International Journal of Engineering Management Science

VOLUME-1 ISSUE-3
ISSN : 2799-18

A FUZZY BASED CLUSTER IMPROVEMENT ANALYSIS BY USING CLUSTERING WITH NEUTROSOPHIC LOGIC

Ms.P.Sowjanya
Malla Reddy Engineering College (Autonomous) Department of Electronics & Communication Engineering

ABSTRACT

Fuzzy C-means has been utilized successfully in a wide range of applications, extending from the clustering capability of the K-means to datasets that are uncertain, vague and otherwise are hard to be clustered. In cluster analysis, certain features of a given data set may exhibit higher relevance in comparison to others. To address this issue, Feature-Weighted Fuzzy C-Means approaches have emerged in recent years. However, there are certain deficiencies in the existing methods, e.g., the elements in a feature-weight vector cannot be adaptively adjusted during the training phase, and the update formulas of a feature-weight vector cannot be derived analytically. In this study, an Improved Feature-Weighted Fuzzy C-Means is proposed to overcome to these shortcomings.

In this work, clustering based image segmentation method used and modified by introducing neutrosophic logic. The clustering technique with neutrosophy is used to deal with indeterminacy factor of image pixels. The approach is to transform the image into the neutrosophic set by calculating truth, falsity and indeterminacy values of pixels and then, the clustering technique based on neutrosophic set is used for image segmentation. The clusters are then refined iteratively to make the image more suitable for the segmentation. This iterative process converges when required number of clusters areformed. Finally, the image in the

1. INTRODUCTION

An image is a2-D array of very small square regions called as pixels. The intensity of pixels in a grayscale image are represented by numeric values that can easily be stored in a 2-D array. Image segmentation is the process of partitioning an image into various regions, each having similar attributes (luminance value for grayscale images and color components for color images). The image segmentation can be used in content based image retrieval, image feature extraction, etc [5]. There are many algorithms for image segmentation based on clustering, which try to partition an image into a number of regions [6]. Clustering is an approach that tries to group a set of data objects in to the same cluster that are similar to each other, and dissimilar data objects are associated into different clusters [7] [12]. The K-Means clustering algorithm tries to group a set of data objects into k clusters based on the distance between the data object and the k centroids selected beforehand [1] [3]. Data objects are associated with the centroid closest to them. The k-means clustering algorithm starts with selecting value of k (number of clusters) and random initial centroids for each cluster. In each iteration, data objects are associated with the closest centroids. Then, centroids are updated based on the distance values of data objects associated with it. After centroid calculation (mean value of all data objects in a cluster), data objects are again compared with new centroids and associated with the closest one. This process continues until termination criteria is met [1] [2]. In the case of image segmentation, a set of image pixels are grouped in a manner such that, pixels within a cluster have more similarity in comparison with pixels in another cluster. The fuzzy logic is proposed to deal with the

VOLUME-1 ISSUE-3
ISSN : 2799-18

vagueness and uncertainty. It has two values associated with each variable, the degree of truthiness and degree of falsehood. It can be represented as FS= {T, F}, where T and F are the degree of truthiness and degree of falsehood of the variable towards the set FS respectively. Neutrosophic logic introduces a new parameter to fuzzy sets, called as indeterminacy. Neutrosophic logic theory considers every possible outcome for a variable X like X, Anti-X and Neut-X which is neither X nor Anti-X [1] [4]. The neutrosophic notation for an element x be x (T, I, F), where T is the degree of membership of x with S, I is the indeterminacy of x with S and F is the degree of non-membership of x with S.

Machine learning is an important branch of artificial intelligence whose main goal is to study and propose those methods which are able to learn from data. In many problems, machine learning may be applied where the process of extracting information from data is complex [1]. These methods have been employed with both supervised and unsupervised learning, where in the case of supervised learning, class labels are used to guide the machine learning algorithm. For several problems, it is not feasible to obtain data labels, therefore approaches for unsupervised learning like clustering are more suitable to solve these problems.

Clustering algorithms play an important role in discovering useful knowledge from large databases. The goal is that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups. Clustering is a mathematical tool that attempts to discover structures or certain patterns in a dataset, where the objects have a certain degree of similarity inside each cluster. It can be achieved by various algorithms. Cluster analysis is a repetitive process of knowledge discovery. It will require modifying parameter and preprocessing until the results achieve the desired properties.
Clustering is a useful tool for understanding and visualizing available structures in data. Fuzzy C-means is one of the commonly used and efficient objective function-which is based on clustering techniques. Data clustering or cluster analysis is an important field in pattern recognition, machine intelligence and computer vision community, that has had numerous applications in the last three decades. Generally, clustering term is known as grouping a set of N samples into C clusters whose members are similar in some sense. This similarity between different samples is either a suitable distance based on numeric attributes, or directly in the form of pair-wise similarity or dissimilarity measurements. With a clustering technique, a collection of objects or feature vectors is partitioned into clusters. In the past few decades, many clustering algorithms have been developed, which mainly contain hierarchical clustering (such as Single Link and Complete Link), partitional clustering (such as k-means, fuzzy C-means, Gaussian Mixture and Density Estimation) and spectral clustering. Moreover, in the last few years, fuzzy C-means (FCM) clustering and spectral clustering algorithms are research focus [2].
Fuzzy clustering introduces the concept of membership into data partition, for this reason that membership can indicate the degree to which an object belongs to the clusters definitely, and actually represents the data partition more clearly. The fuzzy set theory was introduced by Zadeh [3], and it was successfully applied in image segmentation. The fuzzy c-means algorithm was proposed by Bezdek [4], based on fuzzy theory, it is the most widely studied and used algorithm in data clustering for its simplicity and ability to retain more information from images.

2. RELATED WORK

Clustering plays an important role in pattern recognition, image processing, and computer vision. By a clustering technique, a collection of objects or feature vectors are partitioned into clusters. In the past few decades, many clustering algorithms have been developed, which mainly contain hierarchical clustering

VOLUME-1 ISSUE-3
ISSN : 2799-18

(such as Single Link and Complete Link), partitional clustering (such as k-means, fuzzy C-means, Gaussian Mixture and Density Estimation) and spectral clustering. Moreover, in the last few years, fuzzy C-means (FCM) clustering and spectral clustering algorithms are researching focuses.

Unlike the hard clustering techniques (each object is assigned to one and only one cluster), fuzzy C-means clustering allows an object to belong to a cluster with a grade of membership. Moreover, when there is not enough information about the structure of the data, fuzzy C-means clustering algorithm can handle this uncertainty better, and has been widely applied to the data clustering area. However, there are still some open problems in FCM algorithm [5, 6].

In Clustering, one of the most widely used algorithms is fuzzy clustering algorithms. Fuzzy set theory was first proposed by Zadeh in 1965 [3] & it gave an idea of uncertain belonging which was described by a membership function. For each clusters, the data points are assigned for membership values and fuzzy clustering algorithm allow the clusters to grow into their natural shapes. The fuzzy clustering algorithms can be divided into two types: one is Classical fuzzy clustering algorithms and the other is Shape based fuzzy clustering algorithms.

Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. In many cases, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned for membership degrees between 0 and 1 indicating their partial membership. Fuzzy c-means algorithm is the most widely used algorithm in this regard.

As a partitioning clustering method, Fuzzy C-Means (FCM) has been widely studied in many fields [7, 8]. FCM was first proposed by Dunn in 1974 and then it was developed by Bezdek.

In a fuzzy clustering approach for time series based on cepstral coefficients, i.e., based on the fuzzy logic is proposed. Time series were classified in the frequency domain by considering their cepstral representations.

In an improved method for image segmentation using the fuzzy c-means clustering algorithm (FCM) is proposed. They suggested these results for further improvement by acting at three different levels. This algorithm is widely experimented in the field of image segmentation, revealing very successful results.

In recent years, many improved versions of FCM have been proposed. To eliminate the shortcoming caused by the random selection of the initial centers, a new approach to the initialization of the FCM algorithm is proposed. The choice of initialization scheme is of importance because the optima and partitions identified by the FCM algorithm may vary depending on the selected initial cluster centroids. This method is based on the idea that the dominant colors in an image are likely to belong to separate clusters.

The effectiveness of introducing the K-means++ initialization scheme into the context of Fuzzy C-means is investigated empirically. This method improves the way in which Fuzzy C-means initializes its clusters and has several advantages over the discussed methods. This method achieves superior clustering (in terms of validity indexes) compared to a used random initialization like the standard and fewer iterations.

Graves and Pedrycz presented a comprehensive comparative analysis of kernel-based fuzzy clustering and fuzzy clustering. The kernel-based clustering algorithms can cluster specific non-spherical clusters such as the ring cluster quite well, performing FCM and GK for the same number of clusters.

The blind application of the conventional FCM algorithm to image segmentation often performs badly because: (i) FCM is very sensitive to noise and imaging artifacts since segmentation is decided only by pixel intensities, i.e. no spatial information in the image context is considered; (ii) The efficiency of FCM highly depends on the initialization step, because the iterative process easily falls into a locally optimal solution; (iii) The FCM algorithm is based on the Euclidean metric distance, so only it can be

VOLUME-1 ISSUE-3
ISSN : 2799-18

used to detect the data classes with the same super spherical shapes.

For the FCM and its improved versions, it is assumed that all the features of the samples in a given data set make equal contribution while constructing the optimal clusters. However, for certain real-world data sets, some of the features can exhibit higher relevance in the clustering information than others. Thus, the features with higher relevance are more important to form the optimal clustering results than those with lower relevance. Therefore, it is desirable to revisit an FCM method in which different features possess different weights. Recently, several Feature-Weighted Fuzzy C-Means (FWFCM) approaches have been proposed to address the mentioned problem. These variants exhibit two separate stages. At the first stage, the feature-weight vector is determined. Then, at the second stage, the FWFCM is trained by the samples with their features weighted by the obtained feature-weight vector. Unfortunately, the elements of the feature-weight vector are fixed in this second stage, which might not fully reflect their relevance in the clustering process. Therefore, much effort has been devoted to adjust the feature-weight vector during the training course of the FCMs. However, these approaches assign different feature weights for different features of the clusters rather than for different features of the entire data set. The fuzzy c-means algorithm (FCM) is one of the most popular fuzzy clustering algorithms where the membership degrees of the data are obtained through iterative minimization of a cost function, subject to the constraint that the sum of membership degrees over the clusters for each data are equal with 1. The FCM algorithm suffers from several drawbacks: it also tries to minimize the intra-cluster variance as well, and has the same problems like k-means algorithm; the minimum is a local minimum, and the results depend greatly on the initializations In addition, the FCM algorithm is very sensitive to the presence of noise. The membership of noise points might be significantly high. The FCM algorithm cannot distinguish between equally highly likely and equally highly unlikely, and it is sensitive to the selection of distance metric.

3. PROPOSED METHOD

For the large number of clusters, the K-Means clustering algorithm can make several empty clusters. If no points have been allocated to a cluster in the assignment step that can result in noisy image. Therefore, we need a criteria through which we can remove the problem of empty cluster generation. To remove the empty clusters consider each pixel as the data point and then apply clustering algorithm to generate the large number of clusters. Finally merge the similar looking clusters and remove the empty clusters. The approach is to transform an image into neutrosophic set by calculating truth, falsity and indeterminacy values of pixels. Then, a clustering technique based on neutrosophic set is used to segment the image. So, the proposed methodology finds a way to calculate all these values and provides better results than existing K-mean approach discussed so far. The new Image segmentation method using neutrosophic logic consists of following phases as in the fig. 1.

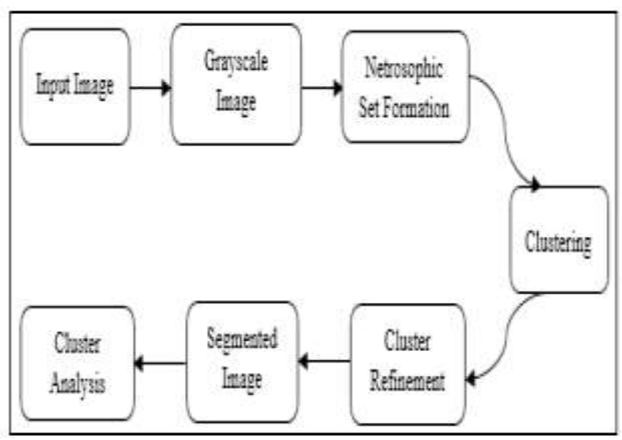

Fig. 1. Flowchart of proposed method *2.1. Converting the input image into grayscale image*

A grayscale image is a simple kind of image that contains one domain, and each pixel in the image can be represented by an integer [0,255] i.e. it carries only intensity information. We have used MATLAB built in function to

VOLUME-1 ISSUE-3
ISSN : 2799-18

calculate the intensity value of each pixel. *2.2. Neutrosophic Set Formation* In this phase, the image is transformed into neutrosophic domain. A pixel P(i, j) of an image is represented in neutrosophic domain using its three components i.e. T(i,j), F(i,j), and I(i,j). This approach considered a pixel in 3*3 window and calculated the mean of the window based on intensity value. The mean value is assigned each pixel in order to calculate truth value.

$$Mean_{(i,j)} = \frac{1}{(2w+1)^2} \sum_{\substack{i=x-w \\ j=y-w}}^{\substack{i=x+w \\ j=y+w}} In(i,j)$$

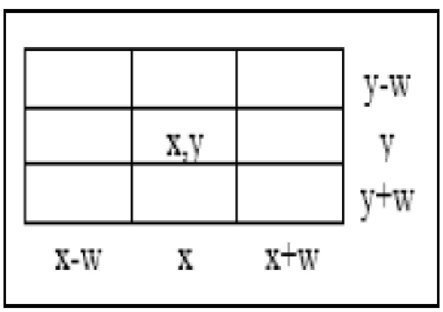

Fig.2: Window Size

Indeterminacy of the pixel means how different the pixel is from its neighboring pixels inside the window. The best way to calculate that is to use standard deviation and difference map of the pixels.Standard deviation calculates the dispersion of a set of pixels from its mean. Mathematically, standard deviation is calculated as the square root of variance.

$$std = \frac{1}{(2w+1)^2} \sum_{\substack{i=x-w \\ j=y-w}}^{\substack{i=x+w \\ j=y+w}} [In(i,j) - Mean(i,j)]^2$$

We compute the maximum difference of pixel intensity from its surroundings pixels as follows:

$$Diff_{(i,j)} = \max\left(In_{(i,j)} - In_{(x,y)}\right)$$

Cluster Refinement

In the refinement stage, the clusters with small number of pixels are merged or grouped with its surrounding or neighboring clusters. Moreover, if cluster Ci has number of pixels less than or equal to minimum cluster size (CZ) then cluster Ci is merged to its surrounding cluster Cj. In this work, minimum cluster size is considered as equals to w*h*0.001, Where w is the width of the image and h is the height of the image. For example if the size of image is 400 × 400 then the minimum cluster size can be calculated as CZ= 400*400/1000=160. Therefore, a minimum of 160 pixels will be grouped in a single cluster.

Pseudo Code

1. For each cluster C_i where having CZ_i < min CZ then
   Merge (C_i, C_j) where C_j is the neighboring cluster.
2. Step 1 is repeated until each cluster size CZ_i >=min CZ

Fig. 3. Pseudo Code for Cluster Refinement

CONCLUSION AND FUTURE WORK

An image can be considered as a 2-D array of very small square regions called as pixels. Images are one of the primary method of information sharing.The image segmentation is an important image processing approach, which analyzes what is inside the image. Image

VOLUME-1 ISSUE-3
ISSN : 2799-18

segmentation can be used in content-based image retrieval, image feature extraction, pattern recognition, etc. Neutrosophy is an enhancement to the fuzzy logic and appends the uncertainty factor in addition to truth and falsity components. The K-Means clustering algorithm tries to group a set of data objects into k clusters based on the distance between the data object and the k centroids selected beforehand. In the case of image segmentation, a set of image pixels are grouped in a manner such that, pixels within a cluster have more similarity in comparison with pixels in another cluster. To start with the proposed method, firstly k-means clustering algorithm is applied on the input image. The output of this step is provided as an input to neutrosophic set by defining a neutrosophic domain. Then,the neutrorosophy isintroduced to image segmentation bycalculating the T, I, F values of each pixel in the image. The K-Means clustering algorithm using neutrosophic logic is implemented to minimize the indeterminacy of the pixels. The resulting clusters are then refined iteratively to make the images suitable for segmentation. The results of the proposed method describes that it can be used to get better results on synthesis image as well as real images with/without noise. The proposed method for image segmentation provides improved results for color images. The future works are described as follows. Neutrosophy can be applied to other image processing problems like feature extraction and classification. Apply neutrosophy to different research area like control theory, artificial intelligence.

REFERENCES

[1] Mitchell, T.M., "Machine Learning. McGraw-Hill",NewYork., 1997.

[2] Dervis Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm". Applied Soft Computing, 2011. 11: p. pp. 652-657.

[3] L.A. Zadeh, "Fuzzy sets". Information and Control 8, 1965: p. pp. 338-353.

[4] Bezdek, J.C., "Fuzzy Mathematics in Pattern Classification". Ph. D. Thesis, Applied Math. Center, Cornell University, Ithaca, 1973.

[5] K.R. Zalik, "Cluster validity index for estimation of fuzzy clusters of different sizes and densities", Pattern Recognition 43 (10) (2010) 3374–3390.

[6] R.K. Brouwer, A. "Groenwold, Modified fuzzy C-means for ordinal valued attributes with particle swarm for optimization", Fuzzy Sets and Systems 161 (13) (2010) 1774–1789.

[7] Yao, H., et al., "An improved k-means clustering algorithm for fish image segmentation". Mathematical and Computer Modelling, 2013. 58(3–4): p. 790-798.

[8] Karami, A. and M. Guerrero-Zapata, "A fuzzy anomaly detection system based on hybrid PSO-Kmeans algorithm in content-centric networks". Neurocomputing, 2015. 149, Part C(0): p. 1253-1269.