Improvement of Fuzzy C-Means Algorithm by using ACO

HARISHBABU.KALIDASU[1], Dr. GUDIPATI MURALI[2]

[1]Research Scholar, [2]Research Supervisor
Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

ABSTRACT

Clustering methods works towards a set of multiple objects under clusters, so the objects in a particular cluster have a highest degree of similarity, however objects belongs to different clusters have a highest degree of dissimilarity. These type of methods are not only major tasks to expose basic structures of given set of data, but favourable tools expose the relations of input output complex system. FCM (Fuzzy C Means) is most used fuzzy cluster algorithms in real-time applications. Whereas the two major restrictions are there in FCM method. The first restriction is default number of clusters stated in-advance. The second one is to insert in minimal solutions. Here we proposed a new algorithm to enhance the clusters with FCM clustering. This algorithm tested with medical domain, results will shows the post processing refining or clarifying the clusters and enhance the quality of cluster. In this chapter presents a new algorithm to enhance the Fuzzy C Means. In this FCM algorithm, applied ant colony optimization algorithm to improve the quality of refining cluster.

Keywords: clustering, ant colony optimization, refining

1. STANDARD FUZZY C-MEANS CLUSTERING

The FCM algorithm is the most used methods in Pattern Recognition [M. Sato, Y. Sato and L. Jain]. This is based on the minimization of below objective function for successfully bring out classifications.

$$F(W, Z) = \sum_{j=1}^{c} \sum_{i=1}^{n} (\mu_{ji})^m \| x_i - v_j \|^2$$

F(W, Z) is square error cluster criteria and the solutions of minimization (1) are less square error at rest points F(W, Z). The following expression X = {x1, x2, x3, x4,….,xn} is gathering some information where n is the number of data points and V = {v1, v2, v3, v4,…vc) is a group of cluster centers in data set X where c is the collection of clusters. μ_{ji} group of data Xi to clustering centre Vj. While μ_{ji} satisfying the following constraints:

$$\mu_{ji} \in [1,0], \forall \iota=1....n, \forall j=1....c$$

$$\sum_{j=1}^{n} \mu_{ji} = 1., \forall \iota = 1... n.,$$

Where F= (μ_{ji}) c*n is fuzzy partitioning matrix ||Xi – Vj|| where Xi, Vj representing distance between Euclidean principle, m is the parameter represents "fuzziness index" and it is used to manage the fuzziness belonging to each data in the range of parameter m $\in [1, \propto]$. In this experiment the parameter value chosen as m=2.0. Even though there is no theory basis of prime selection of parameter 'm', then it has been selected because the values has been applied commonly. The Fuzzy C Means algorithm set out, for example it performs the following steps:

1. First assign the initial cluster centres as V ={v1, v2, v3, v4, v5…Vc}, otherwise assign the matrix μ_{ji} with unspecific values, check the values whether it satisfies above conditions and calculate cluster centres.

2. Compute Fuzzy matrix μ_{ji} by using

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

where, d_{ji} = || Xi -Vi||, for all i= 1 to n., for all j= 1 to c

3. Calculate Fuzzy centres Vj by using the following equation

$$Vj = \frac{\sum_{i=1}^{n} (\mu_{ij})^m x_i}{\sum_{i=1}^{n} (\mu_{ij})^m} \quad \textit{for all } j = 1 \textit{ to } c.$$

4. Now repeat step 2 and step 3 up to minimal value J achieved.

5. Lastly, defuzzification is mandatory to apply each and every data point to a particular cluster. It means by grouping data points to cluster in which the degree of members is maximum.

2. ACO BASED CLUSTER REFINEMENT

A similar refinement approach used here to help the quality of cluster from Fuzzy C means. In this method three ants used to purify the clusters. These are allowed for a casual walk on the particular clusters. When it crossing cluster it picks an item from the refining cluster and to fall into one cluster to another cluster. With drop probability compare the quality of clusters and calculate the cluster valid index. The Partition Entropy and Partition Co-efficient PCoE & PEn are defined as (J. C. Bezdek) :

$$PCoE = 1/\mathrm{n} \sum_{j=1}^{c} \sum_{i=1}^{n} (\mu_{ji})^{2}$$

$$PEn = -1/\mathrm{n} \sum_{j=1}^{c} \sum_{i=1}^{n} (\mu_{ji}) \log_a (\mu_{ji})$$

PCoE and PEn used to calculate fuzziness of partition fuzzy matrix, the underneath fuzziness partition is the larger than PCoE value (or smaller than PEn value). From these partition validity index we can measured drop probability. Here it is:

$$P_d = PEn / PCoE$$

Now the drop probability has minor values than preceding iteration, then drop makes permanent and then next iteration continued with changed cluster index. Apart from the next iteration continued with older cluster indexes'. Then the random walk is repeated for 'N' times. The following steps shown that refine / purity algorithm improves the quality cluster. See the following algorithm:

Step 1: First assign the cluster centre values V as : V = {v1, v2, v3, v4, v5…, Vc}

otherwise assign the matrix μ_{ji} with unspecific values, check the values

whether it satisfies above conditions and calculate cluster centres.

Step 2: Compute Fuzzy matrix μ_{ji}

Step 3: Calculate Fuzzy centres Vj

Step 4: Now repeat step 2 and step 3 up to minimal value J achieved.

Step 5: Lastly, defuzzification is mandatory to apply each and every data point to a particular cluster.

Step 6: Now see the ANT based refinement:

i. Load the cluster values acquired from improved Fuzzy C Means.

ii. For i=1…n then do

   a. Let ants move for random walk and pick items

   b. Leave the items into another cluster.

   c. Although check the quality improved or not by calculating PEn and PCoE

   d. If any improvement in the clusters then drop the items permanently.

iii. Repeat the steps.

3. RESULTS

The cluster validity is a concept that used to assess the clustering quality results. Although the no. of clusters is not knowing in advance to start an algorithm. The cluster validity index used to find out the best number of clusters [J. C. Bezdek]. These index values successfully bring out all the possible clusters with cluster validity index and the best number of clusters could determine to select the least values of the index. A large number of clusters validity indexes have been developed previously.

In these fuzzy methods many people use only membership values of Fuzzy clustering data, so the Partition Co-Efficient (PCoE) and Partition Entropy (PEn). In favour of this type of indexes it is easy to calculate but it is useful only for the minimum number of well good separated clusters. Moreover, the deficiency of extending connection to the properties of geometrical data. While to get the better results Beni and Xie defined " The validity index measured the compactness and separate clusters [G. Beni, X.L. Xie]. So, Beni-Xie index chosen as measuring the cluster validity and it is able to investigate the perfect or exact number of clusters in different experiments as mentioned in [Bezdek and NR Pal]. Beni-Xie validity index is the combination of two functions. At first calculate the data compactness in same cluster and second calculates the data separateness in other clusters.

Let Z represents the final validity index: π be the compactness and 'w' be the fuzzy separation 'c' partition of group data set. Beni-Xie validity index can be shown as:

$$Z = \pi / w$$

Where $$\pi = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ji}^{2} \; || Xi\text{-}Vj||^2 \; / \, n$$

and $$Z = (d_{min})^2$$

Where d_{min} is the shortest distance between cluster centre stated by min_{ji} $\|Vi\text{-}Vj\|$. The least values of π recommends that clusters are very compact and maximum values of Z recommends clusters are separated without any problem. Thus a minimum Z send back that the clusters have maximum separation from one other and they are very compact to each other. The table 5.1 shows the results carried out from different datasets. The estimation similarity study on various measures shown in the graph 1.1

Table 1.1 Refine Fuzzy C Means Clustering Performance Analysis

	K-Means	**K-means refined with ACO**	**Fuzzy C-Means**	**Refined FCM with ACO**
Number of classes	02	02	02	02
Number of clusters	02	02	02	02
PCoE	--	--	0.82	0.98
PEn	--	--	0.29	0.03
Beni-Xie Indexes	--	--	3.38	2.95
Dunnn Index	20.2	21.5	21.7	22.5
DB Index	0.13	0.06	0.12	0.069

(a) WBC Dataset

	K-Means	**K-means refined with ACO**	**Fuzzy C-Means**	**Refined FCM with ACO**
Number of classes	02	02	02	02
Number of clusters	02	02	02	02
PCoE	--	--	0.79	0.91
PEn	--	--	0.19	0.02
Beni-Xie Indexes	--	--	3.21	2.22
Dunnn Index	21.3	21.9	22.2	22.9
DB Index	0.095	0.081	0.096	0.093

(b) Lung Cancer Dataset

	K-Means	**K-means refined with ACO**	**Fuzzy C-Means**	**Refined FCM with ACO**
Number of classes	03	03	03	03
Number of clusters	03	03	03	03
PCoE	--	--	0.89	0.97
PEn	--	--	0.25	0.12
Beni-Xie Indexes	--	--	3.78	2.23
Dunnn Index	26.07	26.2	26.03	26.37
DB Index	0.107	0.080	0.085	0.074

(c) Thyroid disease Dataset

	K-Means	**K-means refined with ACO**	**Fuzzy C-Means**	**Refined FCM with ACO**
Number of classes	02	02	02	02
Number of clusters	02	02	02	02
PCoE	--	--	0.81	0.95
PEn	--	--	0.19	0.03
Beni-Xie Indexes	--	--	3.18	2.56
Dunnn Index	24.5	24.9	25.2	25.4
DB Index	0.119	0.079	0.11	0.10

(d) Hepatitis Datase

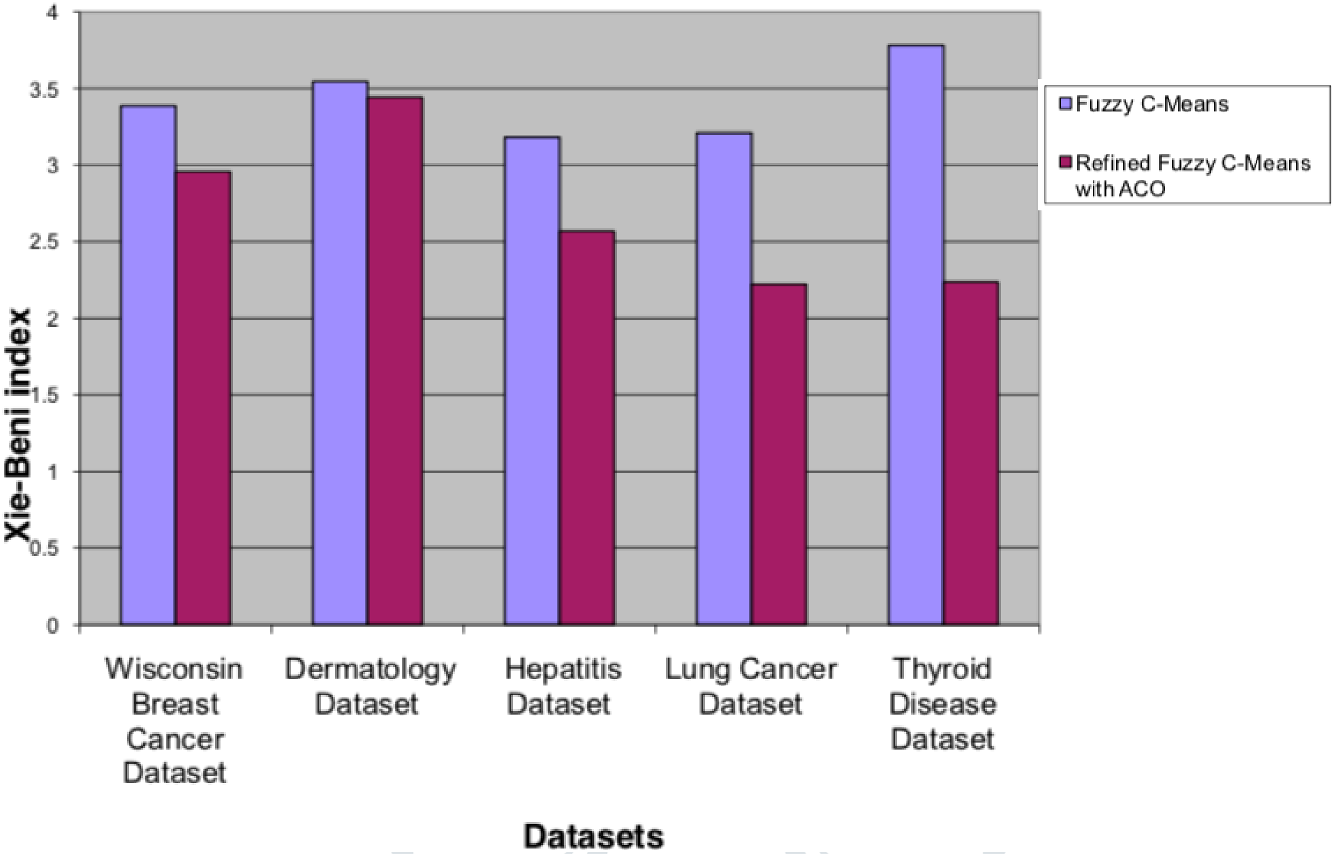

Figure 1.1 (a) Comparison of Xie-Beni Index values on different clustering methods

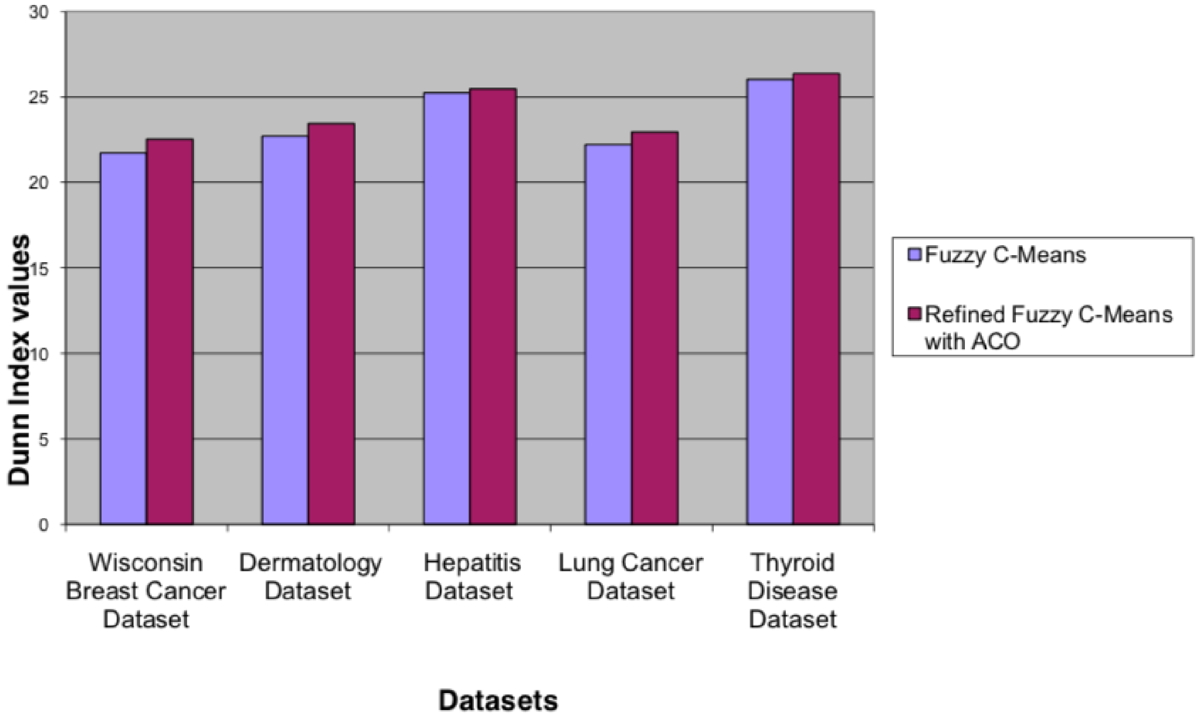

Figure 1.1 (b) Comparison of DunnnIndex values on different clustering methods

Here we compared the performances of proposed method with Refined k-means, Fuzzy C-Means and k-means. The performance of this evaluation based on the validity measures as we are discussed earlier. The continual performance between Refined Fuzzy C-Means (FCM) and Standard Fuzzy C-Means with ACO is compared using PCoE, PEn and Beni-Xie Index values. As we mentioned earlier again the performance is compared with the clustering approach by using Dunnn index and DB index parameters. While this recognition, the largest Dunnn Index and the smallest DB index shows the higher level performances of the Refined FCM with ACO methods. Although here the algorithms are executed 10 times for each dataset and presented the average values of the validity measures.

4. SUMMARY

The cluster analysis is the major tasks in various research areas. Although it presents under different contexts in different names, apart of partitioning in graph theory, taxonomy in biology and pattern recognition. The aim of clustering is to establish and takeout outstanding groups in principal data. It is based on cluster criteria of data that are grouped into data points in a cluster and are more similar to each and other points in different clusters. However the clustering applied in various fields as a numerous clustering techniques and proposed algorithms are available in literature. In this chapter, an ant colony algorithm presented to improvement of cluster from Fuzzy C-Means cluster algorithm. Thus performance compared with different clustering algorithms like k-Means, Standard Fuzzy C-means, Refined k-Means; results shows that proposed method performs better than standard method.

REFERENCES

[1] X. L. Xie and G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 13, pp. 841-847

[2] NR. Pal and JC. Bezdek, ―On cluster validity for the fuzzy c-means model‖, IEEE Trans. Fuzzy Systs., Vol. 3, pp. 370-379,

[3] Deneubourg J.L., Pasteels J.M., Verhaeghe J.C., Probabilistic behavior in Ants: a strategy of errors, Journal of Theoretical Biology, 259.

[4] Dorigo M., Optimization, Learning and Natura Algorithms (In Italia), PhD thesis Dipartimento di Elettronica e Informazione, Politecnico di Milano, IT,

[5] J.C. Dunn, Well separated clusters and optimal fuzzy partitions, Journal of Cybernetics, 4: 95-104.

[6] D.L. Davies and D.W. Bouldin, A Cluster Separation Measure, IEEE Transactions on Pattern Analysis and Machine Learning, 1(2),

[7] J. C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press.

[8] MR Rezaee, BPF. Leieveldt, and JHC. Reiber, A New Cluster Validity Index for the Fuzzy C-Means, Pattern Recognition Letters, Vol. 19, Elsevier, pp. 237-246, 1998.