Gaussian Kernel Based Fuzzy C-Means Clustering Algorithm For Image Segmentation

Rehna Kalam[1], Dr Ciza Thomas[2] and Dr M Abdul Rahiman[3]

[1]Department of Computer Engineering, Kerala University
`rehnakalam@gmail.com`
[2]Department of Electronics Engineering, Kerala University
`cizathomas@gmail.com`
[3]Department of Computer Engineering, Kerala University
`rehman_paika@yahoo.com`

ABSTRACT

Image processing is an important research area in computer vision. clustering is an unsupervised study. clustering can also be used for image segmentation. there exist so many methods for image segmentation. image segmentation plays an important role in image analysis.it is one of the first and the most important tasks in image analysis and computer vision. this proposed system presents a variation of fuzzy c-means algorithm that provides image clustering. the kernel fuzzy c-means clustering algorithm (kfcm) is derived from the fuzzy c-means clustering algorithm(fcm).the kfcm algorithm that provides image clustering and improves accuracy significantly compared with classical fuzzy c-means algorithm. the new algorithm is called gaussian kernel based fuzzy c-means clustering algorithm (gkfcm)the major characteristic of gkfcm is the use of a fuzzy clustering approach ,aiming to guarantee noise insensitiveness and image detail preservation.. the objective of the work is to cluster the low intensity in homogeneity area from the noisy images, using the clustering method, segmenting that portion separately using content level set approach. the purpose of designing this system is to produce better segmentation results for images corrupted by noise, so that it can be useful in various fields like medical image analysis, such as tumor detection, study of anatomical structure, and treatment planning.

KEYWORDS

CLUSTERING, K-MEANS, FCM, KFCM, GKFCM

1. INTRODUCTION

Image segmentation plays crucial role in many applications, such as image analysis and comprehension, computer vision, image coding, pattern recognition and medical images analysis. Many algorithms have been proposed for object segmentation and feature extraction [1]. In this method, a clustering algorithm for medical and other image segmentation will be considered. Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data mining, document retrieval, image segmentation, and pattern classification. However, in many such problems, there is little prior information (e.g., statistical models) available about the data, and the decision-maker must make as few assumptions about the data as possible. It is under these restrictions that clustering methodology is particularly appropriate for the exploration of interrelationships among the data points to make an assessment (perhaps preliminary) of their structure. Clustering is a process of partitioning or grouping a given

set of unlabelled objects into a number of clusters such that similar objects are allocated to one cluster. There are two main approaches to clustering [2].One method is crisp clustering (or hard clustering) ,and the other one is fuzzy clustering. A characteristic of the crisp clustering method is that the boundary between clusters is fully defined. However, in many cases, the boundaries between clusters cannot be clearly defined. Some patterns may belong to more than one cluster. In such cases, the fuzzy clustering method provides a better and more useful method to classify these patterns. The FCM employs fuzzy partitioning such that a data pixel can belong to all groups with different membership grades between 0 and 1.FCM is an iterative algorithm. The aim of FCM is to find cluster centers (centroids) that minimize objective function. The KFCM is derived from the original FCM based on the kernel method [3].KFCM algorithm is extended which incorporates the neighbor term into its objective function [4].Fuzzy clustering is a widely applied method for acquiring fuzzy patterns from data and become the main method of unsupervised pattern recognition. Drawback for FCM algorithm is sensitive to noise or outlier. Drawbacks of FCM were solved by introducing KFCM .In Wu and Gao's paper [5], the Mercer Kernel based method was investigated. They proposed the KFCM algorithm which is extended from FCM algorithm. It is shown to be more robust than FCM .N.A.Mohamed , M.N.Ahmed et al.[6] described the application of fuzzy set theory in medical imaging. In the proposed system, the content set for the various MRI real time images is used to calculate the low intensity area in the in homogeneity form will attain best result for the segmentation and outperforms existing techniques resulting in better accuracy and predicting factor. This method is applicable in different scale of image for different orientation in segmenting the images. Intensity In homogeneity images based Clustering approach is used to overcome the curve in the images, to represent the pure segmented images. Here in previous approach such as Fuzzy c means it fails to target the clustered set point, which fits in the imperfect noisy scaled images in the analysis domain, process of imperfection occurrence in the images due to overlap of the pixel with the different intensity, outcomes lower cluster segmentation in the minimum level for MRI images. GKFCM clustering approach will calculate the estimated parameter automatically for the image data. The clustering process is applied in MRI medical image, for separate group according to their pixel intensity, which is done with the process called Kernel based Fuzzy C means clustering. Where kernel value is selected based on the activities of the membership function. Group of features will selected based on the proper tuning rate of the kernel value, helps in detecting the region separately, level based segmentation is analyzed to detect the intensity region separately, when it comes to in homogeneity Medical images, it is a difficult task for this approach to identify the low intensity region, it can be done by applying the suitable filters to process those images. Initial Impact in clustering of medical images is the drawback in extracting the biological features and it became difficult in identifying the clustered region in similar part of the medical images. Reduction of inhomogeneity in the noisy Medical Images is the extreme end task, and analyzing it feature is open problem and challenging task which yields less attention of approach, which effect the less segmentation accuracy.

2. LITERATURE SURVEY

2.1 K-Means 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 a priori. The main idea is to define k centroids, one for each cluster. These centroids shoud 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 centroid. 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. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 ,$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centres.

ADVANTAGES

1) K-Means algorithm is very fast.

2) It is robust and easier to understand.

3) Relatively efficient in the sense it runs in O(tknd) where t is the number of iterations ,k is the number of

    Clusters ,n is the number of objects and d is the dimension of each object.

DISADVANTAGES

1) K-Means algorithm requires a priori specification of the number of cluster centers.

2) If there are two highly overlapping data then k-means algorithm will not be able to resolve that there are two clusters and is said to be the use of exclusive assignment.

3) It is not invariant to non-linear transformations in the sense we get different results with different representation of data. Data represented in form of cartesian co-ordinates and polar co-ordinates will give different results.

4) It provides the local optima of the squared error function.

5) Randomly choosing of the cluster center cannot lead to the good result.

6) Applicable only when mean is defined.

7) Unable to handle noisy data and outliers.

8) It fails for non-linear data set.

2.2. The Fuzzy C Means Clustering Algorithm(FCM)

The fuzzy c-means (FCM) algorithm is one of the most traditional and classical image segmentation algorithms. The FCM algorithm can be minimized by the following objective function. Consider a set of unlabeled patterns X, let X={x1,,x2,. ..,xN}, x ∈ Rf, where N is the number of patterns and f is the dimension of pattern vectors (features). The FCM algorithm focuses

on minimizing the value of an objective function. The objective function measures the quality of the partitioning that divides a dataset into c clusters. The algorithm is an iterative clustering method that produces an optimal c partition by minimizing the weighted within group sum of squared error objective function. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad , \quad 1 \leq m < \infty$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by

$$U_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right]^{\frac{2}{m-1}}} \qquad\qquad C_j = \frac{\sum_{i=1}^{N} U_{ij}^{m} x_i}{\sum_{j=1}^{N} U_{ij}^{m}}$$

This iteration will stop when $\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| \right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

ADVANTAGES

1) FCM gives best result for overlapped data set.

2) It is comparatively better than k-means algorithm.

3) Data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center whereas in the case of k-means algorithm data point must exclusively belong to one cluster center.

DISADVANTAGES

1) FCM requires a priori specification of the number of clusters.

2) Euclidean distance measures can unequally weight underlying factors.

3) We get the better result with lower value of β but at the expense of more number of iterations.

2.3.The Kernel Fuzzy C Means Clustering Algorithm(KFCM)

The KFCM algorithm adds kernel information to the traditional fuzzy c-means algorithm and it overcomes the disadvantage that FCM algorithm can't handle the small differences between clusters. . The kernel method maps nonlinearly the input data space into a high dimensional feature

space. The essence of kernel-based methods involves performing an arbitrary non-linear mapping from the original d-dimensional feature space Rd to a space of higher dimensionality (kernel space). The kernel space could possibly be of infinite dimensionality. The rationale for going to higher dimensions is that it may be possible to apply a linear classifier in the kernel space while the original problem in the feature space could be highly non-linear and not separable linearly . The kernel method then takes advantage of the fact that dot products in the kernel space can be expressed by a Mercer kernel K. Thus the distance in the kernel space does not have to be explicitly computed because it can be replaced by a Mercer kernel function (typically referred to as a kernel trick). There are two major forms of kernel-based fuzzy clustering. The first one comes with prototypes constructed in the feature space. These clustering methods will be referred to as KFCM-F (with F standing for the feature space). In the second category, abbreviated as KFCM-K, the prototypes are retained in the kernel space and thus the prototypes must be approximated in the feature space by computing an inverse mapping from kernel space to feature space. The advantage of the KFCM-F clustering algorithm is that the prototypes reside in the feature space and are implicitly mapped to the kernel space through the use of the kernel function.

3. GKFCM

Given, $X = \{x_1, \dots x_n\} \subset R^p$, the GKFCM partitions X into c fuzzy subsets by minimizing the following objective function

Equation 1

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

Where c is the number of clusters and selected as a specified value in this paper, n the number of data points, u_{ik} the membership of x_k in class i, satisfying $\sum_{i=1}^{c} u_{ik} = 1$, m the quantity controlling clustering fuzziness, and V the set of cluster centers or prototypes ($v_i \in R^p$). The function J_m is minimized by a famous alternate iterative algorithm.

Now consider the proposed Gaussian kernel fuzzy c-means (GKFCM) algorithm. Define a nonlinear map as

$$\Phi : x \rightarrow \Phi(x) \in F$$

Where $x \in X$. X denotes the data space, and F the transformed feature space with higher even infinite dimension. GKFCM minimizes the following objective function

Equation 2

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

Where

Equation 3

$$\|\phi(x_k) - \phi(v_i)\|^2 = K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i)$$

Where $K(x, y) = \Phi(x)^T \Phi(y)$ is an inner product kernel function. If we adopt the Gaussian function as a kernel function, i.e.,

$K(x, y) = exp(-\|x - y\|^2 / \sigma^2)$, then $K(x, x) = 1$, according to Eqs. (3), Eqs. (2) Can be rewritten as

Equation 4

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

Minimizing Eqs. (4) under the constraint of u_{ik}, we have

Equation 5

$$u_{ik} = \frac{(1/(1 - k(x_k, v_i)))^{\frac{1}{m-1}}}{\sum_{j=1}^{c} (1/(1 - k(x_k, v_i)))^{\frac{1}{m-1}}}$$

Equation 6

$$V_i = \frac{\sum_{k=1}^{n} U_{ik}^m \, K(x_k, v_i) x_k}{\sum_{k=1}^{A} U_{ik}^m \, K(x_k, v_i)}$$

Here we just use the Gaussian kernel function for simplicity. If we use other kernel functions, there will be corresponding modifications in Eqs. (5) and (6).

In fact, Eqs.(3) can be viewed as kernel-induced new metric in the data space, which is defined as the following

Equation 7

$$d(x, y) = \left\| \phi(x_k) - \phi(v_i) \right\| = \sqrt{2(1 - k(x, y))}$$

And it can be proven that d(x, y) defined in Eqs. (7) is a metric in the original space in case that $K(x, y)$ takes as the Gaussian kernel function. According to Eqs. (6), the data point x_k is endowed with an additional weight $K(x_k, v_i)$, which measures the similarity between x_k and v_i, and when x_k is an outlier, i.e., x_k is far from the other data points, then $K(x_k, vi)$ will be very small, so the weighted sum of data points shall be more robust

CLUSTERING ALGORITHM APPLICATIONS

- Clustering Algorithm can be used in Identifying the Cancerous Data Set.
- Clustering Algorithm can be used in Search Engines.
- Clustering Algorithm can be used in Academics.
- Clustering Algorithm can be used in Wireless Sensor Network 's Based Application.
- Clustering Algorithm can be used in Drug Activity Prediction.

4. RESULTS

The experiments and performance evaluation were performed on medical images including a CT image of the MR image of brain The GKFCM clustering and the pro-posed kernel based fuzzy level set method were implemented with Matlab R2013a (MathWorks, Natick, MA, USA) in a Windows 7 System Ultimate. All the experiments were run on a VAIO Precision 340 computer with Intel i3 and 4GB RAM.
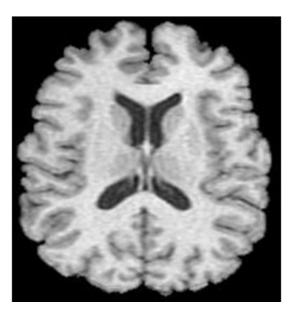

Figure 1 : Original Image

Figure 2 : Cluster 1

Figure 3 : Cluster 2

Figure 4 : Cluster 3

Figure 5 : Cluster 4

5. CONCLUSION

Clustering is one of the efficient techniques in medical and other image segmentation. The primary advantage of the research work is that it includes the kernel method, the effect of neighbour pixel information to improve the clustering accuracy of an image, and to overcome the disadvantages of the known FCM algorithm which is sensitive to the type of noises. The aim of this paper is to propose a new kernel-based fuzzy level set algorithm for automatic segmentation of medical images with intensity in homogeneity. It employs Gaussian kernel-based fuzzy clustering as the initial level set function. It can approximate the boundaries of ROI with parameter estimation simultaneously well. It provides noise-immunity and preserves image details. It can be useful in various fields like medical image analysis, such as tumor detection, study of anatomical structure, and treatment planning

REFERENCES

[1] X. Munoz, J. Freixenet, X. Cufi, and J. Marti, "Strategies for Image Segmentation Combining Region and Boundary Information," PatternRecognition Letters, Vol. 24, No. 1, Pp375–392, 2003.

[2] SteliosKrinidis and VassiliosChatzis," A Robust Fuzzy Local Information C-Means Clustering Algorithm" , IEEE Transactions on Image Processing, Vol. 19, No. 5, MAY 2010.

[3] Gimiami M, "Mercer kernel based clustering in feature space", IEEE Transactions on Neural Networks, Vol. 3, No. 3, Pp780-784, 2002.

[4] Yang Y., Zheng Ch., and Lin P., "Fuzzy c-means Clustering Algorithm with a Novel Penalty Term for Image Segmentation", Opto-Electronics Review, Vol.13, No.4, Pp.309-315, 2005.

[5] Wu Z, Xie,W.X Yu J.P. "Fuzzy C-means Clustering Algorithm Based on Kernel Method" In: Proceedings of Fifth International Conference on Computational Intelligence and Multimedia Applications Pp 49- 56,2003.

[6] Lee Song Yeow, "Spatial Kernel-Based Generalized C mean Clustering for Medical Image Segmentation", School of Computer Sciences, University Sainsmalasia, Dec 2010.

[7] Huynh Van Lung and Jong-Myon Kim, "A generalized spatial fuzzy C-means algorithm for Medical image segmentation", In proc. 18th Int. Conf. on Fuzzy Systems, pp.409-414, 2009.

[8] Deng-Yaun Haung, Ta- Wei Lin, Wu-Chih-Hu," Automatic Multilevel Threshold Based on Two Stage Otsu's Method With Cluster Determination With Valley Estimation" , ICIC International @ 2011 ISSN 1349-4198, pp. 5631-5644.

[9] Chun-yan Yu, Ying Li, Ai-lian Liu, Jing-hong Liu, "A Novel Modified Kernel Fuzzy C-Means Clustering Algorithm on Image Segmentation", IEEE International Conference on Computational Science and Engineering CSE/I-SPAN/IUCC 2011.

[10] F. Gibou and R. Fedkiw, "A fast hybrid k-means level set algorithm for segmentation," in Proceedings of the 4th Annual Hawaii International Conference on Statistics and Mathematics, pp. 281–291, 2002.

[11] T. Saikumar, B. Shashidhar, V. Harshavardhan, and K. S. Rani, "MRI brain image segmentation algorithm using watershed transform and kernel fuzzy C-means clustering on level set method," International Journal on Computer Science and Engineering, vol. 3, pp. 1591–1598, 2011.

[12] G. R. Reddy, K. Ramudu, S. Zaheeruddin, and R. R. Rao, "Image segmentation using kernel fuzzy c-means clustering on level set method on noisy images," in Proceedings of the International Conference on Communications and Signal Processing (ICCSP '11), pp. 522–526, ind, February 2011.

[13] M. Rastgarpour and J. Shanbehzadeh, "Automatic medical image segmentation by integrating KFCM clustering and level set based FTC model," in IAENG Transactions on Electrical Engineering, Special Issue of the International Multi Conference of Engineers and Computer Scientists World Scientific, vol. 1, pp. 257–270, 2013.

[14] M. Rastgarpour, M. Rastgarpour, S. Alipour, and J. Shanbehzadeh, "Improved fast two cycle by using KFCM clustering for image segmentation," in Proceedings of the 7th International Multiconference of Engineers and Computer Scientists, Lecture Notes in Engineering and Computer Science, pp. 678–682, Hong Kong, China, 2012.

[15] Jain, Anil K. "Data clustering: 50 years beyond K-means." Pattern recognition letters 31.8 (2010): 651-666.

[16] Lifflander J, E Meneses, H Menon, P Miller, S Krishnamoorthy, and LV Kale. 2014. "Scalable replay with partial-order dependencies for message-logging fault tolerance." In Proceedings of the 2014 IEEE International Conference on Cluster Computing (CLUSTER), pp. 19-28. September 22-26, 2014, Madrid, Spain. Instituteof Electrical and Electronics Engineers, Piscataway, New Jersey,2014.

[17] Jiang,D.,Tang, C. & Zhang, "A Cluster analysis for gene expression data: A survey". IEEE Transactions on Knowledge and Data Engineering 16,1370-1386,2004.

[18] Prim, H.,Eksioglu,B.,Perkins, A, D. & Yuceer, C."Clustering of high throughput gene expression data."Computers & Operation Research 39,3046-3061 ,2012.

[19] Von Luxburg, U.,Williamson, R. C.& Guyon , I."Clustering : Science or art?" In ICML Unsupervised and Transfer Learning, 65-80,2012.

[20] Arbelaitz, O.,Gurrutxaga, I.Muguerza, J.Perez, J. M.& Perona, I."An extensive comparative study of cluster validity indices" Pattern Recognition 46,243-256,2013

AUTHOR

Rehna Kalam born in 1982 is a full time research scholar at Kerala University. She received the B.Tech degree in Information Technology from Kerala University in 2005 and the M.Tech degree in Computer Science and Engineering from Anna University, Coimbatore in 2011.Co Authors are Dr Ciza Thomas, Professor, Department of Electronics Engineering, College of Engineering, Trivandrum and Dr M Abdul Rahiman, Professor, Department of Computer Science Engineering, LBS Institute of Technology for Women, Poojappura, Trivandrum.