Integrated Intelligent Research (IIR) International Journal of Data Mining Techniques and Applications
Volume: 06, Issue: 02, June 2017, Page No.56-61
SSN: 2278-2419

An Efficient Logical Average Distance Measure Algorithm (LADMA) to Analyse MRI Brain Images

A.Naveen[1], T.Velmurugan[2]
[1]Research Scholar, [2]Associate Professor
PG and Research Department of Computer Science, D. G. Vaishnav College, Arumbakkam, Chennai, India
E-Mail: naveenking@yahoo.co.in, velmurugan_dgvc@yahoo.co.in

Abstract - Malignant and benign types of tumor infiltrated in human brain are diagnosed with the help of an MRI scanner. With the slice images obtained using an MRI scanner, certain image processing techniques are utilized to have a clear anatomy of brain tissues. Some of such data mining technique is k-means and fuzzy C means algorithms. This work proposes a new hybrid algorithm namely LAMDA, which offers successful identification of tumor and perform well for the segmentation of tissue regions in brain. Automatic detection of tumor region in MR (magnetic resonance) brain images has a high impact in helping the radio surgeons assess the size of the tumor present inside the tissues of brain and it also supports in identifying the exact topographical location of tumor region. Experimental results show that the proposed approach reduces the number of features and at the same time it achieves high accuracy level. The observed results to achieve high accuracy level using minimum number of selected features.

Keyword: k-Means Algorithm, Fuzzy C means Algorithm, LAMDA Logical Average Distance Measure Algorithm, and Brain Tissues Segmentation.

I. INTRODUCTION

Data mining (DM) is the discovery of interesting, unexpected or valuable structures in large datasets. Clustering is a division of data into groups of similar objects. Representing the data by fewer clusters necessarily loses certain fine details, but achieves simplification. It models data by its clusters. Data modeling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learning, and the resulting system represents a data concept.

Clustering is a descriptive task that seeks to identify homogeneous groups of objects based on the values of their attributes (dimensions) [24] [25]. Clustering techniques have been studied extensively in statistics [3], pattern recognition [ll] [19], and machine learning [9] [31]. The popular K-means, K-medoid, Fuzzy K Means methods determine K cluster representatives and assign each object to the cluster with its representative closest to the object such that the sum of the distances squared between the objects and their representatives is minimized. Therefore, clustering is unsupervised learning of a hidden data concept. Data mining deals with large databases that impose on clustering analysis additional severe computational requirements. DM of medical image is used to collect effective models, relations, rules, changes, irregularities and common laws from volume of data. This procedure can accelerate the processing accuracy of the diagnosis decisions made by doctors. With the rapid development of digital medical devices, medical information databases have included not only the structured information of patients, but also plenty of non-structured medical image information. These provide a rich data source for medical image data mining. This paper is structured as follows section 2: the concepts of dataset, preprocessing, clustering methods for k-Means, FCM algorithm and LAMDA. Section 3 explains results k-means, FCM and LAMDA our proposed method finally conclusion part is as section 4.

II. METHODOLOGY

A. Dataset

This research work uses MRI brain images taken from normal and abnormal patient images at Swamy Vivekananda Diagnostic Centre (SVDC) Hospital in Chennai at D. G. Vaishnav College campus. An abnormality of the MRI brain images was marked by SVDC head. The MRI brain images in DICOM format is taken for analysis. DICOM means Digital Imaging and Communications in Medicine. It is an International Standard for medical images and related information of DICOM defines the formats for medical images. The DICOM file format support the encapsulation of any information object definition. This type of images can hold images with patient information like age, sex, modality, study description, date of image taken, image size and type etc. Normal and abnormal data are processed in this methodology. The MRI brain images were divided into 8 and 16 clusters by the clustering algorithm.

B. Preprocessing

The preprocessing is an improvement of the image data that suppresses unwanted distortions or enhances some images features important for further processing [9]. It is one of the most critical steps in a data mining process which deals with the preparation and transformation of the initial dataset. Data preprocessing methods are divided into following categories data cleaning, data integration, data transformation and data reduction [1, 2]. The steps involved in this work are applying preprocessing techniques [43] Region of Interest (ROI), Inverse Method and Edge Detection for boundary from the image.

C. k-Means Algorithm

The k-Means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem [15]. Since, k-mean clustering [4] is normally introduced to group a set of data points {x1, x2,……,xn} into k clusters [16]. It has high computational efficiency and can support multidimensional vectors. So it reduces the distortion measurement by minimizing a cost function as:

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

where

$\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster center 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:

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, recalculate 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. The k-means is simple clustering algorithm that has been improved to several problem domains.

D. Fuzzy C Means Algorithm

The fuzzy C-means (FCM) algorithm uses fuzzy clustering method based to minimize the quadratic criterion where clusters are formed by signified a particular centres [3]. Fuzzy clustering is a powerful unsupervised technique to analysis the data and to construct effective models [7]. In several situations, fuzzy cluster algorithm is further regular than hard clustering. The main limitation of FCM algorithm is all about its sensitivity towards noises. FCM implements the clustering task for a data set by minimizing an objective-function subject to the probabilistic constraints that the summation of all membership degrees that result the data point to be focused to one centre. This constraint results in the problem of membership assignment, that is noises are treated the same as points which are close to the cluster centers. The FCM clustering algorithm allows one piece of data to belong to two or more clusters [8]. This method is widely used in pattern recognition. It is based on minimization of the following objective function.

$$F_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \, || xi - c_j ||^2, 1 \le m < \infty \quad (2)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{|| x_i - c_j ||}{|| x_i - c_k ||} \right]^{\frac{2}{m-1}}}, \; cj = \frac{\sum_{i=1}^{N} u_{ij}^{m} . x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \quad (3)$$

Where m is any real number greater than 1, x_i is the i^{th} of d-dimensional measured data, c_j is the d-dimension center of the cluster, u_{ij} is the degree of membership of x^i in the cluster j, and ‖ * ‖ is any norm expressing the similarity between any measured data and the center [17]. 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.

This iteration will stop when $\max \; y\{| u_{ij}^{(k+1)} - u_{ij}^{(k)} |\} < \xi$ where ξ is a termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of Fm. The algorithm is composed of the following steps:

Step 1: Initialize U = [u_{ij}] matrix, $U^{(0)}$
Step 2: At k-step: calculate the centers vectors C(k) = [cj] with U(k)
Step 3: Update U(k), U(k+1)
Step 4: If $||U^{(k+1)} - U^{(k)}|| < \xi$ then STOP; otherwise return to step 2.

E. LAMDA Algorithm

The algorithm designed with two primary categories of algebraic operations AND, OR are applied into clustered image datasets [136]. The corresponding positional pixel values are embedded one with another to apply these operations on cluster results. LADMA method is used for the detection and prediction of optimized significant region in a brain MRI medical dataset. The clusters are formed using range based density values [138]. The logical operations of the two clustering techniques have been examined on the beginning of clustering quality and efficiency of the methods.
This LADMA accumulated using given mathematical equation

$$\text{LCM} \quad I = \sum_{i=0}^{n} \sum_{j=0}^{m} \left| \frac{KI_{ij} \, S_{ij} / F_{ij}}{\left[KI_{11} \, S_{11} / F_{11} + KI_{12} \, S_{12} / F_{12} + \dots\dots\dots \; ..KI_{nm} \, S_{nm} / F_{nm} \right]} \right|^{1/nm} .(3)$$

The LADMA techniques are derived by using above mathematical equation 4. The steps involved to cluster the raw data by the Layer cluster model procedure are given below
Step 1: Initialize $I_{ij} = [I]$ matrix, $I(0)$ from Preprocessed

Step 2: To identify the best cluster point using range based density values $R = \min\left[|I_1|,|I_2|,|I_3|.........\ ...|I_n|\right]$ where I_i represents range index of cluster i.

Step 3: Integrate difference of $D_{ij} = \sum_R \left(\sum_{j=0}^{n} dif\ (I_j) \right)$ image between i^{th} cluster and $(i+1)^{th}$ cluster similarity results

Step 4: Determine the common region based on logical Operation AND

Step 5: Determine the total segment using OR

Step 6: Till $i>n$ then Stop; otherwise return to step 2.

III. RESULTS AND DISCUSSIONS

This research work aimed to determine the significant area of brain tumor using medical image analysis with data mining clustering techniques. This clustering process identifies the significant area on the affected area through measuring the average distance of high density pixels with short time. The current techniques such as k-means and Fuzzy C-Means methods are evaluated and the identification accuracy and time factors are enhanced using the Logical Average Distance Measure Algorithm. The comparative study for the clustering, computation of distance measure and its performance are presented in this result sections. The comparative results variation of cluster analysis and approaches are tabulated in 1. The comparative results shows with respect to approaches are quality based on pixels, memory space and run time.

This algorithm will lead to determine significant area of MRI brain image. A common and significant region is identified by LADMA method and then it produces an output which is better than the other results. Similarly, the process continues until it identifies an appropriate affected image. The process time and space complexity are calculated for the *k*-Means, FCM and LADMA methods. The LADMA method will predict affected region of different pixels are given as number of significant objects. The various steps of entire process are discussed as follows.

Step 1: Given the original image for pre-processing.

Step 2: Apply RoI covariance method.

Step 3: Apply Inverse method to identify the appropriateness of the image

Step 4: Use Boundary method to detect edges.

Step 5: Segment the image using range based on different clustered average density values.

Step 6: Identify the best clustered image, on range based average density values of i^{th} cluster.

Step 7: The best range based average density values of $(i+1)^{th}$ clustering image.

Step 8: Determine the common segment using logical operations.

Step 9: Next, change the number of images into 6 and 8 and continue the process.

Step 10: The results are tabulated in table 1 along with figures.

The clustered image exhibits similar high density region which represents irregular image. The determined common irregular images are unified using logical operators as part of LADMA

Table 1: Comparative Results

Process	Algorithm	Clusters 4	Clusters 6	Clusters 8
Number of Pixels	k-Means	4397	1972	800
	FCM	8946	8349	7417
	LADMA	8886	7960	6733
Memory Size in KB	k-Means	6.40	5.69	4.89
	FCM	6.71	6.64	6.36
	LADMA	6.5	5.96	5.56
Run time in Seconds	k-Means	0.977995	0.847398	1.072045
	FCM	27.44898	42.32934	55.79677
	LADMA	14.33849	21.26686	28.7340

Algorithm	Cluster 4	Cluster 6	Cluster 8

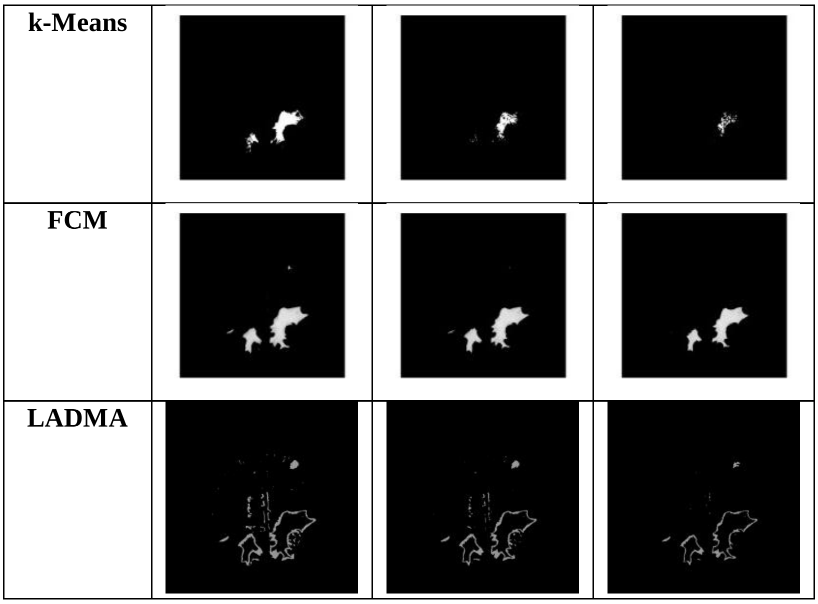

Figure 1: Comparative Output

The figure 1 shows the results of clustering algorithm with respect to segment results produced by the clustering algorithms. The designed model was developed and tested with collected data set. As a first step, the DICOM format of images preprocessed using region of interest, inverse method and boundary detection methods. The second step is used to process the MRI images using the clustering techniques namely k-Means, FCM and LADMA. These methods are applied to find the clusters of MRI images by dividing the image into 4, 6 and 8 groups as number of clustering. The clustering results of abnormal images after the identification of tumor affected regions are done by considering the number of clusters as 4, 6 and 8. One of the best outputs of these three is selected for further use. The results of k-Means, FCM and LADMA methods by taking the number of clustering as 4, 6 and 8 are shown in table 2.

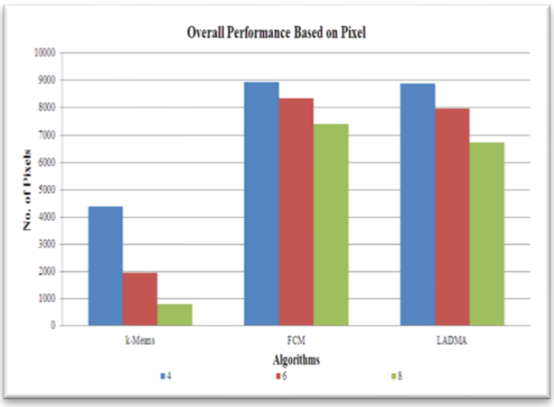

Figure 2: Performance Based on Pixels

It is very trivial to identify the produced output of LADMA is better than the other results. The processing time and space complicity by the k-Means and FCM and LADMA algorithms are tabulated in table 6.5. The pictorial representation is shown in figures 2, figures 3 and in figure 4 for number of pixel, time and memory space respectively. It is easy to identify that the runtime is very less in LADMA algorithm compared with k-Means and FCM algorithm. But, the memory required for the LADMA is little high compared with other two algorithms.

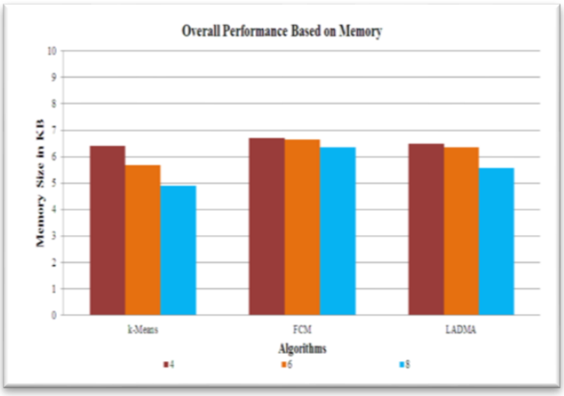

Figure 3: Performance Based on Memory

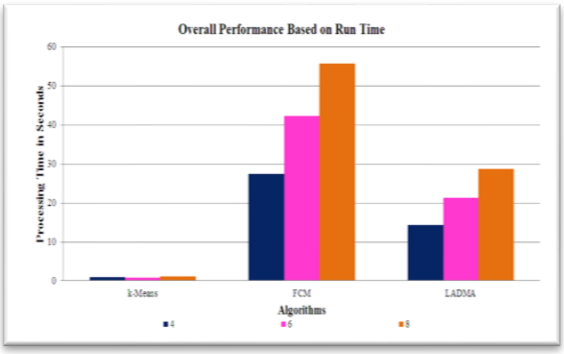

Figure 4: Performance Based on Run Time

The proposed LADMA clustering model is applied on the different clustering with the intension of finding the tumor affected regions in brain images. This method shows tumor area in various clustering distance based on the approach used in LADMA. This process, the resulted output is integrated using LADMA algorithm, from the three outputs; compares the best output which exactly identified the tumor affected area. In this sequence, LADMA algorithm was applied to locate the tumor affected region in MRI brain data. To categorize the significance of affected region in the MRI images, the clustering algorithm produced result is verified by means of identifying tumor region with intensity of images.

IV. CONCLUSION

In this paper provides the concept of clustering algorithms and logical average distance measure algorithm which is designed, developed and evaluated with the k-Means FCM. The significant areas identified in medical images and the clustering quality is discovered based on number of pixel that are computed for the same sources with different logics. It leads to the identification of the common average distance and the existence of the pixel using the logical operator and confirmed it's significant for medical feature analysis.This research technically evaluated the clustering algorithm embedding process for the verification of digital image analysis. The clustering techniques are applied to identify tumor affected region in brain MRI brain images. The major proposed methodology used in this work namely LADMA identifies the tumor affected area in the brain MRI images efficiently. The approaches used here not only evaluate the performance of clustering algorithms, but also measure the clustering quality and other parameters such as execution time and volume complexity. The time taken for the analysis of MRI images by LADMA algorithms is very less when compared with the k-Means and FCM algorithms.

Reference

[1] Jyoti Soni, et al., "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications, 17(8), 2011.

[2] Kaur, H., Wasan, S. K., "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), pp.194-200, 2006.

[3] P. Schrocter and J. Bigun. Hierarchical image segmentation by multi-dimensional clustering and orientation-adaptive boundary refinement. Pattern Recognition, 25(5), pp.695-709, 1995.

[4] Tuan, Tran Manh. "A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation." Expert Systems with Applications 46, pp.380-393, 2016

[5] Fritz H, Garcia-Escudero LA, Mayo-Iscar A. TCLUST: an R package for a trimming approach to cluster analysis. Journal of Statistical Software. 47, pp.1–26, 2012

[6] Nguyen HD, McLachlan GJ, Cherbuin N, Janke AL. False discovery rate control in magnetic resonance imaging studies via Markov random fields. IEEE Transactions on Medical Imaging. 33, pp. 1735–1748, 2014

[7] Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10(2), pp.191–203

[8] Etehadtavakol M, Sadri S, Ng EYK (2010) Application of K- and fuzzy c-means for color segmentation of thermal infrared breast images. J Med Syst 34(1), pp. 35–42

[9] Thanh, Nguyen Dang, and Mumtaz Ali. "A novel clustering algorithm in a neutrosophic recommender system for medical diagnosis." Cognitive Computation 9(4),) pp.526-544, 2017.

[10] Al-Dmour, Hayat, and Ahmed Al-Ani. "MR Brain Image Segmentation Based on Unsupervised and Semi-Supervised Fuzzy Clustering Methods", International Conference on Digital Image Computing: Techniques and Applications, pp. 1-7, 2016.

[11] Li, Hua, Chuanbo Chen, Shaohong Fang, and Shengrong Zhao. "Brain MR image segmentation using NAMS in pseudo-color", Computer Assisted Surgery, 22(1) pp.170-175, 2017.

[12] A.Naveen, T.Velmurugan, "Identification of Calcification in MRI Brain Images by k-Means Algorithm", Indian Journal of Science and Technology, 08 (29), pp. 1-7, 2015.

[13] Wu, Dongfang. "Study on the Technical Architecture of the Computer Data Mining Platform based on the Hadoop Framework", Boletín Técnico, 55(18), 2017

[14] Ding, Jiajun, Zhongtian Chen, Xiongxiong He, and Yizhao Zhan. "Control Conference Clustering by finding density peaks based on Chebyshev's inequality", pp. 7169-7172, 2016.

[15] Gomathy, "Disease Predicting System Using Data Mining Techniques" International Journal of Technical Research and Applications, 1(5), pp. 41–45, 2013

[16] A.Naveen and T. Velmurugan, "A Novel Layer Based Logical Approach (LLA) Clustering Method for Performance Analysis in Medical Images", International Journal of Control Theory and Applications, 9(40), pp. 99-109, 2016.

[17] Ngai, E. W. T., Yong Hu, Y. H. Wong, Yijun Chen, and Xin Sun. "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature", Decision Support Systems, 50(3), pp. 559-569, 2011.

[18] Rygielski, Chris, Jyun-Cheng Wang, and David C. Yen. "Data mining techniques for customer relationship management", Technology in society, 24(4), pp. 483-502. 2002.

[19] Grzymala-Busse, Jerzy W., and Ming Hu. "A comparison of several approaches to missing attribute values in data mining", International Conference on Rough Sets and Current Trends in Computing, pp. 378-385. 2000.

Volume: 06, Issue: 02, June 2017, Page No.56-61
SSN: 2278-2419

[20] Tomar, Divya, and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare', International Journal of Bio-Science and Bio-Technology, 5(5), pp. 241-266, 2013.

[21] Gupta, Anamika, Naveen Kumar, and Vasudha Bhatnagar. "Analysis of medical data using data mining and formal concept analysis", World Academy of Science, Engineering and Technology, 11, pp. 61-64, 2005.

[22] Sundar, N. A., Latha, P. P., and Chandra, M. R. "Performance analysis of classification data mining techniques over heart disease database", International Journal of engineering science & advanced technology, 2(3), pp.470-478, 2012.

[23] Lehmann, Thomas M., Mark O. Güld, Thomas Deselaers, Daniel Keysers, Henning Schubert, Klaus Spitzer, Hermann Ney, and Berthold B. Wein., "Automatic categorization of medical images for content-based retrieval and data mining", Computerized Medical Imaging and Graphics, 29(2), pp. 143-155, 2005.

[24] Khanmohammadi, Sina, Naiier Adibeig, and Samaneh Shanehbandy. "An improved overlapping k-means clustering method for medical applications", Expert Systems with Applications 67, pp.12-18, 2017.