Fuzzy Clustering Algorithm Efficient Implementation Using Centre of Centres

Chundru Raja Ramesh[1*]   Komaragiri Raghava Rao[2]   Gunamani Jena[1]

[1]*Department of Computer Science & Engineering, K L University, India*
[2]*Department of Computer Science & Engineering, BVC Engineering College, India*
* Corresponding author's Email: chrajaramesh@gmail.com

Abstract: Clustering is a procedure of finding similar data items (patterns, documents etc.) and then group the similar data together. Items belongs to different clusters are dissimilar data items, generally cluster values are considered as 1 or 0. The clustering process is not appropriate for all the cases sometimes these values are less than one. In practical situations clusters are not crisp, then it is represented as fuzzy. In order to enhance the clustering rate, two appropriate clustering approaches: K-means clustering and Fuzzy C Means (FCM) are considered. These approaches are modified by minimizing the objective function known as squared error function. The experimental research was performed on the publicly available database (i.e. yeast dataset) to validate its clustering performance in terms of accuracy, specificity, sensitivity and execution time. Experimental outcome shows that the proposed technique improves the accuracy in clustering rate up to 1.5-35% compared to the existing methodologies FCM and k-means approaches.

Keywords: Clustering, Fuzzy c-means, K-means clustering, Squared error function, Yeast dataset.

1. Introduction

In recent decades, the rate of growth of databases are increased exponentially, as a result, the conventional database analysis techniques are fail to find useful information from the large databases [1, 2]. Data clustering plays an important role in such situations, where there is a need to discover the natural grouping structure of data in an unsupervised manner [3]. Clustering is the procedure of organizing objects into teams, whose members are similar in some way. A cluster is a group of objects that similarity measurement is near to each other and the dissimilarity objects are placed in the different clusters [4]. Currently, there are several clustering approaches are available such as, distribution based clustering schemes, density based clustering schemes and centroid based clustering schemes [5]. In this scenario, centroid based clustering methods: FCM, K-means, Fuzzy k means, FCM with optimization methods, etc. are considered for experimental validation. These type of clustering methodologies are demonstrated by a central vector and the clustering procedure is generally defined as optimization, which finds the clusters and assign the items to the closest or most similar cluster such that a certain measure is minimized [6, 7].

Among these available methodologies, FCM and K-means approaches are the well-known centroid based clustering partition method [8]. In K-means approach, the objects are classified based on K-groups. In each cluster, the centroid or mean is taken as the cluster representative. If suppose, the data is real-valued data, the arithmetic mean of the attribute vectors for all objects within a cluster provides appropriate cluster representative [9]. On the other hand, FCM is similar to k-means algorithm, initially selects the number of clusters. Then, assign coefficients for each clusters and compute the centroid of each cluster [10, 11]. One of the major drawback in (K-means and FCM approaches) is very difficult to find the K-value in global dataset and also produces tighter clusters, especially if the clusters are globular. In order to overcome this issue, modified FCM and K-means clustering approaches are developed in this research. These approaches are

modified by minimizing the objective function known as squared error function. To verify the performance of traditional and proposed methods, a publicly available dataset: yeast dataset is utilized to validate the clustering performance in terms of accuracy, specificity, sensitivity and execution time.

This literature is composed as follows. In Section 2, survey several recent clustering strategies. In section 3, modified FCM and K-means strategies are portrayed, to seek better clustering accuracy. In Section 4, the execution of the implemented techniques is assessed by simulation. The conclusion is made in Section 5.

2. Literature review

Several techniques are suggested by the researchers in clustering strategy. In this scenario, brief evaluations of some important contributions to the existing literatures are presented. Kernel FCM is one of the widely used algorithm, which was constantly experimented and modified to give better knowledge about spatial information. F. Zheng, C. Zhang, X. Zhang, and Y. Liu, [12] proposed a fast anti-noise FCM algorithm to reduce the effect of noise by constructing spatial information with the help of pixel value and membership function combination. It gives limited clustering performance, which was made possible by the use of adaptive choose of optimal parameters. In a large dataset, the proposed system failed to achieve better clustering performance by means of accuracy.

P. Sivasangareswari, and K.S. Kumar, [13] made a modification in FCM algorithm based on distance metric to satisfy the problem of intensity in homogeneities and noise. Distance metric was introduced to reduce the random number of variables from the segmented image of modified FCM. The local neighbour relationship was controlled adaptively with a trade-off fuzzy method, but it leads to misclassification of the kernels associated with the image. If the number of samples were low, the FCM algorithm was highly affected in the training set that decreases the performance of clustering.

A. Elazab, C. Wang, F. Jia, J. Wu, G. Li, and Q. Hu, [14] proposed an adaptive regularized kernel based fuzzy c-means clustering for segmentation purpose. It produced enhanced robustness against loss of pixel details with a regularized parameter inclusion, but it requires more number of iterations as it was required to calculate the local coefficient of variance with a square value. The variation of changes between dispersed pixel particles consumes time, thus covariance of weighted image was determined in the proposed algorithm to save time and complexity, which keeps the centre pixel as constant and iterates the neighbour pixels. The system was complex, when the number of features or iteration was more.

E.Nasibov, and C. Kandemir-Cavas, [15] proposed a new methodology called as Optimally Weighted Fuzzy K-Nearest Neighbors (OWFKNN) to predict proteins' subcellular locations based on amino acid composition. The experiment was carried on a publicly available database (i.e., eukaryotic and prokaryotic) to validate its accuracy, robustness and speed. In a few cases, the training data were dependent evaluation or manual adjustment, which was needed to be automated.

S.S. Kumar, and J. Albert, [16] proposed kernel weighted FCM for calculating the area of the tumour and the shape was analysed using edge detection technique. It improves FCM by introducing the weight for pixels within local neighbour windows, but the edge detection was followed by the morphological filtering method, which carries more number of iterations. Thus, proposed method was designed with weighted covariance calculation, which reduces the number of iterations. This technique segment only normal brain tissue, the abnormal tissue classification was absent. In this research, it was so difficult to compare several data, due to lower level alignment concern.

K.G. Satheesh, and A.N.J. Raj, [17] developed a new methodology: Orthogonal Learning Particle Swarm Optimization (OLPSO) technique for tuberculosis detection in lung image. This learning algorithm was used to identify the specific set of features from the image and ranks the features based on decision task equation. Based on which the images were classified. In addition, this paper proposed a hybrid classification method: Multiple Kernel Fuzzy C Means (MKFCM) to differentiate the images as Cavitary TB and Miliary TB. Whereas, the learning algorithm utilized in this research was too sensitive to parameters.

M.G. Alam, and S. Baulkani, [18] developed a web clustering algorithm: Web Document Clustering using K-means and Artificial Bee Colony algorithm (WDC-KABC) for clustering the web documents effectively. The proposed algorithm utilized the features of both K-means and ABC clustering algorithm. In this paper, ABC algorithm was employed as the global search optimizer and K-means was utilized for refining the solutions. Thus, the quality of the cluster was improved. The performance of WDC-KABC was analysed with four different datasets. While performing with semi supervised methodologies, the semantic gap was

maximized between the feature values, which leads to clustering rate.

To overcome the above mentioned drawbacks, effective modified clustering approaches (FCM and K-means) are implemented for enhancing the performance of clustering in various applications.

3. Proposed methodology

In this scenario, for identifying the clustering efficiency, two appropriate clustering methodologies are undertaken such as K-means clustering approach and FCM clustering approach. A brief description about K-means, FCM and their modifications are detailed below.

3.1 K-means clustering approach

K-means clustering is one of the most popular and simple unsupervised learning algorithm that solves the clustering problem. This algorithm procedure is simple and easy way to classify the given data set through a required number of clusters (assume k-clusters) are fixed appropriately define k centres, one for each cluster. These centres should be placed in such a way that the algorithm reaches quick convergence. One such choice is to place them, as much as possible, far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest centre. After assigning each data point to a cluster, the means are re-calculated. After achieve k new centroids, a new binding has to be done between the same data set points and the nearest new centre. As a result of this iteration, the k centres changes their locations after each iteration. The loop will terminate when the changes in the positions of the centres is negligible or less than a small pre-fixed value. Finally, this algorithm aims minimizing the objective function known as squared error function, which is given in the Eq. (1).

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left\| \left(x_i - v_j \right)^2 \right\| \qquad (1)$$

Where,
$\|x_i - v_j\|$ is the Euclidean distance between x_i and v_i
c_j is the number of data points in j^{th} cluster.
c is the number of clusters.

3.2 Proposed modified K-means clustering algorithm

However, it is always not possible or practical to fix the number of clusters. So, k-means algorithm is

modified to achieve stability of the system, at the same time determine the number of clusters dynamically. The algorithm starts with two clusters and increases the number of clusters one at a time until the centre of the centres of the clusters stabilizes or two consecutive 'centres' of centres differ by less than a pre-decided small value. The pseudo code of modified k-means algorithm is represented below.

> Step 1: Set the number of clusters $k = 2$
> Step 2: Find the centre of centres of the cluster
> Step 3: $k = k + 1$
> Step 4: Form k clusters
> Step 5: Find the centre of centres of the clusters
> Step 6: If the distance between two consecutive centre of centres is greater $> \varepsilon$ (a predefined value) go to step 3
> Step 7: output k
> Step 8: Stop

By applying the equation (1) in k-means algorithm, the stability of the system is further achieved. The Matlab code of modified k-means algorithm is represented below.

```
Clear
X=load ('yeastfull.dat');
k=2;
thmin=0.10;
while true
   [idx, C]=kmeans(X, k);
meanc=mean(C);
   sum=0;
   For i = 1: size(C, 1)
  d=distfcm (C (i, :), meanc (1, :));
      d;
      sum + d;
   end
   th=sum/size(C,1)
   if th<thmin
      break;
   else
            k = k + 1;
   end
end k
```

3.3 Fuzzy c-means clustering approach

In this scenario, a brief description about FCM approach is illustrated. Consider a set of unlabelled patterns $X = \{X_1, X_2, \dots X_N\}, X_i \in R^f$. Where, n is the number of patterns and f is the dimension of

pattern vectors (features).The main focus of the FCM algorithm is minimizing the value of an objective function. The objective function measures the quality of the cluster partition the data and divides into c clusters. The FCM algorithm measures the quality of the partition by comparing the distance from pattern point X_i to the current cluster center w_j with the distance from pattern X_i to other cluster centers. FCM algorithm is a method, which allows the data item belongs to two or more clusters. This method is generally used in pattern recognition process. The method developed by Dunn and Bezdeck, which is based on the minimization of the objective function Eq. (2).

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} U_{ij}^m |X_i - c_j|^2, \ i \leq m \qquad (2)$$

Where, m is stated as a real number greater than 1. It was set to 2.00 by Bezdeck, U_{ij} is the degree of membership of X_i in the cluster j, c_j is the center of the $j - th$ cluster and $|| * ||$ is a norm expressing the 'distance'.

Fuzzy clustering is an iterative optimization process with an objective function of membership U_{ij} and the c_j cluster centers in each iteration that is stated in the Eq. (3).

$$U_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{|x_i - c_j|}{|x_i - c_k|} \right]^{\frac{2}{m-1}}} \qquad (3)$$

Step 1: Input threshold value $min - th$.
Step 2: Initialize number of clusters are 2.
Step 3: Initialize the centre C_1 , C_2.
Step 4: Let U_{ij} be the degree of membership of X_i belonging to $j - th$ cluster, where

$$U_{ij} = \frac{1}{m_j} e^{\frac{-(x_i - c_l)^2}{\prod (x_i - c_k)^2}} \quad j \neq i$$

Step 5: At each stage calculate the centre of the clusters as,
$$c_j = \frac{\sum_{i=1}^{N} u_{ij} x_i}{\sum_{i=1}^{N} u_{ij}}$$
Step 6: Update U^k
Step 7: $\left\| c_j^{k+1} - c_j^k \right\| < min - th$, then goto step 8; otherwise return to step 3.
Step 8: Calculate centre of all the centres. If present centre of centres and previous centre of centres are same then stop the process. Otherwise, increase k value by 1 ($k = k + 1$); then goto step 3.
Step 9: Stop.

Where, m is a real number greater than 1 and U_{ij} is the degree of member ship of X_i to cluster j. The values of c_j, centers of the clusters are recursively updated until two successive values differed by pre-fixed small number. After initialization of U_{ij} values, the update of these values is performed by Eqs. (3) and (4)

$$c_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m} \qquad (4)$$

3.4 Proposed modified FCM algorithm

As mentioned earlier, it is neither possible nor practical to fix the number of clusters initially. Initially, fix the number of clusters as 2 and iteratively increase the number of clusters until the distance between two consecutive centres of centres is less than a pre-fixed value. The modified FCM algorithm is described below.

---

1. Initialize $k = 2$.
2. Initialize $U = [uij]$ matrix $U0$
3. At k-step : calculate the centre vectors
$$C(k) = [cj] \, with \, U(k), c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m}$$
4. Update $U^{(k)}, U^{(k+1)}$

5. $u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$
6. $if \, \left\| U^{(k+1)} - U^k < \in \right\|$, then goto the step 6; other-wise return to step 3
7. Calculate centre of all the centres. If the difference between the centre of centres of the last two iterations is less than a pre-defined value, stop the process. Otherwise, increase k value by 1 ($k = k + 1$); then goto step 2.

3.5 FCM Gaussian type member ship function

In FCM algorithm member ship function uses an unknown parameter m which may have a value between 1 and 2. To remove this uncertainty, a Gaussian type member-ship function is used that contains all the properties of fuzzy function. Notice that the degree of membership of ci belongs to a cluster with centers ci is 1 and degree of membership of ci belongs to other cluster is 0. Following algorithm represents the fuzzy Gaussian type member ship function.

3.6 Proposed new Gaussian type member ship function

In the above algorithm, the number of clusters is represented as (k). Now, determine the number of clusters dynamically based on the stability of the system. In proposed algorithm, the k value (number of clusters) is not pre-fixing. Initially, the number of clusters are two and the threshold value is considered as the input, the different threshold values are used as input. In the proposed algorithm, after some specific threshold value, identified the number of clusters obtained as same as existing approach. The modified new Gaussian type member ship function algorithm is described below.

Step 1: Choose the number of clusters K.

Step 2: Initialize the centre $C_1, C_2, C_3 \dots C_k$.

Step 3: Let U_{ij} be the degree of membership of X_i belonging to $j - th$ cluster where

$$u_{ij} = e^{\frac{-\|x_i - c_j\|^2}{\|x_i - c_l\|^2}} \quad j \neq i$$

Step 4: at each stage calculate the centre of the clusters as,

$$c_j = \frac{\sum_{i=1}^{N} u_{ij} x_i}{\sum_{i=1}^{N} u_{ij}}$$

Step 5: Update U_k

Step 6: if $\left\| U_j^{k+1} - U_j^k \right\| < \in$ then next step, otherwise go to step 4.

---

4. Result and discussion

In this scenario, for experimental simulation, MATLAB (version 2017a) was employed on PC with 3.2 GHz with i5 processor. In order to evaluate the efficiency of proposed algorithms, the performance of proposed methods was compared with existing methodologies (k-means, FCM and FCM Gaussian type member ship function) on the reputed database: yeast dataset [19]. The yeast dataset is utilized for identifying the protein localization. It totally consists of 1484 number of instances, ten classes and six number of attributes. The yeast dataset consists of 6554 organic genomes on profiles, each profile records the fluctuation of expression levels over 2 cell cycles that contain seventeen time points. In this experimental research, the performance of the proposed approach was compared in terms of sensitivity, accuracy, specificity and execution time.

4.1 Performance measure

The relationship between the input and output variables of a system understand by employing the suitable performance metrics like sensitivity and specificity. The general formula for calculating the sensitivity and specificity, is given in the Eqs. (5) and (6).

$$Sensitivity = \frac{Number\ of\ TP}{Number\ of\ TP + Number\ of\ FN} \times 100 \quad (5)$$

$$Specificity = \frac{Number\ of\ TN}{Number\ of\ TN + Number\ of\ FP} \times 100 \quad (6)$$

Where, TP is represented as true positive, FP is denoted as false negative, TN is represented as true negative and FN is stated as false negative.

In addition, the accuracy is another suitable evaluation metrics for finding the effectiveness of clustering rate. Accuracy is the measure of statistical variability and a description of random errors. The general formula of accuracy for determining clustering rate is given in the Eq. (7).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (7)$$

4.2 Experimental analysis of k-means and modified k-means

In this experimental analysis, yeast dataset is assessed for comparing the performance evaluation of existing method and the proposed scheme using K-means algorithm. In table 1, sensitivity, specificity, and accuracy value of the proposed and existing approach is compared for different threshold and k-values. The performance measures are calculated by comparing the number of generated clusters with available clusters in yeast dataset. The average accuracy, sensitivity and specificity value of existing methods [19] delivers 55.55%, 47.78% and 63.34%. Similarly, the average accuracy, sensitivity and specificity value of proposed method delivers 68.88%, 67.78% and 70%. The graphical comparison of existing and proposed method using k-means clustering algorithm is represented in the Fig. 1.

Table 1. Performance comparison using K-means algorithm

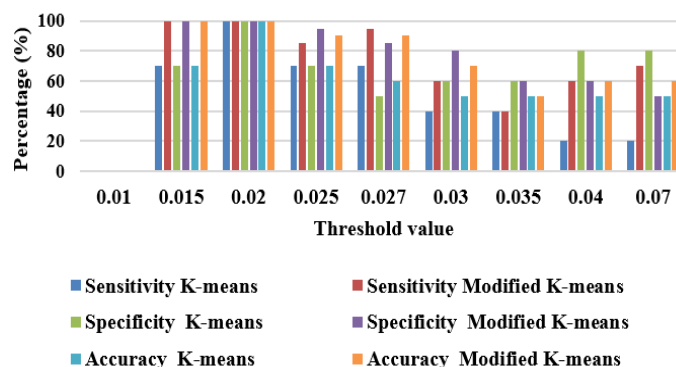| Threshold value	Number of generated clusters		Sensitivity (%)		Specificity (%)		Accuracy (%)	
	K-means [19]	Modified K-means	K-means [19]	Modified K-means	K-means [19]	Modified K-means	K-means [19]	Modified K-means
0.010	39	25	0	0	0	0	0	0
0.015	13	10	70	100	70	100	70	100
0.02	10	10	100	100	100	100	100	100
0.025	7	9	70	85	70	95	70	90
0.027	6	9	70	95	50	85	60	90
0.030	5	7	40	60	60	80	50	70
0.035	5	5	40	40	60	60	50	50
0.040	5	6	20	60	80	60	50	60
0.070	5	6	20	70	80	50	50	60

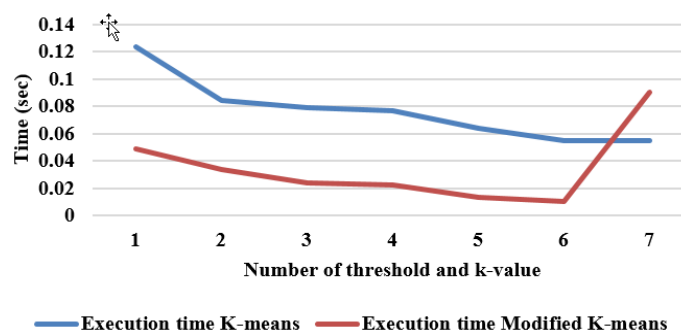Figure. 1 Graphical comparison of existing and proposed approach using k-means algorithm

Figure. 2 Execution time comparison of existing and proposed approach using k-means algorithm

In table 2, the execution time comparison between the proposed and existing methodology is clearly mentioned. The execution time of existing method (K-means) is calculated based on the k-value and the execution time of proposed method (Modified-K-means) is determined based on the threshold value, which is clearly mentioned in the table 2. Threshold value has given input to modified K-means and it generate k number of clusters, so number of cluster for both algorithms are same. The average execution time of the proposed methodology is 0.034 (sec) and the existing methodology delivers 0.0765 (sec) of average execution time. The graphical execution time comparison of existing and proposed method using

k-means clustering algorithm is represented in the Fig. 2.

Table 2. Execution time comparison between existing and proposed method for K- means clustering

Execution time (sec)			
K-value	K-means [19]	Threshold value	Modified K-means
25	0.1232	0.010	0.049
10	0.0840	0.015	0.034
9	0.0788	0.027	0.024
7	0.0764	0.030	0.022
5	0.0635	0.035	0.013
6	0.0549	0.040	0.010
6	0.0549	0.070	0.09

Table 3. Performance comparison using FCM clustering algorithm

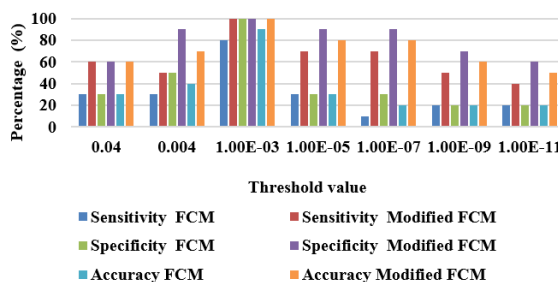| Threshold value	Number of generated clusters		Sensitivity (%)		Specificity (%)		Accuracy (%)	
	FCM [19]	Modified FCM	FCM [19]	Modified FCM	FCM [19]	Modified FCM	FCM [19]	Modified FCM
0.04	3	6	30	60	30	60	30	60
0.004	4	7	30	50	50	90	40	70
1e-3	9	10	80	100	100	100	90	100
1e-5	17	12	30	70	30	90	30	80
1e-7	18	12	10	70	30	90	20	80
1e-9	18	14	20	50	20	70	20	60
1e-11	18	15	20	40	20	60	20	50

Figure. 3 Graphical comparison of existing and proposed approach using FCM algorithm

4.3 Experimental analysis of FCM and modified FCM

In this scenario, yeast dataset is assessed for comparing the performance evaluation of existing method and the proposed scheme using FCM clustering algorithm. In table 3, sensitivity, specificity, and accuracy value of the proposed and existing methodologies compared for different threshold and k-values. The average accuracy, sensitivity and specificity of existing methods [19] delivers 35.71%, 31.42% and 40%. Similarly, the average accuracy, sensitivity and specificity value of proposed method delivers 71.42%, 62.85% and 80%. The graphical comparison of existing and proposed method using FCM clustering algorithm is represented in the Fig. 3.

In table 4, the execution time of existing method (FCM) depends on the k-value and the execution time of proposed method (modified-FCM) depends on the threshold value, which is clearly stated in the table 4. Threshold value has given input to modified FCM and it generate k number of clusters, so number of cluster for both algorithms are same. The average execution time of the proposed methodology is 0.3071 (sec) and the existing methodology delivers 0.3628 (sec) of average execution time. The graphical execution time comparison of existing and proposed method using

Table 4. Execution time comparison between existing and proposed method for FCM clustering

Execution time (sec)			
K-value	FCM [19]	Threshold value	Modified FCM
6	0.14	0.04	0.10
7	0.20	0.004	0.20
10	0.38	1e-3	0.35
12	0.43	1e-5	0.40
12	0.43	1e-7	0.42
14	0.48	1e-9	0.33
15	0.48	1e-11	0.35

FCM clustering algorithm is represented in the Fig. 4.

4.4 Experimental analysis of FCM Gaussian type member ship function and modified FCM Gaussian type member ship function

In this section, experimental analysis for FCM Gaussian type member ship function and modified FCM Gaussian type member ship function is illustrated. In table 5, sensitivity, specificity and accuracy value of the proposed and existing approach is compared for different threshold and k-values. The average accuracy, sensitivity and specificity value of existing methods [19] delivers 68.57%, 72.85% and 64.28%. Similarly, the average accuracy, sensitivity and specificity value of
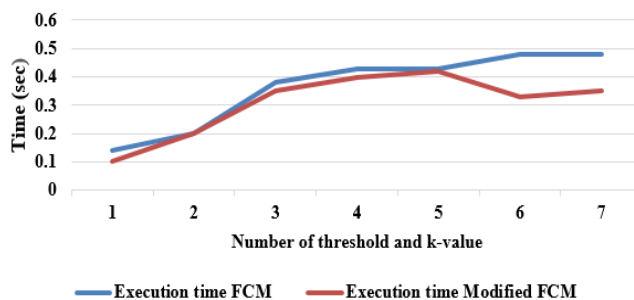
Figure. 4 Execution time comparison of existing and proposed approach using FCM algorithm

Table 5. Performance comparison using FCM Gaussian type membership function

| Threshold value	Number of generated clusters		Sensitivity (%)		Specificity (%)		Accuracy (%)	
	FCM Gaussian [19]	Modified FCM Gaussian	FCM Gaussian [19]	Modified FCM Gaussian	FCM Gaussian [19]	Modified FCM Gaussian	FCM Gaussian [19]	Modified FCM Gaussian
0.04	3	7	30	50	30	90	30	70
0.004	4	7	50	90	30	50	40	70
1e-3	6	6	60	60	60	60	60	60
1e-5	7	8	80	90	60	70	70	80
1e-7	8	10	90	100	70	100	80	100
1e-9	10	11	100	80	100	100	100	90
1e-11	10	10	100	100	100	100	100	100
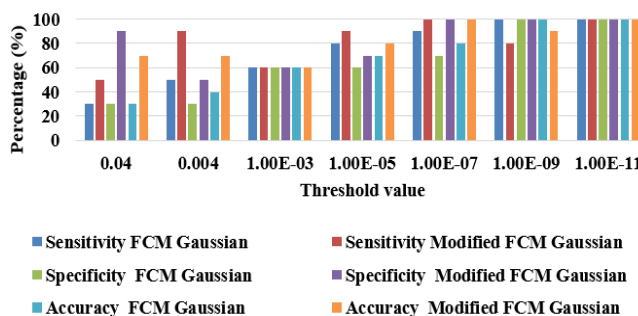
Figure. 5 Graphical comparison of existing and proposed approach using FCM Gaussian type member ship function

proposed method delivers 81.42%, 81.42% and 81.428%. The graphical comparison of existing and proposed method using FCM Gaussian type member ship function algorithm is represented in the Fig. 5.

In table 6, the average execution time of the proposed methodology (modified FCM Gaussian) is 0.198 (sec) and the existing methodology (FCM Gaussian) delivers 0.49 (sec) of average execution time. Threshold value has given input to modified FCM Gaussian and it generate k number of clusters, so number of cluster for both algorithms are same.

The graphical execution time comparison of existing and proposed method using FCM Gaussian type member ship function algorithm is represented in the Fig. 6. The tables 1, 2, 3, 4, 5 and 6 confirmed that the proposed approaches perform effectively compared to the existing methods in the yeast dataset.

Table 6. Execution time comparison between existing and proposed method for FCM Gaussian type member ship function

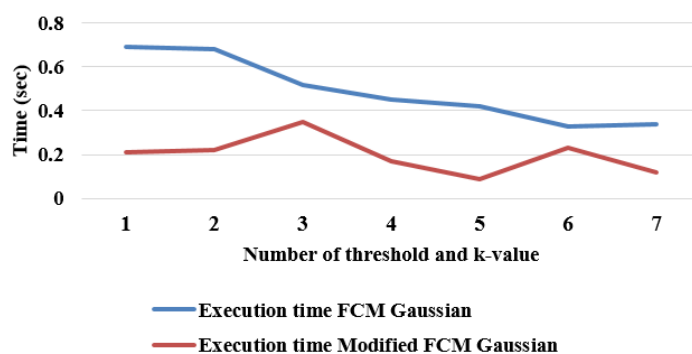| Execution time (sec)			
K-value	FCM Gaussian [19]	Threshold value	Modified FCM Gaussian
7	0.69	0.04	0.21
7	0.68	0.004	0.22
6	0.52	1e-3	0.35
8	0.45	1e-5	0.17
10	0.42	1e-7	0.09
11	0.33	1e-9	0.23
10	0.34	1e-11	0.12

Figure. 6 Execution time comparison of existing and proposed approach using FCM Gaussian type member ship function

Table 7. Comparative analysis

Methods	Modified K-means	Modified FCM	AMOC [20]	Modified FCM Gaussian
Accuracy (%)	68.88	71.42	80.59	81.42

4.5 Comparative analysis

Table 7 presents the comparative study of existing work and the proposed work performance. K.K. Pavan, A.A. Rao, and A.V. Rao, [20] proposed an automatic, simple and effective clustering algorithm:
Automatic Merging for Optimal Clusters (AMOC), which was utilized to generate optimal clusters for the given datasets automatically. This experiment was carried out on yeast dataset to validate its result in terms of accuracy and achieved 80.59%. Whereas, the proposed work achieves 81.42 % of accuracy that was higher than the existing works.

5. Conclusion

Clustering is considered as the most essential unsupervised learning concern, it deals with the identification of structure in an unlabelled data collection. In this research, to improve the clustering rate, two appropriate clustering approaches are modified: K-means and fuzzy c means clustering. The proposed methodology effectively combines the benefits of K-means and fuzzy c means clustering approach. The experimental investigation verified on publicly available database: yeast dataset, which shows a superiority of the proposed methodology. Associated to the other obtainable approaches, the advanced scheme delivered an effective performance by means of accuracy, sensitivity and specificity, around 1.5-35% enhancement than the existing methods like FCM and K-means algorithm. In future work, the proposed clustering methodologies are implemented in specified application with multiple feature extraction approaches and multi class learning classification methodology.

References

[1] M.T. El-Melegy and H.M. Mokhtar, "Tumor segmentation in brain MRI using a fuzzy approach with class center priors", *EURASIP Journal of Image and Video Processing*, No.1, pp.21, 2014.

[2] Y. Gui, X. Zhang, and Y. Shang, "SAR image segmentation using MSER and improved spectral clustering", *EURASIP Journal of Advances in Signal Processing*, No.1, pp.83, 2012.

[3] R. Shang, P. Tian, L. Jiao, R. Stolkin, J. Feng, B. Hou, and X. Zhang, "A spatial fuzzy clustering algorithm with kernel metric based on immune clone for SAR image segmentation", *International Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol.9, No.4, pp.1640-1652, 2016.

[4] S.B. Shokouhi, A. Fooladivanda, and N. Ahmadinejad, "Computer-aided detection of breast lesions in DCE-MRI using region growing based on fuzzy C-means clustering and vesselness filter", *EURASIP Journal of Advances in Signal Processing*, No.1, pp.39, 2017.

[5] V.H. Pham and B.R. Lee, "An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm", *Vietnam Journal of Computer Science*, Vol.2, No.1, pp.25-33, 2015.

[6] D. Pandove, R. Rani, and S. Goel, "Local graph based correlation clustering", *Knowledge-Based Systems*, Vol.138, pp.155-175, 2017.

[7] K.M. Kumar and A.R.M. Reddy, "An efficient k-means clustering filtering algorithm using density based initial cluster centers", *Information Sciences*, Vol.418, pp.286-301, 2017.

[8] H. Yahyaoui and H.S. Own, "Unsupervised clustering of service performance behaviors", *Information Sciences*, Vol.422, pp.558-571, 2018.

[9] M. Śmieja, Ł. Struski, and J. Tabor, "Semi-supervised model-based clustering with controlled clusters leakage", *Expert Systems with Applications*, Vol.85, pp.146-157, 2017.

[10] A. Sokhandan, P. Adibi, and M. Sajadi, "Multitask fuzzy Bregman co-clustering approach for clustering data with multisource features", *Neurocomputing*, Vol.247, pp.102-114, 2017.

[11] C.R. Ramesh, G. Jena, K.R. Rao, and C.V. Sastry, "Fuzzy clustering with Gaussian-type member ship function", In: *Proc. of the International Conference on Intelligent Systems and Signal Processing (ISSP)*, pp. 393-396, 2013.

[12] F. Zheng, C. Zhang, X. Zhang, and Y. Liu, "A fast anti-noise fuzzy C-means algorithm for image segmentation", In: *Proc. of the 20th International Conference on Image Processing*, pp.2728-2732, 2013.

[13] P. Sivasangareswari and K.S. Kumar, "Fuzzy C-Means Clustering with Local Information and Kernel Metric for Image Segmentation", *International Journal of Advanced Research in Computer Science & Technology*, Vol.2, 2014.

[14] A. Elazab, C. Wang, F. Jia, J. Wu, G. Li, and Q. Hu, "Segmentation of Brain Tissues from Magnetic Resonance Images Using Adaptively Regularized Kernel-Based Fuzzy-C-Means Clustering", *Computational and Mathematical Methods in Medicine*, 2015.

[15] E. Nasibov and C. Kandemir-Cavas, "Protein subcellular location prediction using optimally weighted fuzzy k-NN algorithm", *Computational biology and chemistry*, Vol.32, No.6, pp.448-451, 2008.

[16] S.S. Kumar and J. Albert, "Medical Image Segmentation using Kernel Weighted Fuzzy Clustering (KWFC) Algorithm", *International Journal of Inventions in Computer Science and Engineering*, Vol.1, No.9, 2014.

[17] K.G. Satheesh and A.N.J. Raj, "Medical Image Segmentation and Classification Using MKFCM and Hybrid Classifiers", *International Journal of Intelligent Engineering and Systems,* Vol.10, No.6, pp.9-19, 2017.

[18] M.G. Alam and S. Baulkani, "A Hybrid Approach for Web Document Clustering Using K-means and Artificial Bee Colony Algorithm", *International Journal of Intelligent Engineering and Systems,* Vol.9, No.4, pp.11-20, 2016.

[19] P. Ashok, G.M. Kadhar, E. Elayaraja, and V. Vadivel, "Fuzzy based clustering method on yeast dataset with different fuzzification methods", In: *Proc. of the 4th International Conference on Computing, Communications and Networking Technologies*, pp.1-6, 2013.

[20] K.K. Pavan, A.A. Rao, and A.V. Rao, "An automatic clustering technique for optimal clusters", *arXiv preprint arXiv*, pp.1109.1068, 2011.