International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue1 | ISSN : 2456-3307

Tumour Detection Using Clustering Methods

Sukanya Roy*, Prateek Banerjee, Prasun Biswas, Aditi Chatterjee ,Ritwik Goswami, Sagnik Kanta, Suman Paul

Department of CSE, University of Engineering & Management(UEM), Kolkata, Bengal, India

ABSTRACT

As we know that brain tumour detection is a very crucial task, so in this paper we have tried to focus on this matter using image clustering, which is a process of grouping together of pixels having similar attributes. Previously many researches have been done in this field using clustering methods. Here we will use some of the famous clustering methods like the K-means Clustering and the Fuzzy C-means Clustering for the said purpose of tumour detection.

Keywords: Image Segmentation, Pixels, Cluster, K-Means Clustering, Fuzzy C-Means Clustering.

I. INTRODUCTION

Image segmentation is the process of partitioning an image into non-intersecting regions where each region is homogeneous and the union of two adjacent regions is heterogeneous. Every pixel in an image is allocated to one of these categories. A good segmentation is one in which:

1. Pixels In The Same Category Have Similar Grayscale Of Multivariate Values And Form A Connected Region.
2. NeighbouringPixels Which Are In Different Categories Have Dissimilar Values.

This method (image segmentation) is one of the mostly used methods to classify the pixels of an image correctly in a particular manner. It is a valuable tool in many field including health care, image processing, traffic image, pattern recognition etc. There are various methods for image segmentation like threshold based, edge based, cluster based, and neural network based. Amongst these the most efficient means of image segmentation is the cluster based method.

Now, there are different types of cluster based methods like: k-means clustering, mountain clustering, fuzzy c-means clustering, and subtractive clustering. From these different methods the most efficient ones are k-means clustering and fuzzy c-means clustering. K-means method is comparatively simple to implement and computationally faster than the hierarchical clustering. There are other types for the k- means clustering known as the improved k-means clustering and for the fuzzy c-means known as the improved fuzzy c-means clustering.

II. RELATED WORKS

Earlier many works have been done in this particular field i.e. of image segmentation for the sake of brain tumour detection by using different types of image clustering methods. Amongst which many of them have been done based on the different applications of image segmentation. K-means algorithm and Fuzzy C- means are some of the simplest clustering algorithms and many other methods are also being implemented with different approach to initialize the centre. Nowadays many researchers are also trying to produce new methods with the motive of making them more efficient than the existing ones, to produce better segmented results. Some of the existing recent works are discussed here.

Based on the traditional fuzzy c-means method, the neighbourhood attraction is shown to improve the segmentation performance. The neighbourhood attraction depends on two factors on, firstly, relative location and features of neighbouring pixels in the image. The real brain magnetic resonance(MR) images and the simulated ones are segmented to demonstrate the superiority of the proposed method compared to the conventional fuzzy c-means method [1].

RajshekharGhogge proposed brain tumour detection using k-means and fuzzy c-means algorithm under Morphological Image Processing and based on accurate Fast Bounding Box Based Method. In this paper because of the use of the above mentioned methods, it produces reliable results which are also less sensitive to error[2].

Alan Jose, S. Ravi and M. Sambath, proceeded by dividing the process into a total of three parts which are as follows: pre-processing of the image, advanced k-means and fuzzy c-means and at last, the feature extraction. The pre-processing is implemented by using the filter where the quality of the image is improved. Then sequentially the advanced K-means algorithm is used, followed by Fuzzy c-means to cluster the image. Then the resulted image segment is used for the feature extraction for the region of interest. In this process they have used a MRI image for the analysis and calculated the size of the extracted tumour region in the image[3].

In this paper, they proceeded with the approach of template based k-means and modified/improved fuzzy c-means. The template selection is based on convolution between the gray level intensity of a small portion of the brain's image and the brain tumours image. The concept of using the k-means algorithm along with the improved fuzzy c-means algorithm really led fruitful results with less chances of error [4].

An enhanced algorithm is proposed here, to improve the accuracy and efficiency of the k-means clustering algorithm. Here it shows an enhanced k-means algorithm combined with a systematic method consisting of two approaches. First one is finding the initial centroid and second one is assigning the data point to the clusters. Different initial centroids were considered and tested for execution time and accuracy. We can conclude from the results that the proposed algorithm reduced the time complexity without sacrificing the[5].

III. PROBLEM DESCRIPTION

We know that, an image can be defined as a two dimensional function f(x, y), where x and y are planar coordinates and the amplitude of the function f recorded for any pair of co-ordinates say (x, y) is called the intensity level or gray level of the image at that point. The area of interest of the image might get degraded by the impact of any imperfect instrument, the problem with data acquiring process and by interference of natural phenomena. Thus the original image becomes unsuitable for analysis. That's why image segmentation technique is necessary and should be taken as a significant step when processing or analysing an image. The problem addressed here is that, we spilt an image into a number of segmentations, then we choose an appropriate model having better segmentation with reduced computational time. The problem is rectified with minimized computational time and high quality of the results.

IV. CLUSTERING METHODS

A.Fuzzy C-Means Clustering

In Fuzzy C-Means Clustering an iterative search is done for a set of fuzzy clusters and the associated cluster-centres representing the structure of the data in the best manner possible. The algorithm requires the user to specify the number of clusters present in the set of data to be clustered. Say there are c number of clusters, fuzzy c-means partitions the data $X = \{x_1, x_2, \ldots, x_n\}$ into c fuzzy clusters by minimizing the group sum of squared error objective function as follows:

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

where $J_m(U,V)$ is the sum of squared error for the set of fuzzy clusters is represented by the membership matrix U, and V represents the set of associated cluster-centres. $||x_k - v_i||^2$ represents the distance between the data x_k and the cluster centre v_i. The squared error representing the performance index measures the weighted sum of distances between cluster-centres and the elements present in the corresponding fuzzy clusters. The number m controls the influence of membership grades in the performance index. The partition becomes fuzzier with the increase in m and it has been observed that the fuzzy c-means clustering algorithm converges for any $m \in (1,\infty)$. The necessary conditions for Eq. (1) to reach its minimum are:

$$U_{ik} = \left(\sum_{j=1}^{c}\left(\frac{||x_k - v_i||}{||x_k - v_j||}\right)^{2/(m-1)}\right)^{-1} \forall i, \forall k \quad (2)$$

$$\text{and} v_i = \frac{\sum_{k=1}^{n}(U_{ik})^m x_k}{\sum_{k=1}^{n}(U_{ik})^m} \quad (3)$$

In each iteration of the algorithm, the matrix $U$is computed using Eq. (2) and the associated cluster-centres are computed as Eq. (3). This is followed by computing the square error in Eq. (1). The algorithm stops when either the error is below a certain tolerance value or its improvement over the previous iteration is below a certain threshold.

Some of its advantages are as follows:
1. It is unsupervised.
2. Converges.

Some of its limitations are as follows:
1. It takes longer computational time.
2. Sensitivity to the initial guess (speed, local minima) .
3. Sensitivity to noise.

- One expects low or even no membership degree for outliers (noisy points).

B.K-Means Clustering

K-Means Clustering is one of the simplest unsupervised learning algorithm. The key is to find groups of given data set through a certain number of clusters (say K) i.e. we have to define K centres, one for each cluster. These centres should be placed in an appropriate manner because different locations of the centres produces different result. So, it will be handy if we place them far away from each other as much as possible. This algorithm works in an iterative manner to take each point belonging to a given data set and associate it to the nearest centre. Now, we need to re-calculate K numbers of new centroids as barycentre of the clusters resulting from the previous step. After acquiring these K new centroids, a new binding is needed to be done between the same data-set points and the nearest new centre. Because of the loop it is noticed that the K centres keep changing their location until no more changes are done i.e. the centres do not move any more. Finally, this algorithm focuses on minimizing the squared error function which is given by:

$$J(V) = \sum_{i=1}^{c}\sum_{j=1}^{c_i}\left\|x_i - v_j\right\|^2 \quad (4)$$

where,

‖ x_i- v_j ‖ is the Euclidean distance between xi and vj.

'ci' is the no. of data points in ith cluster.

'c' is the no. of cluster centres.

Some of its advantages are as follows:

1. It is efficient in processing large data sets.
2. It takes less computational time in comparison to the other ones.
3. This algorithm performs better when the clusters arein form of compact clouds rather than well separated from each other.

Some of its disadvantages are as follows:

1. The number of clusters are needed to be specified inadvance.
2. This algorithm/method is noise sensitive and alsosensitive to outlier elements.
3. If the clusters are in non-convex shapes then thismethod will not be suitable.
4. This algorithm is applied when the mean/averageof a cluster is defined.

V. PROPOSED METHODS

A. Improved Fuzzy C-Means Clustering

(1) Read the data from the given (MRI) image.

(2) Select the number of clusters and cluster centres randomly.

(3) Initialize the fuzziness factor (say m=3).

(4) for i = 1 to max_iter

- Calculate the distance between pixel and centroid.
- Calculate the membership values (Uik).

End for.

- Update cluster centroid.
- Determine objective function.

(5) Do the segmentation performance threshold = New_Centroid - Old_Centroid

(6) Do the computations of the fuzzy centroids.

(7) Compute the new fuzzy membership functions.

B. Improved K-Means Clustering

K-Means(dataset, k, dissimilarity_measure, replicates)

(1) Read the data from the given (MRI) image.

(2) Select 'K' number of initial centroids randomly.

(3) Repeat, until the number of replicates/centroids do

not change.

- Construct 'K' number of clusters as per the given dissimilarity_measure.
- Compute the centroid of each clusters again.

(4) End;

VI. RESULT AND DISCUSSION

A complete program written using MATLAB (R2017a) programming language, Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz, and a 512 x 416 grayscale MRI brain image have been used.

A. For K-Means Clustering

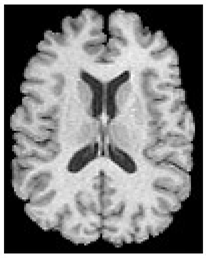
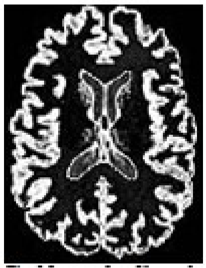

Figure 1. Original MRI image**Figure 2.** Image after Cluster 1


Figure 3. Image after Cluster 2**Figure 4**. Image after Cluster 3

The conventional K-means algorithms produces the result after the completion of 18 iterations.

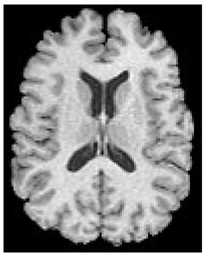
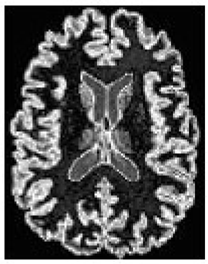

Figure 5. Original MRI image **Figure 6**. Image after Cluster 1

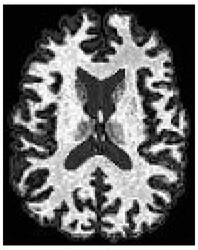

Figure 7. Image after Cluster 2 **Figure 8**. Image after Cluster 3

The conventional Fuzzy C-means algorithm produces the result after the completion of 50 iterations.

After observing the results we can say that K-Means algorithm was much faster compared to the Fuzzy C-Means for the considered datasets containing the clusters in regular or irregular patterns. Fuzzy C-Means is an algorithm which is based on more iterative fuzzy calculations than the K-Means, so its execution was found to be comparatively higher.

VII. CONCLUSION

The proposed work is used for segmentation of brain MRI images and to classify them. Two types of algorithms have been implemented here: K-Means and the Fuzzy C-Means. By observing the results we can conclude that the features used for classification are applicable for both the algorithms, we can also summarize that the segmentation done by means of particular region of interest and the number of iterations can be validated.

Even though these algorithms produce a desired result, there is no such devised algorithm which will prove to be best for all the cases. Thus, to overcome this issue, the only thing that we can do is remain careful while examining the datasets for shapes and scatter of clusters in order to decide a suitable algorithm.

In near future, we hope that researches in MRI segmentation should focus on improving the accuracy, computational speed, and precision of these algorithms along with reducing the amount of manual interactions. This is of utmost importance because, MR imaging is becoming a routine diagnostic procedure in clinical practice. We can now only wait for the time when the clustering algorithms will be so modified that it will produce much more accurate results of the MRI image which will in turn prove to be helpful for diagnosis, and will also assist the doctors in making a decision of, whether to go for a surgery or not.

VIII. REFERENCES

[1]. Z.-P. Liang and P.C. Lauterbur, "Principles of Magnetic Resonance Imaging: A Signal Processing Perspective". The Institute of Electrical and Electronics Engineers Press, 2000.

[2]. Rajshekhar Ghogge, "Brain tumour detection using K-Means and Fuzzy C-Means Clustering Algorithm". International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 7, July 2014.

[3]. Alan Jose, S. Ravi and M. Sambath, "Brain Tumour Segmentation using K –means Clustering and Fuzzy C-means Algorithm and its Area Calculation". International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, issue 2, March 2014.

[4]. Rasel Ahmmed and Md. Foisal Hossain, "Tumour Detection in Brain MRI Image Using Template based K-means and Fuzzy C-means Clustering Algorithm". 2016 International Conference on Computer Communication and Informatics (ICCCI -2016), Jan. 07 – 09, 2016, Coimbatore, INDIA.

[5]. K.A. Abdul Nazeer and M.P. Sebastian, "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm". Proceedings of the World Congress on Engineering, London, WCE, vol. 1, July (2001).

[6]. B. Sathya and R. Manavalan, "Image Segmentation by Clustering Methods: Performance Analysis". International Journal of Computer Applications (0975 – 8887) Volume 29No.11, September 2011.