International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Mining Interaction Patterns among Brain Regions by Clustering Based Interactive K-Means

Amruta S. Chougule[1], S. A. Bajpai[2]

[1, 2]Department of Computer Engineering, JSPM's BSIOTR, Wagholi, Pune, India

Abstract: ***Functional Magnetic Resonance Imaging (FMRI) provides study of brain functions. The information content from that is large in volume and complex and data requires effective and efficient data mining techniques. To understand the complex interaction patterns among brain regions novel clustering technique is proposed. Each subject consider as FMRI image histogram. The objective is to assign objects exhibiting a similar intrinsic interaction pattern to common cluster. To formalize this idea, define a cluster by a set of mathematical models describing the cluster-specific interaction patterns. Based on this novel cluster notion, propose interaction K-means (IKM), an efficient algorithm for partitioning clustering. IKM simultaneously clusters the data and discovers the relevant cluster-specific interaction patterns. The results on two real FMRI studies demonstrate the potential of IKM to contribute to a better understanding of normal brain function and the alternations characteristic for psychiatric disorders.***

Keywords*:* clustering, image histogram, interaction patterns

1. Introduction

Human brain activity is very complex and not fully understood. Many psychiatric disorders like Schizophrenia and Somatoform Pain Disorder are neither identified by biomarkers, nor by physiological abnormalities of the brain. Abnormal brain activity is the only resource to understand psychiatric disorders. Functional magnetic resonance imaging (FMRI) used to study human brain function. The basic signal of FMRI depends on the blood-oxygen-level-dependent (BOLD) effect, which allows indirectly imaging brain activity by changes in the blood flow related to the energy consumption of brain cells. In a typical FMRI experiment, the subject performs some cognitive task while in the scanner [1]. Surprisingly, only about 5% of the energy consumption of the human brain can be explained by the task-related activity. Many essential brain functions, e.g. long-term memory are largely happening during rest. Therefore recent findings support the potential of resting-state FMRI to explore the brain function in healthy subjects and reveal alternations characteristic for psychiatric disorders [2]. In resting state FMRI, subjects are instructed to just close their eyes and relax while in the scanner. The FMRI image is the result of the simplest kind of FMRI experiment.

In brain, Frontal lobe, Temporal lobe, Occipital lobe, Parietal lobes are four regions. Recent findings suggest a modular organization of the brain into different functional modules [3]. To obtain a better understanding of complex brain activity, it is essential to understand the complex interplay among brain regions during task and at rest. A novel technique proposed which is for mining the different interaction patterns in healthy and diseased subjects by clustering. A cluster is defined as a set of subjects sharing a similar interaction pattern among their brain regions. The Interaction K-means (IKM) simultaneously clusters the data and discovers the relevant cluster-specific interaction patterns. Proposed technique clusters FMRI images using IKM. The clustering is based on image histograms. The algorithm IKM is also used for clustering multivariate time series data.

2. Motivation

FMRI data is large in volume and complex. The FMRI image is the result of the simplest kind of FMRI experiment. For mining interaction patterns among brain regions use clustering. And for clustering IKM algorithm is used. Proposed system use IKM algorithm, because IKM yet not applied on FMRI images and also this algorithm simultaneously clusters the data and discovers the relevant cluster-specific interaction patterns. IKM is a partitioning clustering algorithm suitable to detect clusters of objects with similar interaction patterns. The information on interaction patterns provides valuable insights for interpretation. On FMRI data from studies on Somatoform Pain Disorder and Schizophrenia, algorithm detects very interesting and meaningful interaction patterns

3. Related Work

The various techniques are used for clustering of images as well as clustering of time series data. In time series clustering find appropriate similarity measure is difficult task. Compression based similarity measure [8] reduces I/O cost and speed up clustering task but requires statistical conditions for data. Hidden Markov Models (HMM) [9] applied as segmental semi Markov models and viterbi algorithm is applied to compute similarity measure. Also there are Sequence Clustering Refinement Algorithm (SCRA), ICACLUS [10], and Structure Based Statistical Feature Clustering (SF) [11] techniques for clustering multivariate time series data. SF clustering identifies activity patterns from motion stream data with high accuracy. Linear models are recently proposed in [12] [13] to describe interaction patterns of an object for the purpose of efficient compression and classification respectively.

Recently there are few algorithms for clustering FMRI images. An Automatic Unsupervised Classification [4] observed that the grey matter distance can best separate the Alzheimer's disease patients from the cognitively normal control but it takes more time to diagnose the diseases. Hierarchical clustering analyzes micro array data and makes it easier to interpret the results of a cluster analysis. In hierarchical algorithm large number of objects requires huge I/O cost. In clustering of image dataset using k-means and fuzzy k-means algorithm [5], image is a collection of number of pixels. It is difficult to take account of all pixels for clustering. Image segmentation play very useful role in clustering as it save times and it is efficient too. With the use of k-mean and fuzzy k-means algorithm clustering of large data become easy and time saving. Fuzzy clustering algorithms classify the tissues based on single channel MRI images but after clustering noise and misclassification error may present. The K-means algorithm partitions the data into *K* clusters, represented by their centers or means. The center of each cluster is calculated as the mean of all the instances belonging to that cluster. But in k-means algorithm the clustering quality is greatly dependant on the choice of initial centers. Poor choices of the initial centers can degrade the quality of clustering solution and result in longer execution time.

4. Implementation

A. Existing System

In existing system data object is represented by a multivariate time series. Each dimension is a time series corresponding to the FMRI signal of a specific anatomical brain region. IKM is partitioning clustering algorithm and it is applied to detect clusters of objects with similar interaction patterns. This system finds clusters of objects which are represented by multivariate time series sharing a common cluster-specific interaction pattern among the dimensions. It preserves the information on attribute dependencies. FMRI image is result of simplest kind of FMRI experiment. Time series data is not easily implemented in java. Difficult task in clustering time series is to find appropriate similarity measure. So in proposed system FMRI images are used.

B. Proposed System

Problem Definition: Brain Images which have similar interaction patterns assign to common cluster by using IKM

algorithm IKM (data set DS, integer K): Clustering $\mathcal{C}$
  Clustering $bestClustering$;
//initialization
  for $init := 1 \ldots maxInit$ **do**
    $\mathcal{C} := randomInit(DS, K)$;
    for each $C \in \mathcal{C}$ **do**
      $\mathcal{M}_C := findModel(C)$;
    while not converged or iter $<$ **maxIter do**
//assignment
      for each $O \in DS$ **do**
        $O.cid = \min_{C \in \mathcal{C}} \mathcal{E}_{O,C}$
//update
      for each $C \in \mathcal{C}$ **do**
        $\mathcal{M}_C := findModel(C)$;
      if improvement of objective function
        $bestClustering := \mathcal{C}$;
    end while
  end for
return $bestClustering$;

Figure 1: Interaction K-means Algorithm

algorithm and detects specific cluster names as normal or specific diseased using naive algorithm.

1) Algorithm:

In proposed system IKM algorithm is used. The first step of IKM is the initialization. As a common strategy for K-means, propose to run IKM several times with different random initializations and keep the best overall result. For initialization, randomly partition DS into K clusters. For IKM it is favorable that the initial clusters are balanced in size to avoid overfitting. Therefore, partition the data set into K equally sized random clusters and find set of models for each cluster. For model finding apply greedy stepwise algorithm combination with Bayesian Information Criterion (BIC). After initialization, IKM iteratively performs the further two steps until convergence. After assignment, in the update step, the models of all clusters are reformulated. IKM converges as soon as no object changes its cluster assignment during two consecutive iterations. Pseudo code of IKM is provided in Figure 1.

2) System Architecture:

Login:- In proposed system user first login with valid username and valid password for authentication.
Browse Dataset:- In this dataset of FMRI images stored at specific location. Image histograms values are calculated from FMRI images and stored. Then dataset is browse from stored location and show all images of dataset.

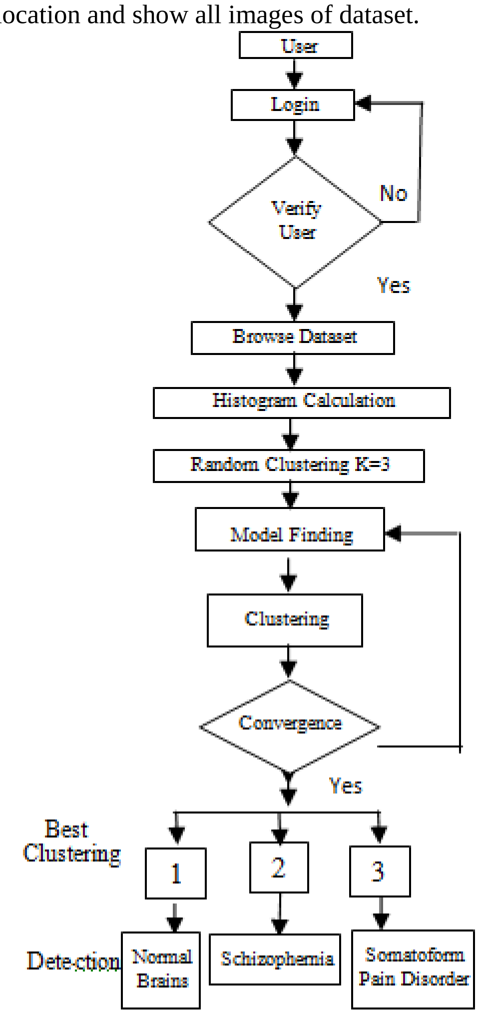

Figure 2: Architecture

IKM algorithm:- Then IKM algorithm is applied on dataset images The image histogram values are taken as input to system.In that first random clustering is done. On that each cluster model finding is applied and clustering is done up to convergence. Then get best clustering. From clustering normal brain images and abnormal brain images evaluates differently. The abnormal brain images consists different clusters of different diseases.

Detection:- To detect which cluster is for which specific cluster comparison is done. The known disease images are stored already as training data which are comparing with clusters and get result. The system architecture shown in figure 2.

5. Mathematical Model

Image clustering using IKM algorithm clusters brain images into normal and diseased.

Assumptions:
L : Login into system
DS: Browse Dataset
HC: Calculate histogram
CL: Make clusters
DT: Detect cluster names

1. User can login into system with valid username and valid password.
2. User can give dataset of brain images as input.
3. Image histogram values are taken as input by system.
4. Output is set of clusters each cluster contain similar interaction pattern images.
5. Considering training data specific cluster names detected.

C. Model Finding

For model finding apply greedy stepwise algorithm combination with Bayesian Information Criterion (BIC). In proposed system linear model used because of images are in sequentially format. They are interpretable and computationally efficient. And also applied for efficient compression and classification respectively.

The greedy stepwise algorithm [6] starts with an empty set of relevant dimensions. In each step, either one dimension is added or removed, depending on which of these two actions is judged superior by the evaluation criterion. The algorithm terminates if none of the two actions leads to a further improvement. BIC [7] which determines a balance between goodness-of-fit and complexity of the model and is defined by:

$$BIC\ (M_a) = -2 \cdot LL\ (a, M_a) + \log(m^*)(|V| + 1).$$

The first term represents the goodness-of-fit, where $LL\ (a, M_a)$ denotes the log-likelihood of dimension a given the model. The second term punishes overly complex models.

D. Detction

The naive algorithm is used for comparison. At the detection known disease structure images are compare with clusters. The type of disease images are predicted by finding similarity difference.

6. Results and Discussion

In proposed system FMRI images dataset is considered. A database contains normal brain images and different diseased brain images for example here proposed system considers Somatoform Pain disorder and Schizophrenia disease brain images. There are two algorithms for clustering K-means algorithm and Fuzzy K-means algorithm. Clustering is either hard clustering or soft clustering. In hard clustering, data is divided into distinct clusters, where

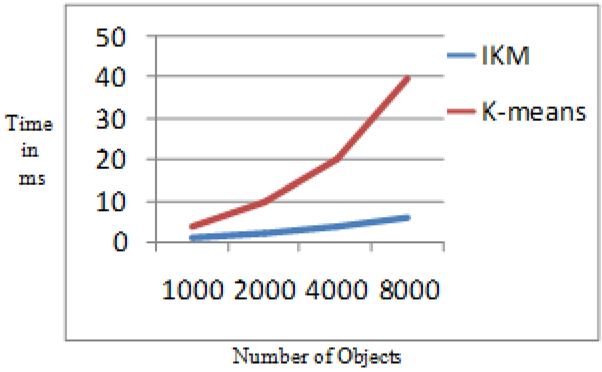

Figure 3: Scalability w.r.t. objects/time

each data element belongs to exactly one cluster. K-means algorithm is hard clustering. In fuzzy clustering it is also referred to as soft clustering, data elements can belong to more than one cluster. K-means clustering produces fairly higher accuracy and requires less computation. Fuzzy K-means clustering produces close results to K-means clustering, but it requires more computation time than K-means. So K-means algorithm is better than Fuzzy K-means.

So K-means algorithm is compared with IKM algorithm. Fig.3 shows comparison between those algorithms.

IKM algorithm achieves significantly faster response time than K-means algorithm. Efficiency is correctness of output. It is number of images which are properly clustered to number of input images. The scalability in terms of number of objects is shown. The number of objects that means images increases then time required for clustering is increases, the clustering time measured in millisecond. Also IKM algorithm achieves better clustering accuracy than K-means algorithm. Best clustering is done using IKM algorithm. For detection training data is compared with clusters using Naive algorithm. It detects normal brains and diseased brains separately. Naive algorithm achieves better accuracy and efficiency.

7. Conclusion

In this paper, novel cluster notion proposed for FMRI image data. Cluster defined as a set of objects sharing a specific interaction pattern among the dimensions. Interaction K-means (IKM), an efficient algorithm for interaction-based clustering is used. The specific cluster for specific diseases detected that means normal brain images, psychiatric disorder images differentiates.

8. Acknowledgment

This is a great pleasure & immense satisfaction to express my deepest sense of gratitude & thanks to everyone who has directly or indirectly helped me in my project work successfully. I express my gratitude towards project guide Prof. S.A.Bajpai, and Prof. G.M.Bhandari, Head of Department of Computer Engineering, Bhivarabai Sawant Institute of Technology and Research College of Engineering, Pune who guided & encouraged me in my project work in scheduled time. I would like to thanks our Principal Dr. T.K.Nagraj, for allowing us to pursue my project in this institute.

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] A. Trouve and Y. Yu, ―Unsupervised clustering trees by nonlinear principal component analysis." Pattern recognition and Image Analysis, vol. 2, pp. 108–112, 2001.
[5] Dehariya, V.K. ; I.T. Dept., S.A.T.I., Vidisha, India ; Shrivastava, S.K. ; Jain, R.C. "Clustering of image dataset using k-means and fuzzy k-means algorithm" Computational Intelligence and Communication Networks (CICN), 2010
[6] D. T. Larose, *Data Mining Methods and Models*. Hoboken, NJ, USA: Wiley, 2006.
[7] E. I. George, "The variable selection problem," *J. Amer. Statist. Assoc.*, vol. 95, no. 452, pp. 1304–1308, 2000.
[8] C. A. Ratanamahatana, E. J. Keogh, A. J. Bagnall, and S. Lonardi, "A novel bit level time series representation with implication of similarity search and clustering," in *Proc. PAKDD*, Hanoi, Vietnam, 2005, pp. 771–777.
[9] X. Ge and P. Smyth, "Deformable Markov model templates for time-series pattern matching," in *Proc. KDD*, New York, NY, USA, 2000, pp. 81–90.
[10] 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.
[11] X. Wang, A. Wirth, and L. Wang, "Structure-based statistical fea-tures and multivariate time series clustering," in *Proc. ICDM*, Omaha, NE, USA, 2007, pp. 351–360.
[12] 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.
[13] C. Böhm, L. Läer, C. Plant, and A. Zherdin, "Model-based classi-fication of data with time series-valued attributes," in *Proc. BTW*, 2009, pp. 287–296.