

CLUSTERING OF BRAIN MRI IMAGE USING DATAMINING ALGORITHM

¹SIJI T.MATHEW, ²NACHAMAI M

¹Ph.D research scholar, Dept. of Computer Science Christ University, Bangalore, ²Associate Professor, Dept. of Computer Science Christ University, Bangalore

E-mail: siji.mathew@res.christuniversity.in, nachimai.m@christuniversity.in

Abstract- Data mining is the exploration and study of large quantities of data in order to discover original, valid, understandable and potentially useful patterns in given data. Image data plays vital role in different systems like business, education, medical, engineering, etc. Image mining comes under data mining where the extraction or mining knowledge is from vast amount of image data. The image mining process involves analyzing the image data from different perception and precise it into useful information by applying algorithms, tools and techniques. These algorithms and tools act as a way to understand and study various relationships, associations and patterns which are hidden in the image data. Medical images contain valuable information about human anatomy. New technologies help to capture and store images in different formats in corresponding databases. These images need further extractions through which useful information can be obtained. The extracted information is valuable as it can affect in an action, a decision, or an outcome. In this paper an attempt is made for clustering brain MRI images using K-Means algorithm. A comparative study on clustering with K-Means algorithm and Fuzzy C-Means algorithm was also done with the MRI image dataset.

Keywords- Fuzzy C-Means , Data mining , image clustering, K-Means, MRI.

I. INTRODUCTION

Data is available in numerous forms in the world. Data contains information in raw format. By arranging data in a systematic and organized way, the uncertainty about the data is reduced and it gives a clear picture of the data for a given context. A vast amount of image data gets generated every day from different medical fields. Medical images are categorized as ultrasound imaging, X-ray, Magnetic Resonance Imaging (MRI), ElectroEncephaloGraphy (EEG), ElectroCardioGraphy (ECG), Computed Tomography (CT), etc.

This work is aimed to compare the existing data mining algorithms for clustering the medical images. K-means algorithm and Fuzzy Cmeans algorithms were used for comparing the clustered output MRI image. The rest of the paper is organized as follows: Section 2 discusses related work in image mining. Section 3 presents the tools and methods that were used for this work. A detailed result and discussion is given in Section 4 and finally, Section 5 concludes the paper.

II. LITERATURE REVIEW

Medical images are having an important role to identify the functional and anatomical information of the human body for various types of diagnosis. These images can be used in medical research and education. An image retrieval system is a computer system for searching, browsing and retrieving images from large databases. The research and development helps to bring out many new technologies in medical field. Nowadays X-rays, ECG, EEG, MRI and CT are available in digital image format also, which is user

friendly for researchers. They have become a standard for diagnosis and treatment in many health related issues. Image mining includes the extraction of knowledge or patterns or relationships, which are not explicitly stored in the images. In image mining two approaches are followed. In the first approach, knowledge is mined from a large collection of images. In the second approach knowledge is mined from the combined collections of images and its associated data. In paper an image mining algorithm in which blobs were used to perform the association mining from the images. In paper author mined the associations between the structures and the functions of human brain using rule mining. Storing and representing the image data takes huge space. The image compression method compress the amount of data required to represent an image and hence reduces the storage space and transmission requirements.

In diagnosing several diseases, a clear visualization of internal structure of human body is required; MRI is used for this purpose. This imaging technique provides high quality images which can be used for various elementary diagnoses of certain diseases. The MRI image texture contains feature information such as brightness, color, slope and size. Diagnosis of brain tumor from brain images in medical field can speedup with the help of (Computer-aided diagnosis) CAD. The human analysis of CT-scan images, without CAD becomes tedious and there is a chance of misinterpretation. In order to avoid human faults and errors in the interpretation of images, the hospitals and medical centers recommend CAD systems.

Collecting all the important objects from the images is done with the preprocessing. In an image

processing system, the preprocessing deals with the noise reduction and enhancement of images. Shape priori algorithm was used for preprocessing of the images. It considered the curve shapes in the images and corrected the image following an iterative step. After the preprocessing the relevant features are extracted using association rule mining. Transactional databases are used to store the extracted features. The decision tree classifier is used to classify the transactional database objects in to three different classes called normal or benign or malignant. In the proposed system for identifying brain tumor the brain images were taken as CT-scan. These images have been proven to be a better way to detect the brain tumor as it gives a good accuracy value than other methods. This had given the efficient features than other preprocessing methods. Association rule-mining algorithm was used for getting the best way of identifying the rule for classification phase. According to the class labels, decision tree classifier classifies the rules and this classification assists the physicians for taking better decisions.

III. METHODOLOGY

In this section a brief explanation about the tool, dataset and the algorithms are given.

A. TOOL

The implementations of K-Means and Fuzzy C-means algorithms were done in MATLAB. MATLAB environment with image processing toolbox supports DICOM files. Development of computer-assisted medical image retrieval is associated with the use of new intelligent capabilities such as multimedia support and data mining in order to discover the relevant knowledge for diagnosis in medical field. If the results of data mining can be communicated or interpreted to humans in an understandable way, then the result can be used as a resource for knowledge. The MRI characteristics will help the physicians to avoid the human error in manual interpretation of medical image content.

B. DATASET

MRI image of brain tumor was downloaded from BRAINIX, an open source database provided by OSIRIX. This database is available open for research works. BRAINIX contains both T1 weighted and T2 weighted MRI images in DICOM format. DICOM is Digital Imaging and Communications in Medicine. It is the international standard for medical images and related information (ISO 12052). DICOM defines the formats for medical images. It contains metadata about the image. DICOM is implemented in almost every medical imaging. The radiotherapy and cardiology imaging device produce DICOM images. Now the X-ray, ultrasound, MRI and CT images are available in DICOM format. DICOM holds more than hundred of different information about the picture in

an invisible form. 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, which are useful for clinical use and research. Everyday thousands of medical images are taken. DICOM has enabled advanced medical imaging and many researches with these types of DICOM images are going on. The BRAINIX contains 22 brain tumor images. This cancer dataset is read and displayed in Matlab. The screen shot of the dataset is shown below

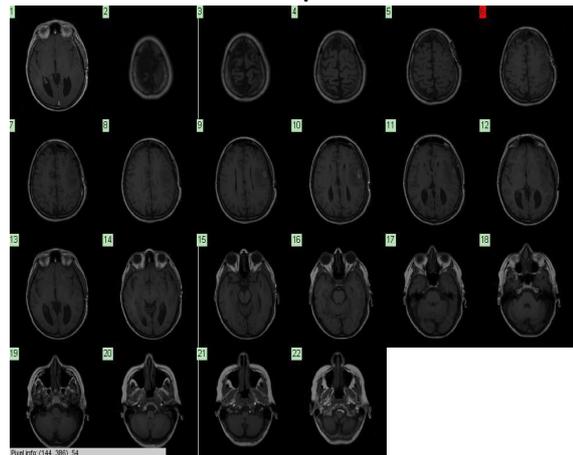


Figure: 1 Dataset Screen shot

C. K-MEANS CLUSTERING ALGORITHM

K-Means algorithm is an unsupervised clustering algorithm. K-Means classifies or partitions the input data points into multiple clusters. First, K-Means algorithm randomly assigns the centroids. Each cluster will have centroids. Then in next iteration the squared euclidean distance between each data point to the centroids are calculated. The distance between the point and each centroid are checked. The minimum distance centroid is taken as new centroids for that data point. After this the position of each cluster is set to the mean of all data points belonging to that cluster. This process will repeat until there is no change in the cluster centroids.

Steps in the algorithm are as follows:

1. Randomly assign K centroids for the clusters
2. Calculate the distance from each data point to the cluster centroids using squared Euclidean distances.
3. Re-assign the data point to another cluster whose distance from the cluster centroids is minimum of all the cluster centroids.
4. Re-calculate the value of cluster centroids.
5. Repeat steps 2 to 4 until there is no change in data points and cluster centroids.

D. FUZZY C-MEANS (FCM)

The Fuzzy c-means (FCM) algorithm is used for data clustering. In FCM dataset is grouped into 'N' clusters by finding a membership relation. The distance between the data point and the clusters are calculated

.The more the data is near to the cluster center the more is its membership towards the particular cluster center. Then the data point is moved to the corresponding cluster which is having minimum distance from that data point.

1. Let D be the dataset which is to be clustered. Each row contains a sample data point.
2. N is the number of clusters.
3. The distance between the data point and the clusters are calculated.
4. The fuzzy membership is assigned to each data point depending on the new calculated distance.
5. New cluster centres are calculated.
6. Move the data points near to the minimum cluster center.
7. Repeat this until a minimum distance membership is obtained.

IV. RESULT AND DISCUSSION

MRI images from BRAINIX were used for clustering. The images are brain tumor MRI. Images are in DICOM format. Reading DICOM images require special type of applications and reader. A DICOM image contains much useful information about the patient and the image. Matlab has function called `dicomread` for reading DICOM images. The primary objective of this work is to segment the given brain images. In Matlab the command `'imcontour'` is used to find the boundaries of the image. This gives a clear separation between the picture components. It draws a contour in the image. Figure shows the output for brain tumor MRI contour segmentation.

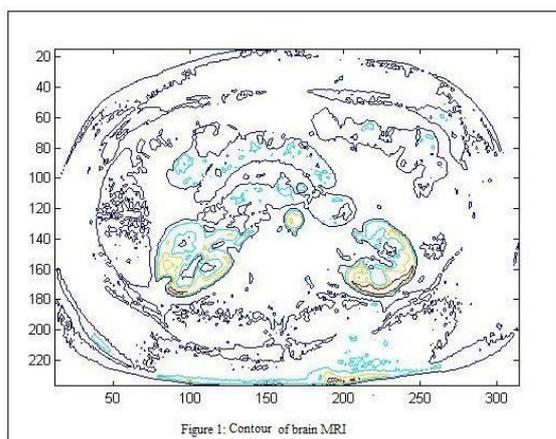


Figure 1: Contour of brain MRI

In this study the clustering of image data was done using both K-Means algorithm and FMC in Matlab. K-Means algorithm clusters the input gray images into different clusters. This clustering helps to separate the data points that are having similarity in their nature. The cluster points are calculated according to the intensity value of the data points. The similar gray matter came in to one cluster and

white matter in to another cluster. The affected part lies between gray matter and white matter. This clustering helps to identify the region of interest (ROI) in the image. From the clustered image the K-means algorithm gives a clear clustering than the FCM. The comparison result is shown in figure. From the study the clustering using K-means gives a clear clustered picture than FCM. The clustering time is also less for K-means algorithm.

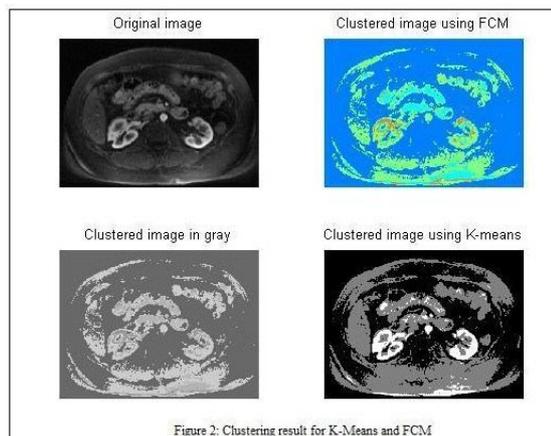


Figure 2: Clustering result for K-Means and FCM

Figure: 3 Comparison result screen in Matlab

Using K-means algorithm the affected area is clearly clustered and is clearly visible in the output image. For future work, the region of interest must be identified and the affected area is to be extracted from the images.

CONCLUSION

The segmentation of brain images using K-means and FCM had been done. This study compared the efficiency of K-means and FCM for clustering MRI images. The segmentation of images using K-means is better than FCM for this dataset. This work is the initial step for developing a system for information retrieval using data mining techniques. The work will continue for larger set of MRI images with the implementation of image mining algorithm.

REFERENCES

- [1] Rahman, M.M.; Tongyuan Wang; Desai, B.C., "Medical image retrieval and registration: towards computer assisted diagnostic approach", Design of Reliable Communication Networks, 2003. (DRCN 2003). Proceedings. Fourth International Workshop , , pp.78-89,1-3 Sept. 2004.
- [2] R.VenkataRamana Chary, Dr.D.Rajya Lakshmi and Dr. K.V.N Sunitha, "Feature extraction methods for color image similarity", advanced Computing: An International Journal Vol.3, No.2, March 2012.
- [3] A.Hema, E.Annasaro, "A Survey In Need Of Image Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 2, February 2013.
- [4] Zhang Ji, Hsu, Mong, Lee, Image Mining: Issues, Frameworks And Techniques, Proceedings of the Second

- International Workshop on Multimedia Data Mining (MDM/KDD'2001), in conjunction with ACM SIGKDD conference. San Francisco, USA, August 26, 2001.
- [5] Megalooikonomou, Davataikos C. and Herskovits, Mining lesion-deficit associations in a brain image database. KDD, San Diego, CA USA, 1999.
- [6] Zaiane, O., Han J., et al. "Mining MultiMedia Data" CASCON'98: Meeting of Minds, Toronto, Canada, November 1998, pp 83-96
- [7] Vamsidhar Enireddy and Kiran Kumar Reddi. Article: A Data Mining Approach for Compressed Medical Image Retrieval. International Journal of Computer Applications 52(5):26-30, August 2012.
- [8] Al-Badarneh, A.; Najadat, H.; Alraziqi, A.M., "A Classifier to Detect Tumor Disease in MRI Brain Images", Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on , 26-29 Aug. 2012 ,pp.784,787
- [9] <http://medical.nema.org/Dicom/about-DICOM.html>
- [10] Rajendran, P.; Madheswaran, M.; Naganandhini, K., "An improved pre-processing technique with image mining approach for the medical image classification", Computing Communication and Networking Technologies (ICCCNT), 2010 International Conference , 29-31 July 2010, pp.1,7
- [11] Abdullah, N.; Ngah, U.K.; Aziz, S.A., "Image classification of brain MRI using support vector machine", Imaging Systems and Techniques (IST), 2011 IEEE International Conference , 17-18 May 2011, pp.242,247
- [12] <http://www.osirix-viewer.com/datasets/>, accessed on 9/6/2014

★ ★ ★