Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 2, February 2015, pg.230 – 235

RESEARCH ARTICLE

Detection Of Brain Disorders using Clustering based Techniques- Expectation Maximization

Amruta Gadekar[1], Prof. Ravi Patki[2]

[1,2]Computer Department & Pune University, India

[1] amruta.gadekar11@gmail.com, [2] dypcoe.ravi@gmail.com

Abstract— Brain is a central as well as an important part of a human body. The Brain activities are very complicated and difficult to understand. Many psychiatric disorders related to the brain are difficult to detect or identify. Functional magnetic resonance imaging or fMRI is a technique which is responsible for measuring a brain activity. The basic signal of fMRI depends upon the blood-oxygen-level-dependent (BOLD) effect which helps to study human brain functions. The purpose of Clustering Technique is to understand the complex interaction patterns among brain regions as well as identify brain disorders. To detect clusters of objects with similar interaction patterns we proposed a partitioning clustering algorithm that is Expectation Maximization (EM) algorithm. The Expectation Maximization (EM) algorithm is Gaussian Mixtures which begins with an initial guess to the cluster centers, and iteratively refines them an efficient algorithm for partitioning clustering.

Keywords— Clustering Technique;Interation pattens; Interation K-Means; Expectation Maximization

I. INTRODUCTION

Brain is the central part of the human body. Brain is responsible to control all the functions of the human body. The functions or the activities related to brain become a very complicated and not easy to understand. If there is any problem in human body related to the brain it will be considered as a brain disorder. Many psychiatric disorders are still not identified. To understand the complex functions and the psychiatric disorders of the brain, we have to first understand the different brain activities. Brain activity is the only resource to understand psychiatric disorders. Functional magnetic resonance imaging (FMRI) helps to study human brain function in a non-invasive way. With fMRI,we capture the signal through pulse sequences which helps to understand the brain activity.The vast majority of functional MR (FMRI) experiments measure the blood oxygen-level dependent (BOLD) signal.

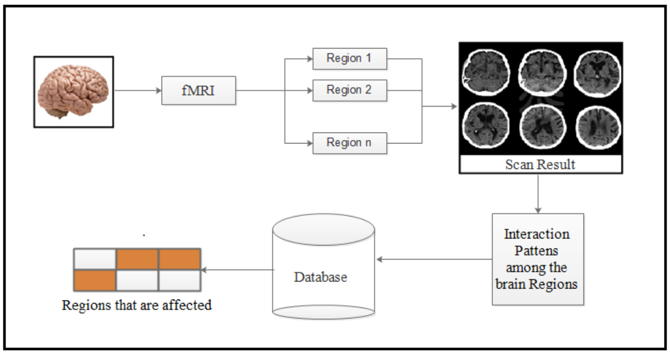

Fig.1:System Architecture

The fig.1. Shows the architecture of identification of brain disorders.With the help of this architecture we can find out whether the brain is affected or not.After scanning the result is compared with the dataset or clusters of normal brain and diseased brain.By comparison, we will find out whether the brain is affected or not.with the help of this architecture it is easy to find out brain disorders.

Clustering and classification are both essential tasks in Data Mining. A cluster is a group of objects which are “similar” among them and are “dissimilar” to the objects belonging to other clusters. So mainly we use clustering technique to detect and make a group of similar patterns. To find the similar interaction patterns among the brain regions we use partitioning clustering technique. With the help of partitioning clustering techniques we capture the different interaction patterns in healthy and diseased subjects.

In every field, measurements are performed over time. A time series represents a collection of values obtained from sequential measurements over time. The purpose of time-series data mining is to try to extract all meaningful knowledge from the unlabeled set of data.

In this paper, we propose the partitioning clustering technique such as Expectation Maximization which is responsible to reduce the time taken to form clusters of an object having a similar interaction pattern. With the Expectation Maximization algorithm we add more efficiency with the correct result, decrease complexity with clustering. The Expectation Maximization algorithm starts with initial guess which defines an iterative process that allows maximizing the likelihood function of the model. In the EM algorithm, we first assign each sample to a component (E step) and then fit (or maximize the likelihood of) each component separately (M step).

II. Basic Concept

This section explains a brief introduction about the basic concept which is clustering technique, FMRI, multivariate time series and the EM algorithm.

A. *Clustering Technique:*

Clustering and classification are both fundamental tasks in Data Mining. A cluster is a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. So mainly we use clustering technique to detect and make a group of similar patterns.Basically clustering technique is divided into partition method and hierarchical method. Hierarchical clustering can be agglomerative or divisive, i.e. bottom-up or top down,respectively. It begins from each element and is intended to form a hierarchical cluster structure i.e.tree structure. Hierarchical clustering methods are categorized into agglomerative (bottom-up) and divisive (top-down).In agglomerative clustering starts with one-point clusters and recursively merges two or more most appropriate clusters. A divisive clustering starts with one cluster of all data points and recursively splits the most appropriate cluster.In partitioning they will create a partition and then sort them by some criteria.

B. *FMRI: Functional magnetic resonance imaging:*

Functional magnetic resonance imaging (FMRI)[1] helps to study human brain function in a non-invasive way. FMRI mechanism is that it detects the changes in blood oxygenation and flow that occur in response to neural activity.When a brain area is more active it consumes more oxygen and to meet this increased demand blood flow increases to the active area. fMRI can be used to create activation maps showing which parts of the brain are involved in a particular mental process.Two general types of pulse sequences are common to understand the brain activity, depending on whether the goal is structural or functional imaging. The goal of structural MR is usually to measure the density of water molecules. The vast majority of functional MR (FMRI) experiments measure the blood oxygen-level dependent (BOLD) signal.

C. Multivariate time series:

A time series[7] represents a collection of values obtained from sequential measurements over time. The purpose of time-series data mining is to try to extract all meaningful knowledge from the unlabeled set of data. Mainly time series clustering is considered in three groups depending upon whether they work directly with raw data, i.e. temporal-proximity based, or indirectly with the features extracted from the raw data, i.e. representation-based clustering and with model build from raw data i.e. Model based. To make an efficient and effective data, we use multivariate time series. Multivariate time series is used when one wants to model and explain the interactions and co-movements among a group of time series variables.

D. Expectation Maximization (EM):

This paper introduces an innovative partitioning clustering technique which is Expectation Maximization. Expectation Maximization finds a similar interaction pattern between the brain regions which helps to detect the brain disorders. The EM algorithm is used to find the maximum likelihood parameters of a statistical model. The EM iteration interchanges between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood calculated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. An EM algorithm gives a guarantee that at each iteration it will increase the likelihood,which will use to find out the similar patterns.

We use Expectation Maximization(EM) because:

- Strong statistical basis.
- Robust against noisy data.
- It can accept a desired number of clusters as an input.
- Increase accuracy.
- Faster than other algorithm.
- More efficient.

III.LITERATURE REVIEW

A. "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging":

M. D. Fox and M. E. Raichle introduces a spontaneous fluctuation in brain activity with functional magnetic resonance imaging (fMRI) blood oxygen level dependent (BOLD) signal method and spatial and temporal properties of spontaneous BOLD fluctuations. According to him modulation of the functional magnetic resonance imaging (fMRI) blood oxygen level dependent (BOLD) signal attributable to the experimental paradigm can be observed in distinct brain regions, such as the visual cortex, allowing one to relate brain topography to function. However, spontaneous modulation of the BOLD signal which cannot be attributed to the experimental paradigm or any other explicit input or output is also present. Because it has been viewed as 'noise' in task-response studies, this spontaneous component of the BOLD signal is usually minimized through averaging.

B. Correspondence of the brain's functional architecture during activation and rest"[2]:

In this paper they explain about the neural connection. FMRI is only the source which helps to study the brain activities.The brain state "active" and "in-rest" is mainly considered to capture the brain activities. In the neural network they concentrate on functional networks, which is related to whether the brain is functionally active or not at the given movement. What will be the result if the brain is in resting state.What brain regions are continuously interacting when the brain is ''at rest.''Accoring to their research functional networks utilized by the brain in action is continuously and dynamically ''active'' even when at ''rest''.They used independent component analysis (ICA), a powerful data-driven approach for finding independent patterns in multivariate data,which allows to identify the major functional networks in the brain to study the brain functions.To study the function network it will use ICA-based analyses of BrainMap and the resting FMRI data.By studying all thses things the result says that Spatial maps from BrainMap and resting FMRI were associated with each other primarily via spatial similarity, by using (Pearson) spatial cross-correlation.

C. Feature Selection for Classification of Variable length Multi-attribute Motions" :

To capture the motion is a new type of multimedia. Recognizing the patterns of human motion we use a 3D camera. The idea of this paper is to capture the data of motions with the multiple attributes. To capture the movements of multiple joints of a subject, having a different lengths for even similar motions. To classify and recognize, multi-attribute motion data of different lengths, Chuanjun Li, Latifur Khan, and Balakrishnan Prabhakaran introduced a new type of multimedia technique which is Singular Value Decomposition(SVD). High accuracy classification is required in effective applications, which extract feature of vectors motions. To obtain feature vectors for motion patterns using the Singular Vector Decomposition (SVD) properties of the motion matrices SVD optimally exposes the geometric structure of a matrix. Different approaches to extracting feature vectors are explored based on how information is to be extracted from SVD. In this paper, they consider geometric structures of motion matrices in a high dimensional space instead of considering motion data rows/frames for motion identification a motion frame has dimension n, then a motion sequence of length m can be taken to be m vectors in an nD space. The distributions of the nD vectors are explored for feature extraction.To obtain feature vectors for motion patterns using the Singular Vector Decomposition (SVD) properties of the motion matrices since SVD exposes the geometric structure of a matrix. Different approaches to extracting feature vectors are explored based on how information is to be extracted from SVD.

D. "Independent component analysis for clustering multivariate time series data" Independent component analysis:

The paper is based on the identification of physical causes of variability of a given dynamical system.To work on identification of physical causes of variability they use classical component extraction techniques. Instead of using classical component techniques,E. H. C. Wu and P. L. H. Yu introduced a new technique - Independent component analysis(ICA).The goal of this technique is to find a linear representation of nongaussian data so that the components are statistically independent, or as independent as possible. Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation.

IV. EXISTING SYSTEM

- The Interaction K-means (IKM) is the existing algorithm which is based on k-means technique.
- IKM is a partitioning clustering used to detect clusters of objects with similar interaction patterns.
- The algorithm IKM is a common technique for clustering multivariate time series using fMRI data.
- IKM takes a more time for clustering so the IKM partitioning clustering is time consuming.
- IKM may be associated with particular regions or functions of the brain or the rest of the nervous system. This is not suitable for different regions of the time series.

V. PROPOSED SYSTEM

The purpose of Clustering Technique is to understand the complex interaction patterns among brain regions as well as identify Brain disorders. To detect clusters of objects with similar interaction patterns we proposed a Partitioning Clustering algorithm that is Expectation Maximization (EM) algorithm.

The basic idea is to make two different clusters from the group of persons. With the Expectation Maximization clustering method we will get the two different clusters, i.e. one is of healthy or normal and another one is deceased. When the patient or a person comes we will take the result of properties of the brain interaction of that person and compare the result with the clusters already we have in the database. In such a way we detect or identify the brain disorders.In such a way we can detect or identify the brain disorders.

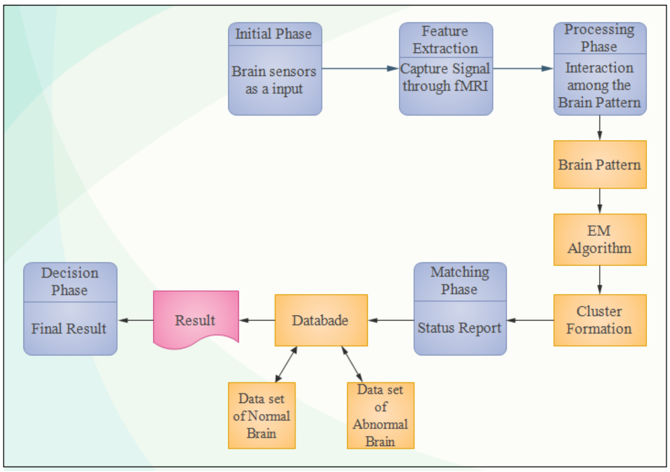

Fig.2:System Flow of Architecture

The System flow is having various phases that will describe as follows:

A. *Initial Phase:*

Initial phase is responsible for taking input for further processing and giving us a desired output.Here we can give a human brain as an input which is going for further processing.With the help of EEG sensor we can record electrical impulses from the brain.The electrodes transfer information from your brain through wires to an amplifier and a machine that measures and records the data.This data will be transferred to feature extraction phase.

B. *Feature Extraction :*

After receiving the data from the initial phase, extraction phase extracts or capture the signal through FMRI. FMRI technique is used to detect the changes in blood oxygenation and flow that occur in response to neural activity.When a brain area is more active it consumes more oxygen and to meet this increased demand blood flow increases to the active area. fMRI can be used to create activation maps showing which parts of the brain are involved in a particular mental process.This Signal result will further send for processing to processing phase.

C. *Processing Phase:*

In this processing phase the FMRI result which scans result will process.Here we are forming a cluster having a similar properties between two brain regions with the help of the EM algorithm.After cluster formation it will send to the matching phase.

D. *Matching Phase:*

In this matching phase, we can compare the clusters with the data sets which is already present in the database. In the database, we already have a data of normal brain and data of diseased brains. So the given data from the processing phase, that we can compare with these data sets.

E. *Decision Phase:*

By comparing,we will get the final result that whether the brain is affected or not. And if it is affected then which region of the brain is affected.

In this way we can understand the complex interaction patterns among brain regions as well as identify the disorders related to brain.

VI. COMPARATIVE ANALYSIS

In this section, we can measure the performance of the algorithms by taking some important parameters for the same system.The comparison is based on results with accuracy, efficiency, noisy data, time consumption.

TABLE I

Name of Algorithm	**Accuracy**	**Efficiency**	**Robust against Noise**	**Time Consuming**
K-Means	Average	Average	Less	More
IKM	Average	Average	Excellent	More
EM	Excellent	Excellent	Excellent	Excellent

The comparison clearly shows that why the Expectation maximization gives better result than any other algorithm in finding brain disorder system. Interaction k-means(IKM) algorithm and the expectation maximization (EM) algorithm are similar in the sense that they allow model refining of an iterative process, but the IKM takes a more time for clustering so the IKM partitioning clustering is time consuming.The EM algorithm is used to provide the functions more effectively than IKM or K-Means.

The Expectation Maximization algorithm starts with initial guess which defines an iterative process that allows maximizing the likelihood function of the model. In the EM algorithm, we first assign each sample to a component (E step) and then fit (or maximize the likelihood of) each component separately (the M step). The EM clustering method having high accuracy of the results and provide a high speed. IKM may be associated with particular regions or functions of the brain or the rest of the nervous system. This is not suitable for different regions of the time series.

With the comparative analysis, EM algorithm is

- Increase accuracy.
- Faster than other algorithm.
- More efficient.
- Robust against noisy data.

VII. CONCLUSION

In this paper, we propose a clustering method which is responsible to find out the similar interaction patterns among the brain region. The Expectation Maximization (EM) algorithm is Gaussian Mixtures which begins with an initial guess to the cluster centers, and iteratively refines them an efficient algorithm for partitioning clustering. EM accomplishes good results on synthetic data and on real world data from various domains, but generally giving a good result on EEG and FMRI data. The EM algorithm works more effectively and provide high accuracy with high speed.

REFERENCES

[1] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging,"Nat. Rev. Neurosci., vol. 8, no. 9, pp. 700–711, 2007.

[2] C. Sorg et al., "Selective changes of resting-state networks in individuals at risk for alzheimer's disease," PNAS, vol. 104, no. 47, pp. 18760– 18765, 2007.

[3] S. Smith et al., "Correspondence of the brain's functional architecture during activation and rest," PNAS, vol. 106, no. 31, p. 13040, 2009.

[4] C. Li, L. Khan, and B. Prabhakaran, "Feature selection for classification of variable length multiattribute motions," in Multimedia Data Mining and Knowledge Discovery, V. A. Petrushin and L. Khan,Eds. London, U.K.: Springer, 2007.

[5] T. W. Liao, "Clustering of time series data – A survey," Patter Recognit., vol. 38, no. 11, pp. 1857–1874, 2005.

[6] T. Oates, L. Firoiu, and P. R. Cohen, "Clustering time series with hidden Markov models and dynamic time warping," in Proc. IJCAI Workshop Neural, Symbolic Reinforcement Learning Methods for Sequence Learning, 1999, pp. 17–21.

[7] X. Wang, A. Wirth, and L. Wang, "Structure-based statistical features and multivariate time series clustering," in Proc. ICDM,Omaha, NE, USA, 2007, pp. 351–360.

[8] E. H. C.Wu and P. L. H. Yu, "Independent component analysis for clustering multivariate time series data," in Proc. ADMA, Wuhan, China, 2005, pp. 474–482.

[9] C. Plant, A. M. Wohlschläger, and A. Zherdin, "Interaction-based clustering of multivariate time series," in Proc. ICDM, Miami, FL,USA, 2009, pp. 914– 919.

[10] F. Mörchen, "Time series feature extraction for data mining using DWT and DFT," Dept. Math. Comput. Sci., University of Marburg, Marburg, Germany, Tech. Rep. 33, 2003.

[11] H.-P. Kriegel, P. Kröger, A. Pryakhin, M. Renz, and A. Zherdin,"Approximate clustering of time series using compact model based descriptions," in Proc. DASFAA, New Delhi, India, 2008,pp. 364–379.

[12] P. Cheeseman and J. Stutz, "Bayesian classification (autoclass): Theory and results," in Proc. Adv. Knowl. Discov. Data Min., Menlo Park, CA, USA, 1996, pp. 153–180.

[13] A. R. McIntosh and B. Miic, "Multivariate statistical analyses for neuro imaging data," Annu. Rev. Psychol., vol.64,no.1,pp.499_525,2013[Online].Available:http://www.annual reviews.org/do i/abs/10.1146/annurev-psych-113011-143804.

