Performance Analysis of Fuzzy c-means, Mountain and Subtractive Clustering Techniques

Jyoti Arora[1#], Meena Tushir[2]

[1]*Research Scholar, GGSIPU.* [1]*Assistant professor, Dept. of Information Technology,*
[2]*Professor, Dept. of Electrical & Electronics Engineering,*
[1,2]*Maharaja Surajmal Institute of Technology, Affiliated to GGSIPU, New Delhi, India*
[#]Corresponding Author, Email: joy.arora@gmail.com

Abstract –The Conventional used Fuzzy c-means clustering technique needs to be initialized manually with the number of clusters present in the data. Mountain clustering and Subtractive clustering overcome this by calculating the number of clusters automatically by analyzing data numerically. The purpose of this paper is to compare these three algorithms namely Mountain Clustering, Fuzzy C-means (FCM) and Subtractive Clustering. The experimental results are carried out on the synthetic datasets with varying distribution. The performances of these algorithms are evaluated on the basis of the regression analysis, position of the center, number of clusters and root mean square error.

***Keywords* – Unsupervised clustering, subtractive clustering, mountain clustering, Fuzzy C-means.**

1. INTRODUCTION

Data mining is a helpful approach for recognizing patterns in the large volume of data. It is basically used to retrieve the relevant information from large data sets for several applications involving business and other real time applications. Clustering as one of the technique of data mining is consider as one of the most valuable tool for data analysis. Clustering can be defined as the process of grouping a set of data in the manner that data within the cluster are more similar as compared to data with the other clusters. Clustering is used as the basic step of processing in many fields for information retrieval like image processing, text mining, web mining and biomedical [1,2].

Cluster analysis is a task to be resolved and not an algorithm. There are a number of algorithms that can perform this task but they all differ in their concept of what are the factors considered while clustering and how to proficiently allocate them. Unsupervised clustering basically involves grouping on the basis of the measure of the similarity between center and data point. In unsupervised clustering, the clustering results correlate with the parameter of the algorithms, and with the initial initialization of some of the parameters [3]. The fuzzy c-means [8] is widely used clustering technique used for this purpose. The outcome of the algorithm largely depends upon the initial values given by the user. These values heavily influence the final solution. Due to the random initialization, sometime may stuck in local optima. Yager and Filev [5] proposed a simple and effective technique for assessing the number of clusters and initial positions of the centers. The method is based on the mechanism of gridding the data space and calculating the potential value for effective grid point based on similarity measurements from the nearest data points. This method is effective but involves high computational cost. Further Chiu [3] proposed subtractive

clustering, an extension of mountain clustering. This method solves the problems related with the mountain clustering. This method takes information from data points as the contenders for cluster centers, in the place of grid points as in the method of mountain clustering. This technique is computationally effective and depends on the size of the data rather than the dimension of the data. The main difficulty with this method is that, sometimes it fails to locate the real cluster centers as the actual cluster centers are not mandatory present at one of the data points. Thus, there is no accurately "correct" algorithm for clustering. The motivation behind this review and detail analysis is to choose the most appropriate algorithm experimentally unless there is a mathematical reason to prefer one over the other. This paper compares most widely used clustering approaches with different concepts. These techniques include Fuzzy C-means, Mountain clustering and Subtractive clustering.

The remaining paper is organized as follows. Section II defines an overview of the concept of the data clustering. Section III presents the mathematical formulation of the FCM, Mountain Clustering and Subtractive Clustering. Section IV discusses the experimental analysis of the three techniques on the basis of position of the center and slope of the regression. Further conclusion is presented in Section V.

2. Overview of Data Clustering

Data clustering involves the grouping of a data set into groups having higher likeliness within a same group than that among other groups. A uniformly distributed algorithms may fail or result in artificially introduced cluster, hence data set to be grouped must have integral grouping to some extent. Sometimes, the overlapping groups reduce the efficiency of the algorithm and this is related to the proportionality of the amount of overlapping [4].

The approach of clustering technique generally involves, random initialization of the cluster centers, followed by optimization process for further refinement. But this can be located mathematically through the mathematical formulation on the distribution of the data points [5,6]. Further through retrieved centers, similarity metrics is calculated between input vector and the centers, belongingness of the particular datapoint to the cluster is calculated. Based on the partitioning procedure clustering is further divided into hard and soft clustering techniques.

- *Hard clustering*- In this technique, data is divided into distinct clusters, where each data element belongs to exactly one cluster [7].
- *Soft clustering*-In soft clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels. This indicate the degree of membership of a particular point with the number of clusters [8].

Some clustering techniques depend on the prior knowledge of the number of clusters so that the data is partitioned in the given number of clusters. The numbers of cluster are given by the user manually. These algorithms are said to be unsupervised clustering. Here the classifications labels are calculated based on the iterative approach of the technique as in FCM [9]. But this is not a necessary condition, the algorithm may start with finding the number of clusters as in Mountain and Subtractive Clustering.

In the datasets where the prior information is not sufficiently available, the concept of mountain clustering can be used for initial determination of cluster centers. Yang and Wu [10] have defined the modified mountain clustering algorithm. The proposed technique automatically initializes all the parameters for the clustering process in accordance with the structure of the dataset. Verma et. al. [11] has given the concept of improved mountain clustering and applied on the gene expression data for analyzing biological information they contain. Further Gong et. al [12] has proposed stream density clustering by using the evolution of the density mountain function. This function evaluates the changes in data distribution through monitoring the changes in the density of the data. Further Subtractive Clustering proposed by Chiu [13] in 1994 is an extension of mountain clustering. The

subtractive clustering includes data points instead of grid points in calculating the centers of the clusters. The proposed method reduced the computational complexity of the mountain clustering. This approach can be used to calculate the number of clusters and position of centers. It shows improved results while working on high dimensional problems. In Subtractive Clustering, the parameter radius is used to be initialized by the user, so determining the radius of each cluster affects the performances of clustering results. To overcome this various hybrid and integrated approaches of subtractive clustering have been proposed. Shieh et. al. [14] proposed two-phase clustering algorithm based on subtractive and k-nearest neighbour. The proposed technique determines the radius of each cluster center using k-nn clustering concept.
Further, the concept of Fuzzy Clustering was integrated with subtractive clustering to improve the performance cluster analysis [15].

3. MATHEMATICAL FORMULATION

This section involves a detailed discussion and mathematical formulation of Fuzzy C-means, Mountain Clustering and Subtractive Clustering.

3.1 Fuzzy C-means Clustering (FCM)

In Fuzzy C-means clustering (FCM) which was proposed by Bezdek [8] in 1973. It is based on the concept of fuzzy partitioning. Here each data point belongs to a cluster to a degree of membership grade. Fuzzy partitioning is employed such that a given data point can belong to several groups with the certain degree of belongingness specified by membership grades between 0 and 1. However, FCM still tries to minimize the cost function while forming clusters.
U is the membership matrix that have the values of the elements in between 0 and 1. But, here membership degree has the constraint that the summation of degrees of membership of a data point to all the clusters is always equal to one:

$$\sum_{i=1}^{c} U_{ij} = 1, \forall_j = 1,....., n \tag{1}$$

Further, the Objective function of FCM is defined as

$$J(U,c) = \sum_{i=1}^{c} u_{ij}^{m} D_{ij}^{2} \tag{2}$$

where c is the number of the cluster which should be more then 1, D_{ij} is the Euclidean distance between the i_{th} cluster center and the j_{th} data point; and m is used for defining degree of fuzziness.
The objective function is minimized and the value of cluster centers and membership matrix is repeatedly calculated using

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{3}$$

$$\text{And } u_{ij} = \sum_{k=1}^{c} \frac{d_{ij}}{d_{kj}}^{-2/(m-1)} \tag{4}$$

The algorithm is executed repeatedly until no more significant improvement is noticed or some maximum number of iteration. The algorithm is carried out in the following steps:-
Step 1: Randomly initialize the membership matrix U with the constraint of eq 1.
Step 2: Specify the number of clusters in which data is to be partitioned and set the value of $m = 2$.
Step 3: Update the distance metrics and eq (3) and (4) till the number of iterations or minimum threshold value.
Step 4: Compute the objective function as in eq (2).

Here the value of membership and centers can have different values due to random initialization and may obstruct in local minima. The number of clusters is also set manually.

3.2 Mountain Clustering

Mountain Clustering proposed by Yager and Filev [5] in 1994 to calculate the number of clusters present in the dat. In a mountain clustering centers of the clusters are formed based on the measure of the density called the mountain function. This method

can be used as a pre-processing step for other unsupervised clustering techniques. The algorithm of the mountain clustering can be defined as:

Step 1: The grids are formed on the data space and the intersections of the grid lines are the potential points to calculate the number of clusters denoted by V as a matrix for centers.

Step 2: A mountain function is constructed, which represents the measure of the density. The height of the mountain function at the grid points $v \in V$ is calculated as:

$$m(v) = \sum_{i=1}^{n} \exp\left(\frac{-\|v - x_i\|}{2\sigma^2} \right) \quad (5)$$

where x_i is the i_{th} and σ is the application specific constant which determines the smoothness and the height of the resultant mountain function. The equation specifies the effect of distribution of data point on the measure of the data density at a point v and this measure of the density is inversely proportional to the distance between the data points x_i and the point under consideration v.

Step 3: Further the cluster centers are updated sequentially from the calculated mountain function. The measure of the greatest density measure at the selected point will allow to update the first center point Obtaining the next cluster center requires eliminating the c_1 effect of the first cluster. This is done by revising the mountain function: a new mountain function is formed by subtracting a scaled Gaussian function centred at c_1:

$$m_{new}(v) = m(v) - m(c_1)\exp\left(\frac{-\|v - c_1\|^2}{2\beta^2} \right) \quad (6)$$

The effect of the first cluster is eliminated by the operation of subtraction. The new mountain function $m_{new}(v)$ reduces to zero at $v = c_1$.

The second cluster center is selected as the point with the greatest value for the new mountain function. The process continues until the process of iteration.

3.3 Subtractive Clustering

The Subtractive Clustering technique proposed by Chiu [13] overcomes the problem of mountain clustering. In mountain clustering, the computation grows exponentially with the increase in the dimension of the data. The mountain function has to be evaluated at each grid point. In subtractive clustering the problem is resolved by using data points as the candidates for the centers of the cluster. This makes the algorithm more efficient as compared to mountain clustering. The computation is done on the basis of size of the problem instead of the problem dimension.

Here each data point participates equally for the candidate of the cluster center. A measure of the density at data point x_i is defined as:

$$D = \sum_{j=1}^{n} \exp\left(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2} \right) \quad (7)$$

where r_a represents positive constant for neighbourhood radius. This signifies with high density will have large number of neighbouring data points.

The density measure of every data point x_i is updated as:

$$D_i = D_i D_{c1} \exp\left(-\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2} \right) \quad (8)$$

Where x_{c1} is the point chosen with the largest density value D_{c1}. r_b represents the positive constant defining neighbourhood measure defining the reductions in the density measure. Further density function is updated, the cluster center is selected with the point having the greatest density value. This process is updated till the number of clusters are obtained.

4. Implementation And Results

In this section, discussions on the experimental evaluations of these techniques have been performed. Implementation of FCM, Subtractive clustering and Mountain clustering have been carried out on the same set of random initialize data. Experiments are performed in MATLAB R 2015.
The data is partitioned into 2 clusters. The Euclidean distance is used to measure the similarity between the input data vector and the center of the cluster.
Each clustering algorithm is passed with the same data, giving the membership of the data associated with the two clusters. The grade of the membership, and the minimization of the objective function is tested to evaluate the results of all the three algorithms. The evaluation of the results has been performed on the measure of the root mean square error (RMSE). It tells the measure of the spread of data around the line of the best fit.
Regression analysis is also performed to estimate the relationship among the variables used to measure the membership grades of the data points in the process of clustering. It helps to analyse the effect on the dependent variable with the change in the values of the independent variables. Further discussions have been performed on the evaluation of the results for each technique.

4.1 Performance on the basis of Position of Centers

FCM is an unsupervised soft clustering technique that allows data to be partitioned with the membership grade allotted with every number of clusters. It alleviates the concept of hard membership partition. FCM involves fuzzy membership measurement as the basis of the calculation of the membership grade and for the identification of the centers of the clusters.
Fig 1 shows the random distribution of the data in the two-dimensional space. In Fig 1(a), the point with red mark signifies the position of the center of two clusters calculated after the optimization of the clustering process using FCM. The effect random initialization of the centers and the membership values does not affect the final values of the result. Fig 1 (a) shows the accurate positioning of the cluster centers as compared to other two techniques.
Mountain clustering relies on dividing the data space into grid points and calculating a mountain function at every grid point. This mountain function is a representation of the density of data at this point. The performance of mountain clustering is strictly affected by the dimension of the problem; the computation needed rises exponentially with the dimension of input data because the mountain function has to be evaluated at each grid point in the data space. For a problem with c clusters, n dimensions, m data points, and a grid size of g per dimension, the required number of calculations is:

$$N = m \times g^n + (c-1)g^n \tag{9}$$

So, mountain clustering is not suitable for problems of dimensions higher than two or three. Fig 1 (b) shows the number of the centers obtained after applying mountain clustering to the data. Fig 1(b) shows the positions of centers are not as accurately plotted as FCM.
Further in Subtractive Clustering, to reduce the number of computations significantly, the density function is computed at every data point rather than grid point, to make the data linearly proportional to the number of input data and avoid being exponentially proportional to its dimension. In subtractive clustering, a problem with c clusters and m data points, the required number of calculations is:

$$N = m^2 + (c-1)m \tag{10}$$

The algorithm of subtractive clustering is stable and does not depend on any randomness; this allows the output to be fixed as compared to mountain clustering. Further, by changing the value of two variables r_a and r_b the performance of the clustering is analysed. These variables generally represent the neighbourhood of the centers to be taken, as the measure of the radius. The measure of these variables diminishes the effect of density function calculated by other data points. Typically, the r_b variable is taken to be as 1.5 r_a The value of these

variables will affect the performance of clustering technique. So, these variables are to be tuned properly. So, a value between 0.4 and 0.7 should be adequate for the radius of neighbourhood.

Fig.1(c) shows the result of subtractive clustering, the fig shows the position of the centers of the two clusters after applying clustering process. The numbers of centers are correctly predicted but the centers are too scattered according to the density of the data points. Fig.1 shows FCM allocates the centers with best accuracy as compared to the other two approaches with respect to the distribution of data points.

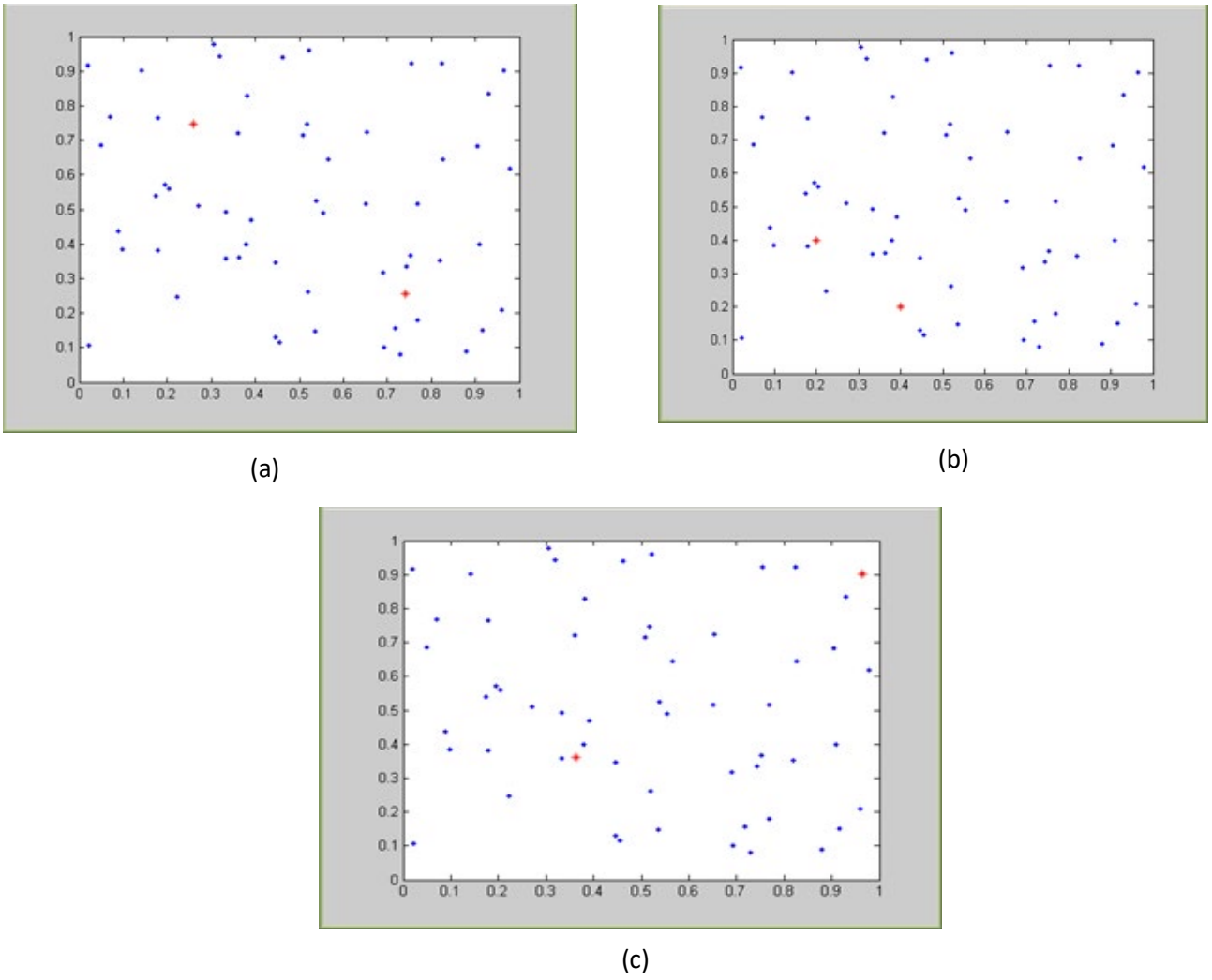

Fig. 1 Performance Analysis of the clustering with respect to the position of the center. (a) FCM Clustering (b) Mountain Clustering (c) Subtractive Clustering

4.2 Performance on the basis of Slope of Regression

Fig.2 shows the performance of FCM, Mountain Clustering and Subtractive Clustering with respect to the slope of regression. The Best linear fit is depicted by the red line and dotted line generalizes the actual spread. The dotted line defines the closeness of the spread with the actual spread. Fig.2

(a) shows the regression slope of the FCM where the interpolated spread is close to the actual spread between the two clusters. Fig 2(b) shows the regression slope of mountain clustering, where observed results are not as good as FCM and subtractive clustering. The result shows the poor performance of the clustering techniques with the widely spread data points. Fig 2 (c) shows the result of the subtractive clustering, where the spread of

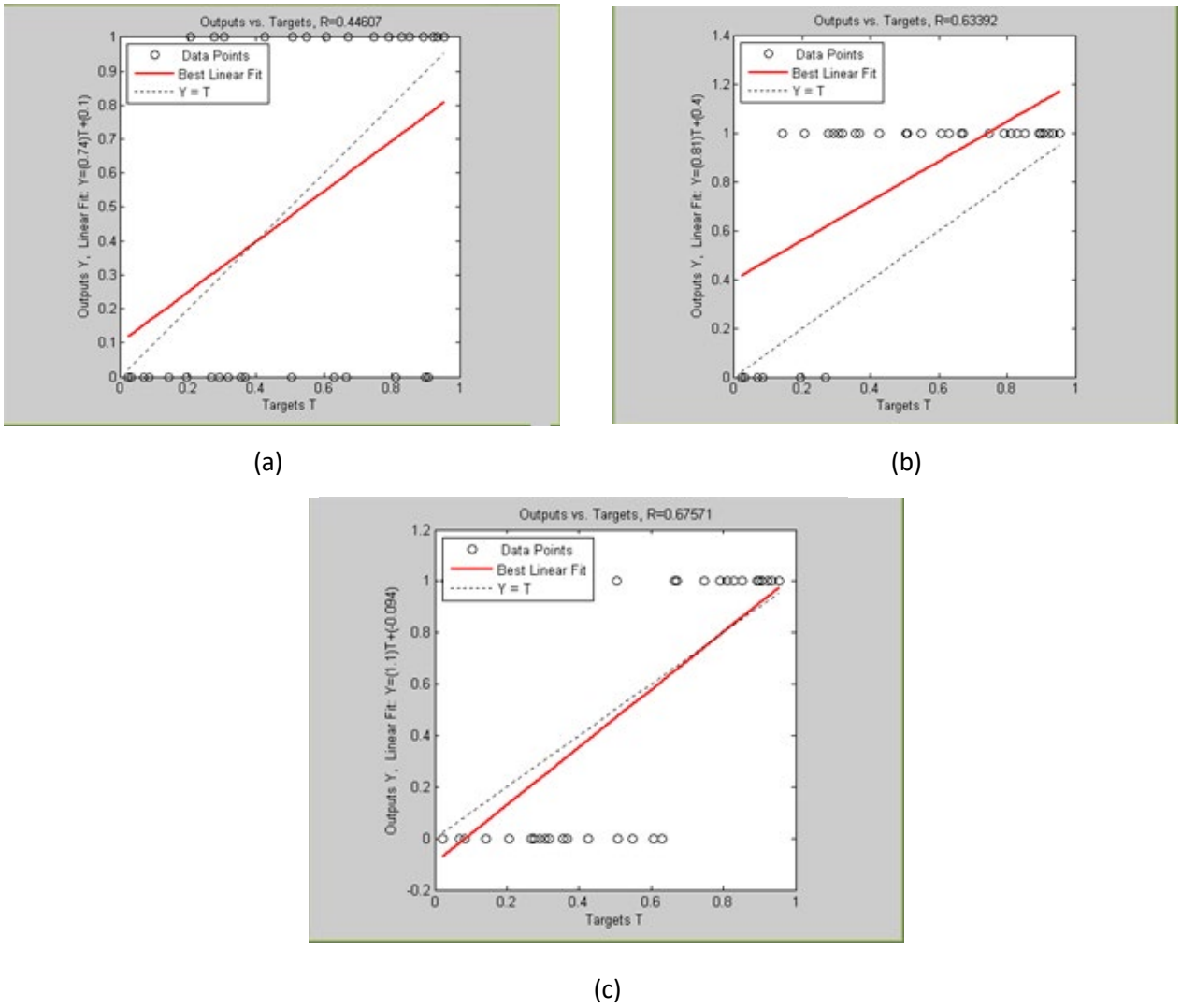

Fig.2 Performance Analysis of the clustering with respect to slope of regression. (a) FCM Clustering (b) Mountain Clustering (c) Subtractive Clustering

two clusters are not uniformly distributed around the actual line of slope. Here the variable defining the radius (ra and rb) are need to be properly tuned in order to balance the effect of neighbouring data points. Too small values of the radii will result in

ignoring the effect of neighbouring data points and increasing the value will alter the effect.
Further Table 1. represents performance evaluation of all the three clustering techniques on the basis of Root mean square error (RMSE) and regression line slope [9]. Table 1 shows the performance of FCM is better as compared to subtractive and mountain clustering.

FCM requires less computation as compare to other techniques. FCM shows better results with the datasets of higher dimension and is more efficient if the knowledge of the number of clusters is known in advance or can be predicted from the datasets.

Table 1. Performance analysis on the Basis of RMSE and Regression Line Slope

Algorithm	**Comparison aspect**	
	RMSE	**Regression line slope**
FCM	0.455139	0.446
Mountain	0.428146	0.634
Subtractive	0.371574	0.676

Mountain clustering requires large number of computations with poor performance and degrades with increase in the dimensionality of the data set. Mountain clustering is suitable with the datasets of two or three dimensionality.

5. CONCLUSION

The paper reviews the three widely used clustering techniques, namely Fuzzy C-means, mountain clustering and subtractive clustering. These techniques are unsupervised clustering techniques where unlabelled data is clustered with the defined number of clusters. The mountain clustering is used to calculate the number of clusters based on the measure of the density of the data points. The clusters are so formed that the similarity in each cluster is larger than inter clusters. The three techniques have been implemented and tested against a random distribution of the dataset. The results conclude, the performance of the mountain clustering decreases with the increase in the dimensionality due to its exponential proportionality to the dimension of the problem. However, these are used where the number of clusters are not known, but if the knowledge about the number of clusters is known then FCM is widely used with efficient results and reduced complexity. Subtractive clustering seems to be a better substitute to mountain clustering. But here the radius variable is need to be tuned properly for the effective results. Finally, clustering techniques can be used in the number of application areas where one technique can be nested with other and can be used in conjunction with other machine learning techniques to increase the overall system performance.

REFERENCES

[1] Jain, A. K., Data Clustering: 50 years beyond k-means. Pattern Recognition Letters, 31(8), 651–666 ,2010.

[2] Xu, R. and Wunsch, D. C., Clustering algorithms in biomedical research: a review. IEEE Reviews in Biomedical Engineering, 3, 120–154 ,2010.

[3] Hathaway, R., Bezdek, J., and Hu, Y., Generalized fuzzy c-Means clustering strategies using Lp norm distances. IEEE Transaction on Fuzzy Systems, 8(5), 576–582 2000.

[4] Xu, R. and Wunsch, D., Survey of clustering algorithms. IEEE Transaction on Neural Networks, 16(3), 645-678 ,2005.

[5] Yager, R. and Filev, D., Generation of Fuzzy rules by mountain clustering. Journal of Intelligent and Fuzzy Systems, 2(3), 209-214 1994.

[6] Wang, J., Liu, J. and Liu, L., A mountain means clustering algorithm. 7th World Congress on Intelligent Control and Automation, Chongqing, 5045-5049 ,2008.

[7] Li, Y. and Wu, H., A Clustering method based on K-means Clustering. Physics Procedia, 25, 1104-1109 ,2012.

[8] Bezdek, J. C., Pattern recognition with fuzzy objective function. New York, 1981.

[9] Yadav, R. S. and Singh, V. P., Modelling Academic performance evaluation using Fuzzy c means clustering techniques. International Journal of Computer Application, 60(8), 2012.

[10] Yang, M. S. and Wu, K. L., A modified mountain clustering algorithm. Pattern Analysis and Applications, 8, 125-138, 2005.

[11] Verma, N. K., Roy, A. and Cui, Y., Improved Mountain Clustering for Gene Expression Data Analysis. Journal of Data Mining and Knowledge Discovery, 2 (1), 30-35 ,2011.

[12] Gong, S., Zhang, Y. and Yu, G., Clustering Stream Data by Exploring the Evolution of Density Mountain. Proceedings of the VLDB Endowment, 11(4), 393-405 ,2017.

[13] Chiu, S. L., Fuzzy model identification based on cluster estimation. J. Intell Fuzzy systems, 2, 267-268 ,1994.

[14] Shieh, H. L., Kuo, C. C. and Chen, F. H., Two-phases clustering algorithm based on subtractive clustering and k-nearest neighbours, International Conference on Machine Learning and Cybernetics, 2013.

[15] Widodo, I. D., Fuzzy Subtractive clustering-based prediction model for brand association analysis. MATEC Web of Conferences 154,1-6, 01082 ,2018.