ISSN No. 0976-5697

Volume 3, No. 2, March-April 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Performance Evaluation of K-Means and Fuzzy C-Means Clustering Algorithms for Identification of Hematoma in Brain CT scan Images

Bhavna Sharma*
Department of Informatics and Computational Sciences
MLSU,Udaipur,India
bhavna.jecrc@gmail.com

Prof. K. Venugopalan
Department of Informatics and Computational Sciences
MLSU, Udaipur,India
venumlsu@gmail.com

Abstract: Clustering is the assignment of a set of observations into subsets, called clusters so that observations in the same cluster are similar in some sense. This research paper deals with two of the most delegated clustering algorithms namely centroid based K-Means and Fuzzy C-Means for identification of hematoma in brain CT scan. The behaviour of both the algorithms depends upon the brain CT images as well as on the number of clusters. The performance of both the algorithms is investigated during different executions on the input images. The execution time for each algorithm is also analyzed and the results are compared with one another.

Keywords: Fuzzy C-Means Clustering, K-Means Clustering.

I. INTRODUCTION

Segmentation is required as a preliminary step in the analysis of medical images for computer aided diagnosis due to the fact that brain has a complex structure and its precise segmentation is very important for detecting tumors, edema, hematoma or any other abnormality. Segmentation is the process of partitioning a digital image into multiple segments based on pixels [1]. It might be used for object recognition, image compression, image editing, etc. The quality of the segmentation depends upon the digital image [2]. Brain image segmentation can be further aided by clustering algorithms. Clustering is a concept to determine the pattern through map and analysis of available data set according to the need such that data belonging to a same cluster are similar and data from different clusters are dissimilar. Clustering can either be supervised clustering including the hierarchical approaches such as relevance feedback techniques [3] or unsupervised clustering including the density based clustering methods. In supervised clustering method, grouping is done according to user feedback whereas in unsupervised clustering, the images with high features similarities to the query may be very different in terms of semantics [4].

Medical imaging techniques for brain scan like MRI, CT scan, PET scan etc. are the tools used for extracting information by the radiologist. As compared to the other techniques used, CT scan is preferred because of wide availability, low cost and better contrast. Majority of medical images are gray scale representation, so segmentation is based on the gray level value of pixels. In Brain CT scan images for identification of hematomas [5] automatic and fast segmentation of image is a preliminary step and thus choosing the accurate segmentation from various available methods is a crucial task. This paper compares two standard K-Means and Fuzzy C-Means clustering algorithms in order to find out the best technique for identifying brain hematoma in CT scan images.

II. METHODOLOGY

The two unsupervised clustering methods, namely K-Means and Fuzzy C-Means are based on dividing the image in clusters according to the distance between data points and cluster centers. Number of clusters is chosen randomly by user.

A. K-Means Algorithm:

K-Means is one of the simplest unsupervised clustering algorithms that solve the well known problem [6].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. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. At this point it is necessary to re-calculate k new centroids as bar centers of the clusters resulting from the previous step. After obtaining 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, one 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.

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

Where $\left\|x_i^{(i)} - c_j\right\|^2$ is a chosen distance measure between a data point and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centers. The algorithm is composed of the following steps [7]:

Step 1: Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.

Step 2: Assign each object to the group that has the closest centroid.

Step 3: When all objects have been assigned, again calculate the positions of the K centroids.

Step 4: Repeat Steps 2 and 3 until the centroids no longer move.

This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

B. Fuzzy C-Means Algorithm:

Fuzzy C-Means is an unsupervised clustering algorithm that has been applied to wide range of problems involving feature analysis, clustering and classifier design [8]. Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. In this algorithm, data are bound to each cluster by means of a membership function which represents the fuzzy behavior of the algorithm. The algorithm builds an appropriate matrix named U whose factors are numbers between 0 and 1, and represent the degree of membership between data and centers of clusters. It is based on minimization of the following objective function:

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

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:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}} \qquad (3)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \qquad (4)$$

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 and k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m. The algorithm is composed of the following steps [9]:

Step 1: Initialize U= [u_{ij}] matrix, $U^{(0)}$

Step 2: At k-step: calculate the centers vectors $C^{(k)} = [c_j]$ with $U^{(k)}$

Step 3: Update $U^{(k)}$, $U^{(k+1)}$

Step 4: If $\| U^{(k+1)} - U^{(k)}\| < \varepsilon$ then STOP otherwise return to step 2.

III. RESULT AND DISCUSSION

This paper portrays the results of an objective evaluation of two popular segmentation techniques: K-Means Clustering and Fuzzy C-Means Clustering based segmentation algorithm. The key factors in determining the performance of the segmentation algorithm in object detection system is Correctness and Completeness [10].Another important factor is Stability. Segmentation methods are implemented in MATLAB 7.1 on window XP computer with 2.80 GHz CPU and 1 GB of RAM. Input images are publicly available brain CT scan images and all the images are of same quality. Segmentation was done manually on the original image by the radiologist. The correct boundary of the brain hematoma is identified to obtain a ground truth. A logical AND operation was performed between the ground truth (GT) and the resultant image obtained after segmentation to obtain true positive image (TP). Next the difference between segmented image and true positive image was taken as the false positive image (FP). The difference between manual and automatic segmentation gives False negative (FN), it is the number of pixels that are not covered by the automatic segmentation. Original image of brain CT scan and results of applying K-Means clustering and Fuzzy C-Means clustering with 3 clusters are shown in Fig.1.

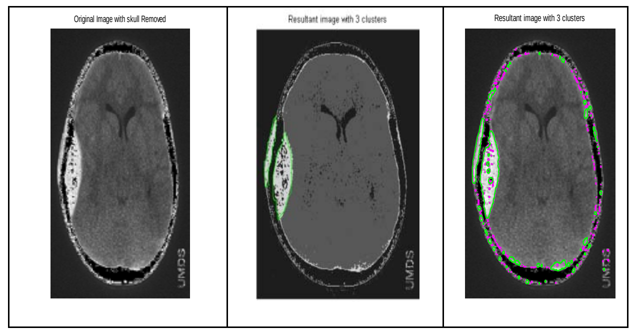

Figure.1: Segmentation of original Brain CT scan image using K-Means and Fuzzy C-Means Algorithm

The first comparison performed is by considering the correctness of the algorithms .Correctness can be defined as the percentage of correctly segmented region by the segmentation algorithm and can be calculated as

$$\text{Correctness (CR)} = \frac{TP}{TP+FP} \qquad (5)$$

Values of CR can be ranged from 0 to 1. The value 1 implies that the region extracted by the segmentation algorithm is 100% correct. CR values less than 1, denotes that the algorithm has under-segmented the region by (100-CR) %. On an average in terms of correctness, K-Means algorithm performed much better than the Fuzzy C-means algorithm as shown in Table I.

The second comparison performed is considering the Completeness, it can be defined as the percentage of the true region (GT) extracted by the segmentation algorithm and can be calculated as

$$\text{Completeness (CM)} = \frac{TP}{TP+FN} \qquad (6)$$

Values of CM can be ranged from 0 to 1. The value 0 represents that none of the pixels belongs to GT region is properly segmented by the algorithm, and 1 represent that all the pixels belongs to the GT region are segmented. In the context of CM, values less than 1 denotes that the algorithm has over-segmented the region by (100-CM) %. In terms of correctness, K-Means algorithm performed slightly better than the Fuzzy C-means algorithm shown in Table I.

A more important performance evaluation parameter of segmentation algorithms is stability. If an algorithm gives reasonably correct segmentations on average, but is wildly unpredictable on any given image or with any given parameter set, it will be useless as a preprocessing step. There are two basic types of stability, one is stability with respect to parameters and another is stability across images. Stability with respect to parameters refers to achieving consistent results on the same image given different parameter inputs to the algorithm. Stability across images refers to achieving consistent results on different images given the same set of parameters.

The third comparison that we performed was considering stability with respect to parameters (number of clusters). In this comparison, K-Means algorithm performs well (up to a limit on number of clusters) than the Fuzzy C-Means algorithm. K-Means showed less variability but stability decrease rapidly in Fuzzy C-means with changing values of k (number of cluster). This comparison can be seen in Figure 2-Figure 7, where each cluster's pixel values are shown as in the original image and rest of the image is painted black. The brightest cluster is hematoma region and by further analysis of that cluster, the type of hematoma can be detected by its shape. Finally, we compared the stability of a particular parameter choice over the set of images i.e. fixing cluster size k=3.Fuzzy C-Means and K-Means segmentation methods are much closer here, both performing significantly well on set of images with cluster size k=3.

The fourth comparison performed was the computation time taken by each algorithm. Generally the computation time varies from processor to processor. In this study, we have used, i3 processor, MATLAB 7.3 with 1GB of RAM. On an average applying both the algorithms on number of images of same quality but hematoma of different types, K-Means clustering algorithm takes 2-3 seconds to perform and segments the image in prescribed number of clusters but Fuzzy C-Means clustering takes 1-3 minutes to perform segmentation, moreover this time increases with increasing the number of clusters.

Table I: Parameters for comparing K-Means and Fuzzy C-Means Clustering Algorithm

Segmentation method	Correct-ness (CR)	Complete-ness (CM)	Stability	CPU time
K-Means	95%	98%	Stability decreases slowly with increasing no of clusters	2-3 sec
Fuzzy C-Means	75%	88%	Stability decreases rapidly with increasing no of clusters	1-3 min

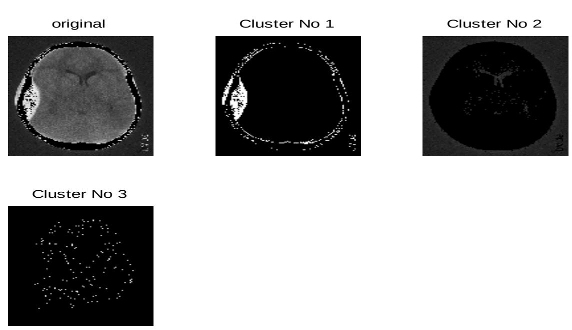

Figure 2: Segmentation of original brain CT scan image using K-Means having cluster size k=3

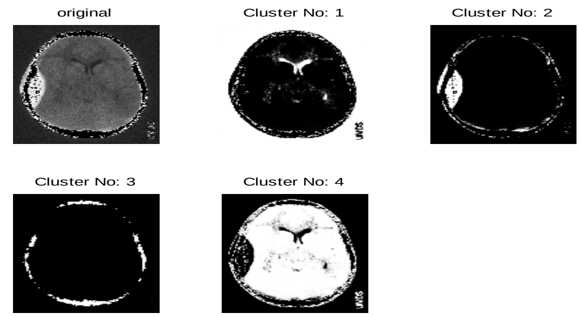

Figure 3: Segmentation of original brain CT scan image using K-Means having cluster size k=4

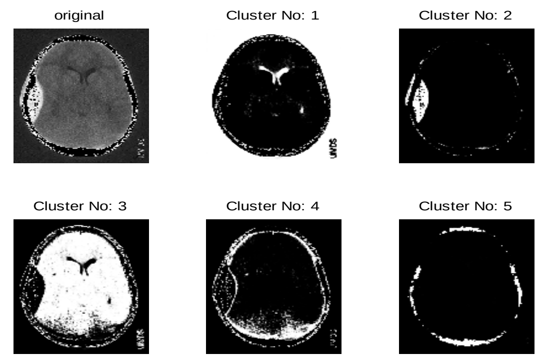

Figure 4: Segmentation of original brain CT scan image using K-Means having cluster size k=5

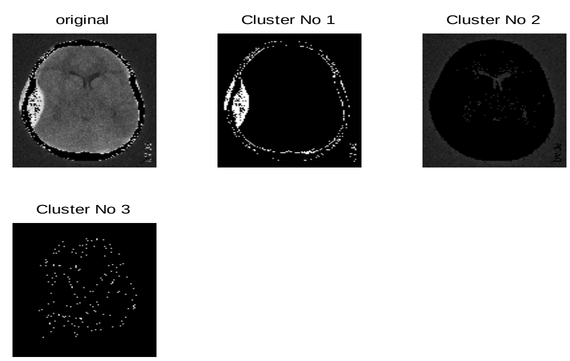

Figure 5: Segmentation of original brain CT scan image using Fuzzy C-Means having cluster size k=3

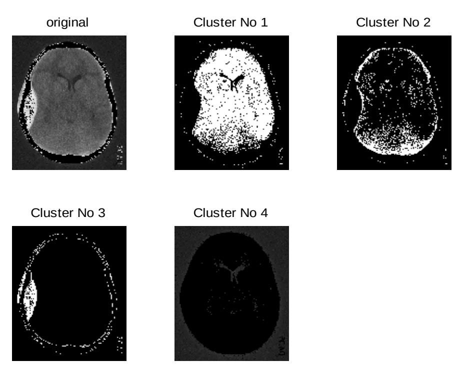

Figure 6: Segmentation of original brain CT scan image using Fuzzy C-means having cluster size k=4

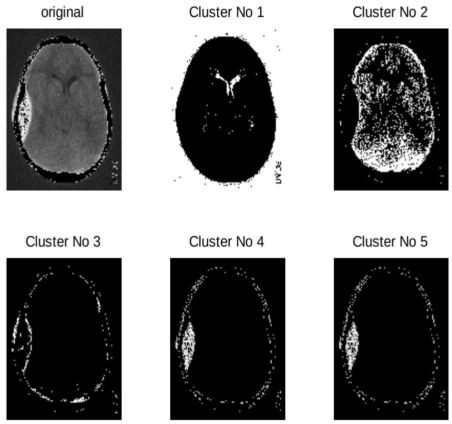

Figure 7: Segmentation of original brain CT scan image using fuzzy C-Means having cluster size k=5

IV. CONCLUSION

It is very evident from the results that the performance of the K-Means algorithm is better than that of FCM for identification of hematoma in brain images because it gives clean clusters and segments the image correctly and completely.FCM produces close results to K-Means clustering, yet it requires more computation time because of the fuzzy measures calculations involved in the algorithm. Moreover a defuzzification step is required to get clean clusters. Both the algorithms are sensitive to initial number of clusters chosen, therefore our future research can be directed towards the automatic detection of the exact number of clusters from intensity values in the original image and then applying K-Means clustering algorithm to get a correct, complete and stable segmented image of brain hematoma.

V. ACKNOWLEDGMENT

We would like to thank Dr. Kapil Vyas, Radiologist, Geetanjali Medical College, Udaipur and Dr. Shailendra Pareek, Fortis Hospital, Jaipur, for their support lent to us during this study.

VI. REFERFENCES

[1] Rafael C. Gonzalez, Richard E. Woods, 2007. Digital Image Processing, 2nd ed., Beijing: Publishing House of Electronics Industry.

[2] N.R. Pal, S.K. Pal,“A Review on Image Segmentation Technique ”, Pattern Recognition 1993,26(9), pp. 1277-1294..

[3] D.L. Pham, C. Xu, J.L. PRINCE, “Current Methods in Medical Image Segmentation”, Annual Review of Biomedical Engineering 2000 2(1), pp. 315-338.

[4] W.L. Zhang, X. Z. Wang, “Extraction and Classification for Human Brain CT Images”, Proceedings of Sixth International Conference on Machine learning and cybernetics 2007, IEEE, 1155-1156.

[5] Alexander Rakhlin , Andrea Caponnetto, “Stability of K-Means Clustering”, Advances in Neural Information Processing Systems 12, MIT Press, cambridge, MA, 2007, pp. 216–222.

[6] Zhou Liping and Gao Xinbo, “Image Segmentation via fast Fuzzy C-Means Clustering,” Journal of Computer Engineering and Application 2004, vol. 40(8), pp.68–70.

[7] T. Velmurugan, T. Santhanam, "Performance Evaluation of K-Means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points”, European Journal of Scientific Research 2010 vol.46 no.3, pp.320-330.

[8] Yong.Y, Z. Chongxun, L. Pan, “A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding”, Measurement Science Review 2004, vol 4(1).

[9] I. Shanthi, M. L.Valarmathi, "Comparison of K-Means and Fuzzy C-Means Clustering Algorithms for SAR Image Despeckling using Edge Detection”, European Journal of Scientific Research 2011,vol.62, no.3, pp.426-463.

[10] Gajanayake,G.M.N.R. ,Yapa,R.D.,Hewawithana,B.,”Comparison of Standard Image Segmentation Methods for Segmentation of Brain Tumors from 2D MR Images”, In Proceedings of the 4th International Conference on Industrial and Information Systems, ICIIS 2009,IEEE, University of Peradeniya, Sri Lanka, pp. 301-305.