SURVEY ON MEDICAL IMAGE SEGMENTATION USING ENHANCED K-MEANS AND KERNELIZED FUZZY C- MEANS

Gunwanti S. Mahajan & Kanchan S. Bhagat.
Dept of E &TC, J. T. Mahajan C.o.E Faizpur, India

ABSTRACT

Diagnostic imaging is an invaluable tool in medicine. Magnetic resonance imaging (MRI), computed tomography (CT), digital mammography provide an effective means for noninvasively mapping the anatomy of a subject. With the increasing size and number of medical images, the use of computers in facilitating their processing and analyses has become necessary. More recently clustering is an effective tool in segmenting medical images for further treatment plan. In order to solve the problems of clustering performance affected by initial centers of clusters, a new technique is introduces a specialised center initialization method for executing the proposed algorithms in segmenting medical images. Clustering is the process of organizing data objects into a set of disjoint classes called clusters. The objective of this paper is to make survey of enhanced k-means and Kernelized fuzzy c-means for a segmentation of brain magnetic resonance images.

KEYWORDS: *Enhance k-means, Kernelized fuzzy c- means, Magnetic resonance imaging, centre initialisation, clustering algorithm.*

I. INTRODUCTION

Segmentation is an important step in the analysis of medical images for computer-aided diagnosis and therapy [1]. Medical image segmentation separates the image into distinct classes such as brain tumours, and necrotic tissues, etc. It provides an appropriate therapy prescription by quantifying tissue volumes and detecting tumours, and necrotic tissues. As a result, this technique is currently a crucial diagnostic imaging technique for early detection of abnormal changes in tissues and organs. Many image processing techniques have been proposed for brain MRI segmentation including thresholding, region growing, and clustering [2],[3]. However, these techniques based on pixel attributes lead to inaccuracy with segmentation because medical images are limited spatial resolution, poor contrast, noise, and non-uniform intensity variation.

Clustering is the process of organizing data objects into a set of disjoint classes called clusters. Clustering aims to analyze and organize data into groups based on their similarity. Clustering is an example of unsupervised classification. Classification refers to a procedure that assigns data objects to a set of classes [4]. Unsupervised means that clustering does not depends on predefined classes and training examples while classifying the data objects. Cluster analysis seeks to partition a given data set into groups based on specified features so that the data points within a group are more similar to each other than the points in different groups. Therefore, a cluster is a collection of objects that are similar among themselves and dissimilar to the objects belonging to other clusters.

Fuzzy c-means of unsupervised clustering techniques which used on established outstanding results in automated segmenting medical images in a robust manner. Fuzzy c-means clustering is successfully applied in many real world problems such as astronomy, geology, medical imaging, target recognition, and image segmentation [4]. Among them, fuzzy c-means segmentation method has considerable benefits, because they could retain much more information from the original image than hard segmentation methods. But as we know that the clustering depends on choice of initial cluster centre hence we have used the cluster centre initialisation algorithm in order to mprove the performance of k-means and fuzzy C-means in addition with Kernelized fuzzy c-means and Enhanced k means. In this work mainly confined K-means, Fuzzy c-means, Kernelized fuzzy c-means, and Enhanced k means.

The rest of this paper is organized as follows: Section II literature survey describes different segmentation methods for MRI images. Sections III, IV give conclusion and future work respectively.

II. LITERATURE SURVEY

Numerous methods are available in medical image segmentation. These methods are chosen based on the specific applications and imaging modalities. Imaging artifacts such as noise, partial volume effects, and motion can also have significant consequences on the performance of segmentation algorithms. Some of these methods with their idiosyncrasies are thresholding Method Classifiers ,Markov Random Field Models ,Artificial Neural Networks , Atlas-Guided Approaches, Deformable Models Clustering Analysis, Fuzzy C-Means ,Clustering K-means clustering [5].

2.1. K-Means Clustering

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 a priori. The main idea is to define k centroids, one for each cluster[5]. These centroids 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. The next step is to take each point belonging to a given data set and associate it to the nearest centroids. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more [5].

Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is given as

$$\boldsymbol{J_w(U,V) = \sum_{i=1}^{k}\sum_{j=1}^{n}\left\|x_j - v_i\right\|^2} \qquad (2.1.1)$$

Where $\left\|x_j - v_i\right\|$ is a chosen distance measure between a data point X_j and the cluster centre V_i, is an indicator of the distance of the *n* data points from their respective cluster centers.

A novel initialization algorithm of cluster centres for *K*-means algorithm has been proposed by S. Deelers et al[7]. The algorithm was based on the data partitioning algorithm used for colour quantization. A given data set was partitioned into *k* clusters in such a way that the sum of the total clustering errors for all clusters was reduced as much as possible while inter distances between clusters are maintained to be as large as possible[7]

2.2 Fuzzy C-means Clustering

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters[8]. The traditional FCM algorithm has been used with some success in image segmentation. The FCM algorithm is an iterative algorithm of clustering technique that produces optimal *c* partitions, centres $V=\{v_1, v_2,\dots, v_c\}$ which are exemplars, and radii which defines these *c* partitions. Let unlabelled data set $X=\{x_1, x_2,\dots, x_n\}$ be the pixel intensity where *n* is the number of image pixels to determine their memberships. The FCM algorithm tries to partition the data set *X* into *c* clusters [8]. The standard FCM objective function is defined as follows

$$\boldsymbol{J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n}U_{ik}^{m}\left\|x_k - v_i\right\|^2} \qquad (2.2.1)$$

Where $\|\boldsymbol{x_k - v_i}\|^2$ represents the distance between the pixel x_k and centroid v_i, along with constraint$\boldsymbol{\sum_{i=1}^{c} U_{ik} = 1}$, and the degree of fuzzification $m\geq 1$.

A data point x_k belongs to a specific cluster v_i that is given by the membership value U_{ik} of the data point to that cluster. Local minimization of the objective function $J_m(U,V)$ is accomplished by repeatedly adjusting the values of U_{kj} and v_i according to the following equations[3].

$$\mathbf{U_{ik}} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2}\right)^{\frac{1}{m-1}}} \tag{2.2.2}$$

$$\mathbf{V_i} = \frac{\sum_{k=0}^{n} U_{ik}^{m} X_k}{\sum_{k=0}^{n} U_{ik}^{m}} \tag{2.2.3}$$

Starting with an initial guess for each cluster centre, the FCM converges to a solution for $\mathbf{V_i}$ representing the local minimum or a saddle point of the cost function. Convergence can be detected by comparing the changes in the membership function or the cluster centre at two successive iteration steps.[8].

Keh-Shih Chuang etal[8] proposed spatial FCM that incorporates the spatial information into the membership function to improve the segmentation results. The membership functions of the neighbours centred on a pixel in the spatial domain are enumerated to obtain the cluster distribution statistics. These statistics are transformed into a weighting function and incorporated into the membership function. This neighbouring effect reduces the number of spurious blobs and biases the solution toward piecewise homogeneous labelling. The new method was tested on MRI images and evaluated by using various cluster validity functions. Preliminary results showed that the effect of noise in segmentation was considerably less with the new algorithm than with the conventional FCM.

2.3 Kernelized fuzzy C-means

The standard FCM objective function for partitioning a dataset $\{X_k\}_{k=1}^{N}$ into *c* clusters is given by

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} U_{ik}^{m}\|x_k - v_i\|^2 \tag{2.3.1}$$

Where $\{V_i\}_{i=1}^{C}$ are the centers or prototypes of the clusters and the array $\{U_{ik}\} = U$ represents a partition matrix satisfying

$$U \in \left\{u_{ik} \in [0,1] \middle| \sum_{i=1}^{c} u_{ik} = 1, \forall k \text{ and } 0 < \sum_{k=1}^{N} u_{ik} < N, \forall i\right\} \tag{2.3.2}$$

The parameter *m* is a weighting exponent on each fuzzy membership and determines the amount of fuzziness of the resulting classification. In image clustering, the most commonly used feature is the gray-level value, or intensity of image pixel [14]. Thus the FCM objective function is minimized when high membership values are assigned to pixels whose intensities are close to the centroid of its particular class, and low membership values are assigned when the point is far from the centroid[14].

From the discussion, we know every algorithm that only uses inner products can implicitly be executed in the feature space *F*. This trick can also be used in clustering, as shown in support vector clustering and kernel (fuzzy) c-means algorithm[5]s. A common ground of these algorithms is to represent the clustering centre as a linearly-combined sum of all Φ (xk), *i.e.* the clustering centres lie in feature space. In this section, we construct a novel Kernelized FCM algorithm with objective function as following

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} U_{ik}^{m}\|\phi(X_k) - \phi(V_i)\|^2 \tag{2.3.3}$$

Where Φ is an implicit nonlinear map as described previously. Unlike, $\Phi(V_i)$ here is not expressed as a linearly-combined sum of all $\Phi(X_k)$ anymore, a so-called dual representation, but still reviewed as an mapped point (image) of i v in the original space, then with the kernel substitution trick, we have

$$\begin{aligned}
\| \phi(x_k) - \phi(x_k) \|^2 &= \left(\phi(x_k) - \phi(x_k)\right)^T \left(\phi(x_k) - \phi(x_k)\right) \\
&= \phi(x_k)^T\phi(x_k) - \phi(v_i)^T\phi(x_k) - \phi(x_k)^T\phi(v_i) + \phi(v_i)^T\phi(v_i) \\
&= K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i)
\end{aligned} \tag{2.3.4}$$

Below we confine ourselves to the Gaussian RBF kernel, so $K(x, x) = 1$. From Eq. (2.3.4), Eq.9(2.3.3) can be simplified to

$$J_m = 2\sum_{i=1}^{c}\sum_{k=1}^{N} u_{tk}^{m}\left(1 - K(x_k, v_i)\right) \tag{2.3.5}$$

Formally, the above optimization problem comes in the form

$$\min_{U,\{v_i\}\, i=1}^{\quad c} J_m, \tag{2.3.6}$$

In a similar way to the standard FCM algorithm, the objective function *Jm* can be minimized under the constraint of *U*. Specifically, taking the first derivatives of *Jm* with respect to *uik* and *vi* , and zeroing them respectively, two necessary but not sufficient conditions for *Jm* to be at its local extreme will be obtained as the following

$$\boldsymbol{u_{tk}} = \frac{\left(\mathbf{1}-\boldsymbol{K}(\boldsymbol{x_k},\boldsymbol{v_i})\right)^{-1/(m-1)}}{\sum_{j=1}^{c}\left(\mathbf{1}-\boldsymbol{K}(\boldsymbol{x_k},\boldsymbol{v_j})\right)^{-1/(m-1)}} \qquad (2.3.7)$$

$$\boldsymbol{V_i} = \frac{\sum_{k=1}^{n} \boldsymbol{u_{tk}^{m}}\,\boldsymbol{K}(\boldsymbol{x_k},\boldsymbol{v_i})\boldsymbol{x_k}}{\sum_{k=1}^{n} \boldsymbol{u_{tk}^{m}}\,\boldsymbol{K}(\boldsymbol{x_k},\boldsymbol{v_i})} \qquad (2.3.8)$$

E.A. Zanaty etal had presented [9**]** alternative Kernelized FCM algorithms (KFCM) that could improve magnetic resonance imaging (MRI) segmentation. Then they implemented the KFCM method with considering some spatial constraints on the objective function. The algorithms incorporate spatial information into the membership function and the validity procedure for clustering. We use the intra-cluster distance measure, which is simply the median distance between a point and its cluster centre. The number of the cluster increases automatically according the value of intra-cluster, for example when a cluster is obtained, it uses this cluster to evaluate intra-cluster of the next cluster as input to the KFCM and so on, stop only when intra-cluster is smaller than a prescribe value. The most important aspect of the proposed algorithms is actually to work automatically. Alterative is to improve automatic image segmentation[9]

2.4 Enhanced K-means

Although *K*-means is simple and can be used for a wide variety of data types, it is quite sensitive to initial positions of cluster centres. The final cluster centroids may not be optimal ones as the algorithm can converge to local optimal solutions. An empty cluster can be obtained if no points are allocated to the cluster during the assignment step. Therefore, it is quite important for *K*-means to have good initial cluster centres. The algorithms for initializing the cluster centres for *K*-means have been proposed a new cluster centre initialization algorithm. Hence the enhanced k-means algorithm will be as follows.

1. Read the input image.
2. Decide the number cluster and initialize the cluster centre obtained from cluster centre initialization algorithm.
3. Partitioning the input data points into k clusters by assigning each data point to the closest cluster centroid using the selected distance measure,
4. Computing a cluster assignment matrix U.
5. Re-computing the centroids.
6. If cluster centroids or the assignment matrix does not change from the previous iteration, stop; otherwise go to step 2.

In Research of Shreyansh Ojha it has proven that the Enhanced k-means algorithm is better than the conventional K-Means Clustering Algorithm for colour image segmentation, the validity measure of nearly all the images has been better than the conventional K-Means clustering algorithm, the conventional K-means algorithm uses user defined number of cluster which use to cause noisy image, but in the proposed algorithm, it uses the method for determining the number of optimal cluster. It also removes the problem of empty clusters problem from conventional K-Means clustering algorithm where there was issue that if no pixel is assigned to a cluster then that cluster remains empty.[10]

III. CONCLUSION

Medical image segmentation is fascinating and very important as well. Fuzzy C-Means, K-Means and ,Kernelized Fuzzy C-Means ,Enhanced K-means clustering algorithms have been considered so far they have been seen effective in the image segmentation. They are easy to use unlike some other methods in existence. But there are still limitations that like k-means segmenting with predetermined number of clusters Fuzzy C-means generating an overlapping results and not being able to segment coloured images until they are converted into grey scale .To improve the performance of k-means and fuzzy C-means , Kernelized fuzzy c-means cluster centre initialisation algorithm is used in Enhanced K means algorithm.

IV. FUTURE WORK

In general the clustering algorithm chooses the initial centres in random manner. In future a new centre initialization algorithm for measuring the initial centres of clustering algorithms is used. This algorithm is based on maximum measure of the distance function which is found for cluster centre detection process. The future implementation will be focusing on comparison of different parameters like silhouette score, mean square error, peak signal to noise ratio of these four algorithms K means, fuzzy c mean, Kernelized fuzzy c means , and Enhanced k means.

REFERENCES

[1]. S. Shen, W. Sandham, M. Granat, and A. Sterr, "MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 3, pp. 459-467, 2005

[2]. R. Gonzalez, and R. Woods, *Digital image processing*, 2nd ed., Prentice hall, Upper Saddle River, New Jersey, 2001

[3]. K. S. Fu and J. K. Mu, "A Survey on Image Segmentation," *Pattern Recognition*, vol. 13, pp. 3-16, 1981

[4]. A Generalized Spatial Fuzzy C-Means Algorithm for Medical Image Segmentation Huynh Van Lun and Jong-Myon Kim, Member, *IEEE*, Korea, August 20-24, 2009978-1-.

[5]. Fuzzy k-c-means Clustering Algorithm for Medical Image Segmentation", Ajala Funmilola A, Oke O.A, Adedeji T.O, Alade O.M, Adewusi E.A ,Department of Computer Science and Engineering, LAUTECH Ogbomoso, Oyo state, Nigeria. Journal of Information Engineering and Applications ISSN 2224-5782 (print) ISSN 2225-0506 (online)Vol 2, No.6, 2012

[6]. S. Deelers, and S. Auwatanamongkol, "Enhancing K-Means Algorithm with Initial Cluster Centres Derived from Data Partitioning along the Data Axis with the Highest Variance*", International Journal of Electrical and Computer Engineering 2*:4 2007.

[7]. Fuzzy c-means clustering with spatial information for image segmentation, Keh-Shih Chuang a, Hong-Long Tzeng a,b, Sharon Chen a, Jay Wu a,b, Tzong-Jer Chen c.

[8]. A Kernelized Fuzzy C-means Algorithm for Automatic magnetic Resonance Image Segmentation", E.A. Zanaty ,Sultan Aljahdali ,Narayan Debnath.

[9]. Enhanced K-Means Clustering Algorithm for Color Image Segmentation", Thesis submitted by Shreyansh Ojha at *School Of Mathematics Andcomputer Applications, Thapar University Patiala* – 147004.

[10]. N. Hema, R.Bhavani, "Enhancing K-means and Kernelized Fuzzy C-means Clustering with Cluster Center Initialization in Segmenting MRI brain images", *2011 IEEE*

[11]. Vinod D, S Shrivastava, R. C. Jain, "Clustering Of Image Data Set Using K-Means And Fuzzy K-Means Algorithms", 2010 *International Conference on Computational Intelligence and Communication Networks.*

[12]. Martin Tabakov, "A Fuzzy Clustering Technique for Medical Image Segmentation", *International Symposium on Evolving Fuzzy Systems*, 2006.

[13]. D-Q ZHANG, S-C CHEN, "Kernel-Based Fuzzy Clustering Incorporating Spatial Constraints For Image Segmentation", *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, 2003.

[14]. J.M. Pena ,J.A. Lozano, "An Empirical Comparison of Four Initialization Methods for the K-Means algorithm", *Pattern Recognition Letters* 20 (1999)

[15]. Stephen L. Chiu , "Fuzzy Model Identification Based on Cluster Estimation", *Journal of Intelligent and Fuzzy Systems 2* (1994) 267–278.

[16]. http://crsouza.blogspot.in/2010/03/kernel-functions-for-machine learning. html# kernel_methods.

[17]. Shehroz S.Khan, Amir Ahmad, "Cluster center initialization algorithm for K-means clustering", *Pattern Recognition Letters* 25 (2004) 1293– 1302.

[18]. S.R. Kannan, S. Ramathilagam, "Robust kernel FCM in segmentation of breast medical images", Elsevier Expert Systems with Applications 38 (2011) 4382–4389

[19]. Image Segmentation using k-means clustering, EM and Normalized CutsSuman Tatiraju Department of EECS", Avi Mehta Department of EECS.

[20]. "Kernel-based fuzzy and possibilistic c-means clustering", Dao-Qiang Zhang and Song-Can Chen, Department of Computer Science Nanjing University of Aeronautics and Astronautics Nanjing, 210016, People's Republic of China.

[21]. T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, " k-means clustering algorithm: Analysis and implementation", *IEEE Conf. Computer Vision and Pattern Recognition*, pp.881-892.

AUTHORS BIOGRAPHY

Gunwanti Subhash Mahajan was born in Bhalod, India, in Year 1982. She received the Bachelor in Electronics and Telecommunication degree from the University of Pune, in Year 2005. She is currently pursuing the Master degree with the Department of Electronics And Telecommunication Engineering, Faizpur. Her research interests include Image processing and information theory.

Kanchan S. Bhagat was born in Jalgaon, India, in Year 1974. He received the Bachelor in Electronics and Telecommunication degree from the University of Marathwada, Aurangabad, in Year 1997 and the Master in Electronics and Telecommunication degree from the University of Amravati, in Year 2010, both in Electronics and Telecommunication engineering. His research interest include Image processing.