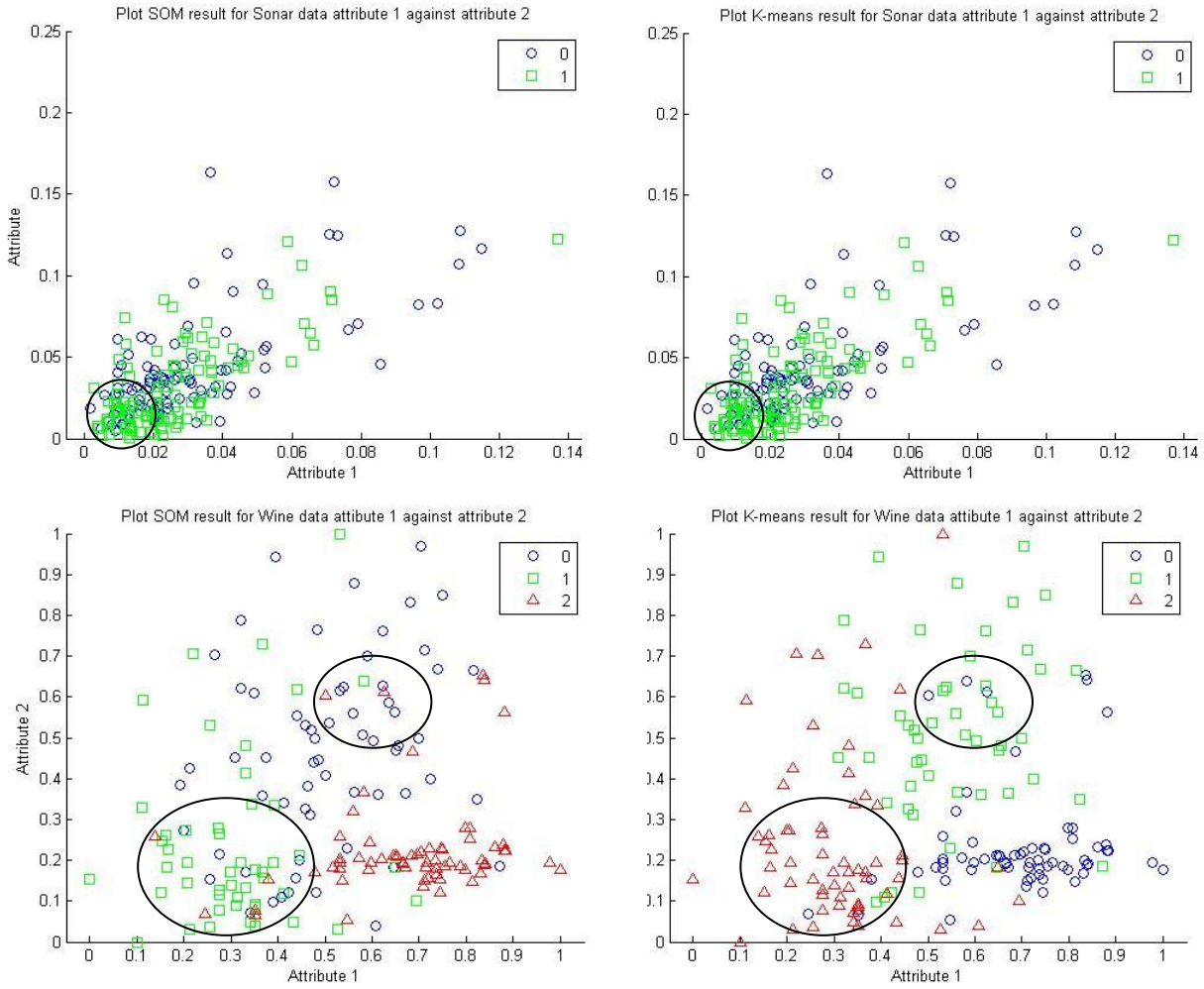


Figure 2. SOM (left) and K-means (right) output visualization for attribute 1 against attribute 2

Through these output visualization, it can be indicated that K-means is slightly better than SOM. In SOM there are more cluster mix to each other than in K-means. The circle shows the mixed clusters. In K-means, each cluster can be indentified more clearly. For iris and Sonar data, this condition is difficult to be visualized because of the data separation, but the differences between SOM and K-means results for Wine is easier to be illustrated.

3.3 Output visualization through PCA

Another test to compare SOM and K-means clustering results are conducted through PCA. These analyses also use visualization but the graphic is not comparing one attribute with another attribute separately, but through dimensional reduction. PCA is one of family techniques for taking high dimensional data, and using the dependencies between the variables to represent it in a more tractable, lower dimensional form, without losing too much information. It starts with *p*-dimensional feature vectors and wants to summarize them by projecting down into a *q*- dimensional subspace. The summary will be the projection of the original vectors on to *q* directions, the principal components (Shalizi, 2010). Through this technique, the result is easier to be visualized in low dimensional output. Figure 2 illustrate the biplot, which plot data and the corresponding clusters from SOM and K-means. The horizontal axis shows projections on to the first principal component, while the vertical axis shows the second component. It can be indicated from the figure that K-means is slightly better than SOM.There are more data mixed in SOM than in K-means in each data.

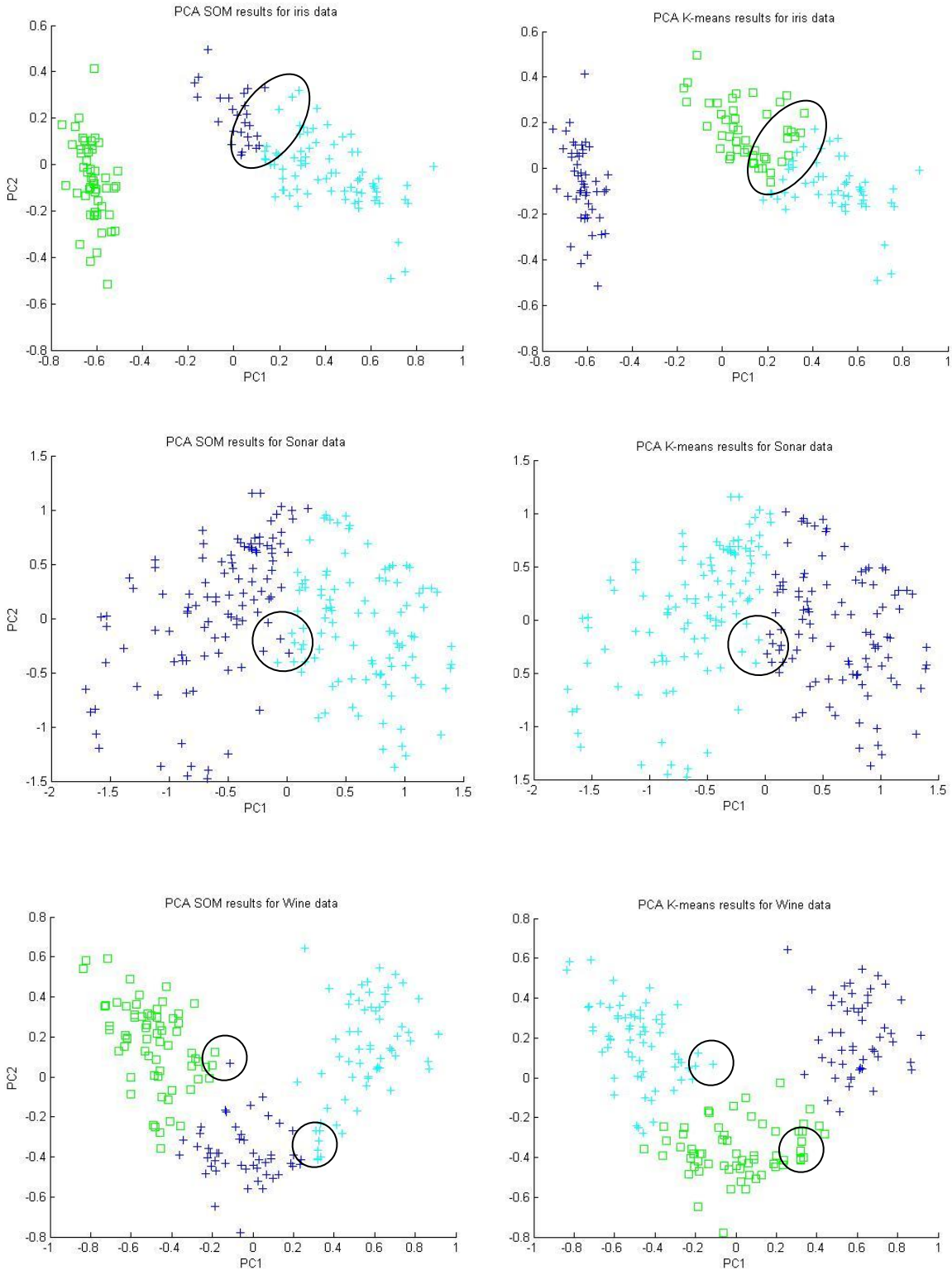


Figure 3. Biplot of Iris, Sonar and Wine data for PCA results

4. Conclusion

In this study, it can be conclude that percent misclassified analysis indicates K-means is outperform SOM, and this is confirmed by the results that shown by output visualization graph (separately and simultaneously with PCA). This result is the same with what have been

mentioned by Balakrishnan et al. (1994). However, as mention by other scholar, there is no exact solution which one is better between SOM and K-means. Both are popular and powerful clustering methods. Their results are depended on the data set, parameter setting and also the programming language. It is better to conduct further analysis using another data set and it will be much better if the analysis is preceded by a proper parameter setting, for example using Experimental Design, before conducting the training process. Moreover, it is better to conduct the training process several times to get the optimum solution.

References

Bação, F., Lobo, V., and Painho, M.(2005). Self - organizing Maps as Substitutes for K-means Clustering.*Proceeding of ICCS*,476-483. Berlin.

Balakrishnan, P.V., Cooper, M.C., Jacob, V.S., and Lewis, P.A.(1994). A study of the classification capabilities of neural networks using unsupervised learning: A comparison with K-means clustering.*Psychometrika*, 59 (4), 509-525.

Beale, M.H., Hagan, M.T., and Demuth, H.B., (2011).*Neural Network Toolbox User's Guide*.Massachusetts: The Math Works Inc.

Cai, Z., Gong, W., Ling, C.X., and Zhang, H.(2010). A clustering-based differential evolution for global optimization.*Applied Soft Computing*, 11, 1363-1379.

Das, S., Abraham, A., and Konar, A.(2008). Automatic Clustering Using an Improved Differential Evolution Algorithm.*IEEE Transaction on System, Man, and Cybernetics*, 38 (1).

Flexer, A.(1999). On the Use of Self-organizing Maps for Clustering and Visualization.*Principles of Data Mining and Knowledge Discovery*, 1704, 80-88.

Han, J., and Kamber, M. (2006).*Data Mining: Concepts and Techniques, 2nd edition.*Morgan Kaufmann.

Huang, Z.(1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorial Values.*Data Mining and Knowledge Discovery*,2,283-304.

Jain, A.K.(2010). Data clustering: 50 years beyond K-means.*19th International Conference in Pattern Recognition (ICPR)*, 31 (8), 651-666.

Jain, A.K., Murty, M.N., and Flynn, P.J.(1999). Data clustering: a review.*ACM Computer Survey*, 31 (3), 264-323.

Kwedlo, W.(2011). A clustering method combining differential evolution with the K-means algorithm.*Pattern Recognition Letters*, 32, 1613-1621.

Nugraha, Y.(2010).*Aplikasi Jaringan Syaraf Tiruan Bertipe Kohonen dalam Pengelompokkan Negara-negara anggota ASEAN untuk Mengetahui Pemetaan Pasar Ragional dan Posisi Strategis Indonesia di Kawasan ASEAN*,Bachelor Thesis,Universitas Gadjah Mada, Yogyakarta.

Shalizi, C.(2010).*Principle Component Analysis.*. Online access 18 November, 2013, fromwww.stat.cmu.edu/~cshalizi/490/pca/pca-handout.pdf,.

Tan, P.N., Steinbach, M., and Kumar, V. (2006).*Intoduction to Data Mining*. Massachusetts: Perason Education.

Vesanto, J., and Alhoniemi, E. (2000). Clustering of Self-Organizing Map.*IEEE Transaction on Neural Networks*, 11 (3), 586-600.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J.(1999). Self-organizing map in MATLAB: the SOM Toolbox. *Proceedings of the Matlab DSP Conference*,35-40. Espoo, Finland

Waller, N.G., Kaiser, H.A., Illian, J.B., and Manry, M.(1998). A Comparison of the Classification Capabilities of the 1-dimensional Kohonen Neural Network with Two Partitioning and Three Hierarchichal Cluster Analysis Algorithm.*Psychometrika*, 63 (1), 5-22.