A Survey on Fuzzy C-means Clustering Techniques

[1]Sandhya Prabhakar H, [2]Prof Sandeep Kumar
[1] Mtech Student, [2]Associate Professor
[1]Computer Science and Engineering,
[1]Cambridge Institute of Technology, Bengaluru, India

_____

Abstract— **This paper reviews different Fuzzy C-Means Clustering techniques. Here we mainly discuss few Fuzzy C-Means clustering techniques like Conventional Fuzzy C-Means (FCM), Fast Fuzzy C-Means (FFCM) , Feature weighted Fuzzy C-Means (FWFCM) , Weighted image patch-based Fuzzy C-Means (WIPFCM) , Kernel Based Fuzzy C-Means (KFCM) , Multiple Kernel Fuzzy C-Means (MKFCM) and Kernel-based Fuzzy C-Means Clustering based on Fruit Fly Optimization Algorithm (FOAKFCM).**

Index Terms— **Segmentation, Clustering, Fuzzy C means (FCM), Fruit fly Optimization, Multiple Kernel, Weighted Image Patch.**

_____

I. INTRODUCTION

Image analysis generally refers to preparing of images by computer with the objective of discovering what objects are exhibited in the image [1].Image segmentation is one of the fundamental and difficult tasks in many of the image and vision applications. It has been studied extensively over the past several decades with a huge number of segmentation algorithms being published in the literature. Those image segmentation approaches can be divided broadly into four categories: thresholding, clustering, edge detection and region extraction.

Clustering is a process of assigning data points with similar properties to the same group and dissimilar data points to different groups [2]. Members within a cluster exhibit similar characteristics than the members of other clusters. The clustering process is usually based on a proximity measure or, on the properties that data points share. Based on the way of data organization, clustering can be taught of three types: hierarchical, partitioning and mixture model methods.

Generally, there are two main clustering approaches: hard clustering (crisp clustering) and soft clustering (fuzzy clustering). In the crisp clustering method a data point can belong to only one cluster. However, in many real cases, the boundaries between clusters cannot be clearly defined. Some objects/datapoints may belong to more than one cluster. In such cases, the fuzzy clustering method provides a better and more useful method to cluster these objects [3].In Fuzzy clustering a datapoint can belong to more than one cluster. The degree of belongingness plays a vital role in Fuzzy Clustering and provides more flexibility. Cluster analysis has been widely used in a variety of areas such as data mining and pattern recognition.

A common approach to image clustering involves addressing the following issues [4]:
- Image features – how to represent the image.
- Organization of feature data – how to organize the data.
- Classifier – how to classify an image to a certain cluster.

II. FUZZY C MEANS CLUSTERING TECHNIQUES

I. Conventional Fuzzy C-Means Clustering

The Fuzzy C-Means algorithm (FCM) , is one of the best known and the most widely used fuzzy clustering algorithms. It was first proposed by Dunn and promoted as the general FCM clustering algorithm by Bezdek. The main purpose of the FCM algorithm is to make the vector space of a sample point be divided into a number of sub-spaces in accordance with a distance measure.

FCM has a wide domain of applications such as agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis, and target recognition [5].

Major advantages of FCM are [6]:
- Straightforward implementation.
- Fairly robust behaviour.
- Applicability to multidimensional data.
- Ability to model uncertainty within the data.

However, the FCM algorithm does not take the local spatial property of images into consideration, and hence suffers from high sensitivity to noise.

In the FCM algorithm the data items are assigned to more than one cluster with membership values between 0 and 1 [7].

The center is initialized and the count t, for the number of iterations is initialized to zero. Then the membership function is found using the equation 1. Then the value of t is incremented by 1 and new centers are found using equation 2. Till convergence the second and third steps are run. The definition of membership function is :

$$\mu ij = \frac{1}{\sum_{m=1}^{c} \frac{\|Xi - Cj\|}{\|Xi - Cm\|^{\frac{2}{(k-1)}}}} \qquad (1)$$

The expression for the centre of the cluster is:

$$c_i = \frac{\sum_{j}^{N} \mu ij^{k} x_i}{\sum_{j=1}^{N} \mu ij} \qquad (2)$$

II. Fast Fuzzy C-Means (FFCM) Clustering

The FFCM algorithm features several improvements over the FCM. One of its important features is decreasing the number of calculations by checking the membership value for each point and eliminating these points with membership values smaller than a threshold value. The choice of the appropriate threshold is based on experimentations [6].

FFCM aims at decreasing the number of distance calculations of the FCM by computing the distances between data points and the nearest cluster centres for points with membership values greater than a threshold, T, where the value of T is less than 1 and greater than 0. In this case, there is no need to calculate distances for points with membership values less than T since these values do not severely affect the results and therefore, some distance calculations can be saved.

Assume that we want to determine a fuzzy partition with two clusters (i.e., C = 2). Assume also that we choose $T = 0.5$, then we obtain part of the U matrix shown in Table 1. Here, we don't compute distances between the cluster centers and the points for U values less than T (the shaded values in Table 1).

For example, the distance between cluster $C2$ and point X (1) in not computed, and hence, some time savings can occur.

Table 1 Fuzzy partition, when $T = 0.5$ and $C = 2$

	X(1)	X(2)	X(3)	...	X(139)	X(140)
C1	0.99	0.01	0.10	...	0.06	0.13
C2	0.01	0.99	0.90	...	0.94	0.87

It is expected that more time savings can be obtained for a larger number of clusters.

III. FEATURE WEIGHTED FUZZY C -MEANS CLUSTERING

In the field of cluster analysis, most of existing algorithms assume that each feature of the samples plays a uniform contribution for cluster analysis. Considering different features with different importance, feature-weight assignment can be regarded as a special case of feature selection. That is, the feature assigned a value in the interval [0, 1] indicating the importance of that feature, we call this value "feature-weight".

A key question that arises here is how to determine the importance of each feature. In other words, how to assign a weight to each feature so that the weight of each feature determines the *FWA* of it [8].

To determine the *FWA* of features of a data set two major approaches can be adopted: **Human-based approach** and **Automatic approach**. In human-based approach the *FWA* of each feature is determined based on negotiation with an expert individual who has enough experience and knowledge in the field that is the subject of clustering. In automatic approach we use the data set itself to determine the *FWA* of its features [8].

The algorithm for FWFCM [9]:

Algorithm - Feature -Weighted Fuzzy C-Means

Input   : Dataset
Output   : Final fuzzy partition matrix;
            Final center matrix;
            Final feature-weight vector;

Begin algorithm
    Initialize number of clusters C, fuzzifiction exponent m and fuzzy partition matrix
    Initialize Feature-Weight vector using normalized Term Variance:
     While (not achieve termination condition)
        Update the cluster centers
        Calculate the distances
        Update the fuzzy partition matrix;
        Update the elements in the feature-weight vector;
     End while
End algorithm

IV. Weighted Image Patch-Based Fuzzy C-Means Clustering

Image patches have been widely used in image denoising, especially for the non-local based algorithms, which use the local information embedded in image patches to measure image similarity. Generally, image patch-based denoising methods perform much better than those based on pixels, since an image patch contains more information than a single pixel and can better describe the properties of the image. The FCM-based image segmentation algorithm can be improved by replacing each pixel used in constructing the objective function with the corresponding image patch, in which all pixels are weighted adaptively. Different from other FCM-based algorithms, the weighted image patch-based FCM (WIPFCM) algorithm views each image patch, instead of each pixel, as the basic unit to be clustered. Thus, the spatial constraints are incorporated intrinsically into the clustering process without supplementing a penalty term [10].

V. Kernel Based Fuzzy C-Means Clustering

Kernel FCM is an extension of FCM, and in this extension, the original inputs are mapped into a much higher dimensional space by some function. In the new space, the examples are more easily to be separated or clustered [11]. The KFCM algorithm uses kernel function in place of Euclidian distance in traditional Fuzzy c means algorithm, exchanging scale, the sample is needed to map onto high dimension space by kernel function, in order to increase the differences among cluster center, and this can overcome the disadvantage that traditional FCM do not do well in dealing with fuzzy cluster center [12].

VI. Multiple Kernel Based Fuzzy C- Means Clustering

Recently, developments on kernel methods and their applications have emphasized the need to consider multiple kernels instead of a single fixed kernel [13]. With the application of multiple kernels, the kernel methods obtains more flexibility and also reflects the fact that practical learning problems often involve data from multiple heterogeneous or homogeneous sources [13-14, 15, 16, 17, 18, 19, 20]. Specifically, in the image segmentation problem, the input data are the properties of image pixels, and they could be derived from different sources. For example, the intensity of a pixel is directly gained from itself and some complicated texture information of a pixel maybe gained from some wavelet filtering of the image [21]. Combining multiple kernels into the kernel methods provides us a great tool to fusion information from different sources [17].

By using multiple kernels and automatically adjusting the kernel weights, MKFC is more important to ineffective kernels and irrelevant features. It makes the choice of kernels less crucial [22].

VII. Kernel Based Fuzzy C-Means Clustering Based on Fruit Fly Optimization Algorithm

A new optimization algorithm called the Fruit Fly Optimization Algorithm or Fly Optimization Algorithm (FOA) was proposed by Pan [24]. Fruit fly Optimization algorithm simulates the foraging behaviour of fruit flies for searching the global optimum. Such an optimization algorithm has advantages such as a simple computational process, ease of transformation of such concept into program code and ease of understanding, etc. It has been widely used in continuous function optimization problem [25], annual power, load forecasting [26], joint replenishment problems [27], and multidimensional knapsack problem [28] and so on.

The kernelized fuzzy c-means algorithm uses kernel methods to improve the clustering performance of the well known fuzzy c means algorithm by mapping a given dataset into a higher dimensional space non-linearly. To further improve the clustering performance and to overcome the drawbacks of the traditional algorithms such as, sensitivity to initialization, trapping into local minima, a new clustering method (FOAKFCM) based on kernelized fuzzy c-means algorithm and a based on fruit fly optimization algorithm (FOA), is proposed.

In FOAKFCM algorithm the fruit fly optimization algorithm and the kernel technique are combined. Firstly, the fruit fly algorithm is used to optimize the initial clustering\ center. Then, the KFCM algorithm is availed to guide the categorization, so as to improve the clustering performance of the FCM algorithm.

III. COMPARATIVE STUDY OF DIFFERENT FUZZY C-MEANS CLUSTERING TECHNIQUES

Table 2 Comparative Study of different Fuzzy C-means Clustering Techniques

Serial No	Research Paper	Focus	Limitations
1	JIA-YIN KANG, LE-QUAN MIN, QING-XIAN LUAN, XIAO LI, JIN-ZHU LIU", DENTAL PLAQUE QUANTIFICATION USING FCM-BASED CLASSIFICATION IN HSI COLOR SPACE", Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov. 2007.	This paper proposed an approach to quantify the dental plaque based on FCM clustering algorithm incorporated with color information in HSI color space. Experimental results shown that the dental plaque quantifications based on our approach are automated, objective and quantitative, whereas those indexed by traditional plaque indices are manual, subjective and semi-quantitative.	A major disadvantage of conventional FCM is its need for a large amount of time to converge.
2	MOH'D BELAL AL- ZOUBI, AMJAD HUDAIB, BASHAR AL-SHBOUL," A Fast Fuzzy Clustering Algorithm", Department of Computer Information Systems, University	In this paper, a Fast Fuzzy C-Means (FFCM) algorithm is proposed. One important feature of the FFCM algorithm is decreasing the number of	In case of FFCM the better performance is obtained when a threshold value is within

		distance calculations required by the FCM algorithm. This is done by checking the membership value for each point and eliminating these points with membership values smaller than a threshold value.	the range (0.28 - 0.5). However, the quality of the results is degraded, especially when the number of clusters decreases.
3	Huaiguo Fu, Ahmed M. Elmisery, "A NEW FEATURE WEIGHTED FUZZY C-MEANS CLUSTERING ALGORITHM", Telecommunications Software & Systems Group Waterford Institute of Technology, Waterford, Ireland.	This paper proposed a new feature weighted fuzzy c-means clustering algorithm in a way which this algorithm be able to obtain the importance of each feature, and then use it in appropriate assignment of feature-weight.	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.
4	Zexuan Ji, Yong Xia , Qiang Chen, Quansen Sun, Deshen Xia, David Dagan Feng , "Fuzzy c-means clustering with weighted image patch for image segmentation", Applied Soft Computing,2012.	This paper describes that the proposed WIPFCM algorithm can segment images more accurately, and is more robust to noise than other state-of-the-art FCM-based segmentation algorithms.	This method increase computational burden because each pixel need to be represented by a patch and additional step of calculation of their weights is required.
5	Jinxian Lin, Shuangyang Zheng, "An optimizing search based on Kernel-based Fuzzy C-Means Clustering", China.	In this paper it is shown how to optimize search based on Kernel-based Fuzzy C-Means. It also shows how KFCM algorithm can further improve search performance.	Not Flexible enough to support data from different heterogeneous sources. Flexibility can be improved with the application of multiple kernel based Fuzzy C means.
6	Long Chen, Mingzhu Lu, C. L. Philip Chen "Multiple Kernel Fuzzy C-means based Image Segmentation", China.	In this paper, the multiple kernel fuzzy c-means is introduced and applied into image segmentation problems. Multiple kernel fuzzy c-means algorithms use the kernel function combined by multiple kernels. These kernels are selected for different information or properties.	It can be further optimized with several other optimization techniques.
7	Qiuping Wang, Yiran Zhang , Yanting Xiao , Jidong Li ,"Kernel-based Fuzzy C-means Clustering Based on Fruit Fly Optimization Algorithm", China.	This paper describes that the FOAKFCM algorithm proposed overcomes FCM's defects efficiently and improves the clustering performance greatly.	Pure FOA algorithms yields low convergence precision, easily trapped in a local optimum value at the later evolution stage. Hence hybrid versions are better.

IV. CONCLUSION

The quality of the clustering algorithm is determined by the quality of the clusters produced. Hence different clustering algorithms are being implemented and used based on different applications and the areas of interest. One has choose the best suited clustering algorithm to obtain the intended results.

V. ACKNOWLEDGMENT

REFERENCES

[1] Y. J. Zhang, "A survey on evaluation methods for image segmentation", Pattern Recognition, Vol. 29, No. 8, pp. 1335-1346, August 1996.

[2] Han J., Kamber M., 2001 *"Datamining: Concepts and Techniques"*. Morgan Kaufmann Publishers, San Francisco.

[3] Hung M, D. ang D, 2001 *"An efficient fuzzy c-means clustering algorithm"*. In Proc. the 2001 IEEE International Conference on Data Mining.

[4] Y. J. Zhang & J.J. Gerbrands, "Objective & quantitative segmentation evaluation & comparisons", Signal Processing, Vol. 39, pp. 43-54, 1994.

[5] Hopner , K, R., Runkler, 1999 *"Fuzzy Cluster Analysis"*, John Wily & sons.

[6] MOH'D BELAL AL- ZOUBI, AMJAD HUDAIB, BASHAR AL-SHBOUL,"*A Fast Fuzzy Clustering Algorithm"*. Department of Computer Information Systems, University of Jordan, Amman.

[7] S.Bharathi, Dr.P.Venkatesan, "A survey on Image Segmentation by Fuzzy C Means Clustering Techniques", SCSVMV University, Enathur, Kanchipuram, Tamilnadu, India.

[8] Huaiguo Fu, Ahmed M. Elmisery, "A NEW FEATURE WEIGHTED FUZZY C-MEANS CLUSTERING ALGORITHM", Telecommunications Software & Systems Group, Waterford Institute of Technology, Waterford, Ireland.

[9] Amirhadi Jahanbakhsh Pourjabari† , Mojtaba Seyedzadegan ," An improved method of fuzzy c-means clustering by using feature selection and weighting", 64 IJCSNS International Journal of Computer Science and Network Security, VOL.16 No.10, October 2016.

[10] Zexuan Ji, Yong Xia , Qiang Chen, Quansen Sun, Deshen Xia, David Dagan Feng , "Fuzzy c-means clustering with weighted image patch for image segmentation", Applied Soft Computing,2012.

[11] Long Chen, Mingzhu Lu, C. L. Philip Chen "Multiple Kernel Fuzzy C-means based Image Segmentation", China.

[12] Jinxian Lin, Shuangyang Zheng, "An optimizing search based on Kernel-based Fuzzy C-Means Clustering", China.

[13] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," Proceedings, Twenty First International Conference on Machine Learning, ICML 2004. pp. 41-48.

[14] Z. Wang, S. C. Chen, and T. K. Sun, "Multi K-MHKS: A novel multiple kernel leaming algorithm," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 348-353, Feb, 2008.

[15] S. A. Rojas, and D. Fernandez-Reyes, "Adapting multiple kernel parameters for support vector machines using genetic algorithms," 2005 IEEE Congress on Evolutionary Computation, Vols 1-3, Proceedings, pp. 626-631 2871,2005.

[16] G. Camps-Valls, L. Gomez-Chova, 1. Munoz-Mari et aI., "Composite kernels for hyperspectral image classification," IEEE Geoscience and Remote Sensing Letters, vol. 3, no. 1, pp. 93-97, Jan, 2006.

[17] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari et aI., "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 6, pp. 1822-1835, Jun, 2008.

[18] S. Sounenburg, G. Ratsch, C. Schafer et aI., "Large scale multiple kernel learning," Journal of Machine Leaming Research, vol. 7, pp. 1531-1565, 2006.

[19] A Apatean, A Rogozan, and A Bensrhair, "Obstacle recognition using multiple kernel in visible and infrared images," 2009 IEEE Intelligent Vehicles Symposium, Vols I and 2, pp. 370-375 1424, 2009.

[20] M. Lu, C. L. P. Chen, J. Huo et aI., "Optimization of combined kernel function for SVM based on large margin learning theory," Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. pp. 353-358.

[21] K. Muneeswaran, L. Ganesan, S. Arumugam et aI., "Texture image segmentation using combined features from spatial and spectral distribution," Pattern Recognition Letters, vol. 27, no. 7, pp. 755-764, May, 2006.

[22] M. Ganesh1 and V. Palanisamy, "A MULTIPLE KERNEL FUZZY C-MEANS CLUSTERING ALGORITHM FOR BRAIN MR IMAGE SEGMENTATION", International Journal of Advances in Engineering & Technology, Nov. 2012. ©IJAET ISSN: 2231-1963.

[23] Dhanalakshmi L, S.Ranjitha, Dr Suresh H N , "Image Processing Using Modified Multiple kernel fuzzy c- means clustering (MMKFCM) Technique" , Bangalore.

[24] Wen-Tsao Pan, "A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example," Knowledge-Based Systems,   vol.26, pp. 69–74, 2012.

[25] Quan-Ke Pan, Hong-Yan Sang, Jun-Hua Duan, Liang Gao, "An improved fruit fly optimization algorithm for continuous function optimization problems," Knowledge-Based Systems, vol.62, pp. 69–83,2014.

[26] Hong-ze Li, Sen Guo, Chun-jie Li, Jing-qi Sun, "A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm," Knowledge-Based Systems, vol.37, pp. 378–387, 2013.

[27] Lin Wang, Yuanlong Shi, Shan Liu, "An improved fruit fly optimization algorithm and its application to joint replenishment problems," Expert Systems with Applications, vol. 42, pp. 4310–4323, 2015.

[28] Ling Wang, Xiao-long Zheng, Sheng-yao Wang, "A novel binary fruit fly optimization algorithm for solving the multidimensional knapsack problem," Knowledge-Based Systems, vol. 48, pp. 17–23, 2013.

[29] Qiuping Wang, Yiran Zhang , Yanting Xiao , Jidong Li ,"Kernel-based Fuzzy C-means Clustering Based on Fruit Fly Optimization Algorithm", China.

[30] Mahipal Singh Choudhry and Rajiv Kapoor, "Performance Analysis of Fuzzy C-Means Clustering Methods for MRI Image Segmentation", Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016).