EXPERIMENTS ON HYPOTHESIS "FUZZY K-MEANS IS BETTER THAN K-MEANS FOR CLUSTERING"

Srinivas Sivarathri[1] and A.Govardhan[2]

[1]Department of Computer Science,
Acharya Nagarjuna University, Guntur, Andhra Pradesh, India
[2]School of Information Technology,
Jawaharlal Nehru Technological University, Hyderabad, Telangana, India

ABSTRACT

Clustering is one of the data mining techniques that have been around to discover business intelligence by grouping objects into clusters using a similarity measure. Clustering is an unsupervised learning process that has many utilities in real time applications in the fields of marketing, biology, libraries, insurance, city-planning, earthquake studies and document clustering. Latent trends and relationships among data objects can be unearthed using clustering algorithms. Many clustering algorithms came into existence. However, the quality of clusters has to be given paramount importance. The quality objective is to achieve highest similarity between objects of same cluster and lowest similarity between objects of different clusters. In this context, we studied two widely used clustering algorithms such as the K-Means and Fuzzy K-Means. K-Means is an exclusive clustering algorithm while the Fuzzy K-Means is an overlapping clustering algorithm. In this paper we prove the hypothesis "Fuzzy K-Means is better than K-Means for Clustering" through both literature and empirical study. We built a prototype application to demonstrate the differences between the two clustering algorithms. The experiments are made on diabetes dataset obtained from the UCI repository. The empirical results reveal that the performance of Fuzzy K-Means is better than that of K-means in terms of quality or accuracy of clusters. Thus, our empirical study proved the hypothesis "Fuzzy K-Means is better than K-Means for Clustering".

INDEX TERMS

Data mining, K-Means, Fuzzy K-Means, unsupervised learning, similarity measure

1. INTRODUCTION

K-Means has been around for many years to discover patterns by grouping objects based on some similarity measure. It is faster and simple. However, it takes uniform clusters and needs to know the number of clusters beforehand. Another important feature of K-Means is that it keeps an object into a specific cluster. However, in the real world an object might be closer to more than one cluster. The K-Means clustering is also known as hard clustering. To overcome the limitations of K-Means, Fuzzy K-Means came into existence which is known as soft clustering approach. Though both are unsupervised learning algorithms the significant difference is that the

DOI : 10.5121/ijdkp.2014.4502

Fuzzy K-Means is flexible enough and can allow an object to belong to more than one cluster. In the literature it is found that Fuzzy K-Means has better utility in the real world applications than K-Means with respect to the quality of clusters. This is the reason behind this research work. We made an empirical study besides review of literature to prove that the Fuzzy K-Means exhibits better clustering performance than K-Means. The literature on these two and their comparison besides other derivatives of them[1], [2], [3], [4], [5],[6], [7], [8], [9], [10], [11], [12], [13],[14], [15], [16], [17], [18], [19],[20], [21], [22], [23] and [24] can be found in section IV.

Our contributions in this paper include the study of K-Means and Fuzzy K-Means algorithms through literature and empirical study to know whether the hypothesis "Fuzzy K-Means is better than K-Means for Clustering"holds true. The empirical results revealed that the clustering performance of the Fuzzy K-Means is better than that of K-Means in terms of accuracy and quality of clusters. The remainder of the paper is structured as follows. Section II describes K-means algorithm. Section III provides details about the Fuzzy K-Means. Section IV reviews related literature. Section V described the proposed methodology to work on the hypothesis "Fuzzy K-Means is better than K-Means for Clustering". Section VIpresents experimental results while section VII concludes the paper.

2. K-MEANS ALGORITHM

K-Means [25] is one of the top ten clustering algorithms which are widely used in real world applications. It is a very simple unsupervised learning algorithm that discovers actionable knowledge by grouping similar objects into various clusters. However, it needs the number of clusters to be known priori. That is nothing but the value of K. With K value known, it defines the number of centroids required. The centroids are to be taken carefully to ensure the cluster quality. After making the centroids, the algorithm takes data points from data source and associates them with the nearest centroid. This process is done until no data point is left ungrouped. After completion of this early grouping k new centroids are computed and then the objects are bound with the nearest centroid. The process of centroid changing its location takes place until there are no more changes needed. As a final step, the K-Means algorithm minimizes an objective function. The objective function is known as the sum squared error function as given below.

$$J=\Sigma^{K}_{j=1}\Sigma^{n}_{i=1}||x_i^{(j)} -- C_j||^2$$

Between a data point and cluster center $||x_i^{(j)} -- C_j||^2$is the chosen distance measure. Figure 1 illustrates the steps in the K-Means algorithm.

As can be seen in Figure 1, the algorithm has the following steps precisely.

1. Form initial centroids based on the number of clusters (K)
2. Assign each object taken from the data set to the nearest centroid to complete the initial grouping process.
3. Then re-compute the positions of K centroids
4. Repeat the steps 2 and 3 until there is no need for the centroids to be adjusted. Thus, the final clusters are formed.

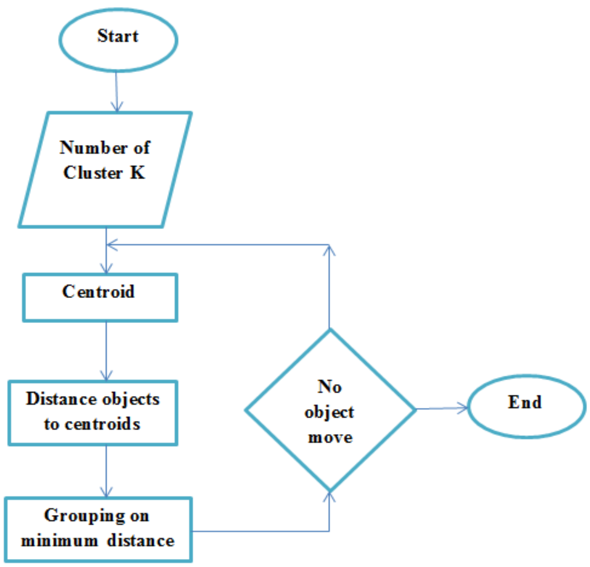


Figure 1 –Illustrates the flow of K-Means Algorithm

3. Fuzzy K-Means Algorithm

Fuzzy K-Means [26] is an improved form of K-Means algorithm which allows the degree of belonging. It does mean that an object can belong to more than one cluster in some degree. In case of K-Means it is not possible. Generally the points or objects which are on the edge of cluster might have less degree of belonging while the objects in the center might have higher belongingness. Coefficients are used to provide the degree of belongingness and they are defined as follows.

$$\forall x \sum_{k=1}^{num.clusters} \mathrm{u}(\mathrm{x}) = 1$$

In case of Fuzzy K-Means the mean of all points constitute the centroid. The objects are weighted by the degree in which they belong to a particular cluster.

$$\text{center}_k = \frac{\sum_x \mathrm{uk(x)mx}}{\sum_x \mathrm{uk(x)m}}$$

The inverse of distance to the cluster has inverse relationship with the degree of belonging. This is computed as follows.

$$\mathrm{u}_{k(x)} = \frac{1}{d(\text{Centerk,x})}$$

Afterwards, the coefficients are normalized besides fuzzification. The fuzzification uses real parameter m>1. Thus the sum is computed as 1. Therefore the following equation arrives.

$$\mathrm{u}_k(\mathrm{x}) = \frac{1}{\sum_j \left(\frac{d(\text{Centerk,x})}{d(\text{Centerj,x})}\right)^{2/(m-1)}}$$

When the coefficients are normalized, they make the sum as 1. When m value is 1 or closer to 1, it does mean that the point is closest to cluster center and more weight is given to that point. Figure 2 shows the flow of the algorithm.

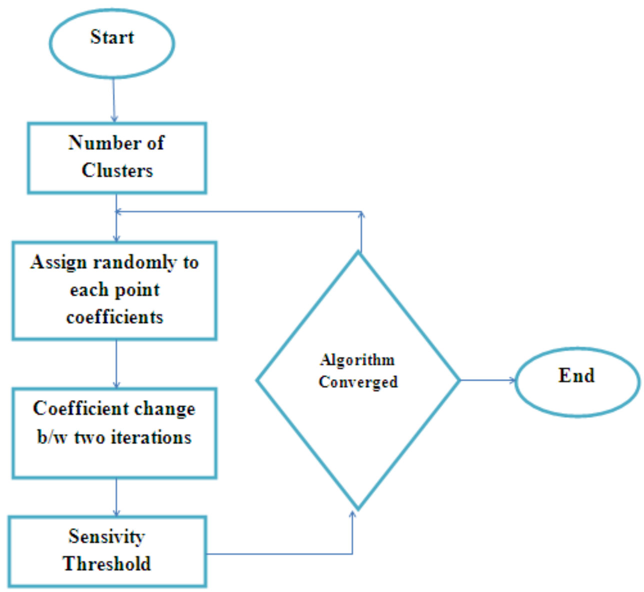


Figure 2 – Flow of Fuzzy K-Means Algorithm

As can be seen in Figure 2, it is evident that the algorithm has the following steps in execution.

1. Choosing number of clusters
2. Assigning coefficients of points randomly for being in the clusters
3. Every time the coefficients' change between two iterations is observed and the sensitivity threshold is considered.
4. This process continues until the convergence of the algorithm.

In [4] Fuzzy K-Means is also used to cluster biomedical sample characterization where the fuzzy logic model is used. The fuzzy logic model is as shown in Figure 3.

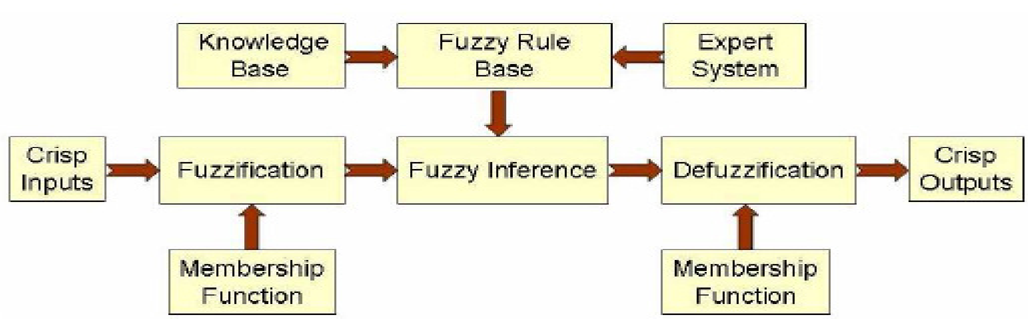


Figure 3 – Fuzzy logic model [4]

As shown in Figure 3, it is evident the fuzzy logic model makes use of the fuzzy filtering concept which takes crisp inputs and produce crisp results. The intermediate steps include fuzzification, fuzzy inference, and defuzzification. Knoeland expert system is used in the fuzzy rule base which in turn used for fuzzy inference. Membership function is used in both fuzzification and defuzzification. The whole process is a non-linear mapping between inputs and the outputs. The convergence with respect to experimental results of this approach is satisfactory [4]. Fuzzy K-Means clustering algorithm can be used for getting what we want from the Internet. The search engine results can be clustered with satisfactory performance. The clustering takes place based on sentence similarity [10].

4. RELATED WORKS

This section provides review of literature on K-Means, Fuzzy K-Means and their applications in the real world. Newton and Mitra [1] used fuzzy clustering along with a neural network thus making a hybrid architecture known as the Adaptive Fuzzy Leader Clustering (AFLC). In the control structure of the neural network the Fuzzy K-Means learning algorithm is embedded. The empirical results revealed that the hybrid architecture is capable of arbitrary data patterns. Watanabe [2] focused hyperbolic Fuzzy K-Means for fuzzy modeling. Kau [3] applied Fuzzy K-Means Clustering for lossless image coding. The non linear predictor proposed by this author performed well when compared with linear predictor. Ye et al. [4] applied Fuzzy K-Means on biomedical sample characterization which helped in medical diagnosis and decision making. To improve the quality of clustering and time performance Yu et al. [5] applied Fuzzy K-Means to form a new sampling technique based on clonal optimization which provided quality results with huge databases. Du et al. [6] used Fuzzy K-Means algorithm for fabric handle clusters that relay on mechanical and physical properties. The results proved that the algorithm is reliable. Xiong et al. [7] studied K-Means in data – distribution perspective. They described many validation measures such as CV, purity, entropy and F-measure. They came to know that K-Means produces clusters of uniform size. The true cluster sizes might be slightly different even.

Gour et al. [8] compared three clustering techniques, namely K-Means, Self Organizing Map (SOM) and ART neural network. They could find that neural network approach provided best results when compared to the other two techniques. Zhu et al. [9] opined that K-Means convergences faster. They also proved it by initializing it using affinity propagation to provide better results than normal K-Means. De-Sheng and Ming-Qin [11] applied Fuzzy K-Means algorithm for getting required information from the Internet. Yang and Lin [11] compared block Fuzzy K-Means with their own algorithm known as Fuzzy K-Modes. Their proposed method proved to give better performance. Yang and Li [12] introduced kernel K-Means which is fuzzy classifier. The results of their experiments proved that its performance is higher. Zhang et al. [13] proposed a new flavor of Fuzzy K-Means known as Agglomerative FKM with a feature that facilitates automatic selection of cluster number. This algorithm provided accurate results besides supporting the cluster number automatically.

Hoang et al. [14] proposed a new Fuzzy K-Means Clustering protocol for wireless sensor networks. This improved energy efficiency of the network. Sulaiman and Isa [15] presented a novel clustering algorithm known as Adaptive Fuzzy K-Means for image segmentation with better visual quality. Dehariya et al. [16] presented experimental results of both K-Means and Fuzzy K-Means for image segmentation and proved that Fuzzy K-Means is better than K-Means. Rajini and Bhavani [17] focused on improving K-Means and Fuzzy K-Means and applying them to segment MRI brain images. The results reveal that the new algorithms outperformed their original counterparts. Zhang et al. [18] proposed speech feature extraction based on Kernel Fuzzy K-Means Clustering algorithm. The algorithm showed high rate in speech recognition. Yan [19] presented many derivatives of Fuzzy K-Means. They are Rational Fuzzy C-Means, Any Rational Fuzzy C-Means, and None Euclidian Relational Fuzzy C-Means for sensitivity and efficiency in clustering. Vargas et al. [20] presented the comparison results of K-Means, Fuzzy K-Means and ckMeans. The ckMeans is the hybrid of the first two and provides better performance. Zhang et al. [21] built decision cluster classifiers using an agglomerative Fuzzy K-Means which proved to be effective in building classifiers. Esteves and Rong [22] compared K-Means and Fuzzy C-Means in the cloud environment for clustering Wikipedia's latest articles. They compared the results with Apache Mahout and proved that Mahout provides better performance. Honda et al. [23] improved K-Means algorithm and named it as Fuzzy C-Lines with a robust distance measure that provides comparable improvement in results. Svetlova [24] proposed an improved K-Means algorithm known as Minkowski Metric Fuzzy Weighted K-Means for clustering high-dimensional data.

Ghosh and Dubey [28] provided a comparative study of K-means algorithm and its improved form namely fuzzy k-means. The k-means is centroid based while the fuzzy c-means is a representative object based. Based on the efficiency of clustering, the algorithms are compared. In FCM an object can belong to more than one cluster resulting in soft clustering while an object belong to only one cluster in case of K-means algorithm. Another significant difference between them is that the K-means needs less computational power while the FCM needs more computational power. This is due to the processing of clusters based on the representative object in FCM. Change et al. [29] proposed cluster center displacement concept in which an FCM algorithm is achieved. Their proposed method is known as CDFKM. When this algorithm is compared with the FCM, the computing time of the CDFKM is reduced by a factor of 3.2 to 6.5 as it makes use of the concept of Gauss Markov sequence. The number of distance calculations is also significantly different so as to reduce the computational complexity further in case of CDFKM. The CDFKM can be extended further using the distortion measure. For instance, it can be improved using Humming distance measure. With such measure, it can provide better performance at less cost.

Gharehchopogh et al. [30] proposed algorithms that can perform clustering. However, their experiments are based on intrusion detection systems that can prevent intrusions or attacks launched by adversaries. Interestingly K-means and fuzzy k-means algorithms can be used to overcome the intrusions effectively. This is achieved by clustering the data available and making well informed decisions. Two groups of intrusion detection techniques are studied. They are known as supervised and unsupervised learning mechanisms with both k-means and fuzzy k-means algorithms. The uncertain quality parameter is explored by using the clustering algorithms. DoS attacks are explored in communication networks with respect to preventing intrusions by using clustering mechanisms effectively. The experiments revealed that the fuzzy k-means has better utility in detecting intrusions accurately.

Mingoti and Lima [31] employed SOM neural network and the improved form of K-means known as FCM for comparison of algorithms. The experiments proved that FCM has better comparable performance on all parameters. Outliers and overlapping have got their influence on the performance of all clustering algorithms. FCM is no exception for this. With respect to traditional hierarchical clustering approach K-means has shown similar kind of performance.

FCM has provided stable results and the SOM neural network have got performance issue and it makes use of more computational power.

Barathi et al. [32] applied FCM for images and them the results are compared with K-Means. Segmentation of color images is done by both the algorithms and the performance is compared. The empirical results revealed that the FCM has provided more accurate results. However, it consumed more computational power. This is due to fuzzy measures involvement in the process of clustering. Harima and Hanai [33] applied the FCM to gene expression analysis. When correctness ratios are compared, the FKM and fuzzy ART algorithms have shown the best performance. As far as accuracy is concerned they are good and of course they achieve this at the cost of computational complexity.

Burrough et al. [34] employed FKM to climate data for scientific studies. It was part of forest mapping research in USA. Here also it is proved that FKM has sensible approach in making clusters. It does mean that its accuracy is more when compared to the traditional K-means. For quantification of results also FCM is using representative approach and provide more accuracy. The clustering results reveal that morphological features are same with landscapes when compared with large areas. Ray and Turi [35] proposed an algorithm based on K-means for clustering. They applied clustering for color image segmentation. The experiments were done on both synthetic and real world images. A validity measure was employed to know the number of clusters known with respect to natural images. The validity measure makes use of the inter-cluster measure automatically and helps in obtaining the best results. Gasch et al. [2002] employed FKM for experiments on yeast gene expression. Especially they studied the conditional coregulation in yeast. However, they used the modified version of FKM that is based on heuristics. From the experiments it is understood that FKM is an analytical tool for deriving biological insights from gene expressions. It brings about specific and latent coregulation dynamics involved in yeast gene expressions.

Yang [37] has made an extensive survey of cluster analysis that involved both K-means and FKM algorithms. It does mean that it explored both soft and hard clustering. From the review, they concluded that the fuzzy clustering algorithms will have obvious performance gains over their counterparts that do not use fuzzy logic. However, they also consume more computational resources of the system.

5. METHODOLOGY

The aim of this paper is to know whether the hypothesis "Fuzzy K-Means is better than K-Means for Clustering". Towards this end diabetes dataset is obtained from the UCI repository [1]. A prototype application is built to demonstrate the efficiency of the algorithms in terms of fixed clusters, best separation clusters, single pass clusters, multi-pass clusters, random fixed clusters with single pass, random fixed clusters with multi-pass, exchange best separation clusters, exchange fixed clusters, with other factors such as number of clusters, fuzziness, maximum iterations, precision, CPU time, and compactness. The empirical results are presented the ensuing section.

6. EXPERIMENTAL RESULTS

We built a prototype application to make experiments on real data sets. The application is built using Java platform. JDBC API is used for interacting with the data set. MY SQL and text based files are used as backend. JfreeChart and JGraph API are used to generate graphs. Swing API is used to build graphical user interface. The environment includes a PC with Core 2 Dual processor;2 GB RAM running Windows XP operating system.

Dataset

The real dataset on diabetes was obtained from the UCI repository [27]. The dataset has four important attributes such as date, time, code and value. The dataset is basically containing insulin administered to diabetes patients. The date column holds the date on which insulin is administered. The time attribute holds the time at which the insulin is administered to respective patient. The code attribute is very important in this research which represents the code of various activities and insulin doses. The actual codes and description of them are presented in Table 1.

Table 1 –Codes representing various actions and insulin doses

Code	Description
33	Regular insulin dose
34	NPH insulin dose
35	UltraLente insulin dose
48	Unspecified blood glucose measurement
57	Unspecified blood glucose measurement
58	Pre-breakfast blood glucose measurement
59	Post-breakfast blood glucose measurement
60	Pre-lunch blood glucose measurement
61	Post-lunch blood glucose measurement
62	Pre-supper blood glucose measurement
63	Post-supper blood glucose measurement
64	Pre-snack blood glucose measurement
65	Hypoglycemic symptoms
66	Typical meal ingestion
67	More-than-usual meal ingestion
68	Less-than-usual meal ingestion
69	Typical exercise activity
70	More-than-usual exercise activity
71	Less-than-usual exercise activity
72	Unspecified special event

The value attribute holds the actual value of various measures pertaining to a patient. For instance, post breakfast blood – glucose measurement, post – supper blood glucose measurement and so on. These values play a pivotal role in understanding the vital signs of patients in the real world. However, in this paper our research is limited to the grouping of objects which are similar using both clustering algorithms such as K-Means and Fuzzy K-Means. The focus of this paper is to work on the hypothesis ""Fuzzy K-Means is better than K-Means for Clustering".

Results

The experimental results with both K-Means and Fuzzy K-Means algorithms are presented in Table 2. The experiments are made in terms of fixed clusters, best separation clusters, single pass clusters, multi-pass clusters, random fixed clusters with single pass, random fixed clusters with multi-pass, exchange best separation clusters, exchange fixed clusters, with other factors such as number of clusters, fuzziness, maximum iterations, precision, CPU time, and compactness.

Table 2 –Experimental Results

ALGORITHM	FEATURE	# of Clusters	Fuzziness	Max Iterations	Precision	CPU Time	Compactness
K-MEANS	Single Pass	3	1.7	200	0.01	51ms	9.541E
	Multi Pass	3	1.7	200	0.01	80ms	6.304E
	Best Separation	2	1.7	200	0.01	76ms	6.324E
	Fixed	3	1.7	200	0.01	69ms	6.314E
FUZZY K-MEANS	Fixed	3	1.7	200	0.01	81ms	5.641E
	Best Separation	2	1.7	200	0.01	396ms	6.641E

As can be seen in Table 2, the accuracy or quality of clusters with respect to Fuzzy K-Means is more when compared to that of K-Means. However, K-Means is faster than Fuzzy K-Means and consumes less CPU cycles. The results are presented as a series of graphs as shown below.

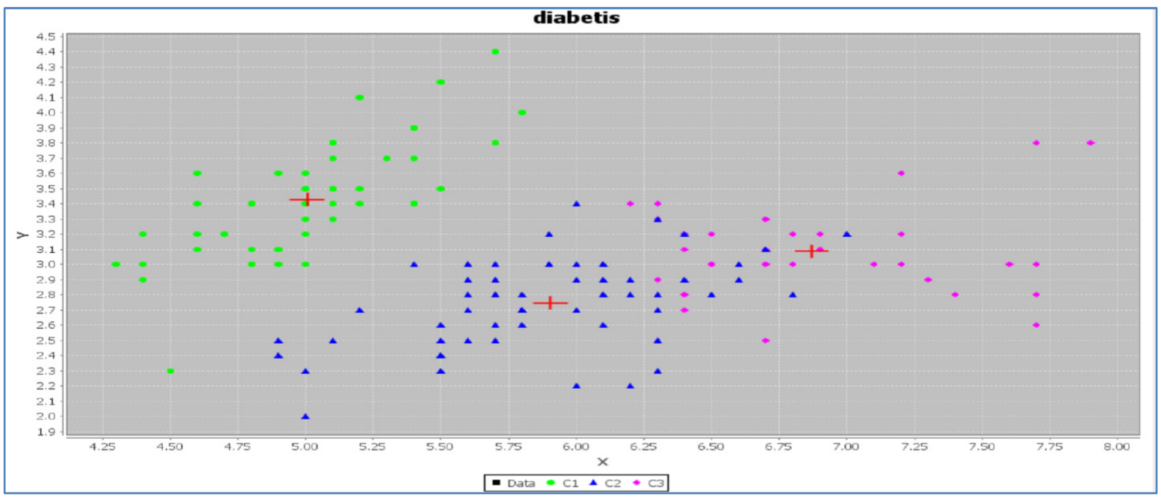


Figure 4 –Results of K-Means with Random Fixed Clusters and Single Pass

As can be seen in Figure 4, it is evident that the horizontal axis represents the classes of cluster labels while the vertical axis shows value. As visible in the graph there are three clusters formed. Each cluster is represented with different color. The cluster center is also presented. The results reflect the performance of K-Means with random fixed clusters in a single pass.

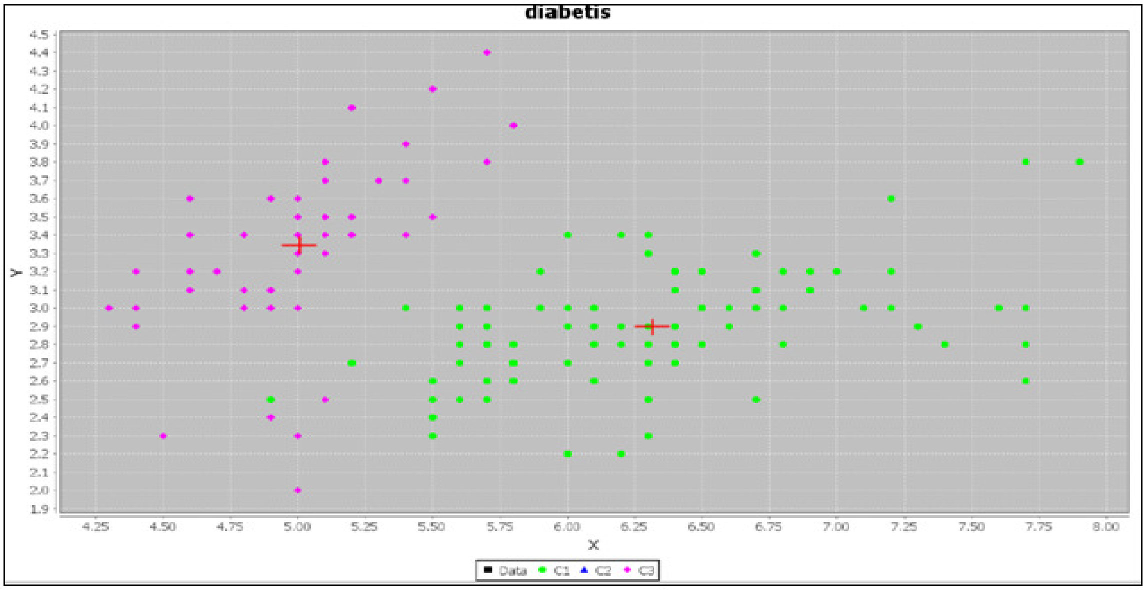


Figure 5 – Results of K-Means with Random Fixed Clusters and Multi Pass

As can be seen in Figure 5, it is evident that the graph shows the classes of cluster labels. As visible in the graph there are two clusters formed. Each cluster is represented with different color. The cluster center is also presented. The results reflect the performance of K-Means with random fixed clusters and multi-pass.

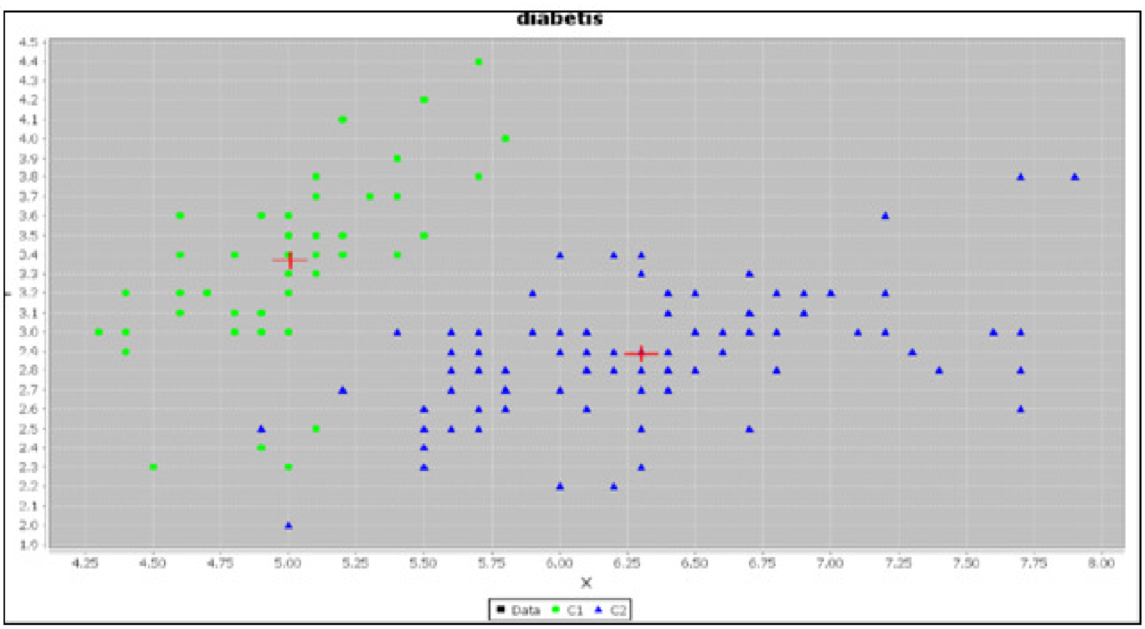


Figure 6 – Results of K-Means with Exchange Best Separation Clustering

As can be seen in Figure 6, it is evident that the graph shows the classes of cluster labels. As visible in the graph there are two clusters formed. Each cluster is represented with different color. The cluster center is also presented. The results reflect the performance of K-Means with exchange best separation clustering.

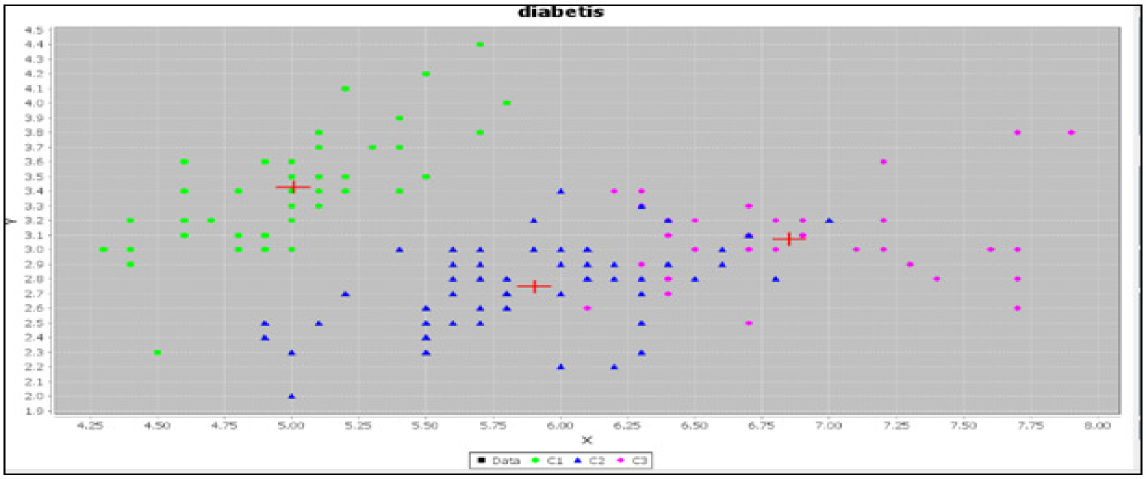


Figure 7 – Results of K-Means with Exchange Fixed Clustering

As can be seen in Figure 7, it is evident that the graph shows the classes of cluster labels. As visible in the graph there are three clusters formed. Each cluster is represented with different color. The cluster center is also presented. The results reflect the performance of K-Means with exchange and fixed clustering.

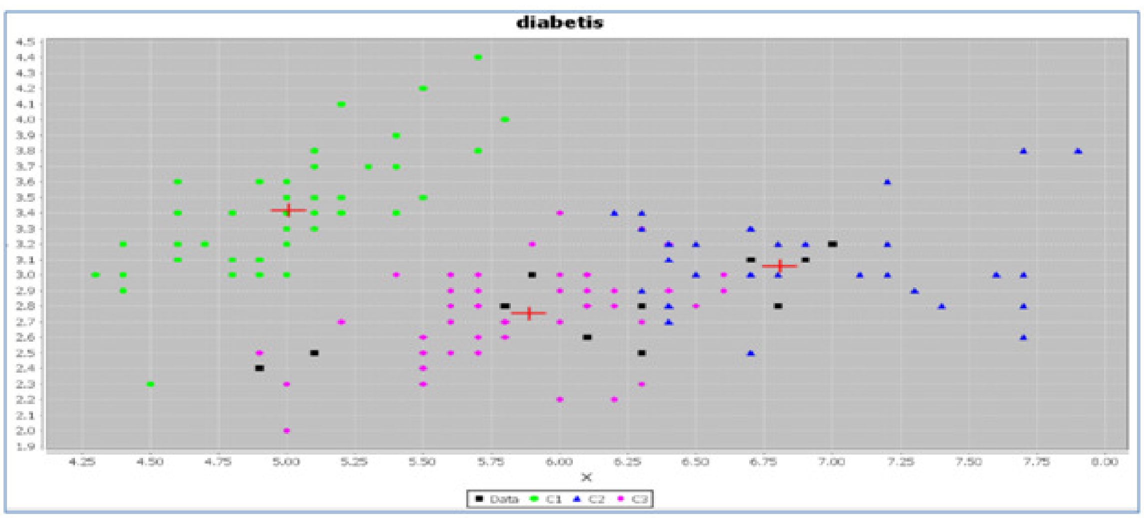


Figure 8 – Results of Fuzzy K-Means with Fixed Clusters

As can be seen in Figure 8, it is evident that the graph shows the classes of cluster labels. As visible in the graph there are two clusters formed. Each cluster is represented with different color. The cluster center is also presented. The results reflect the performance of K-Means with fixed clusters.

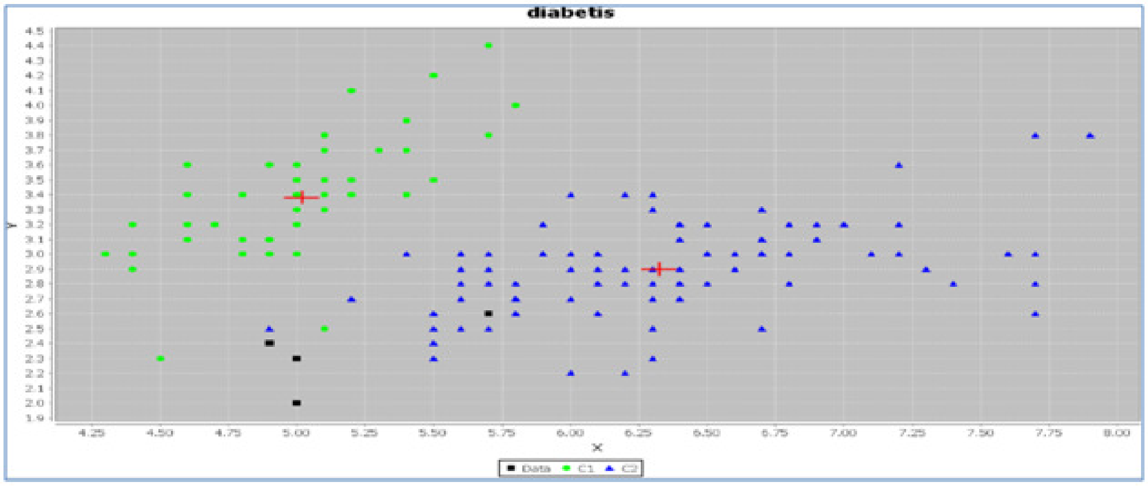


Figure 9 – Results of Fuzzy K-Means with Best Separation Clusters

As can be seen in Figure 9, it is evident that the graph shows the classes of cluster labels. As visible in the graph there are two clusters formed. Each cluster is represented with different color. The cluster center is also presented. The results reflect the performance of K-Means with exchange best separation clustering.

4. CONCLUSION

In this paper, we studied the two widely used clustering algorithms for data mining purposes such as K-Means and Fuzzy K-Means. From the literature a hypothesis is conceived such as "Fuzzy K-Means is better than K-Means for Clustering". We made experiments to know whether the hypothesis holds true. Experiments are made on the real data sets obtained from the UCI repository. The data sets used are related to diabetes disease. We built a prototype application to demonstrate the experiments with the two algorithms in terms of fixed clusters, best separation clusters, single pass clusters, multi-pass clusters, random fixed clusters with single pass, random fixed clusters with multi-pass, exchange best separation clusters, exchange fixed clusters, with other factors such as number of clusters, fuzziness, maximum iterations, precision, CPU time, and compactness. The empirical results revealed that the Fuzzy K – Means take more iterations and the CPU time when compared to that of K-Means. However, the accuracy and quality of the clusters made by Fuzzy K-Means is more comparatively. Thus our experiments proved the hypothesis "Fuzzy K-Means is better than K-Means for Clustering".

REFERENCES

[1] Scott C. Newton and S. Mitra. (1992). AN ADAPTIVE FUZZY SYSTEM FOR CONTROL AND CLUSTERING OF ARBITRARY DATA PATTERNS. IEEE. p383-370.

[2] Norio WATANABE. (2002). Fuzzy Modeling by Hyperbolic Fuzzy bMeans Clustering. IEEE. p1528-1531.

[3] Lih-Jen Kau. (2003). Adaptive Predictor with Dynamic Fuzzy K-Means Clustering for Lossless Image Coding. IEEE. p944-949.

[4] Zhengmao Ye, Yongmao Ye, Habib Mohamadian Pradeep Bhattacharya and Kai Kang. (2005). Fuzzy Filtering and Fuzzy K-Means Clustering on Biomedical Sample Characterization. IEEE. p90-95.

[5] Haiqing Yu, Ping Li and Yugang Fan. (2006). Sampling Fuzzy K-Means Clustering Algorithm Based on Clonal Optimization. IEEE. p6102-6105.
[6] Zhaoqun Du, Weidong Yu, Ye Yu, Lei Chen, Yan Ni, Jiaqi Luo and Xiaohan Yu. (2008). Fabric Handle Clusters Based On Fuzzy Clustering Algorithm. IEEE. p546-550.
[7] Hui Xiong, Junjie Wu and Jian Chen. (2009). K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective. IEEE. 39 (2), p318-331.
[8] Bhupesh Gour, T. K. Bandopadhyaya and Sudhir Sharma. (2008). ART Neural Network Based Clustering Method Produces Best Quality Clusters of Fingerprints in Comparison to Self Organizing Map and K-Means Clustering Algorithms. IEEE. p282-286.
[9] Yan Zhu, Jian Yu and Caiyan Jia. (2009). Initializing K-means Clustering Using Affinity Propagation. IEEE. p338-343.
[10] Zhu De-Sheng and Zhou Ming-Qin2. (n.d). Get What You Want from Internet Using Fuzzy k-means Clustering Algorithm. IEEE. p1-4.
[11] Miin-Shen Yang and Chih-Ying Lin. (2009). Block Fuzzy K-modes Clustering Algorithm. IEEE. p384-389.
[12] Aimin Yang, Qing Li and Xinguang Li. (2009). A Constructing Method of Fuzzy Classifier Using Kernel K-means Clustering Algorithm. IEEE. p73-76.
[13] Yanfeng Zhang, Xiaofei Xu and Yunming Ye. (2010). NSS-AKmeans: An Agglomerative Fuzzy K-Means Clustering Method with Automatic Selection of Cluster Number. IEEE. p32-38.
[14] D.C. Hoang, R. Kumar and S.K. Panda. (2010). Fuzzy C-Means Clustering Protocol for Wireless Sensor Networks. IEEE. p3477-3482.
[15] Siti Noraini Sulaiman and Nor Ashidi Mat Isa. (2010). Adaptive Fuzzy-K-means Clustering Algorithm for Image Segmentation. IEEE. p2661-2668.
[16] Vinod Kumar Dehariya, Shailendra Kumar Shrivastava and R. C. Jain. (2010). Clustering Of Image Data Set Using K-Means And Fuzzy K-Means Algorithms. IEEE. p386-391.
[17] N. Hema Rajini and R.Bhavani. (2011). Enhancing K-means and Kernelized Fuzzy C-means Clustering with Cluster Center Initialization in Segmenting MRI brain images. IEEE. p259-263.
[18] Junchang Zhang, Yuanyuan Chen and Jian Zhang. (2011). Speech Feature Extraction of KPCA Based on Kernel Fuzzy K-means Clustering.IEEE. p756-759.
[19] Chunjuan Yan. (2011). Derivatives of Fuzzy C-means Method and Their Application comparisons. IEEE. p326-329.
[20] Rog´erio R. de Vargas, Benjam´ın R. C. Bedregal and Eduardo Silva Palmeira. (2011). A comparison between K-Means, FCM and ckMeans algorithms. IEEE. p32-38.
[21] Yanfeng Zhang• Xiaofei Xu • Yingqun Liu • Xutao Li • Yunming Ye. (2011). An Agglomerative Fuzzy K-means Approach to Building Decision Cluster Classifiers. IEEE. p379-382.
[22] Rui Máximo Esteves and Chunming Rong. (2011). Using Mahout for clustering Wikipedia's latest articles. IEEE. p565-569.
[23] Katsuhiro Honda, Sakuya Nakao, Akira Notsu, Hidetomo Ichihashi. (2012). Alternative Fuzzy c-Lines and Comparison with Noise Clustering in Cluster Validation. IEEE. p1-6.
[24] L. Svetlova, B. Mirkin and H. Lei. (2013). MFWK-Means: Minkowski Metric Fuzzy Weighted K-Means for high dimensional data clustering.IEEE. p692-699.
[25] R.Xu, D.Wunsch A.Jain, M. Murty, P. Flynn, "Data clustering: Areview", ACM Computing Surveys, 31, 1999,pp- 264
[26] YE Ping, Fuzzy K-means algorithms based on membership functionimprovementChangchun Institute of Technology(Natural SciencesEdition),,2007,(01)
[27] Michael Kahn. Diabetes Data Set. Available: http://archive.ics.uci.edu/ml/datasets/Diabetes. Last accessed17/02/2014.
[28] Soumi Ghosh and Sanjay Kumar Dubey, Comparative Analysis of K-Means and Fuzzy C-Means Algorithms, IJACSA, vol. 4, No. 4, 2013, p1-5.
[29] CHIH-TANG CHANG, JIM Z. C. LAI AND MU-DER JENG, A Fuzzy K-means Clustering Algorithm Using Cluster Center Displacement, JISE, 2011, p1-15.

[30] Farhad Soleimanian Gharehchopogh, Neda Jabbari, Zeinab Ghaffari Azar, Evaluation of Fuzzy K-Means And K-Means Clustering Algorithms In Intrusion Detection Systems, IJSTR, Vol. 1, 2012, p1-7.

[31] Sueli A. Mingoti *, Joab O. Lima, Comparing SOM neural network with Fuzzy c-means,K-means and traditional hierarchical clustering algorithms, Science Direct, 2006, p1742-1759.

[32] Mrs. Bharati R.Jipkate and Dr.Mrs.V.V.Gohokar, A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms, IJCER, 2012, 737-739.

[33] Chinatsu Arima and Taizo Hanai, Gene Expression Analysis Using Fuzzy K-Means Clustering, Genome Informatics, 2003, p334-335.

[34] Peter A. Burrough1∗, John P. Wilson2, Pauline F.M. van Gaans1 & Andrew J. Hansen, Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA, Landscape Ecology, 2011, p-523-546.

[35] Siddheswar Ray and Rose H. Turi, Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation, p1-7.

[36] Audrey P Gasch* and Michael B Eisen, Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering, genomebiology.com, 2013, p1-22.

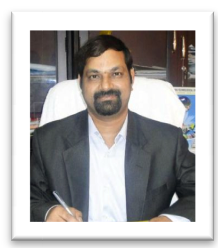
[37] M.-S. YANG, A Survey of Fuzzy Clustering, Mathl. Comput. Modelling Vol. 18, No. 11, pp. 1-16, 1993, p1-16.

AUTHORS

SRINIVAS SIVARATHRI

Srinivas Sivarathri started his professional career in 1998 at CBIT, one of the top listed engineering institutes in India. As an Assistant Professor, he was responsible for teaching engineering graduates and post-graduates on various subjects. He has 14+ years of experience in teaching globally, guiding IT projects in several technologies and managed different educational events. He has a graduation in computer science, post graduation in Computer Applications and post graduation in Computer Technology from Osmania University (a five star rated University in India). He is also a Prince2 certified professional in project management (APMG group, UK).

A.GOVARDHAN



Dr.A.Govardhan is presently a Professor of Computer Science & Engineering & Director at School of Information Technology and Executive Council Member, Jawaharlal Nehru Technological University Hyderabad (JNTUH), India. He served and held several Academic and Administrative positions including Director of Evaluation, Principal, Head of the Department, Chairman and Member of Boards of Studies and Students' Advisor. He did his 10th Standard from Z.P. High School, Choutuppal in 1986, Intermediate from Andhra Pradesh Residential Junior College (APRJC), Nagarjuna Sagar in 1988, B.E.(CSE) from Osmania University College of Engineering, Hyderabad in 1992, M.Tech from Jawaharlal Nehru University(JNU), New Delhi in 1994 and Ph.D from Jawaharlal Nehru Technological University, Hyderabad in 2003. He is the recipient of 23 International and National Awards including A.P. State Government Best Teacher Award, Pride of Asia International Award, Best Principal, Bharat Seva Ratna Puraskar, CSI Chapter Patron Award, Bharat Jyoti Award, International Intellectual Development Award and Mother Teresa Award for Outstanding Services, Achievements, Contributions, Meritorious Services, Outstanding Performance and Remarkable Role in the field of Education and Service to the Nation. He is a Chairman and Member on several Boards of Studies of various Universities. He is the Immediate Past Chairman of CSI Hyderabad Chapter. He is a Member on the Editorial Boards for Ten International Journals.