A Comparative Study of Color Image Segmentation Using Hard, Fuzzy,Rough Set Based Clustering Techniques

Venkateswara Reddy Eluri, Dr.E.S.Reddy
Research scholar, Acharya Nagarjuna University, Guntur.
evr.eluri@gmail.com
Prinicipal,University College of Engineering, ANU, Guntur.
edra_67@yahoo.com

ABSTRACT

Image segmentation is the process of subdividing an image into its constituent parts and extracting these parts of interest, which are the objects. Colour image segmentation emerges as a new area of research. It can solve many contemporary problems in medical imaging, mining and mineral imaging, bioinformatics, and material sciences. Naturally, color image segmentation demands well defined borders of different objects in an image. So, there is a fundamental demand of accuracy. The segmented regions or components should not be further away from the true object than one or a few pixels. So, there is a need for improved image segmentation technique that can segment different components precisely. Image data may have corrupted values due to the usual limitations or artifacts of imaging devices. Noisy data, data sparsity, and high dimensionality of data create difficulties in image pixel clustering. As a result, image pixel clustering becomes a harder problem than other form of data.

Taking into account all the above considerations we propose an unsupervised image segmentation method using Rough-Fuzzy C-Mean a hybrid model for segmenting RGB image by reducing cluster centers using rough sets and Fuzzy C-Means Method, and also compare the effectiveness of the clustering methods such as Hard C Means (HCM), Fuzzy C Means (FCM), Fuzzy K Means (FKM), Rough C Means (RCM) with cluster validity index such as DB Index, XB Index and Dunn Index. A good clustering procedure should make the value of DB index as low as possible, for Dunn Index high value, and for XB Index low value.

Indexing terms/Keywords

Rough Set, Image Segmentation, Fuzzy C-Means, Rough Fuzzy C Means Algorithm.

SUBJECT CLASSIFICATION

Computer Science, Digital Image Processing, Computer Vision, Graphics

TYPE (METHOD/APPROACH)

Ali Salem Bin Samma [5] proposed an adaptive k-means clustering algorithm for image segmentation. In this he tries to develop K-means algorithm to obtain high performance an efficiency. This method proposes initialization step in K-mean algorithm. In [6], Chih-Tang Chang and Jim Z.C.LAI suggested a fuzzy k-means clustering algorithm using the cluster center displacement between successive iterative processes to reduce the computational complexity of conventional fuzzy k-means clustering algorithm.

1. INTRODUCTION

Clustering analysis has been an emerging research issue in data mining due its variety of applications. With the advent of many data clustering algorithms in the recent few years and its extensive use in wide variety of applications, including image processing, computational biology, mobile communication, medicine and economics, has lead to the popularity of this algorithms. Main problem with the data clustering algorithms is that it cannot be standardized. Algorithm developed may give best result with one type of data set but may fail or give poor result with data set of other types. Although there has been many attempts for standardizing the algorithms which can perform well in all case of scenarios but till now no major accomplishment has been achieved. Many clustering algorithms have been proposed so far. However, each algorithm has its own merits and demerits and cannot work for all real situations. Before exploring various clustering algorithms in detail let's have a brief overview about what is clustering. Clustering is a process which partitions a given data set into homogeneous groups based on given features such that similar objects are kept in a group whereas dissimilar objects are in different groups. It is the most important unsupervised learning problem.

The goal of clustering is to group similar objects in one cluster and dissimilar objects in different clusters. Probably the most frequently used clustering algorithm is the classic k-means with applications in virtually any real life domain. The k-means clustering is characterized by non-overlapping, clearly separated ("crisp") clusters with bivalent memberships: an object either belongs to or does not belong to a cluster. Another example for the need for overlapping clusters is as follows. For diagnosis of a flu a general practitioner normally requires the temperature, of a patient and whether she has headache and cough (example inspired by Grzymała-Busse [2]). In

such a case classic k-means with a bivalent classification (Flu=yes or Flu=no) is fully sufficient. However, for some special cases these features might not be sufficient to decide if theses patients suffer from flu or not (e.g. further, more detailed diagnoses are required). In such cases rough clustering is an appropriate method since it separates the objects that are definite members of a cluster from the objects that are only possible members of a cluster. In our case most patients are assigned to the clusters Flu=yes or Flu=no. But some, the tricky ones, are labeled as the "we do not know yet" patients.

In this paper, we propose a hybrid algorithm, termed as rough-fuzzy c-means, based on rough sets and fuzzy sets. While the membership function of fuzzy sets enables efficient handling of overlapping partitions, the concept of lower and upper approximations of rough sets deals with uncertainty, vagueness, and incompleteness in class definition. Each partition is represented by a set of three parameters, namely, a cluster prototype (centroid), a crisp lower approximation, and a fuzzy boundary. The lower approximation influences the fuzziness of the final partition. The cluster prototype (centroid) depends on the weighting average of the crisp lower approximation and fuzzy boundary. Several quantitative measures are introduced based on rough sets to evaluate the performance of the proposed algorithm. The effectiveness of the proposed algorithm, along with a comparison with crisp, fuzzy, possibilistic, and rough c-means, has been demonstrated on a set of benchmark data sets.

The structure of the rest of this paper is as follows. Section 2 describes the Hard C Means (K-Means) Traditional Clustering algorithm. In Section 3, we describe fuzzy C Means Clustering algorithm, Section 4 describes , proposed rough-fuzzy c-means(RFCM) clustering algorithm based on the theory of rough sets and fuzzy c-means. Several quantitative performance measures are introduced in Section 5 to evaluate the quality of the proposed algorithm. Concluding remarks are given in Section 6.

2. k-means clustering algorithm

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriority. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other.

Algorithmic steps for k-means clustering

Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$ be the set of data points and $P = \{p_1, p_2, \ldots, p_c\}$ be the set of centers.

1) Randomly select 'c' cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..

4) Recalculate the new cluster center using:

$$P_i = (1/C_i)\sum_{j=1}^{C_i} x_i$$

Where, 'c_i' represents the number of data points in i^{th} cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3).

3. Fuzzy C-Means Algorithm (FCM)

The algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center clearly summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right]^{\frac{2}{m-1}}}$$

Algorithm steps for Fuzzy C-Means Clustering

Step 1: Initialize U=[μ_{ij}] matrix, u[0]

Step 2: At k-step : calculate the center vectors c(k) =[c_j] with U(K).

$$c_j = \frac{\sum_{i=1}^{n}\mu_{ij}^m x_i}{\sum_{i=1}^{n}\mu_{ij}^m}$$

Step 3: Update U(K), U(k+1)

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right]^{\frac{2}{m-1}}}$$

Step 4: if

$$\|U(K+1) - U(K)\| < \text{Threshold then STOP; otherwise return to step2}$$

4. Rough-set theory

Pawlak introduced rough set theory as a new framework for dealing with imperfect knowledge. Rough set theory provides a methodology for addressing the problem of relevant feature selection, by selecting a set of information rich features from a data set that retains the semantics of the original data and requires no human inputs unlike statistical approaches. It is often possible to arrive at a minimal feature set (called reduct in rough set theory) that can be used for data analysis tasks such as classification and clustering [Lingras and West, 2004], [Mitra et al., 2006]. When feature selection approaches based on rough sets are combined with an intelligent classification system like those based on fuzzy systems or neural networks, they retain the descriptive power of the overall classifier and result in simplified system structure which enhances the understandability of the resultant system [Shen, 2007].

The core idea is to separate discernible from indiscernible objects and to assign objects to lower and upper approximations of a set $(\underline{B}(X), \overline{B}(X))$. In rough clustering we are not considering all the properties of the rough sets [4].

However, the family of upper and lower approximations are required to follow some of the basic rough set properties such as

1. An object v can be part of at most one lower approximation. This implies that any two lower approximations do not overlap.

2. An object v that is member of a lower approximation of a set is also part of its upper approximation $(v \in \underline{B}(X_i) \to v \in \overline{B}(X_i))$. This implies that a lower approximation of a set is a subset of its corresponding upper approximation $(\underline{B}(X_i) \subseteq \overline{B}(X_i))$

3. If an object v is not part of any lower approximation it belongs to two or more upper approximations. This implies that an object cannot belong to only a single boundary region.

Rough Fuzzy C Means (RFCM)

1. Choose initial means M_i , $1 \le i \le k$ for the k clusters.

2. Compute u_{ij} by equation 3.9 for k clusters and n data objects.

3. Assign each data point x_j to the lower approximation $\underline{L}U_i$ or upper approximation $\overline{L}U_i$, $\overline{L}U_{i'}$ of cluster pairs U_i , $U_{i'}$ by computing the difference membership $u_{ij} - u_{i'j}$.Let u_{ij} be maximum and $u_{i'j}$ be the next to maximum. If $u_{ij} - u_{i'j}$ is less than some threshold then $x_j \in \overline{L}U_i$ and $x_j \in \overline{L}U_{i'}$ and x_j can't be a member of any lower approximation else $x_j \in \underline{L}U_i$ such that the membership u_{ij} is maximum over the k clusters.

4. Compute the new mean M_i for each cluster as,

$$M_i = \begin{cases} \dfrac{\sum_{x_j \in \overline{L}U_i - \underline{L}U_i} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in \overline{L}U_i - \underline{L}U_i} \mu_{ij}^{\alpha}} \\ W_{low} \times \dfrac{\sum_{x_j \in \underline{L}U_i} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in \underline{L}U_i} \mu_{ij}^{\alpha}} + W_{up} \times \dfrac{\sum_{x_j \in \overline{L}U_i - \underline{L}U_i} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in \overline{L}U_i - \underline{L}U_i} \mu_{ij}^{\alpha}} \\ \dfrac{\sum_{x_j \in \underline{L}U_i} \mu_{ij}^{\alpha} x_j}{\sum_{x_j \in \underline{L}U_i} \mu_{ij}^{\alpha}} \end{cases}$$

5. Repeat steps 2-5 until convergence, i.e., there are no more new Assignments, or upper limit on the number of iterations are reached.

5. Results & Discussion

The performance of proposed RFCM algorithm is compared extensively with that of different clustering algorithms such as Hard C Means (K-Means), Fuzzy C –Means (FCM), Fuzzy K-Means and Rough C-Means. All the methods are implemented in Matlab 9.0 and run in windows environment and also it can be allowed in LINUX environment.

To analyze the performance of the proposed method, the experimentation has been done in two parts In the first part, we present the results on segmentation of color images which are taken from Corel Dataset and Berkeley Segmentation Dataset (BSDS). In the second part we have used Iris, Wine data sets from UCI Machine Learning repository. The major metrics for evaluating the performance of different algorithm are Davies-Bouldin (DB) Index, Xie-Beni (XB) Index and Dunn Index, which are described next.

Table 1.1 Performance of K-means, FKM, FCM, RCM, RFCM on Image-86000 BSDS

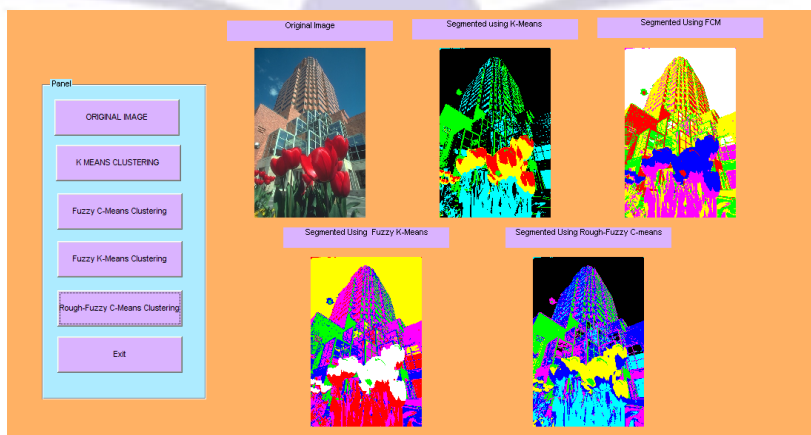|	DB Index	Dunn Index	XB Index
K-means	1.85	1.04	1.76
Fuzzy K Means	1.63	1.54	1.54
Fuzzy C Means	1.58	2.09	1.23
Rough C Means	1.22	2.34	0.86
Rough-Fuzzy C Means	0.96	3.17	0.64

Figure 1.1 IMAGE-86000 (BSDS) : Original and segmented versions of K-means, FCM, FKM, RFCM

Consider Fig. 1.1 as an example, which represents a color image of BSDS (IMAGE-86000) along with the segmented images obtained using different clustering algorithms such as K-means, Fuzzy C Means(FCM), Fuzzy K Means (FKM)and Rough Fuzzy C Means (RFCM). Each Image is of size 321x 481. Table 7.1 depicts the values of Davies-Bouldin(DB) index, Dunn index, and Xie-Beni(XB) index of K-Means, FKM, FCM, RCM, RFCM clustering algorithms. The results reported here with respect to DB, XB and Dunn index confirm that RFCM achieve their best results for c=6.
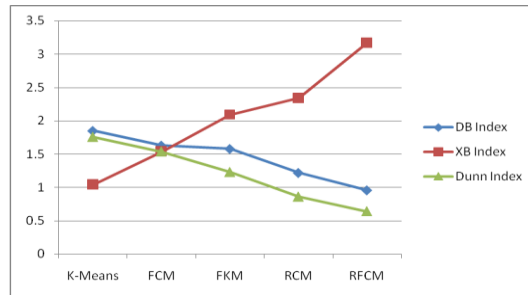

Fig 1.2 Cluster Validity Indexes of Different Clustering Techniques for RGB IMAGE-86000

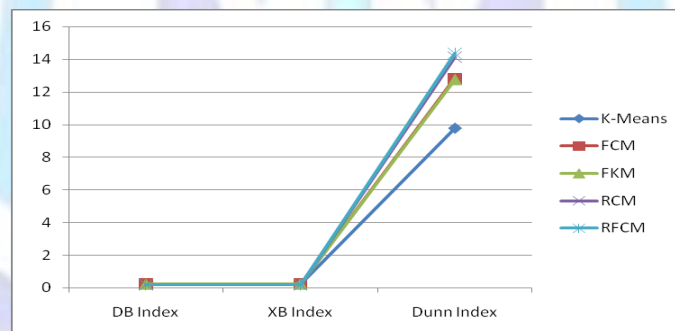| Algorithms	DB Index	XB Index	Dunn Index
K-Means	0.23116	0.23387	9.78226
Fuzzy C Means	0.21578	0.22659	12.82379
Fuzzy K Means	0.22659	0.23389	12.77125
Rough –C Means	0.19346	0.19879	14.13384
Rough-Fuzzy C Means	0.17236	0.18457	14.34478

Table 1.2 Results on IRIS Data Set (C=2)

Fig 1.3 Cluster Validity Indexes of Different Clustering Techniques for Real data set: Iris

6.Conclusion

The developed approach contributes a hybrid methodology, which integrates judiciously rough sets and fuzzy c-means algorithm. This formulation is geared towards maximizing the utility of both rough sets and fuzzy sets with respect to knowledge discovery tasks. Rough fuzzy c-means algorithm is an extension of fuzzy c means algorithm, which incorporates fuzzy membership of datapoints in each cluster. In each iteration, cluster centers are updated and data point is assigned to lower approximations or upper approximation of a cluster. This process is repeated until convergence criterion is met. Finally, the effectiveness of the proposed algorithm is demonstrated, along with a comparision with other related algorithms, on several real data sets from UCI Machine Learning Repository and Images from Corel, BSDS data set.

Although our methodology of integrating rough sets, fuzzy sets, and c-means algorithm has been efficiently demonstrated for real data sets, along with the segmentation of color images from BSDS data set. The concept can be applied to the other unsupervised classification problems. Some of the cluster validity indices such as DB index, XB index and Dunn Index for evaluating the quality of the proposed algorithm and compared to other unsupervised Clustering algorithms such as Hard C Means, Fuzzy C Means, and Rough C Means clustering algorithms.

REFERENCES

[1] A. Papoulis, Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill,   1984.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", J. R. Statist. Soc., Series B., no.1, pp. 1-38,1977.

[3] C. Chesnaud, V. Boulet, and P. Refregier, "Statistical Region Snake-Based Segmentation Adapted to Different Physical Noise Models", IEEE Trans.Pattern Anal. Machine Intell., vol. 21, pp. 1145-1157, Nov. 1999.

[4] M., Mitra, S., Pal, S. K.: Rough-Fuzzy MLP: Knowledge Encoding and Classification, IEEE Transactions on Neural Networks, 9(6), 1998, 1203–1216.

[5] P. Dhanalakshmi& T.Kanimozhi : Automatic Segmentation of Brain Tumor using K-Means Clustering and its Area Calculation,IEEE Transsactions, 2013, 2278-2285.

[6] Asharaf, S., Murty,M. N.: A Rough Fuzzy Approach toWeb Usage Categorization, Fuzzy Sets and Systems, 2004, 119–129.

 [7] Barni, M., Cappellini, V., Mecocci, A.: Comments on A Possibilistic Approach to Clustering, IEEE Transactions on Fuzzy Systems, 4(3), 1996, 393–396.

[8] Bezdek, J. C.: Pattern Recognition with Fuzzy Objective Function Algorithm, New York:Plenum, 1981.

E. Venkateswara Reddy doing his PhD from Acharya Nagarjuna University and he done his M.Tech from Satyabhama University India in 2003. His research interests are in Image processing and Data Mining He is a life member for Computer Society of India (CSI). Currently he is working as a Head of Department for Computer Science and Engineering, Amara Institute of Engineering &Tech, Guntur. He has authored a good number of research papers published in International Journals and Conference Proceedings like IEEE ,Springer, IJCA and Global Journals Inc, USA.

Dr. E. S. Reddy Principal, University College of Engineering, Acharya Nagarjuna University, Nagarjuna Nagar,Guntur,AP , Received PhD from Acharya Nagarjuna University, he done his M.S from BITS,Pilani in 1997 and M.Tech from Visveswaraya Technological University, Karnataka,India in 2000. His research interests are in Image processing and Signal Processing He is a life member for Computer Society of India (CSI), IETE, ISTE. He has authored a good number of research papers published in International Journals and Conference Proceedings like IEEE ,Springer, Elsevier etc,.