International Journal of Advanced Science and Technology
Vol.62, (2014), pp.43-54
http://dx.doi.org/10.14257/ijast.2014.62.04

A New Initialization Method to Originate Initial Cluster Centers for K-Means Algorithm

Yugal Kumar and G. Sahoo

Department of Information Technology, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India

yugalkumar.14@gmail.com, gsahoo@bitmesra.ac.in

Abstract

K means algorithm is most popular partition based algorithm that is widely used in data clustering. A Lot of algorithms have been proposed for data clustering using K-Means algorithm due to its simplicity, efficiency and ease convergence. In spite this K-Means algorithm has some drawbacks like initial cluster centers, stuck in local optima etc. In this study, a new method is proposed to address the initial cluster centers problem in K-Means algorithm based on binary search technique. Binary search technique is a popular searching method that is used to find an item in given list of array. So in proposed method, the initial cluster centers have obtained using binary search property and after that K-Means algorithm is applied to gain optimal cluster centers in dataset. The performance of the proposed algorithm is tested on the two benchmark dataset which are downloaded from the UCI machine learning repository and compared with Random, Hartigan and Wang, Ward, Build, Astrhan and Minkowaski ward methods. The proposed method is also applied on the Minkowaski weighted K-Means algorithm to prove its significance and effectiveness.

Keywords: *Clustering, Cluster centers, K-Means, Binary Search, Ward, Hartigan*

1. Introduction

Data clustering is an important technique for data analysis which can be used to discover the similarity or dissimilarity between groups of items in a dataset such that items in one group are more similar than other groups and vice versa [11]. Mathematically the clustering problem can be defined by number of attributes and number of partitions in the dataset.

For a given dataset (X, P), X is number of attributes can be defined as:

$$X = \{x_1, x_2, x_3 \dots \dots x_m\} \text{ where } x_i \in R^N$$

and P is number of partitions on the dataset can be defined as:

$$P = \{P_1, P_2, P_3 \dots \dots P_k\} \text{ where } \forall_{i \neq j} P_i \cap P_j = \emptyset, \bigcup_{i=1}^{k} P_i = X, \qquad \forall_i P_i = \emptyset$$

So, the clustering problem can be viewed as searching problem that can searches a particular partition with minimum criterion function. Sum of squared error is the most

ISSN: 2005-4238 IJAST

common criterion function that can be used to search a particular partition in clustering task. Sum of squared error function can be defined as:

$$SSE(X,P) = \sum_{i=1}^{k} \sum_{x_j \in P_i} \left\| x_j - p_i \right\|^2$$

Large numbers of algorithms have been developed by various researchers for data clustering task. In the clustering domain K-Means is the oldest and probably the most popular algorithm proposed is proposed by [35]. It is easy to implement and it is fast and sensitive. However the K-Means algorithm has some drawbacks [21, 10, 12]. These are

- Lack of knowledge how to treat with inappropriate and clatter attributes.
- Lack of universal method how to choose the initial location of cluster centroids.
- No information about number of clusters in the dataset.
- Stuck in local optima.

To overcome the drawbacks of K-Means algorithm a lot of works have been done by various researchers. To enhancement and feature weighting in K-Means algorithm, a lot of work has been done by Modha and Spangler [39]; Dhillon and Modha [5]; Modha and Spangler [18]. To overcome the attribute selection problem in K-Means algorithm Huang *et al.,* [33] have proposed an automated weighted method for attribute selection and called it weighted K-Means algorithm. The sum of within cluster dispersion is used to calculate the weight of attributes and the attributes with lower weight is removed from the set of attributes. Another issue related to the K-Means algorithm is how many numbers of clusters exist in a dataset and initialization of initial cluster centers. In real life clustering problems it is quite difficult to choose the number of clusters present in final result [27, 7]. A large numbers of procedures have been developed to determine the number of clusters present in the dataset [6, 37, 20]. These procedures are divided in various categories such as variance based approach [26, 13, 23], structural approach [9, 17, 22], consensus distribution approach [19, 28, 14], hierarchical approach [25, 38, 34, 32] and resampling approach [28, 4, 40].

In this study, a new method is proposed to enhance the initialization problem of K-Means algorithm because the convergence result of K-Means algorithm is highly dependent on the initial cluster centers. If the initial cluster centers are not chosen properly then the local optimum problem will be exist in K-Means. The good convergence result is directly proportional to the good cluster centers. Hence the proposed method addresses the initialization as well as local optimum issues of K-Means. Rest of paper is structured as background and related work in Section 2; Section 3 describes the proposed algorithm; result and discussion in Section 4 and conclusion of the paper is summarized in Section 5.

2. Background and Related Work

De Amorim and Mirkin [3] have developed Minkowaski Weighted K-Means (MW K-Means) algorithm and intelligent Minkowaski Weighted K-Means (iMWK-Means) algorithm to address the attribute selection and initial location of cluster center problems in which minkowski metric is used as distance measure and initial cluster center is specified using anomalous clusters. De Amorim and Komisarczuk [27] have used six different centroid initialization methods with minkowski weighted K-Means (MWK-Means) algorithm to

evaluate which method gives better performance. These methods are compared on the behalf of the accuracy and processing time in which Ward method provides good results. To reduce the dependency of the K-Means on initial centroid Chan *et al.*, [2] (2006) have applied the greedy elimination method with K-Means to generate consistent and optimal clusters center in gene expression data and this method produces better results as compare to standard K-Means and fast greedy incremental method. To initialize the k clusters points in K-Means algorithm Bradley and Fayyad [29] have used minimizing concave function as bilinear program to evaluate the k number of clusters centers. A heuristic function based on the mode of the joint probability density function has applied by Bradley *et al.*, [30] with K-Means algorithm to generate good initial clusters points. Cao *et al.*, [1] have proposed a new method based on the cohesion and coupling degree between adjacent items using rough set model to determine the initial clusters center using K-Means. To overcome the dependency of K-Means algorithm on the initial clusters centroids (instead of random generation) Likas *et al.*, [15] have developed a global K-Means algorithm based on the deterministic global optimization and K-Means in which K-Means implemented as local search algorithms. To generate good initial cluster centers Lu *et al.*, [16] have applied the hierarchical clustering approach with K-Means algorithm and this method required less iteration time and higher convergence speed but the method has some drawback such that the values of attributes must be numeric if the values are non-numeric then these values must be converted into numeric values. According to Meila & Heckerman [36] and De Amorim & Komisarczuk [31], there are large number of methods exist to refine and initialization of the clusters centers in K-Means but at present there is not a single method that can be recognized as universal method to generate initial cluster centers. Table 1 provides the list of well-known different initialization methods that are used with K-Means algorithm as well as criteria to generate the initial cluster centers.

Table1. Different initialization method for K-Means Algorithm

Initialization Method	Cluster center
Random	A random function is used to generate the cluster center, *i.e.*, rand ().
Hartigan	Generate the cluster center using the following equation: 1+(k-1)*[N/K] Where, k=1, 2, 3 …….. K N = numbers of instance and the numbers of instance must be sorted according to center of gravity.
Ward	Generate the initial center based on the ward criteria: $dist(a,b) = \frac{n_a n_b}{n_a + n_b} \|x_a - x_b\|^2$ Where , n_a= number of points in cluster a, n_b = number of points in cluster b x_a =cluster points of cluster a,x_b = cluster points of cluster b

Build Algorithm	Generate the initial center based on the following equation: $M_y = \sum_{y \epsilon (D-C)} d(j, c_1) - d(y, j)$ Where , C=$\{c_1\}$ is the close to median, D= set of all instances M_y= Maximum value of data instance for c_1
Astrahan	Generate the initial cluster using the density of data. The density of data is calculated by given equation in the dataset. $d_i = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \|y_i - y_j\|$

3. Proposed Initialization Method

In this section an initialization method is proposed for K-Means algorithm. The proposed method is used to generate the initial cluster centers rather than random or user specified cluster centers. The proposed initialization method is based on the algorithm developed by Hatamlou [8] to obtain most favorable cluster points but this method is used here to generate the initial cluster centers for K-Means algorithm. In our proposed method, the initial cluster points are generated by using the unique property of binary search algorithm to find the value of middle item in a given list, *i.e.*,

$$\mathrm{A[mid]} = \frac{\mathrm{A[beg]} + \mathrm{A[end]}}{2} \quad (1)$$

The above property of binary search is modified to generate the initial cluster point for K-Means algorithm.

- A [beg] is replaced by A [max]
- A [end] is replaced by A [min]
- 2 is replaced by K , numbers of clusters
- A [mid] is replaced by any variable such as M
- Plus symbol is replaced by minus symbol

Now, the equation (1) is formulated in another equation as given below

$$\mathrm{M} = \frac{\mathrm{A(max)} - \mathrm{A(min)}}{\mathrm{K}} \quad (2)$$

The generalization of the equation (2) can be written as

$$M_i = \frac{\max(A_i) - \min(A_i)}{K} \quad (3)$$

The equation (3) is used to calculate the value of the variable M that specifies the range of initial cluster centers but not give the cluster centers. The cluster centers for K-Means algorithm are generated using given equation.

$$C_k = \min(A_i) + (k-1)M \quad (4)$$

Consider an example dataset D that is given in Table 2. The given dataset is applied with proposed method to get the initial cluster points. This dataset is consist total number of instances (N) = {14}, no. of attributes (i) = {2} and no. of Clusters (K) = {3}. The working of proposed method is given below:

Table 2. Example dataset D to generate the initial cluster center

Objects	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
A	1.1	1.3	1.2	3.2	2.8	2.9	2	1.9	2.2	8	7.4	7.2	9	8.8
B	4.3	3.9	3.8	4.8	3.9	3.7	3.6	3.3	3.2	3.2	5.9	3.8	4	6.9

- Calculate the maximum and the minimum values of each attribute in the dataset.

$$\text{maximum} = (9, 6.9) \text{ and } \text{minimum} = (1.1, 3.2)$$

- Calculate the value of M as

$$M = \left\{\frac{(9-1.1)}{3}, \frac{(6.9-3.2)}{3}\right\}$$

$$M = (2.63, 1.23)$$

- Generate the initial cluster centers for initialization as

$$C_1 = \big(1.1 + \big((1-1) * 2.63\big), 3.2 + ((1-1) * 1.23)\big)$$

$$= (1.1, 3.2)$$

$$C_2 = (1.1 + ((2-1) * 2.63), 3.2 + ((2-1) * 1.23))$$

$$= (3.73, 4.43)$$

$$C_3 = (1.1 + ((3-1) * 2.63), 3.2 + ((3-1) * 1.23))$$

$$= (6.36, 5.66)$$

The newly generated cluster centers (1.1, 3.2), (3.73, 4.43) and (6.36, 5.66) are used as initial cluster centers for K-Means algorithm. The steps of the proposed method are given below.

Algorithm: Input (Dataset (D) and k)

Step 1: Set the number of clusters (k) where k=1, 2, 3… m.

Step 2: Generate the range of the initial centroids using following:

$$M = (\max (D_j) - \min (D_j))/k \text{ where } j = 1, 2, 3 \ldots\ldots\ldots n.$$

Step 3: Obtain the initial cluster centers using the following equation:

$$C_k = \min (A_i) + (k-1)M$$

Step 4: Calculate the Euclidean distance as similarity measure of each attribute x j and assigned to cluster center Ck using following equation:

$$\text{Dist.} = \min (\| X_i - C_k \|^2)^{1/2}$$

Step 5: Recalculate the centers for each cluster centers using the equation in step 4 until the cluster centers are changed.

Step 6: Quit and return the final cluster centers.

4. Results and Discussion

The results have taken on a system with Intel core i3 processor and windows 8 operating system. Matlab 2010 (a) (Math works) environment is used to code the proposed initialization method with simple K-Means and Minkowaski weighted K-Means (WKM) as well as Random, Hartigan, Ward, Build algorithm and Astrahan initialization methods with fifty runs of each algorithm. To ensure the efficiency of the proposed initialization method Iris, wine and Pima Indian diabetic datasets [24] are used and the performance of proposed method is evaluated using accuracy, intra cluster distance and processing time parameters.

Accuracy: Accuracy is the ratio of correctly predicted instances divided by total number of instances. In clustering problem accuracy is defined as:

$$\text{Accuracy} = \sum_{i=1}^{k} \frac{a_i}{N}$$

Intra cluster distance: It is a quality parameter for clustering problem that can be defined as the sum of distances between instances within a cluster to the center points of cluster. Minimum sum of intra cluster distance indicates the good quality of cluster.
Processing time: Processing time is defined as the amount of time required to execute a given algorithm.

a. Dataset Information

- **Iris dataset**

Iris dataset contains the three class of iris flower: setosa, versicolour and virginica. This dataset contains 150 instances and three classes. In iris dataset, each class contains 50 instances with four attributes: sepal length, sepal width, petal length, and petal width.

- **Wine dataset**

This dataset contains the chemical analysis of wine in the same region of Italy but three different cultivators. The dataset contains 178 instances and three classes with 13 attributes. The attributes of dataset are alcohol, phenols, proanthocyanins, color intensity malic acid, ash, alcalinity of ash, magnesium, total phenols, hue, OD280/OD315 of diluted wines flavanoids, nonflavanoid and proline.

- **Pima Indians diabetes dataset**

This dataset contains 768 records of female diabetic patients which belong to indian origin. The dataset contains 768 data instance and two classes (tested positive for diabetes and tested negative for diabetes) with 8 attributes. Each class contains 500 and 268 data instances respectively. The attributes of datasets are Number of times pregnant, Plasma glucose concentration a 2 hours in an oral glucose tolerance test, Diastolic blood pressure, Triceps skin fold thickness, 2-Hour serum insulin, Body mass index, Diabetes pedigree function and Age.

Table 3 provides the list of initial cluster centers calculated by the proposed method for iris, wine and diabetic datasets which are used as initial clusters points for K-means algorithm instead of random cluster centers. The Figure 2 shows the distribution of data instances within three classes of iris dataset using proposed method with K-Means.

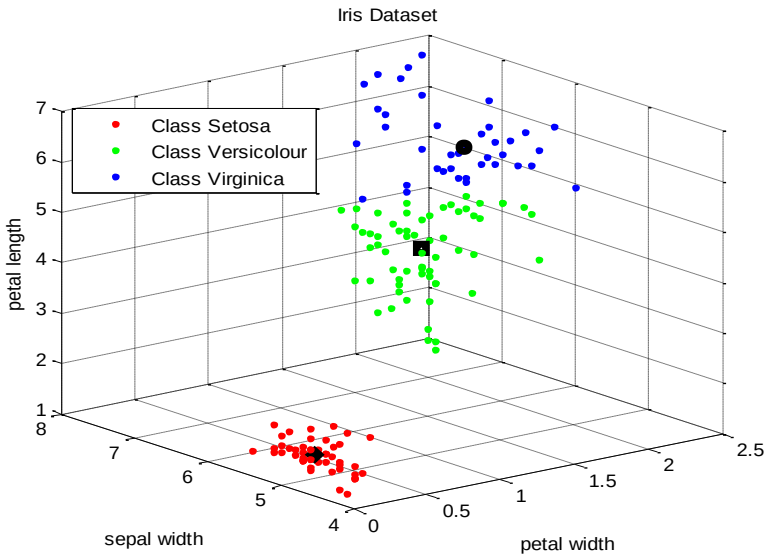

Figure 1. Representation of data instance into different classes of Iris dataset using proposed method

Table 3. List of initial cluster centers computed by proposed method for K-Means Algorithm

Dataset	Cluster 1	Cluster 2	Cluster 3
Iris	4.3	5.5	6.7
	2	2.8	3.6
	1	2.9667	4.9333
	0.1	0.9	1.7
Diabetic	Cluster 1	Cluster 2	
	0	8.5	
	0	99.5	
	0	61	
	0	49.5	
	0	423	
	0	33.55	
	0.078	1.249	
	21	51	

Dataset	Cluster 1	Cluster 2	Cluster 3
Wine	11	12.3	13.6
	0.7	2.4	4.1
	1.4	2	2.6
	10.6	17.1	23.5
	70	100.7	131.3
	1	1.9	2.9
	0.3	1.9	3.5
	0.1	0.3	0.5
	0.4	1.5	2.5
	1.3	5.2	9.1
	0.5	0.9	1.3
	1.3	2.2	3.1
	278	745.3	1212.7

Table 4 shows the comparison of proposed method and other five initialization methods with simple K-Means algorithm using accuracy and intra cluster distance parameters. From the table 4, it is noticed that the proposed approach is obtained high accuracy rate (82.93, 68.94, and 67.7) and low intra cluster distance (105.72, 18059.81 and 52071.4) for iris, wine and pima Indian diabetic datasets that shows the significance of proposed method to initialization of centroid for K-Means algorithm. On the analysis of other methods, it is found that the Build and Hartigan & Wang methods provide less accuracy with iris, diabetic (81.67 and 65.8) and wine (68.27) datasets respectively and these methods also have high intra cluster distance (18062.37, 52083.36 for wine and diabetic datasets, 107.46 for iris dataset) as compared to others methods. From the above discussion, it is concluded that proposed method has significant merit than other methods.

Table 4. Comparisons of different initialization methods and proposed initialization method for K-Means Algorithm

Methods	Iris		Wine		Pima Indian diabetic	
	Accuracy	Intra Cluster Distance	Accuracy	Intra Cluster Distance	Accuracy	Intra Cluster Distance
Random	82.6	106.5	68.88	18061	67.4	52073.89
Hartigan and Wang	81.92	107.46	68.27	18061.46	66.5	52077.52
Minkowaski Ward	82.43	106.26	68.43	18061.73	66.9	52076.91
Build	81.67	106.89	68.56	18062.37	65.8	52083.36
Artshan	82.46	106.43	68.63	18061.21	66.1	52078.53
Proposed method	**82.93**	**105.72**	**68.94**	**18059.81**	**67.7**	**52071.12**

Table 5. Comparisons of different initialization methods and proposed initialization method for IWKM Algorithm

Methods	Iris			Wine			Pima Indian diabetic		
	Accuracy	P	Processing Time (s)	Accuracy	P	Processing Time(s)	Accuracy	P	Processing Time(s)
Random	96.5± 0.63	2	0.40 ± 0.18	96.1 ±0.18	2	2.40 ± 1.0	70.1± 0.32	2	11.05 ± 1.8
Hartigan and Wang	95.3	2	*0.54*	94.6	2	2.34	69.5	2	10.31
Minkowaski Ward	96.4	2	0.43	95.3	2	1.19	69.3	2	13.51
Build	95.8	2	1.4	94.9	2	2.43	68.9	2	15.52
Artshan	96.1	2	1.7	95.5	2	3.29	69.2	2	16.48
Minkowaski WKM	96.7	2	0.51	95.4	2	2.16	69.8	2	13.28
Proposed method	**96.7**	2	**0.46**	**95.8**	2	**1.46**	**70.3**	2	**12.8**

To ensure the effectiveness of the proposed method, it is also tested on the Minkowaski weighted K-Means algorithm. Table 5 provides the comparison of proposed method and six other initialization methods with Minkowaski WKM algorithm using accuracy, value of P and time parameters. From the Table 5, it is observed that proposed method obtains high accuracy rate (95.8 and 70.3) with wine dataset using Minkowaski WKM algorithm while with iris dataset all initialization algorithms exhibits same behavior in terms of accuracy. On the analysis of time parameter, it is concluded that minkowaski ward method require less processing time (0.43 and 1.19) among all methods for iris and wine datasset while hartigan and wang method require less processing time (10.31) for diabetic dataset. While, the Arthsan method takes more time (1.7, 3.29 and 16.48) with all of datasets. From this table, it is also noticed that proposed method takes less time as compare to most of algorithms for execution. Hence from the above study, it is concluded that proposed approach provides good results among all other initialization methods with simple K-Means as well as Minkowaski WKM and from the above discussion, it is also stated that the proposed approach is applied with partition based algorithms for initialize the initial clusters points instead of random initialization.

6. Conclusion

This study focuses on the initialization problems in K-Means algorithm. The initialization problem of K-Means algorithm is formulated by two ways; first, how many clusters required for clustering task and second, how to initialize initial cluster centers for K-Means algorithm. This paper addresses the second issue of the initialization problem. To resolve it, a binary search based initialization method is proposed to initialize the initial cluster points for K-Means algorithm. In proposed algorithm, initial cluster centers are obtained with the help of binary search based method and after that K-Means algorithm is applied. Performance of the proposed algorithm is evaluated with some datasets that are downloaded from UCI repository

and compared with five well known cluster centers initialization methods for K-Means algorithm. The performance of the proposed algorithm is better than all other methods. The binary search based method is also used with Minkowaski WKM algorithm to generate the initial cluster points and the performance of this method is compared with six cluster centers initialization methods in which proposed method performs better.

References

[1] F. Cao, J. Liang and G. Jiang, "An initialization method for the K-Means algorithm using neighborhood model", Comput. Math. Appl., vol. 58, **(2009)**, pp. 474-483.
[2] Z. S. H. Chan, L. Collins and N. Kasabov, "An efficient greedy K-Means algorithm for global gene trajectory clustering", Expert Syst. Appl., vol. 30, **(2006)**, pp. 137-141.
[3] R. C. De Amorim and B. Mirkin, "Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering", Pattern Recognit., vol. 45, **(2012)**, pp. 1061-1075.
[4] S. Dudoit and J. Fridlyand, "A prediction-based resampling method for estimating the number of clusters in a dataset", Genome biology, vol. 3, no. 7, **(2002)**: research0036.
[5] I. S. Dhillon and D. S. Modha, "Concept decompositions for large sparse text data using clustering", Mach. Learn., vol. 42, **(2001)**, pp. 143-175.
[6] R. Dubes and A. K. Jain, "Validity studies in clustering methodologies", Pattern Recognit., vol. 11, **(1979)**, pp. 235-254.
[7] B. S. Everitt, "Unresolved problems in cluster analysis", Biometrics, vol. 35, **(1979)**, pp. 169-181.
[8] A. Hatamlou, "In search of optimal centroids on data clustering using a binary search algorithm", Pattern Recognit. Lett., vol. 33, **(2012)**, pp. 1756-1760.
[9] L. J. Hubert and J. R. Levin, "A general statistical framework for assessing categorical clustering in free recall", Psychol. Bull., vol. 83, **(1976)**, pp. 1072-1080.
[10] A. K. Jain, "Data clustering: 50 years beyond K-means", Pattern Recogn. Lett., vol. 31, **(2010)**, pp. 651-666.
[11] A. K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: A review", ACM Comput. Surveys, vol. 31, **(1999)**, pp. 264-323.
[12] Y. T. Kao, E. Zahara and I. W. Kao, "A hybridized approach to data clustering", Expert Syst. Appl., vol. 34, **(2008)**, pp. 1754-1762.
[13] W. J. Krzanowski and Y. T. Lai, "A criterion for determining the number of groups in a data set using sum-of-squares clustering", Biometrics, vol. 44, **(1988)**, pp. 23-34.
[14] L. I. Kuncheva and D. P. Vetrov, "Evaluation of stability of k-means cluster ensembles with respect to random initialization", IEEE Trans. Pattern Anal. Machine Intell., vol. 28, **(2006)**, pp. 1798-1808.
[15] A. Likas, N. Vlassis and J. J. Verbeek, "The global k-means clustering algorithm", Pattern Recognit., vol. 36, **(2003)**, pp. 451-461.
[16] J. F. Lu, J. B. Tang, Z. M. Tang and J. Y. Yang, "Hierarchical initialization approach for K-Means clustering", Pattern Recogn. Lett., vol. 29, **(2008)**, pp. 787-795.
[17] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set", Psychometrika, vol. 50, **(1985)**, pp. 159-179.
[18] D. S. Modha and W. S. Spangler, "Feature weighting in k-means clustering", Machine Learn., vol. 52, **(2003)**, pp. 217-237.
[19] S. Monti, P. Tamayo, J. Mesirov and T. Golub, "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data", Machine Learn., vol. 52, **(2003)**, pp. 91-118.
[20] C. Perruchet, "Les epreuves de classifiabilite en analyses des donnees [Statistical tests of classifiability]", Tech. Rep. NT/PAA/ATR/MTI/810). C.N.E.T. Issy-Les-Moulineaux, France, **(1983)**.
[21] S. Z. Selim and M. A. Ismail, "K-means type algorithms: A generalized convergence theorem and characterization of local optimality", IEEE Trans. Pattern Anal. Mach. Intellg., vol. 6, **(1984)**, pp. 81-87.
[22] J. Shen, S. I. Chang, E. S. Lee, Y. Deng and S. J. Brown, "Determination of cluster number in clustering microarray data", Applied Math. Comput., vol. 169, **(2005)**, pp. 1172-1185.
[23] C. A. Sugar and G. M. James, "Finding the number of clusters in a dataset: An information-theoretic approach", J. Am. Stat. Assoc., vol. 98, **(2003)**, pp. 750-763.
[24] www.archive.ics.uci.edu/ml/datasets.html.
[25] R. Duda and P. Hart, "Pattern Classification and Scene Analysis", Wiley Interscience Publication, New York, **(1973)**.
[26] J. A. Hartigan, "Algorithm CHAID. Clustering Algorithms", John Wiley and Sons, New York, **(1975)**.
[27] P. H. A. Sneath and R. R. Sokal, "Numerical Taxonomy", W. H. Freeman, San Francisco, **(1973)**.
[28] B. Mirkin, "Clustering for Data Mining: A Data Recovery Approach", Chapman and Hall, London, **(2005)**.

[29] P. S. Bradley and U. M. Fayyad, "Refining initial points for K-means clustering", proceedings of the 15th International Conference on Machine Learning, **(1998)** July 24-27, Morgan Kaufmann, San Francisco, pp. 91-99.
[30] P. S. Bradley O. L. Mangasarian and W. N. Street, "Clustering via Concave Minimization", In: Advances in Neural Information Processing Systems, Mozer, M.C., M.I. Jordan and T. Petsche (Eds.), MIT Press, Cambridge, MA, USA, **(1997)**, pp. 368-374.
[31] R. C. De Amorim and P. Komisarczuk, "On initializations for the minkowski weighted K-means", proceedings of the 11th International Conference on Advances in Intelligent Data Analysis XI, **(2012)** October 25-27, 2012, Helsinki, Finland, pp. 45-55.
[32] Y. Feng and G. Hamerly, "PG-Means: Learning the Number of Clusters in Data", Advances in Neural Information Processing Systems, Scholkopf, B., J.C. Platt and T. Hofmann (Eds.), MIT Press, Cambridge, MA, USA, ISBN-13: 9780262195683, **(2007)**, pp. 393-400.
[33] J. Z. Huang, J. Xu, M. Ng and Y. Ye, "Weighting Method for Feature Selection in K-Means", In: Computational Methods of Feature Selection, Liu, H. and H. Motoda (Eds.). Chapman and Hall, New York, **(2008)**, pp. 193-209.
[34] T. Ishioka, "An expansion of X-means for automatically determining the optimal number of clusters", Proceedings of the International Conference on Computational Intelligence, **(2005)** July 4-6, Calgary, AB., Canada, pp. 91-96.
[35] J. MacQueen, "Some methods for classification and analysis of multivariate observations", Proc. Berkeley Symp. Math. Statist. Prob., vol. 1, **(1967)**, pp. 281-297.
[36] M. Meila and D. Heckerman, "An experimental comparison of several clustering and initialization methods", Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, **(1998)** July 24-26, Morgan Kaufmann, San Francisco, CA, pp. 386-395.
[37] G. W. Milligan, "A discussion of procedures for determining the number of clusters in a data set", Proceedings of the Classification Society Meeting, **(1981)** May 31-June 2, Toronto, Canada.
[38] B. Minaei-Bidgoli, A. Topchy and W. F. Punch, "A comparison of resampling methods for clustering ensembles", Proceedings of the International Conference on Machine Learning: Models, Technologies and Application, **(2004)** June 21-24, Las Vegas, USA, pp. 939-945.
[39] D. S. Modha and W. S. Spangler, "Clustering hypertext with applications to web searching", Proceedings of the 11th ACM on Hypertext and Hypermedia, **(2000)** May 30-June 3, San Antonio, TX, pp. 143-152.
[40] G. B. Mufti, P. Bertrand and L. E. Moubarki, "Determining the number of groups from measures of cluster stability", Proceedings of International Symposium on Applied Stochastic Models and Data Analysis, **(2005)** May 17-20, Brest, France, pp. 405-413.

