

Image segmentation using K-mean clustering for finding tumor in medical application

Piyush M. Patel¹, Brijesh N Shah², Vandana Shah³,

¹PG student, EC Dept, CSPIT, CHARUSAT – Changa,

²Prof. & HOD., EC Dept, CSPIT, CHARUSAT – Changa,

³ Asst. professor, Department of EC Engineering ,SCET- Surat,

Abstract — Clustering algorithms have successfully been applied as a digital image segmentation technique in various fields and applications. However, those clustering algorithms are only applicable for specific images such as medical images, microscopic images etc. In this paper, we present a new clustering algorithm called Image segmentation using K-mean clustering for finding tumor in medical application which could be applied on general images and/or specific images (i.e., medical and microscopic images), captured using MRI, CT scan, etc. The algorithm employs the concepts of fuzziness and belongingness to provide a better and more adaptive clustering process as compared to several conventional clustering algorithms.

Keywords — *clustering methods; K-means segmentation; partitioning methods; hierarchical methods; density-based methods; centroid method.*

INTRODUCTION

Clustering can be considered the most important unsupervised learning problem [12], so, it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other

clusters. Clustering algorithms may be classified as listed below :

- Exclusive Clustering
- Overlapping Clustering
- Hierarchical Clustering
- Probabilistic Clustering

In the first case data are grouped in an exclusive way, so that if a certain datum belongs to a definite cluster then it could not be included in another cluster. On the contrary the second type, the overlapping clustering, uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, data will be associated to an appropriate membership value. A hierarchical clustering algorithm is based on the union between the two nearest clusters. The beginning condition is realized by setting every datum as a cluster. After a few iterations it reaches the final clusters wanted.

MATERIALS AND METHODS

K-means segmentation

The K-means algorithm implements a divisive clustering and was first discussed by Duda and Hart [2]. The algorithm uses a similarity metric to assign all documents to one of k clusters. The clusters are represented as an average of all documents

contained within the cluster. This average can be thought of as the centroid of the cluster.

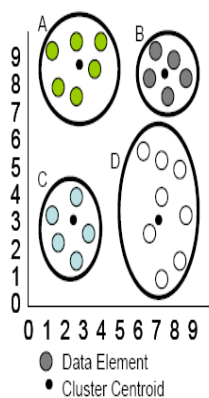


Fig.1 Example of K-means output with K equal to four in two dimensions

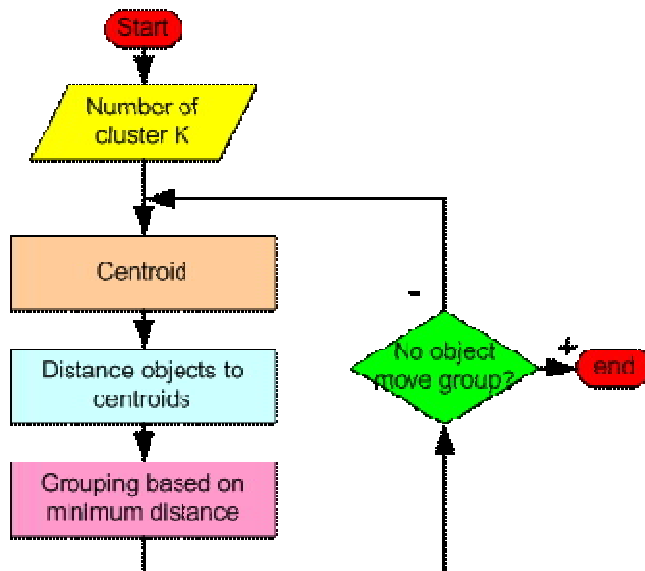
A simple two dimensional case for K-means clustering is shown. The K-means algorithm set with $k = 4$ results in four clusters represented by A, B, C, and D. The K-means algorithm operates as follows:

1. Assign document vectors, $d_i \in D$, to a cluster using an initial seed.
2. Initialize cluster centroids, C , from initial document assignments.
3. For each document $d \in D$ (a) Recalculate distances from document d_i to centroids (C_1, C_2, \dots, C_k), and find the closest centroid C_{min} .
(b) Move document d from current cluster C_k into new cluster C_{min} and re-calculate the centroid for C_k and C_{min} .
4. Repeat step 3 until either the maximum epoch limit is reached or an epoch passes in which no changes in document assignments are made. An epoch is a complete pass through all documents.

The initial seed clusters can be either assigned or generated by randomly assigning documents to clusters. K-means has been used in the clustering of images. A hardware implementation of K-means to cluster hyper spectral images was created

by Estlik et. al. and eeser et. al. Hyper spectral images have 224 16-bit channels which can be thought of as features to cluster per pixel .

Proposed flow diagram



Proposed approach

Original K-means algorithm choose k points as initial clustering centers, different points may obtain different solutions. In order to diminish the sensitivity of initial point choice, we employ a mediod [11], which is the most centrally located object in a cluster, to obtain better initial centres. The demand of stochastic sampling is naturally bias the sample to nearly represent the original dataset, that is to say, samples drawn from dataset can't cause distortion and can reflect original data's distribution. In order to lessen the influence of sample on choosing initial starting points, following procedures are employed. First, drawing multiple sub-samples (say J) from original dataset (the size of each sub-sample is not more than the capability of the memory, and the sum for the size of J sub-samples is as close as possible to the size of original dataset) . Second, use K-means for each sub-sample and producing a group of medioids respectively. Finally, comparing J solutions and

choosing one group having minimal value of square-error function as the refined initial points.

To avoid dividing one big cluster into two or more ones for adopting square-error criterion, we assume the number of clustering is K' ($K > K'$, K' depends on the balance of clustering quality and time). In general, bigger K' can expand searching area of solution space, and reduce the situation that there are not any initial value near some extremum. Subsequently, re-clustering the dataset through K-means with the chosen initial conditions would produce K' mediods, then merging K' clusters (which are nearest clusters) until the number of clusters reduced to k .

Results and Discussion:

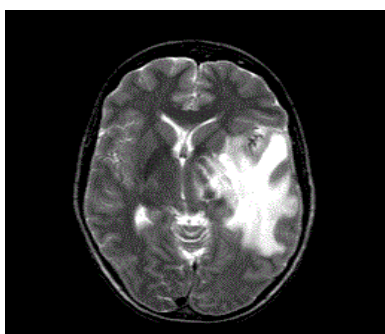


Fig 2: Input image of Brain

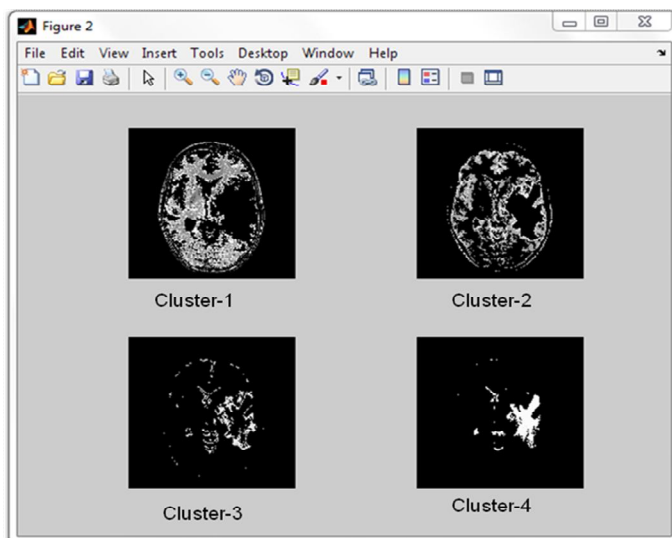


Fig 3: Analysis of brain tumor using K-mean algorithm

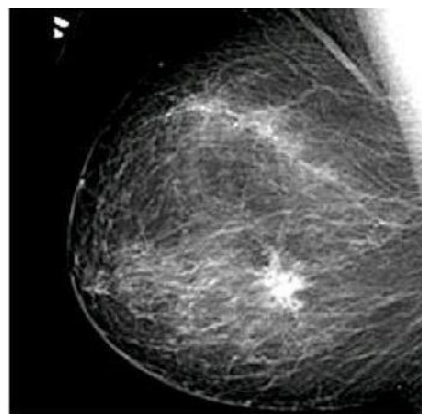


Fig 4: Input image of Breast

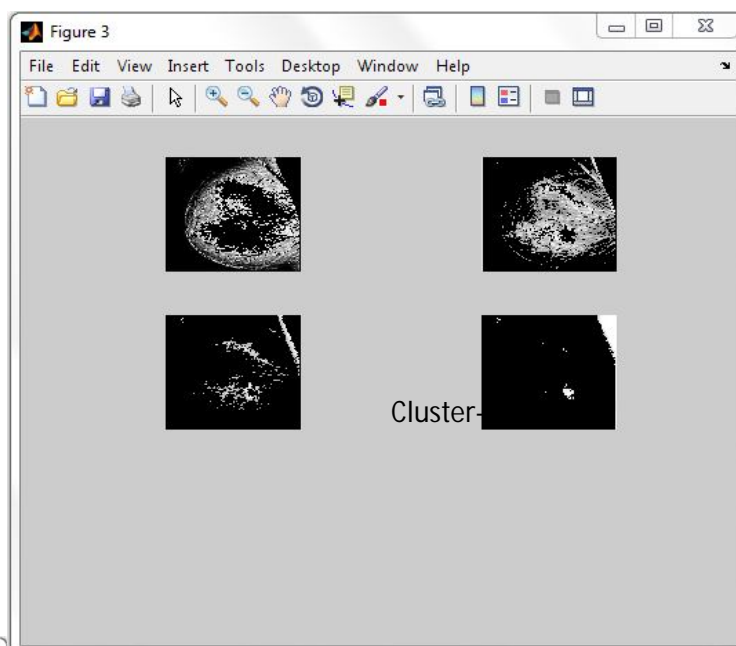


Fig 5: Analysis of breast tumor using K-mean algorithm

Conclusion:

K-means algorithm is a popular partition algorithm in cluster analysis, which has some limitations when there are some restrictions in computing resources and time, especially for huge size dataset. The improved K-means algorithm presented in this paper is a solution to handle large scale data, which can select initial clustering centre purposefully, reduce the sensitivity to isolated point, avoid dissevering big cluster, and overcome deflexion

of data in some degree that caused by the disproportion in data partitioning owing to adoption of multi-sampling.

REFERENCES

- [1] M. Masroor Ahmed and Dzulkifli Bin Mohamad "Segmentation of Brain MR Images for Tumor Extraction by Combining Kmeans Clustering and Perona-Malik Anisotropic Diffusion Model" International Journal of Image Processing, Volume (2) Issue(1),20 10.
- [2] Arash Azim Zadeh Irani and Bahari Belaton "A K-means Based Generic Segmentation System" 2009 Sixth International Conference on Computer Graphics, Imaging and Visualization
- [3] Ming-Ni Wu, Chia-Chen Lin and Chin-Chen Chang "Brain Tumor Detection Using Color-Based K-Means Clustering Segmentation" International Conference on Computer Graphics, 2009
- [4] S.Sathish Kumar, M.Moorthi, M.Madhu and Dr.R.Amutha "An Improved method of Segmentation using Fuzzy-Neuro Logic" Second International Conference on Computer Research and development,2009
- [5] S. Murugavalli and V. Rajamani " An improved Implementation of Brain tumor Detection Using Segmentation Based on Neuro Fuzzy Technique", Journal of Computer Science 3 ,2007 october.
- [6] Kharrat, A. Benamrane, N. Ben Messaoud, M. Abid, Comput. & Embedded Syst. Lab. (CES), Nat. Eng. Sch. of Sfax, Sfax, Tunisia, Nov. 2009

Fig.1 Example of K-means output with K equal to four in two dimensions

Fig 2: Input image of Brain

Fig 3: Analysis of brain tumor using K-mean algorithm

Fig 4: Input image of Breast

Fig 5: Analysis of breast tumor using K-mean algorithm

LIST OF FIGURES