Enforcement of Rough Fuzzy Clustering Based on Correlation Analysis

Revathy Subramanion[1],Parvathavarthini Balasubramanian[2],and Shajunisha Noordeen[3]
[1]Research Scholar,Sathyabama University, India
[2]Department of Master of Computer Applications, Anna University, India
[3]Post Graduate Scholar,Sathyabama University, India

Abstract: *Clustering is a standard approach in analysis of data and construction of separated similar groups. The most widely used robust soft clustering methods are fuzzy, rough and rough fuzzy clustering. The prominent feature of soft clustering leads to combine the rough and fuzzy sets. The Rough Fuzzy C-Means (RFCM) includes the lower and boundary estimation of rough sets, and fuzzy membership of fuzzy sets into c-means algorithm, the widespread RFCM needs more computation. To avoid this, this paper proposes Fuzzy to Rough Fuzzy Link Element (FRFLE) which is used as an important factor to conceptualize the rough fuzzy clustering from the fuzzy clustering result. Experiments with synthetic, standard and the different benchmark dataset shows the automation process of the FRFLE value, then the comparison between the results of general RFCM and RFCM using FRFLE is observed. Moreover, the performance analysis result shows that proposed RFCM algorithm using FRFLE deals with less computation time than the traditional RFCM algorithms.*

Keywords: *Software clustering, FCM,RCM, RFCM, FRFLE.*

Received March 27, 2014; accepted May 12, 2014

1. Introduction

An essential technique of data mining is unsupervised clustering. It deals with finding a structure in ensemble of unlabeled data. The clustering approach [1,2,22] can be classified into two classifications such as soft and hard clustering. Each object assigns to exactly one cluster based on the hard clustering. Though they are simple to implement, it has demerits of being sensitive to outliers and also difficult to handle ambiguous, uncertainty and moreover, it cannot deal with objects which are close to two clusters. These problems were conquered by the soft clustering algorithm. It defines that an object can belong to more than one cluster. Soft clustering algorithms have board categories like Fuzzy C-Mean (FCM), Rough C-Mean (RCM) and Rough FCM (RFCM). FCM algorithm [8,7,18] permits each data objects to cluster according to the membership of fuzzy sets, which is capable of handling the overlapping data objects and FCM results can be evaluated using many popular indices[20]. FCM algorithm has been descended by the characteristics such as too descriptive, slow to converge. RCM algorithm [11, 15,16,17] uses the concept of lower and upper estimation of rough set which is used to effectively handle the uncertainties. Ideas of both fuzzy and rough set are integrated into the RFCM algorithm, such that the RFCM algorithm establishes the crisp lower and fuzzy boundary proposition.RFCM algorithm has been widely used inmany applications [10,12,13,5].It is also used for intrusion detection[21].

Traditional RFCM algorithm is designed using several procedures and framed through the concept of collaboration. Novel collaborative clustering [19] is developed using the RCM and RFCM algorithms introduced by Mitra*et al.*[6] The collaboration concept is incorporated by exchanging information between the modules regarding the local partitions. It includes two phases as with and without collaboration. Hybrid algorithm [13] established by association of both rough and rough fuzzy concept. Rough-Fuzzy Possibilistic C-Means (RFPCM) [14] where the rough fuzzy is based on both probabilistic and possibilistic membership to avoid the problems such as noise sensitivity of the FCM and the coincident clusters of PCM. The concept of crisp lower and fuzzy boundary of each cluster is introduced in the RFPCM by Maji and Sankar [13]these conventional RFCM algorithms face the following downsides including more computation to converge and time taken is high. Even though the rough correlation factor is proposed by Joshi *et al.* [3] to overcome these issues which is used only to translate the FCM into RCM not for RFCM and also its not automated for different real world datasets. Rough fuzzy clustering algorithm using fuzzy rough correlation factor by Revathy and Parvathavarthini [19], though it translates the FCM into RFCM not automated the factor value for different benchmark datasets. This paper puts forward the Fuzzy to Rough Fuzzy Link Element (FRFLE) to conceptualize the RFCM algorithm. This Scheme is used

to establish the RFCM algorithm by yielding less computation and time consumption.FRFLE is determined based on the Degree of Fuzziness Ratio (DFR).The clear process of obtaining FRFLE, automation of FRFLE for various benchmark datasets and performance analysis of RFCM using FRFLE based on the computation time are also been discussed.

2. Literature Review

In this section the traditional algorithms such as FCM, RCM and RFCM have been discussed.

2.1. Fuzzy C-Means

According to Bezdek's (1981) FCM concept let $R=\{R_1, R_2, ..., R_n\}$be the set of N objects and $V=\{v_1, v_2, ..., v_i, ...,v_c\}$be the set of centroids, and $V_i \in R$. It segments R_k into c clusters based on the degree of membership value, where $1 \leq m < \infty$ is the fuzzifier, V_i is the i^{th} cluster center,$u_{ik} \in [0, 1]$ is the membership of the k.

$$V_i = \frac{\sum_{k=1}^{N} (u_{ik})^m R_k}{\sum_{k=1}^{N} (u_{ik})^m} \quad (1)$$

The process begins by randomly choosing k objects as the centroid of the c clusters. The memberships are calculated based on the relative distance of the objectsR_k to the centroids $\{V_i\}$ by Equation 1.

$$u_{ik} = \frac{1}{\sum_{i=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}} \quad \forall i \quad (2)$$

Where $d_{ik} = \|R_k - V_i\|^2$ subjectto, $\sum_{i=1}^{c} u_{ik} = 1, \quad \forall k$,and $0 < \sum_{k=1}^{N} u_{ik} < N \ , \forall i$. After computing membership of all the objects, the new centroids of the clusters are calculated as per Equation 1. The process continues until the objective function converges, i.e., the centroids have identical value.

2.2. Rough C-Means

In real life situations, uncertainty may arise from incompleteness in class definition. This type of uncertainty can be handled by rough sets. The rough set concept was introduced by Pawlak [15], according to that each set consists of two parts, as lower and boundary region.

Properties of rough sets such as:

- An object $r_k$can be part of at most one lower bound.
- $r_k \in \underline{O}(U_i) \Rightarrow r_k \in \overline{O}(U_i)$.
- An object $r_k$is not part of any lower bound $\Rightarrow r_k$belongs to two or more upper bounds.

RCM and evaluationaryrough k-means were introduced by Lingras and West [4]. Here, each cluster consists of two parts, namely a crisp lower approximation, and crisp boundary approximation. It adds the idea of lower $\underline{O}(U_i)$ and upper $\overline{O}(U_i)$ estimation which segments the object regions as lower and boundary of clusterU_i. The boundary region of cluster $U_i$is denoted as $A(U_i) = \{\overline{O}(U_i) - \underline{O}(U_i)\}$. The 3 clusters separation example for four data elements using RCM is shown in Figure 1, thus the element x2 presents in the upper approximations of two clusters c1 and c2.

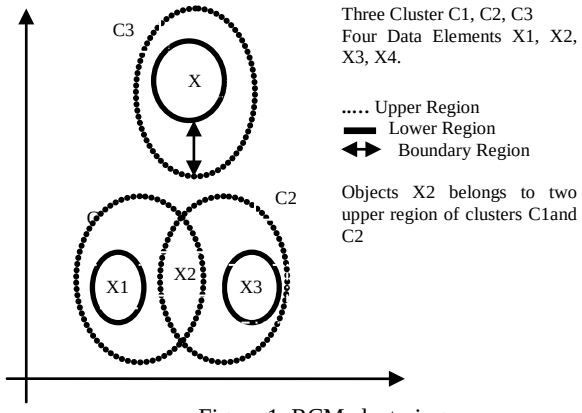

Figure 1. RCM clustering.

Calculation of the centroid is modified as below Equation 3 to include the effects of lower as well as upper bounds. In RCM, for each dataset choose appropriate threshold value. Compactness is based on the threshold, Importance of lower and upper approximation w_{low} andw_{up} values.

$$v_i = \begin{cases} w_{low} \frac{\sum_{R_k \in \underline{O}U_i} R_k}{|\underline{O}U_i|} + w_{up} \frac{\sum_{R_k \in (\overline{O}U_i - \underline{O}U_i)} R_k}{|\overline{O}U_i - \underline{O}U_i|}, if\ \underline{O}U_i \neq \varphi \cap \overline{O}U_i - \underline{O}U_i \neq \varphi \\ \frac{\sum_{R_k \in (\overline{O}U_i - \underline{O}U_i)} R_k}{|\overline{O}U_i - \underline{O}U_i|}, if\ \underline{O}U_i = \varphi \cap \overline{O}U_i - \underline{O}U_i \neq \varphi \\ \frac{\sum_{R_k \in \underline{O}U_i} R_k}{|\underline{O}U_i|}, otherwise \end{cases} \quad (3)$$

2.3. Rough Fuzzy C-Means

Rough fuzzy collaborative clusteringwas developed by Mitra*et al.* [6] and rough fuzzy cmedoids algorithm was introduced by Maji*et al.*[9], to handle the overlapping segments efficiency by both fuzzy and rough sets concepts. Each cluster consists of three parameters, namely a cluster centroid, a crisp lower approximation, and fuzzy boundary. The overlapping data objects handled through the fuzzy set theory and the concept of lower and upper estimation of rough sets deals with lack of certainty, ambiguous, and not completeness in class definition. The three clusters separation example for four data elements using RFCM is shown in Figure 2, thus the object x2 belongs to the both clusters C1 andC2 with their corresponding membership value range from [0 to 1].

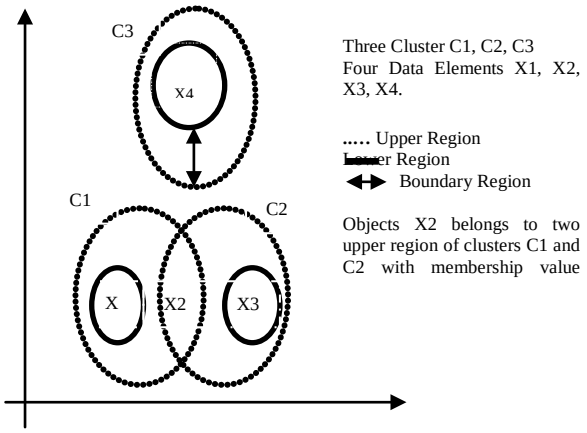

Figure 2. RFCM clustering.

The importance of lower and upper approximationw_{low}=1-w_{up}, 0.5<w_{low}<1, fuzzifier value m=2. Calculation of the new centroid is based on lower and upper approximation. Allocate each data object (pattern)$R_k$to the lower $\underline{O}U_i$ or upper$\overline{O}U_i$ approximation according to the threshold (here threshold (δ) value is various depending upon dataset) , if u_{ik}-$u_{jk}$is less than threshold,then $R_k \in \overline{O}U_i$ and$R_k \in \underline{O}\,U_i$and $R_k$cannot be a part of any lower estimation or else$R_k \in \underline{O}\,U_i$and $U_i$will be themaximum membership value. The centroid v_i calculation for rough fuzzy c-means is given as below:

$$v_i = \begin{cases} w_{low}\dfrac{\sum_{r_k \in \underline{O}U_i} u_{ik}^m R_k}{\sum_{r_k \in \underline{O}U_i} u_{ik}^m} + w_{up}\dfrac{\sum_{r_k \in (\overline{O}U_i - \underline{O}U_i)} u_{ik}^m R_k}{\sum_{r_k \in (\overline{O}U_i - \underline{O}U_i)} u_{ik}^m}, & if\ \underline{O}U_i \neq \varphi \cap \overline{O}U_i - \underline{O}U_i \neq \varphi \\ \dfrac{\sum_{r_k \in (\overline{O}U_i - \underline{O}U_i)} u_{ik}^m R_k}{\sum_{r_k \in (\overline{O}U_i - \underline{O}U_i)} u_{ik}^m}, & if\ \underline{O}U_i = \varphi \cap \overline{O}U_i - \underline{O}U_i \neq \varphi \\ \dfrac{\sum_{r_k \in \underline{O}U_i} u_{ik}^m R_k}{\sum_{r_k \in \underline{O}U_i} u_{ik}^m}, & otherwise \end{cases} \quad (4)$$

According to Maji and Sankar [9] RFCM algorithm the centroid is computed same as mention in above Equation 4 and the modified threshold for RFCM is given by:

$$\delta = \frac{1}{N}\sum_{k=1}^{N}\left(u_{ik} - u_{jk}\right) \quad (5)$$

Where N is the total number of objects,u_{ik} and$u_{jk}$is the maximum and second maximum membership value.

3. RFCM Algorithm using FRFLE

In this paper the RFCM algorithm is conceptualized using the FRFLE. The FRFLE is proposed for the correlation of FCM and RFCM results to convert the FCM into RFCM. This is then extended to the automation of the FRFLE for several real data sets and the performance analysis has been done based on the execution time. Figure 3 shows the architecture of the proposed system.

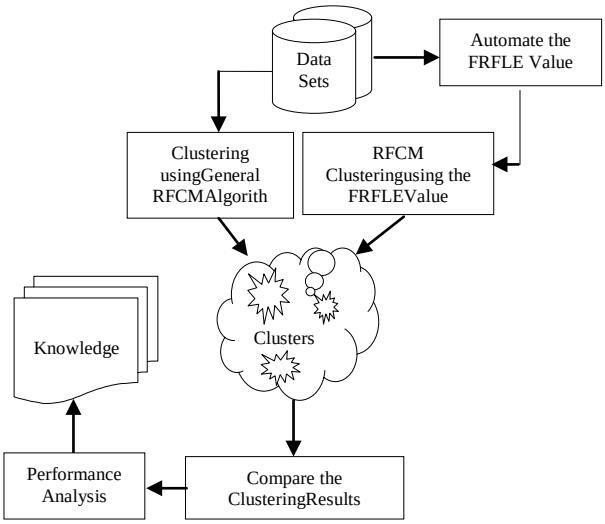

Figure 3.Architecture for proposed system.

3.1.Automation of the FRFLE Value

Algorithm 1 for computing FRFLR value is given below:

Algorithm 1: FRFLE computation.

Input: *The Dataset*
Output: *The FRFLE Value*

- *Step* 1: *Select the Dataset*.
- *Step* 2: *Compute the original membership matrix* m_1*using FCM*.
- *Step* 3: *Acquire the DFR Matrix* m_2 *based on the equation given below*:

$$m_2(k,i) = max(u_k) / u_{ik}\ ; \quad \forall \text{clusters } i=1,2, \ldots,c, \text{ objects } k=1, 2, \ldots, n$$

Where u_k *is the maximum fuzzy membership value of* k^{th}*objectfor all clusters and*u_{ik} *is the fuzzy membership value of corresponding cluster*.

- *Step* 4: *Compute the membership matrix* m_3 *using RFCM*.
- *Step* 5: *Obtain the membership value and correlate the matrix* m_2, m_3 *results as following procedure*:

1. *Obtain the membership value based on the following condition*:

if $\left(m_3(k,i)^1 < 1 \,\&\, \&m_3(k,i) == 0\right)$

[*lower*(1), *boundary*(<1) *estimation*]

thenm_4 (k, i)= $m_3(k, i)$ ∀*clustersi*=1,2, …, *c*, objects *k*=1,2, …, *n*.

2. *Obtain those elements DFR value from the matrix*m_4.
3. *Formulate the matrix* m_4.

- *Step* 6: *Retrieve the minimum DFR value other than* (0 *and* 1) *in matrix* m_4*and Obtain the FRFLE value fac by*:

fac value=(*minimumvalue* -0.0001)

3.2. RFCM using the FRFLE Value

Algorithm 2 for RFCM using FRFLE is given below:

Algorithm 2:Computing rough fuzzy clusters using FRFLE.

Input: *The Dataset*
Output: *The RFCM Clustering Result*.

- *Step* 1: *Select the dataset*.

- *Step* 2: *Compute the original membership matrix* m_1 *using FCM*.
- *Step* 3: *AcquiretheDFRmatrix*m_2.
- *Step* 4: *Assignthedataobjecttothelowerandboundaryestimationbythefollowingcondition*.
  Ifm_2 (*k*, *i*)>*facthen*m_2 (*k*, *i*)=1 [*lowerestimation*] ∀*clustersi*=1,2, …, *c*, *objectsk*=1,2, …, *n*.
  elsem_2 (*k*, *i*)=*fuzzymembershipvalue*[*boundaryestimation*]
  end.
- *Step* 5: *ObtaintheRFCMClusteringresult.*

4. Exploratory Data Sets

Different synthetic data set and benchmark datasets for the endorsement. Obtaining the RFCM results using FRFLE value with a specific procedure for each dataset is observed.

4.1. Synthetic Data Set

The synthetic data setwhich is shown in Table 1 has been developed for a clear evaluation of FRFLE automation and RFCM result and its underlying cluster structure.

Table 1. Synthetic dataset.

	Attributes	
R1	13	13
R2	14	14
R3	15	15
R4	45	45
R5	46	46
R6	47	47
R7	65	65
R8	64	64
R9	67	67
R10	55	55

4.2. Benchmark Data Sets

Discrete benchmark data sets like: Lenses, wine, balloon, seeds, iris and teaching assistant evaluation form the universal repository were used for the experimental explanation in this paper.

4.2.1. Lenses Dataset

Lenses dataset contain 24 samples with four attributes. The attributes information is as follows: Age of the patient (young, pre-prebyopic, presbyopic), spectacle prescription (myope, hypermetrope), astigmatic (no or yes), tear production rate:Reduced and normal. It also contains three classes: As the patient should be fitted with hard contact lenses, the patient should be fitted with soft contact lenses, not be fitted with contact lenses.

4.2.2. Wine Dataset

Wine dataset has 178 samples with 13 attributes. The attributes information is as follows: Alcohol, malic acid, ash, alcalinity of ash, magnesium,total phenols,flavanoids, nonflavanoid phenols, proanthocyanins, color intensity, hue, OD280/OD315 of diluted wines, proline. It also contains three classes based on the analysis determined by the quantities of 13 constituents present in each type of wine. Figure 4 shows the scatter representation between the alcohol and alkalinity of ash attributes in the wine dataset.

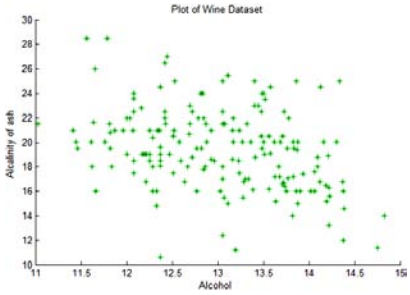

Figure 4.Wine dataset.

4.2.3. Balloons Dataset

Balloons dataset contain 16 instances and 4 attributes. The attribute information such as color (yellow=1, purple=2), size (small=1, large=2), action (stretch=1, dip=2), age (adult=1, child=2) and it also contains two classes: Inflated (T, F) if adult and stretch, then true else false. Figure 5 shows the linear representation of the action and age attributes in the balloons dataset.

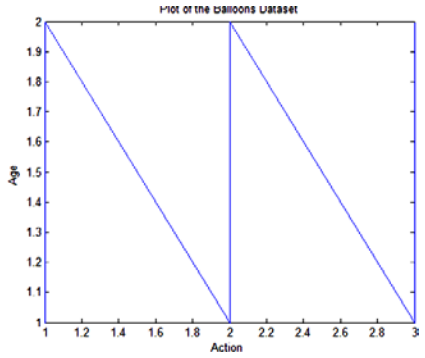

Figure 5.Balloons dataset.

4.2.4. Iris Dataset

Iris dataset has 150 samples in four dimensional measurement spaces. Iris consists of two or three clusters because of the substantial overlap of two of the clusters. It consists of four attributes which includes sepal length in cm, sepal width in cm, petal length in cm and petal width in cm. It consists of three classes such as: Iris setosa, iris versicolour, and iris virginica.

4.2.5. Seeds Dataset

Seeds dataset contain 210 samples with 7 attributes.The attribute details as follows: Seven geometricparameters of wheat kernels were measured like area *A*, perimeter *P*, compactness *C*=4*pi*A/P^2, length of kernel and groove, width of kernel, asymmetrycoefficient, Figure 6 shows the stem plot representation of the area and perimeter attributes in the seeds dataset.

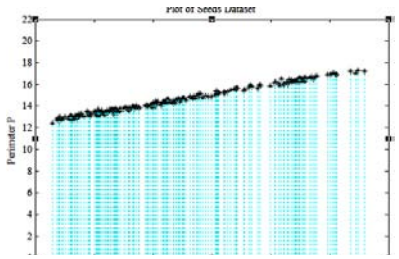

Figure 6. Seeds dataset.

4.2.6. Teaching Assistant Evaluation

It contain 151 samples with 5 attribute which includes whether or not the TA is a native English speaker, course instructor, course, summer or regular semester, class size. It also contains three classes such as low, medium, high.

5. Results and Discussions

This segment gives the detailed strategy about the process of preset FRFLE value and also concedes the RFCM clustering from the FCM result using the FRFLE value for the all above mentioned datasets. For both algorithms a standard fuzzifier value m=2 is used for the membership computation. Initially the dataset is loaded into the MATLAB software. Here, the following synthetic data is considered for the experimental evaluation; it contains 10 objects and 2 attributes.

5.1. Automation of the FRFLE Value

Following Figure 7 manifested the process of automating the FRFLE value.

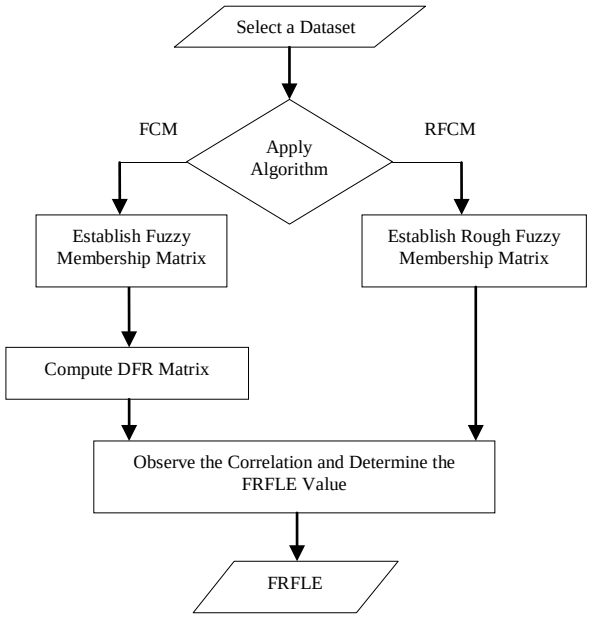

Figure 7.Flow diagram for automation.

4.2.1. FRFLE Automation Results for Synthetic Dataset

For the selected dataset apply the MATLAB standard function FCM to acquire the original membership matrixm_1. For experimental calculation, the total number of cluster is taken as 3, the Table 2 shows the matrixm_1 value.

Table 2.Membership matrix m_1 using FCM.

	C1	C2	C3
R1	0.9987	0.0009	0.0004
R2	1.0000	0.0000	0.0000
R3	0.9986	0.0010	0.0004
R4	0.0038	0.9868	0.0094
R5	0.0008	0.9967	0.0025
R6	0.0000	1.0000	0.0000
R7	0.0000	0.0001	0.9999
R8	0.0003	0.0022	0.9975
R9	0.0017	0.0118	0.9865
R10	0.0226	0.5829	0.3945

The membership matrix m_3 for RFCM algorithm, formulated by Maji and Sankar[10] RFCM algorithm (mention in the above section). Table 3 shows the membership value of sample data set using RFCM algorithm.

Table 3.Membership matrixm_3 using RFCM.

	C1	C2	C3
R1	1.0000	0.0000	0.0000
R2	1.0000	0.0000	0.0000
R3	1.0000	0.0000	0.0000
R4	0.0000	0.0000	1.0000
R5	0.0000	0.0000	1.0000
R6	0.0000	0.0000	1.0000
R7	0.0000	1.0000	0.0000
R8	0.0000	1.0000	0.0000
R9	0.0000	1.0000	0.0000
R10	0.0000	0.4625	0.4475

FRFLE value is the threshold value for DFR. The DFR is used to decide which cluster's characteristics are dominantly present in the object. The DFR matrixm_2 is computed based on the membership matrix $m_1$value as the ratio between the maximum membership value for each object for all clusters and the corresponding membership value of each cluster. Table 4 shows the DFR matrix m_2 value.

Table 4.DFR matrixm_2.

	C1	C2	C3
R1	1.0000	1135.3	2646
R2	1.0000	21427000	5096.8
R3	1.0000	1034.4	2516.3
R4	256.937	1.0000	104.9256
R5	1174.5	1.0000	405.8173
R6	246.937	1.0000	71732
R7	68839	8641.0	1.0000
R8	3850.6	448.7437	1.0000
R9	583.2062	83.6237	1.0000
R10	25.825	1.0000	1.4777

After attaining both matrixes DFR matrixm_2 RFCM membership matrix m_3. Find each element DFR value based on following condition as if the element belongs to the lower region then select the membership values of other clusters, or else if the element belongs to the(two or more)boundary region then select the membership values of other clusters from the RFCM membership

matrix m_3 and compute the comparison matrix m_4 as given in the Table 5.

Table 5.Compatrison matrixm_4.

	C1	C2	C3
R1	0.0000	1135.3	2646
R2	0.0000	21427000	5096.8
R3	0.0000	1034.4	2516.3
R4	256.937	1.0000	0.0000
R5	1174.5	1.0000	0.0000
R6	246.937	1.0000	0.0000
R7	68839	0.0000	1.0000
R8	3850.6	0.0000	1.0000
R9	583.2062	0.0000	1.0000
R10	25.8255	0.0000	0.0000

According to the comparison matrix m_4 the FRFLE value is the minimum value of the object (other than zero and one) which fit in the upper approximation. For above synthetic dataset example the FRFLE value *fac* is obtained as:

$$fac = (25.8255 - 0.0001) = 25.8254$$

5.2. RFCM Result using the FRFLE Value

The RFCM algorithm using the FRFE value process is given in the following flow diagram Figure 8.

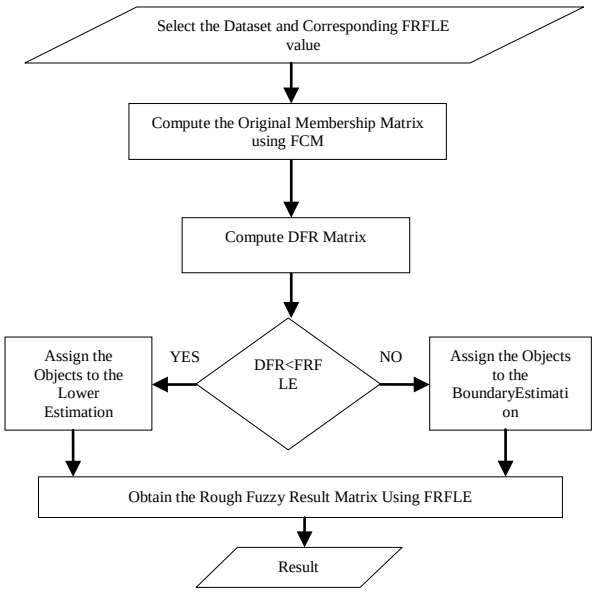

Figure 8. Flow diagram for RFCM result using FRFLE value.

The conceptualization of the RFCMalgorithm using FRFLE has less computation when compared to the traditional RFCM algorithm and also it's very easier to understand. Each data object is clustered based on the fixed lower (1) and fuzzy boundary estimation (membership value), which leads to handle the lack of certainty, ambiguous, and incompleteness in the class definition.

5.2.1. RFCM Result for Synthetic Dataset using FRFLE Value

Initially the dataset and the corresponding FRFLE value for each selected datasets are obtained. For example the FRFLE value of synthetic dataset here used is*fac*= 25.8249. By applying the FCM function in MATAB, the original membership matrix m_1 forthe selected dataset is occurred at the beginning as shown in above Table 2.

After that the DFR matrix $m_2$is also calculated as shown in the previous section Table 5. The RFCM matrixRFCMis formulated using FRFLE value based on the condition (m_2<*fac)* assign the element to the lower (1) or else to the boundary estimation with their membership as shown in Table 6.

Table 6.RFCM result matrixRFCMusing FRFLE.

	C1	C2	C3
R1	1.0000	0.0000	0.0000
R2	1.0000	0.0000	0.0000
R3	1.0000	0.0000	0.0000
R4	0.0000	0.0000	1.0000
R5	0.0000	0.0000	1.0000
R6	0.0000	0.0000	1.0000
R7	0.0000	1.0000	0.0000
R8	0.0000	1.0000	0.0000
R9	0.0000	1.0000	0.0000
R10	0.0000	0.4625	0.7992

The automated FRFLE values for the different benchmark datasets and the count of the objects that belongs to the lower and the boundary estimation of the each cluster result in the datasets has been given in the below Table 7. The membership value can be computed using the FCM algorithm as shown before. Here, in Table 7 C denotes the total number of clusters, L denotes lower and B denotes boundary region elements count. C(1, ..., 5) identifies the clusters individually.

Table 7.Result for various benchmark datasets.

Dataset	**C**	**FRFLE Value**	**Lower Objects Count and Boundary Objects Count Value**				
			C1	**C2**	**C3**	**C4**	**C5**
Lenses	2	3.9123	L(8),B(8)	L(8),B(8)	-	-	-
	3	1.0081	L(8)	L(8)	L(8)	-	-
	4	1.0070	L(7)	L(5)	L(7)	L(5)	-
Balloons	2	2.0044	L(8),B(8)	L(4),B(8)	-	-	-
	3	1.0135	L(4),B(2)	L(8),B(2)	L(8)	-	-
	4	0.9999	L(5)	L(4)	L(3)	L(8)	-
Wine	2	1.0000	L(115), B(16)	L(47), B(16)	-	-	-
	3	1.1857	L(43), B(3)	L(61), B(5)	L(69), B(5)	-	-
	4	1.0440	L(23)	L(33)	L(57), B(2)	L(65), B(2)	-
	5	1.0254	L(20)	L(55)	L(29)	L(27)	L(47)
Teaching Assistant Evaluation	2	1.5910	L(80), B(24)	L(47), B(24)	-	-	-
	3	1.0081	L(64)	L(26)	L(61)	-	-
	4	1.0017	L(42)	L(40)	L(24)	L(46)	-
	5	1.0665	L(26)	L(23)	L(42)	L(26)	L(34)
Iris	2	0.9999	L(90), B(10)	L(50), B(10)	-	-	-
	3	2.8974	L(35), B(22)	L(38), B(21)	L(50), B(3)	-	-
	4	1.1384	L(50)	L(27)	L(29), B(1)	L(43), B(1)	-
	5	1.0000	L(50)	L(11)	L(39)	L(25)	L(25)
Seeds	2	1.5832	L(74), B(20)	L(116), B(20)	-	-	-
	3	1.2932	L(63), B(2)	L(75), B(5)	L(66), B(7)	-	-
	4	1.0067	L(58)	L(71)	L(33)	L(48)	-
	5	1.0861	L(26)	L(47)	L(50)	L(37), B(2)	L(48),B(2)

6. Performance Analysis

In this paper, performance of the RFCM clustering is examined based on the computation time. The RFCM using FRFLE value takes less time for the execution, because the traditional RFCM algorithm has more iteration in calculating the centroid values and the objective function are slow to converge. The comparison chart of algorithm according to their time of execution is shown in Figure 9.

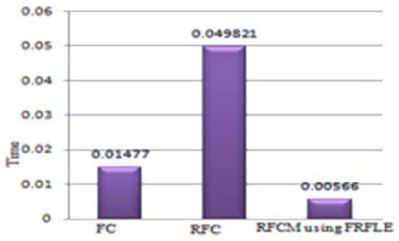

Figure 9.Comparison chart.

7. Conclusions

Thus, the hybridized RFCM reduces the fuzziness of FCM and handles vagueness, incompleteness of RCM efficiently. This paper puts forward the FRFLE to conceptualize the RFCM clustering from the FCM result and also includes the process of FRFLE value automation for various benchmark datasets which efficiently overcomes the general RFCM algorithm problems such as high time complexity, more computation.

FRFLE value for the fresh data set can be computed using original RFCM, then data can be clustered using obtained FRFLE value. Only for the first time this will be happened. Once FRFLE is computed then data can be clustered directly by using FRFLE by using less time. For the first time this process will be taking more time. Then, for the subsequent times the speed for obtaining rough fuzzy clusters will be high.

Even though the advanced scheme of automating the FRFLE is well organized for synthetic and benchmark datasets to built the RFCM algorithm, the idea can also be appeal to other unsupervised classification issues.

References

[1] Gordon A.,Classification Monographs on Statistics and Applied Probability,Amazon, 1981.

[2] Jensen R. and Shen Q., "Fuzzy-Rough Attribute Reduction with Application to Web Categorization," *Fuzzy Sets and Systems*, vol. 141, no. 3, pp. 469-485, 2004.

[3] Joshi M., Lingras P., and Rao C., "Correlating Fuzzy and Rough Clustering,"*Journal FundamentaInformaticae-Rough Sets and Knowledge Technology*, vol. 115, no. 2-3, pp.233-246, 2012.

[4] Lingras P., "Evolutionary Rough K-Means Clustering,"*Rough Sets and Knowledge Technology*, vol. 5589, pp. 68-75, 2009.

[5] Mitra S., and Pradipta M., "A New Rough-Fuzzy Clustering Algorithm and Its Applications,"*in Proceedings of 2nd International Conference on Soft Computing for Problem Solving*, India, pp. 1245-1251, 2012.

[6] Mitra S., Banda H., and Pedrycz W., "Rough-Fuzzy Collabrative Clustering," *IEEE transactions on systems, man and cybernetics part cybernetics*, vol. 36, no. 4, pp. 795-803, 2006.

[7] Murugappan I. and Vasudev M., "PCFA: Mining of Projected Clusters in High Dimensional Data Using Modified FCM Algorithm," *The International Arab Journal of Information Technology*, vol.11, no. 2, pp. 168-177, 2014.

[8] Pradipta M. and Sankar K., "RFCM: A Hybrid Clustering Algorithm using Rough and Fuzzy Sets,"*Journal FundamentaInformaticae*, vol. 80, no. 4, pp. 475-496, 2008.

[9] Pradipta M. and Sankar K., "Rough-Fuzzy C-Medoids Algorithm and Selection of Bio-Basis for Amino Acid Sequence Analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 6, pp. 859-872, 2007.

[10] Pratipta M. and Sankar K., Rough Fuzzy Image Analysis, CRCnetBASE, 2010.

[11] PradiptaM. and Sankar K., "Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices,"*IEEE Transactions on Systems, Man and Cybernetics*, vol.37,no 6., pp. 1529-1540, 2007.

[12] PratiptaM. and Sushmita P., "Robust Rough-Fuzzy C-Means Algorithm: Design and Applications in Coding and Non-coding RNA Expression Data Clustering,"*Journal FundamentaInformaticae-Cognitive Informatics and Computational Intelligence: Theory and Applications*, vol.124, no. 1-2, pp.153-174.

[13] PratiptaM. and Sushmita P., Rough-Fuzzy Clustering for Grouping Functionally Similar Genes from Microarray Data, IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol.10, no. 2, pp.286-299, 2012.

[14] Pal S.,"Case Generation Using Rough Sets With Fuzzy Representation," *IEEE Transaction on Knowledge and Data Engineering*,vol.16, no. 3, pp. 293-300, 2004.

[15] Pawlak. Z., "Rough sets,"*International Journal of Information Computer Science*,vol. 11,no. 5, pp. 145-172, 1982

[16] Peters G., "Some Refinements of Rough k-Means Clustering,"*Pattern Recognition*, vol. 39, no. 8, pp. 1481-1491, 2006.

[17] Peters G,Crespo F., and Weber R., "Soft Clustering-Fuzzy and Rough Approaches and their Extensions and Derivatives,"*International Journal of Approximate Reasoning*, vol.54, no.2,pp. 307-322, 2013.

[18] Revathy S. and Parvathavarthini B., "Comparison of FCM and RFCM on Iris Data Set Using Matlab,"*in Proceeding of International Conference on Computer Networks and Information Technology*, Bangkok, 2012.

[19] RevathyS. and Parvathavarthini B., "Integrating Rough Clustering with Fuzzy sets,"*in Proceeding of International Conference on Sustainable Energy and Intelligent Systems*,India, pp. 865-869, 2011.

[20] Wang W. and Zhang Y., "On Fuzzy Cluster Validity Indices,"*Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2095-2117, 2007.

[21] Witcha C., Abdul-Hanan A., Mohd N., Siriporn C., and Surat S., "A Rough-Fuzzy Hybrid Algorithm for Computer Intrusion Detection," *TheInternational Arab journal of Information Technology*, vol. 4, no. 3, pp. 247-254.

[22] Xu R., "Survey of Clustering Algorithms,"*IEEE Trans on Neural Networks*, vol.16, no.3, pp. 645-678, 2005.

RevathySubramanion has received B.E in Computer Science and Engineering from ArulmiguKalasalingam College of Engineering and ME from Sathayabama University. She is doing research in data mining in Sathyabama University. Her research interest includes Data Clustering, Decision Theory and Bayesian classifiers.

Parvathavarthini Balasubramanianreceived MSc and MPhil degree in 1988 and 1989 respectively. She received MBA and ME degree in 1998 and 2000 respectively. She received PhD degree in 2008. She is working as Head and Professor of the department of Master of Computer Applications at St.Joseph's College of Engineering. Her research includes Computer Networks, Security, multimedia applications and Graphics. She has published fifty seven research papers in National/International Conferences and International Journals.

ShajunishaNoordeen received the B.E Information Technology from AvinashilingamUniversity, Coimbatore. She is currently pursuing Post Graduate in MTech Information Technology at Sathyabama University, Chennai. Her research interests include data mining and rough set theory.