International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-3 Issue-6, January 2014

Hybrid Fuzzy Data Clustering Algorithm using Different Distance Metrics: A Comparative Study

O. A. Mohamed Jafar, R. Sivakumar

Abstract*— *Clustering is the process of grouping a set of objects into a number of clusters. K-means and Fuzzy c-means (FCM) algorithm have been extensively used in cluster analysis. However, they are sensitive to noise and do not include any information about spatial context. A Penalized Fuzzy c-means algorithm (PFCM) was developed to overcome the drawbacks of FCM algorithm. Euclidean distance measure is commonly used by many researchers in traditional clustering algorithms. In this paper, a comparative study on hybrid fuzzy data clustering algorithm using different distance metrics such as Euclidean, City Block and Chessboard is proposed. The K-means, FCM and hybrid K-PFCM algorithms are experimented and tested on five real-world benchmark data sets from UCI machine learning repository. The experimental results show that FCM and hybrid K-PFCM algorithms report good performance for Chessboard distance. The hybrid K-PFCM algorithm shows best objective function value than K-means algorithm. The performance of the algorithms is also evaluated through standard cluster validity measures. The Hybrid K-PFCM algorithm is effective under the criteria of PC, PE and intra-cluster distance.

Index Terms*— *Data clustering, K-means, Fuzzy c-means, Penalized Fuzzy c-means, Hybrid K-PFCM, Distance metrics, Cluster validity measures.

I. INTRODUCTION

Most organizations generate enormous amount of data and store them in files, databases or other repositories. In recent days, the data continues to grow at a phenomenal rate. The huge amount of stored data contains valuable information, which could be used to increase the efficiency, to provide valuable services to customers and to gain the competitive advantage. Extracting information and knowledge from a large database is a difficult task. Hence, a process for converting huge volume of data to knowledge will become invaluable. The area of knowledge discovery in databases (KDD) has arisen over the last decade to address this challenge.

Data mining is one of the steps in KDD process. It is defined as finding meaningful information in a database. It is the process of extracting previously unknown, valid and potentially useful information from large databases. There are many data mining functions including association rules, classification, clustering, regression, summarization, sequence discovery and time series analysis [1].

Clustering is a process of grouping a set of data elements into a number of clusters such that data elements within the same cluster are similar to one another and are dissimilar to the elements in other clusters.

There are many methods for solving clustering problems in the literature. K-means and Fuzzy c-means algorithm are the important partitioning methods and are widely used for data clustering and image segmentation. However, they are sensitive to noise and are easily struck at local optima. They do not incorporate any information about spatial context. To overcome the drawbacks of fuzzy c-means algorithm, many derivatives were proposed in the literature. Penalized FCM algorithm (PFCM) is one such algorithm to produce effective and robust result. Distance metrics have played key role in data clustering problems. Distance metric is used to determine the similarity between data points. Euclidean distance is commonly used in many classical fuzzy clustering algorithms. In this paper, we study the performance of K-means, FCM and hybrid K-PFCM algorithms for data clustering problems by using different distance metrics including Euclidean, City Block and Chessboard. We evaluate the algorithms through benchmark data sets such as Wine, Teaching Assistant Evaluation, Mammographic Mass, Image Segmentation, and Glass.

The rest of this paper is organized as follows. Section 2 briefly describes the background which includes the clustering algorithms and the related works of K-means, FCM and PFCM algorithms. In section 3, Methodology is explained. Experimental Results and Discussions are given in section 4. Finally, conclusions are drawn in section 5.

II. BACKGROUND

In this section, we discuss the basic concepts of clustering algorithms and related works of K-means, FCM and PFCM.

A. Clustering Algorithms

Clustering is a technique for the discovery of knowledge from a data set. It is used to group the data items into clusters such that similar items are placed in the same cluster while dissimilar items are placed in different clusters. The groupings of data items are based on similarity, dissimilarity metrics or distance metrics. Clustering of data is generally based on two methods, namely, hierarchical methods and partitioning methods [2][3].

Hierarchical methods: These methods create a hierarchical decomposition of the objects. They can be categorized as bottom-up (agglomerative) and top-down (divisive). Bottom-up methods start with each object forming a separate group. They successively merge the objects that are close to one another, until all the groups are merged into one or until a termination condition holds.

Manuscript received January 15, 2014.

O.A. Mohamed Jafar, Research Scholar, P.G. & Research Department of Computer Science, A.V.V.M. Sri Pushpam College (Autonomous), Poondi, Thanjavur, Tamil Nadu, India.

R. Sivakumar, Associate Professor, P.G. & Research Department of Computer Science, A.V.V.M. Sri Pushpam College (Autonomous), Poondi, Thanjavur, Tamil Nadu, India.

Top-down methods begin with one object in a single cluster and then split the cluster into smaller groups until each object is in one cluster.

Partitioning methods: These methods partition the database into predefined number of clusters. Given a database of 'N' objects, they attempt to determine 'K' clusters or groups, which satisfy the following conditions: each object must belong to exactly one group and each group must contain at least one object.

One of the most important fuzzy clustering algorithms is fuzzy c-means algorithm. The problem of fuzzy clustering is to partition a collection of n data elements $X=\{x_1,x_2,...x_n\}$ into a collection of c cluster centers $Z=\{z_1,z_2,...z_c\}$ with respect to minimization criterion. In this algorithm, the membership values are given by $U=u_{ij}\in[0,1],\ \forall i=1,2,...n;\ \forall j=1,2...c$ where each membership value u_{ij} indicates the degree to which element x_i belongs to cluster z_j.

Clustering has been used in many application areas, including marketing, customer segmentation, image segmentation, pattern recognition, image processing, document retrieval, supplier selection and microarray gene expression [4][5].

B. Related Works

K-means algorithm is the center-based hard partitional clustering method [1] [6] [7]. Fuzzy c-means algorithm is developed by Dunn [8] in 1973 and improved by Bezdek [9] in 1981. K-means and FCM are widely used to solve data clustering problems and are also applied in many real-world applications. LiXiang Jun et al. [10] proposed fuzzy c-means clustering algorithm for macro-economic forecast. Singh Yadav et al. [11] proposed fuzzy c-means clustering technique for student academic performance evaluation. The drawbacks of K-means and FCM algorithms are sensitive to initial cluster centers and converge to local optimal solution. Miin-Shen Yang [12] made the fuzzy extension of the classification maximum likelihood (CML) procedure in conjunction with fuzzy c-partitions and called it a class of fuzzy CML procedures. They derived a generalized type of fuzzy c-means clustering algorithms, called the penalized FCM (PFCM) algorithms. Miin-Shen Yang and Chen-Feng Su [13] described proposed three approaches, namely Wolfe EM, FCM and PFCM to estimate parameters of normal mixtures. Yong Yang and Shuying Huang [14] presented a novel extended FCM algorithm for image segmentation.

III. METHODOLOGY

A. Hybrid K-PFCM Algorithms

Hybrid algorithms are based on the combination of two or more algorithms. Recently hybrid algorithms are mainly used for improving the performance of clustering results. In this paper, K-means and PFCM algorithms are integrated for solving data clustering problems. The algorithm is experimented with benchmark data sets for different distance metrics including Euclidean, City Block and Chessboard.

The K-PFCM Algorithm is described below:

Step 1: Let $X=\{x_1,x_2...x_n\}$ be the set of data points and $a=\{a_1,a_2...a_c\}$ be the set of cluster centers.

i. Select 'c' cluster centers randomly from the data set.

ii. Determine the distance between the data points and cluster centers.

iii. Assign each data point x_i to its nearest cluster center a_j

iv. Recalculate the cluster center a_j using $a_j=\dfrac{\sum_{x_i\in j} x_i}{n_j}$

$$ \tag{1} $$

where n_j is the number of data points belonging to the j-th cluster

v. Repeat the steps (ii) to (iv) until convergence is reached.

vi. Return the final cluster centers.

Step 2: Calculate the distance using centers from step 1 and obtain the membership values by using

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}}\ ; \tag{2}$$

Step 3: Set the iteration counter t = 0 and the initial values $u_{ij}^{(0)}, i=1,2...n,\ j=1,2...c$ from Step 2.

Step 4: Find $\alpha_j^{(t+1)}$ using

$$\alpha_j=\frac{\sum_{i=1}^{n}u_{ij}^m}{\sum_{j=1}^{c}\sum_{i=1}^{n}u_{ij}^m} \tag{3}$$

Step 5: Find the centroid $a_j^{(t+1)}$ by using

$$a_j=\frac{\sum_{i=1}^{n}u_{ij}^m x_i}{\sum_{i=1}^{n}u_{ij}^m} \tag{4}$$

Step 6: Calculate the distance d_{ij}^2

Step 7: Calculate the objective function value $J_{PFCM}(U^{(t+1)},a)$ by using

$$J_{PFCM}(U^{(t+1)},a)=\sum_{j=1}^{c}\sum_{i=1}^{n}u_{ij}^m d_{ij}^2-\omega\sum_{j=1}^{c}\sum_{i=1}^{n}u_{ij}^m\ln\alpha_j^{(t+1)} \tag{5}$$

where, $\omega\geq 0, \alpha_j\geq 0$, $\sum_{j=1}^{c}\alpha_j=1$, $\forall j$ and

$d_{ij}=\|x_i-a_j\|$,

Step 8: Find the new membership values $u_{ij}^{(t+1)}$ by using

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$u_{ij} = \frac{(d_{ij}^2 - \omega \ln \alpha_j)^{-1/(m-1)}}{\sum_{k=1}^{c} (d_{ik}^2 - \omega \ln \alpha_k)^{-1/(m-1)}} \quad (6)$$

Step 7: If $\max | u_{ij}^{(t+1)} - u_{ij}^{(t)} | < \varepsilon$ then stop otherwise go to step 4

B. Distance Metrics

Clustering methods use distance metrics to determine the similarity or dissimilarity between any pair of objects. The distance between data and centroid can be measured using distance metrics. The following are the important properties of distance metrics [15][16][17][18]:

1. $d(x, y) \geq 0$ for every x and y
2. $d(x, y) = 0$ only if x = y
3. $d(x, x) = 0$ for every x
4. $d(x, y) = d(y, x)$ for every x and y
5. $d(x, z) \leq d(x, y) + d(y, z)$ for every x, y and z

In this paper, we have studied the performance of K-means, FCM and K-PFCM algorithms using Euclidean, City Block and Chessboard distance metrics.

Euclidean Distance

This distance metric is commonly used for clustering applications. It is also called L_2 distance or Pythagorean metric. It is calculated by the following formula:

$$d(x, z) = \| x - z \| = \sqrt{\sum_{i=1}^{n} (x_i - z_i)^2} \quad (7)$$

City Block Distance

This distance metric is also called L_1 distance or Manhattan distance. The city block distance between data vector x and the centroid z is given by

$$d(x, z) = \sum_{i=1}^{n} | x_i - z_i | \quad (8)$$

Chessboard Distance

This distance metric is also called Chebyshev norm or L_∞ norm. It is named after Pafnuty Lvovich Chebyshev. The Chebyshev distance between data vector x and the centroid z is given by

$$d(x, z) = \max_{i=1,2 \ldots n} | x_i - z_i | \quad (9)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The objective of this paper is to study the performance of K-PFCM algorithm to data clustering problems using different distance metrics. The performance is measured by the objective function value and cluster validity measures such as PC, PE and intra-cluster distance. The algorithms are implemented using Java program. We used a PC Pentium IV by considering the maximum of 100 iterations and 10 independent test runs for K-means, FCM and K-PFCM with the parameter setting of stop criterion $\epsilon = 0.00001$, weighting exponent m = 2 and ω =0.5. The weighting exponent m and the parameter ω have played key role in FCM and K-PFCM algorithms respectively. The objective function value of PFCM depends on the parameter ω. When ω = 0, J_{PFCM} becomes J_{FCM}.

The following five real-world data sets from UCI machine learning repository have been considered:

Data set 1: The wine data set has 178 data points with 13 attributes such as alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, prothocyanins, color intensity, hue, OD280/OD315. There are three categories in the data set: class 1(59 instances), class 2 (71 instances) and class 3 (48 instances).

Data set 2: The teaching assistant evaluation data set consists of 151 objects and 3 different types of classes characterized by 5 attributes.

Data set 3: The mammographic mass data set, which consists of 961 instances and 2 different types characterized by 5 attributes.

Data set 4: The training image segmentation data set has 210 instances, which consist of seven classes namely brick face, sky, foliage, cement, window, path and grass characterized by 19 attributes.

Data set 5: The glass data set has 214 instances with 9 attributes and 6 classes.

Table I summarizes these five data sets with their characteristics. The comparison of K-means and K-PFCM with respect to the best objective function value is given in Table II. The best, average and worst objective function values (OFV) are shown in the Table III for Euclidean, City Block and Chessboard distance metrics. Both FCM and K-PFCM algorithms based on Chessboard distance shows better objective function values than Euclidean distance and City Block distance for all the five data sets. Figures 1-3 show the best OFV of K-PFCM of three different distance measures on various data sets. The quality of clustering is evaluated using cluster validity measures such as PC, PE and intra-cluster distance [19]. Table IV summarizes the cluster validity measures.

The *Maxdis* and *Mindis* of K-PFCM Algorithm for Chessboard distance are shown in Table V. The *Maxdis* represents the maximum distance between the cluster centers and *Mindis* represents the minimum distance between the cluster centers. Intra-cluster distance is average of the sum of all the distances between the objects within a cluster and the centroid of the cluster. The smaller intra-cluster value has the higher quality of clustering. It is observed that K-PFCM based on Chessboard distance has maximum PC, minimum PE and minimum intra-cluster distance. Table VI shows the results of cluster validity measures of FCM and K-PFCM algorithms.

V. CONCLUSIONS

Clustering is an important technique for grouping of data sets into group of clusters or classes. The fuzzy clustering problem is one of the difficult problems to solve optimally in nature. Many classical clustering algorithms were based on Euclidean distance metric. In this paper, we have made an attempt to study the performances of K-means, FCM and K-PFCM with different distance metrics such as Euclidean, City Block and Chessboard.

Five real-world data sets from

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

TABLE I REAL-WORLD UCI REPOSITORY DATA SETS AND THEIR CHARACTERISTICS

Data set	No. of attributes	No. of classes	No. of instances (with size of classes)
Wine	13	3	178 (59, 71, 48)
Teaching Assistant Evaluation	5	3	151 (49, 50, 52)
Mammographic Mass	5	2	961 (516, 445)
Image Segmentation	19	7	210 (30, 30, 30, 30, 30, 30, 30)
Glass	9	6	214 (70, 76, 17, 13, 9, 29)

TABLE II COMPARISON OF K-MEANS AND K-PFCM ALGORITHMS

		Best OFV	
Data set	**Distance Metrics**	**K-means**	**K-PFCM**
Wine	Euclidean	2370689.687	1796158.191
	City Block	2938172.782	2216038.760
	Chessboard	2341963.645	1775196.519
Teaching Assistant Evaluation	Euclidean	16784.152	10602.595
	City Block	40730.695	25616.939
	Chessboard	11447.982	7401.498
Mammographic Mass	Euclidean	74722.541	59571.483
	City Block	119517.686	93436.237
	Chessboard	71237.745	56590.693
Image Segmentation	Euclidean	963775.201	418697.591
	City Block	4591640.196	2210375.506
	Chessboard	427696.709	185402.238
Glass	Euclidean	338.922	217.590
	City Block	1490.553	750.994
	Chessboard	164.439	135.660

UCI machine learning repository, including Wine, Teaching Assistant Evaluation, Mammographic Mass, Image Segmentation, and Glass were chosen to validate the performance of the algorithms. The experimental results showed that FCM and K-PFCM algorithms reported good results for Chessboard distance. K-PFCM algorithm showed minimum objective function value than K-means algorithm for all the distance metrics. The results also revealed that the K-PFCM gave the better PC, PE and intra-cluster distance.

TABLE III OBJECTIVE FUNCTION VALUES (OFV) OF FCM AND K-PFCM ALGORITHMS

Data set	Method	Distance Metrics	Best OFV	Average OFV	Worst OFV
Wine	FCM	Euclidean	1796083.172	1958653.486	9900718.661
		City Block	2215967.401	2363699.836	10652568.984
		Chessboard	1775124.133	1906082.667	9483145.687
	K-PFCM	Euclidean	1796158.191	1778323.630	1802787.774
		City Block	2216038.760	2193848.082	2219640.872
		Chessboard	1775196.519	1757551.893	1781262.361
Teaching Assistant Evaluation	FCM	Euclidean	10558.080	10808.791	21782.797
		City Block	25575.604	26151.897	48938.856
		Chessboard	7358.160	7487.883	16679.034
	K-PFCM	Euclidean	10602.595	10498.794	10729.418
		City Block	25616.939	25275.459	25640.027
		Chessboard	7401.498	7313.556	7401.498
Mammographic Mass	FCM	Euclidean	59302.592	61120.728	137052.345
		City Block	93177.376	96022.174	195190.552
		Chessboard	56319.471	58129.925	133765.427
	K-PFCM	Euclidean	59571.483	58976.507	59622.366
		City Block	93436.237	92505.689	93636.909
		Chessboard	56590.693	56028.089	56809.278
Image Segmentation	FCM	Euclidean	677266.977	707550.886	2460303.656
		City Block	2743121.046	2932244.536	17359843.776
		Chessboard	430780.089	444819.228	1573868.257
	K-PFCM	Euclidean	418697.591	414792.324	430385.599
		City Block	2210375.506	2194187.651	2268071.037
		Chessboard	185402.238	183520.968	185860.362
Glass	FCM	Euclidean	154.146	162.185	656.036
		City Block	659.105	693.329	2801.969
		Chessboard	78.431	82.251	309.171
	K-PFCM	Euclidean	217.590	216.057	240.449
		City Block	750.994	750.054	799.831
		Chessboard	135.660	134.713	160.156

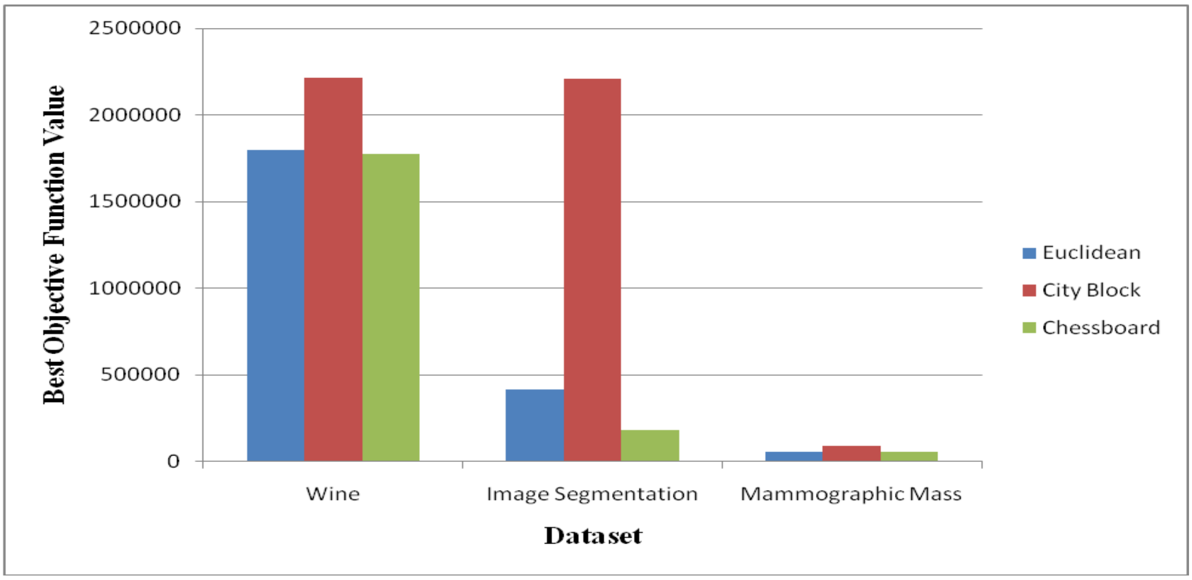

Fig. 1 Best Objective Function Value of K-PFCM algorithm for Wine, Image Segmentation and Mammographic Mass data sets

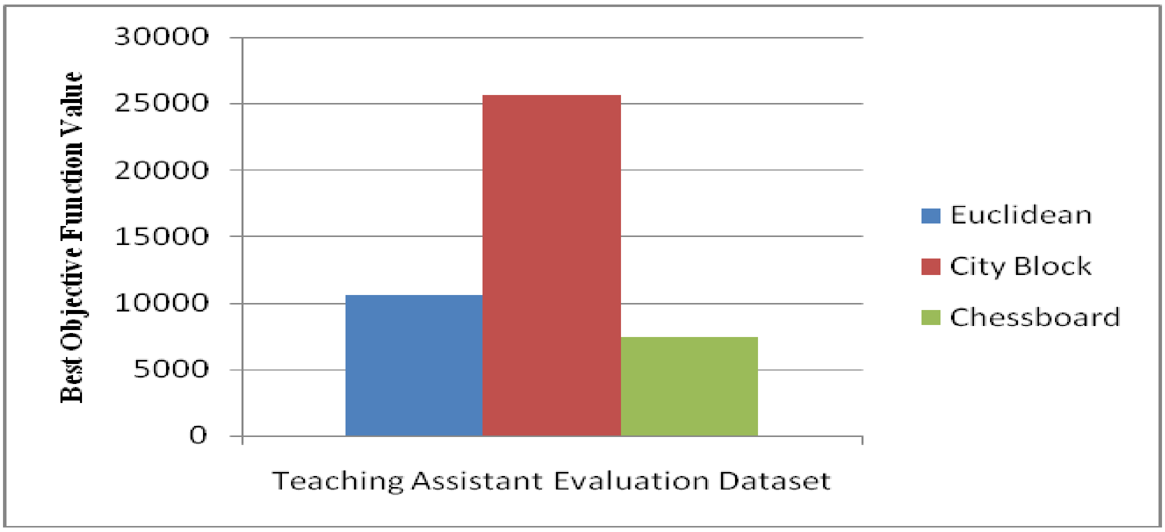

Fig. 2 Best Objective Function Value of K-PFCM algorithm for Teaching Assistant Evaluation data set

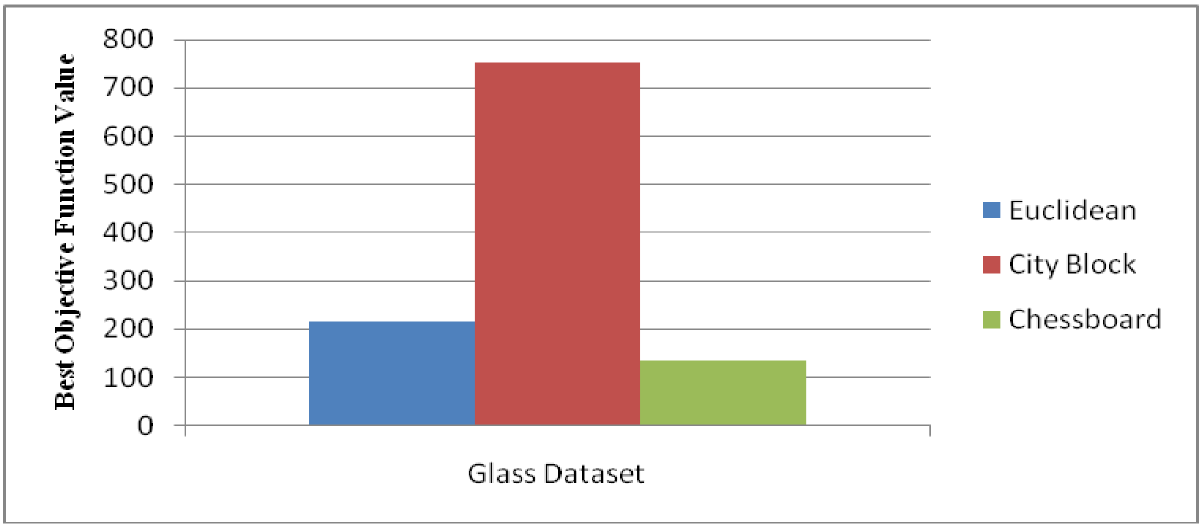

Fig. 3 Best Objective Function Value of K- PFCM algorithm for Glass data set

TABLE IV CLUSTER VALIDITY MEASURES FOR FUZZY CLUSTERING

Cluster Validity Measure	Description & Best Result
Partition Coefficient (PC)	$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{2}$ Maximum
Partition Entropy (PE)	$-\frac{1}{n}\left\{\sum_{i=1}^{n}\sum_{j=1}^{c}[u_{ij}\log_2 u_{ij}]\right\}$ Minimum
Intra-cluster distance	$\frac{1}{n}\sum_{k=1}^{K}\sum_{x\in C_k}\lVert x-m_k\rVert^2$ m_k - centroid of k-th cluster Minimum

TABLE V *MAXDIS* AND *MINDIS* OF K-PFCM ALGORITHM FOR CHESSBOARD DISTANCE

Data set	*Maxdis*	*Mindis*
Wine	761.633	283.125
Teaching Assistant Evaluation	27.156	11.253
Mammographic Mass	24.299	24.299
Image Segmentation	1507.082	81.047
Glass	6.365	2.940

TABLE VI RESULTS OF CLUSTER VALIDITY MEASURES OF FCM AND K-PFCM ALGORITHMS

Data set	Method	Distance Metrics	PC	PE	Intra-cluster Distance
Wine	FCM	Euclidean	0.7909	0.5488	56.9890
		City Block	0.7670	0.6100	76.7059
		Chessboard	0.7934	0.5418	56.8599
	K-PFCM	Euclidean	0.7909	0.5488	56.9883
		City Block	0.7670	0.6101	76.7048
		Chessboard	0.7934	0.5418	56.8592
Teaching Assistant Evaluation	FCM	Euclidean	0.5586	1.0983	0.4690
		City Block	0.4968	1.2346	0.9170
		Chessboard	0.5391	1.1277	0.6105
	K-PFCM	Euclidean	0.5572	1.1015	0.4692
		City Block	0.4962	1.2360	0.9131
		Chessboard	0.5373	1.1315	0.6116
Mammographic Mass	FCM	Euclidean	0.8195	0.4285	0.0177
		City Block	0.7872	0.4992	0.0257
		Chessboard	0.8270	0.4100	0.0174
	K-PFCM	Euclidean	0.8187	0.4304	0.0177
		City Block	0.7868	0.5003	0.0257
		Chessboard	0.8270	0.4113	0.0173
Image Segmentation	FCM	Euclidean	0.4239	1.8348	19.1562
		City Block	0.4009	1.8739	118.680
		Chessboard	0.4269	1.8060	3.813
	K-PFCM	Euclidean	0.4505	1.6812	19.1938
		City Block	0.4453	1.6737	149.059
		Chessboard	0.4804	1.5966	3.6891
Glass	FCM	Euclidean	0.4930	1.4373	0.003
		City Block	0.4558	1.5750	0.011
		Chessboard	0.5097	1.3737	0.002
	K-PFCM	Euclidean	0.3741	1.8	0.003
		City Block	0.4067	1.6	0.003
		Chessboard	0.4373	1.6	0.003

REFERENCES

1. J. Han and M. Kamber, "Data mining: concepts and techniques," Morgan Kaufmann, San Francisco, 2001.
2. A.K. Jain, M.N. Murty, P.J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, pp. 264-323, 1999.
3. C. Fraley and A.E. Raftery, "How many clusters? Which clustering method? Answers via Model-Based Cluster Analysis," Department of Statistics, University of Washington, Tech. Rep. No. 329, 1998.
4. R. Xu and D II Wunsch, "Survey of clustering algorithms," *IEEE Trans Neural Netw*, vol. 16, Issue 3, pp. 645-678, 2005.
5. P. Berkhin, "Survey clustering data mining techniques," Accrue Software, San Jose, California, Tech. Rep., 2002.
6. J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281-297, 1967.
7. L. Kaufman and P. Rousseeuw, "Finding groups in data: an introduction to cluster analysis," Wiley, New York, 1990.
8. J.C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybernet*, vol 3, pp. 32-57, 1973.
9. J.C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," Plenum Press, New York, 1981.
10. LiXiang Jun, Wu You Xin, Qiu Ji Guang and Wan Li Hui, "The applications of fuzzy c-means clustering in macro-economic forecast," Second International Symposium on Electronic Commerce and Security, vol. I, pp. 609-611, 2009.
11. Singh Yadav, Ramjeet, Pradap Singh and Vijendra, "Modeling academic performance evaluation using fuzzy c-means clustering techniques," *International Journal of Computer Applications*, vol. 60, pp. 15-23, 2012.
12. Miin-Shen Yang, "On a class of fuzzy classification maximum likelihood procedure," *Fuzzy Sets and Systems*, vol. 57, pp. 365-375, 1993.
13. Miin-Shen Yang and Chen-Feng Su, "On parameter estimation for normal mixtures based on clustering algorithms," *Fuzzy Sets and Systems*, vol. 68, pp. 13-28, 1994.
14. Yong Yang and Shuying Huang, "Image segmentation by fuzzy c-means clustering algorithm with a novel penalty term," Computing and Informatics, vol. 26, pp. 17-31, 2007.
15. K.V. Mardia, J.T. Kent and J.M. Bibby, "Multivariate analysis," Academic Press, 1979.
16. G.A.F. Seber, "Multivariate observations," Wiley, 1984.
17. M. Mimack, Gillian, J. Mason, Simon, S. Galpin, and Jacquelin, "Choice of distance matrices in cluster analysis: Defining regions," Journal of climate, vol. 4, Issue 12, pp. 2790-2797, 2001.
18. R.K. Bock and W. Krisher, "The data analysis brief book," Springer-Verlag, New York, 1998.
19. Maria Halkidi, Yannis Batistakis, Michalis Vazirgiannis, "On Clustering Validation Techniques," *Journal of Intelligent Information Systems*, vol. 17:2/3, pp. 107-145, 2001.

AUTHORS PROFILE

Mr. O.A. Mohamed Jafar obtained his M.Phil. in Computer Science from Bharathidasan University Tiruchirappalli in 1998. Currently he is working as Associate Professor at PG & Research Department of Computer Science, Jamal Mohamed College (Autonomous), Tiruchirappalli, Tamil Nadu, India. He is doing research in Computer Science in the Department of Computer Science, A.V.V.M. Sri Pushpam College (Autonomous), Poondi, Thanjavur, Tamil Nadu, India under Bharathidasan University, Tiruchirappalli. His areas of interests include Data Mining, Clustering, Swarm Intelligence and Evolutionary Algorithms.

Dr. R. Sivakumar obtained his Ph.D. in Computer Science from Bharathidasan University, Tiruchirappalli. Currently he is working as Associate Professor and Head of the Department at PG & Research Department of Computer Science, A.V.V.M. Sri Pushpam College (Autonomous), Poondi, Thanjavur, Tamil Nadu, India. His areas of interests include Human-Computer Interaction, Neural Networks, Artificial Intelligence, Swarm Intelligence and Evolutionary Algorithms.