International Journal of Advanced Research in

Computer Science Engineering and Information Technology

Volume: 2 Issue: 2 08-Apr-2014,ISSN_NO: 2321-3337

# MODEL BASED ESSENTIAL INTERACTIONS CLUSTER MINING IN MULTIVARIATE TIME

V.SARAVANAN[1], S.CHITRA[2],

PG Scholar, Dept of Computer Science and Engineering, Manakula Vinayagar Institute of Technology, India [1].
Asst.Professor, Dept.of Information Technology, Manakula Vinayagar Institute of Technology, India[2]
vsaravananmtech@gmail.com[1],
chitrashanmougam@gmail.com[2]

.

***ABSTRACT—****This Functional magnetic resonance imaging or functional MRI (fMRI) is a functional neuroimaging procedure using MRI technology that measures brain activity by detecting associated changes in blood flow. The goal of fMRI data analysis is to detect correlations between brain activation and a task the subject performs during the scan. It also aims to discover correlations with the specific cognitive states, such as memory and recognition, induced in the subject. In this system, we propose a novel framework for clustering the essential fMRI signals based on their interactions and also correlation which is generated in a multivariate time series. To formalize this framework we cluster only Important Interactions based on the patient's medical records with the help of Essential Clustering Algorithm. The Essential clusters (EC) are then clustered again based on their dependencies on various brain regions. These EC's are grouped under specific models. The changes detected are mined based on the type of cluster grouped under a certain model. Our method shows that certainly increases the efficiency of the system along with increases in the effectiveness with minimal resource utilization.*

**Keywords— Clustering, Dependencies, Brain Region, Efficiency.**

## 1, INTRODUCTION

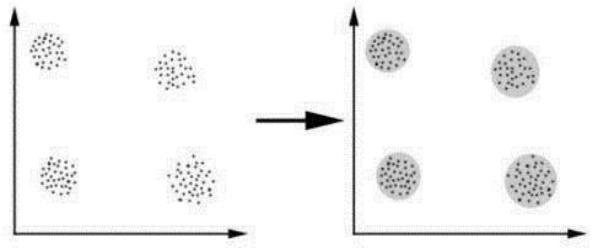
The Human brain activity is very complex and far from being fully understood. Many psychiatric disorders like Schizophrenia and Somatoform Pain Disorder can so far neither be identifies by biomarkers, nor by physiological or histological abnormalities of the brain. Aberrant brain activity often is the only resource to understand psychiatric disorders. Functional magnetic resonance imaging (fMRI) opens up the opportunity to study human brain function in a non-invasive way. The basic signal of fMRI relies 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.

### A. Brain Functions

In a typical fMRI experiment, the subject performs some cognitive task while in the scanner. Recently, resting-state fMRI has attracted considerable attention in the neuroscience community. 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, most of them without consciousness of the subject and many of them are still not well understood. Therefore recent finding support the potential of resting-state fMRI to explore the brain function in healthy subjects and reveal alternations characteristic for psychiatric disorders. In resting state fMRI, subjects are instructed to just close their eyes and relax while in the scanner. fMRI data are time series of 3-dimensional volume images of the brain. The data is traditionally analyzed within a massunivariate framework essentially relying on classical inferential statistics, e.g. contained in the software package SPM. A typical statistical analysis involves comparing groups of subjects or different experimental conditions based on univariate statistical tests on the level of the single 3-d pixels called voxels. Data from fMRI experiments are massive in volume with more than hundred thousands of voxels and hundreds of time points. Since these data represent complex brain activity, also the information content can be expected to be highly complex.

**B. Clustering Important**

The Clustering aims at dividing the data set into groups (clusters) where the inter-cluster similarities are minimized while the similarities within each cluster are maximized. A partitioning method was one of the earliest clustering methods to be used in Date usage mining.



An incremental algorithm that produces high quality clusters is used. Each user session is represented by an n-dimensional feature vector, where n is the number of Web pages in the

session. The value of each feature is a weight, measuring the degree of interest of the user in the particular Brain.

**C. Contributions**

The major contributions of this paper can be summarized as follows:

We use Interaction K-means (IKM), a partitioning clustering algorithm suitable to detect clusters of objects with similar interaction patterns. We find the essential clustering from the interaction clustering using Essential Clusters Identification (ECI) algorithm, essential clustering is reducing the amount of the data from the dataset it will be useful for the time consuming. We generate the clustering models based on dependencies of clustering and perform mining in easy manner and show the result which Decreases the mining time thus making the approach more real-time, Increases the efficiency by clustering the important interactions into their individual models.

## II. RELATED WORK

The most difficult task in clustering time series is to find an appropriate similarity measure. Many approaches rely on feature transformation and dimensionality reduction. Features derived by Discrete Wavelet Transform and the Discrete Fourier Transform, as well as obtained by Principle Component Analysis, have been successfully applied for clustering. Alternative approaches to feature extraction include e.g. the method of multi-resolution piecewise aggregate approximation presented in.

Each dimension or attribute ai ∈ A is a time series with m time points, i.e. ai = {t1,i, ..., tm,i}.We also use _ai to denote the m time points of dimensional as a column vector. We use italics to denote sets, e.g. A denotes the set of attributes of DS and O a set of objects. Capital letters denote matrices composed by column vectors of dimensions. We further denote by m∗ the overall number of time points considering one distinct dimension of some fixed set of objects.

We consider a clustering C as a non-overlapping partitioning of DS into K clusters, i.e. DS = _1≤j≤K Cj and drop the indices whenever non ambiguous [1].The remainder of this paper is organized as follows. In the next section, we briefly survey related work. In Section 3 we introduce the Model based Essential interactions cluster. Section 4 Discuss about Interaction K-Mean cluster to cluster the brain regions. In Section 5 we propose the Essential Cluster identification model and Section 6 contains an Essential Cluster model grouping method. In section 7 explaining about Interaction among brain regions and Section8 concludes the paper.

### A. Time Series Data Mining

**Indexing** (Query by Content): Given a query time series Q, and some similarity/dissimilarity measure D(Q,C), find the nearest matching time series in database DB.

**Clustering**: Find natural groupings of the time series in database DB under some similarity/dissimilarity measure D (Q, C).

**Classification:** Given an unlabeled time series Q, assign it to one of two or more

predefined classes.

**Segmentation:** Given a time series Q containing n data points, construct a model Q, from K piecewise segments (K << n) such that Q closely approximates Q [2].

## III. MODEL BASED ESSENTIAL INTERACTIONS CLUSTER

We propose a system for reduce the time taken and complexity of clustering, so we introduce the essential clustering technique based on their interactions and also correlation which is generated in a multivariate time series using Essential Clusters Identification algorithm(ECI).

Then implement Essential Cluster Model Grouping (ECMP) algorithm for make the models that clusters based on the relation between the clusters and their dependencies. Using DepMine algorithm performing mining Clusters based on their Dependencies and their Models from previous Sample Models. Results used in Identification of Brain Activities and to identify Disorders.

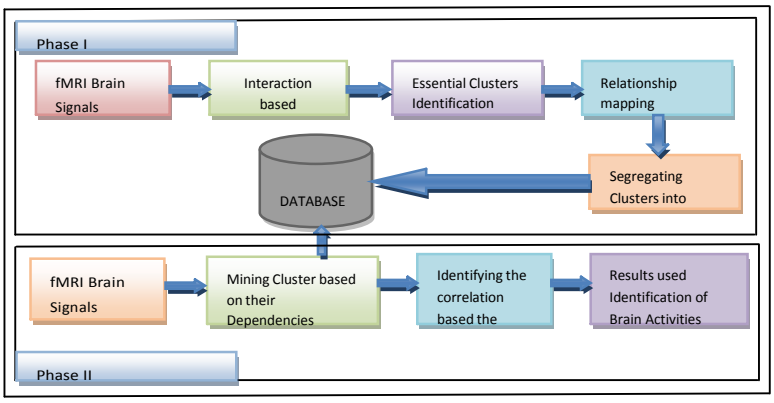



Figure.2. System Architecture: In Our System The Each Step Of Process Are Mentioned And Discussed In This Section.

### A. Interaction-Based Cluster

In this section, we elaborate our cluster notion based on characteristic interaction patterns. We want to find clusters of objects which are represented by multivariate time series

sharing a common cluster-specific interaction pattern among the dimensions. The dimensions of this object exhibit a simple interaction pattern: The time series of dimension dim12 can be expressed by a linear combination of some other dimensions: dim12:= 2 • dim4 + dim5 + dim6. Typically, not all dimensions of an object are interacting. For simplicity, only the dimensions involved in the interaction pattern [1].

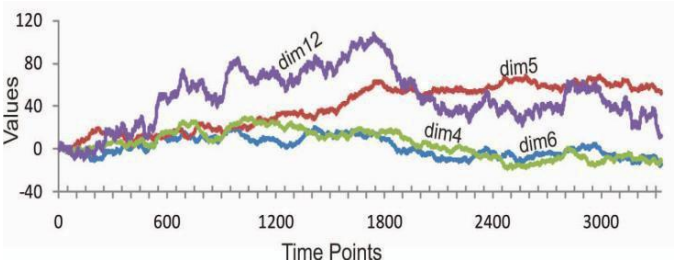



Figure.3. Interaction Pattern Within A Multivariate Time Series: The Signal Of Dim12, In Our Application Representing A Specific Region Of The Brain, Is Provided By A Linear Combination Of Some Other Dimensions, Here By Dim4 To Dim6, Representing An Interaction Pattern Among A Set Of Brain Regions

Before addressing the problem of how to find the clusters, we need to describe how the set of models MC can be computed from the set of objects OC which are associated to a cluster C. Since we focus on linear models, this involves solving d regression problems. Multiple least square regressions can be applied to derive the models. However, a common problem is over fitting. The more dimensions are included into the model, the more variance is explained and thus the smaller is the error term. Models involving a large part or even all dimensions are not generalizable and hard to interpret. Therefore, the set of explanatory variables of each model needs to be carefully selected. To determine the really relevant dimensions, we apply a greedy stepwise algorithm for model finding [3] in combination with the Bayesian Information Criterion (BIC) [4] as evaluation criterion. The greedy stepwise algorithm is an established technique for variable selection in regression problems. This algorithm 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. As evaluation criterion, we apply BIC which determines a balance between goodness-of-fit and complexity of the model and is defined

by: BIC(Ma) = -2 • LL(a,Ma) + log(m∗)(|V| + 1) The first term represents the goodness-of-fit, where LL(a,Ma) denotes the log-likelihood of dimension a given the model.

The d

ata are grouped into clusters so the data which are similar will be in one group. In this project first we have performed clustering from that lot of cluster group formed using the datasets so the mining become slower, low efficient. So we are using Essential Clusters Identification method. In the Essential Clusters Identification we are finding the essential

clusters which are used more frequently, for identifying the essential clustering Essential Clusters Identification (ECI) algorithm used.

Time-series are common in many recent applications, e.g., stock quotes, e-commerce data, system logs, network traffic management, etc. Compared with traditional datasets, streaming time-series pose new challenges for query processing due to the streaming nature of data which constantly changes over time. Clustering is perhaps the most frequently used data mining algorithm. Surprisingly, clustering streaming time-series still have not explored thoroughly, to the best of our knowledge [10].

**B. Cluster Model Grouping**

The ES clusters are arrange as group and form the model, to from the model we have to find the relation between the each clusters based on the relation the hierarchy, We find the relation between the ES clusters, that relation is based on the certain clusters then we create the models in the hierarchy based, we arrange the clusters it can easy for mining the data in the brain regions. We can orderly mining the data based on the value of input, the system can easily mining and give the effective time consuming output for user.

**C. Mining Clusters and Dependencies**

The system arrange the ES clusters in the model, when the system can get the input from the user it can start mining process based upon the dependencies. The system can compare the data's and finding the output. The dependencies are easy to mine data in the hierarchy model. The dependencies can mine data in the model when it find the result it can give the output to the user, this system can provide the time consuming output to the user.

## IV. INTERACTION K-MEAN CLUSTERING

In this section, we use the algorithm interaction- K-means (IKM) which minimizes the clustering objective function provided in Definition 3. Similar to classical K-means [5], IKM is an iterative algorithm which efficiently converges towards a local minimum of the optimization space.

Algorithm IKM. Analogously to K-means, the first step of IKM is the initialization. As a common strategy for K-means, we propose to run IKM several times with different random

initializations and keep the best overall result. For initialization, we randomly partition DS into K clusters. For IKM it is favorable that the initial clusters are balanced in size to avoid over fitting. Therefore, we partition the data set into K equally sized random clusters and find a set of models for each cluster as described in the previous section. After initialization, IKM iteratively performs the following two steps until convergence: In the

assignment step, each object O is assigned to the cluster which the error is minimal; IKM converges as soon as no object changes its cluster assignment during two consecutive iterations. Usually, a fast convergence can be observed, but there are some rare cases in which IKM does not converge. Analogously to standard K-means, it can be straightforwardly proven that the assignment and the update step strictly decrease the objective function provided in Definition 3. However, due to the greedy stepwise algorithm applied for model finding. O.cid = minC∈C EO, C. It is easy to see that this minimizes the objective function in Definition 3. After assignment, in the update step, the models of all clusters are reformulated. Pseudo code of IKM is provided in Figure 2. As an iterative partitioning clustering algorithm, IKM follows a similar algorithmic paradigm as K-means. However, note that there are significant differences: Our cluster notion requires a similarity measure, which is very different to LP metric distances. The similarity measure applied in IKM is the errors with respect to the set models of a cluster. This similarity measure is always evaluated between an object and a cluster, and not between two objects, In contrast to K-means or K-medoid algorithms.

```
algorithm IKM (data set DS, integer K):
Clustering C Clustering bestClustering;
//initialization
for init := 1 . . . maxInit do
C := randomInit(DS,K);
for each C ∈ C do
MC := findModel(C);
while not converged or iter < maxIter do
//assignment
for each O ∈ DS do
O.cid = minC∈C EO,C
//update
for each C ∈ C do
MC := findModel(C);
if improvement of objective function
```

Figure.4. Algorithm Interaction K-Means.

## V. ESSENTIAL CLUSTER IDENTIFICATION

In this section, we use the algorithm Essential Clusters Identification (ECI) which minimizes the clustering amount. ECI is an algorithm which efficiently provides the mining steps to the system.

Algorithm ECI. The first step of ECI is the initialization. We initialize the k to number of clusters it analyses the maximum number of clusters. The C is finding the essential clusters and storing it in continually for that it should be initialized in algorithm. Ci from that it selecting the clusters from the initial process from that it will continues analysing the all the clusters till it reach all the clusters.

For covering all the clusters we use the distance calculation and we use ei for that it calculating distance. The ci will be updating when the distance ei will increase. When it reaches the maximum all the distance are covered and find the essential clusters it end the process. The more dimensions are included into the model, the more variance is explained and thus the smaller is the error term. Models involving a large part or even all dimensions are not generalizable and hard to interpret. Therefore, the set of explanatory variables of each model needs to be carefully selected [1].

```
Algorithm ECM Grouping
Input E={e1,e2….en}(set of
entities to be  clustered)
//initialization
K(number of clusters)
MaxIters(limit of iterations)
Output:C={c1,c2,….ck}(set of
clustered  centroids)
L= {l (e)l e=1,2,….n}(set of
cluster labels of E)
For each ci €{1…….k}
I ci  = cj € E(e.g random
selection )
End
For each ei€Edo
 I l(e1) = argmin
Distance(e1,cj)j€{1……..k}
End
changed = false;
```

Figure5. Algorithm Essential Cluster Identification.

In this section, we use the algorithm Essential Clusters Model Grouping which minimizes the clustering mining time. ECMG is an algorithm which efficiently provides the mining steps to the system.

Algorithm ECMG. The first step of ECMG is the initialization.  We initialize the k to number of clusters it analyses the maximum number of clusters. The αnxm is for memory of the particular module running. r1….rk. Is a portioning the clustering into the models. The distance all will be calculated for the grouping process using c1……ck. The index will

be formed for the every model in the hierarchy model then the system process easy mining with user input and provide effective output in time.

Considering a pair of clusters, we first generate the models of each individual cluster of from the training data. Then we compute the error of the test objects all models and sum up all errors. To obtain a ranking of the models regarding their ability to discriminate among the clusters, we consider errors the correct cluster of the test object with a positive sign (these errors should be small) and errors the other cluster with a negative sign, respectively. Finally, we sort all models ascending according to the error. The top ranked models best discriminate among the clusters.

```
Algorithm ECMG
Input: anxm: dataset k: number of
clusters
//Integer number
Input K; Memory - αnxm
//Partition dataset into k clusters
r 1 (1.nm/k) □α[(1……..n/k).m)]
r2 (1.nm/k)
□α[(1/k+1……..2n/k).m)]
rk (1……..nm/k) □α[(n(k-
1)/k+1…….2n/k).m)]
//select a center at random
c1 randomize (r1)
 c2 randomize (r1)
 c3 randomize (r1) While(c==q)do
For y=1 to k do py=∑(r1-Cy)2
//Vector for Euclidian distance
```

Figure.6. Algorithm Essential Cluster Model Grouping

## VII. INTERACTIONS AMONG BRAIN REGIONS

We obtained data sets DS11 and DS12 from functional MRI experiments. Functional MRI generates a series of 3-D volume images of the brain. Each image consists of about 60,000 voxels and the interval between time points is about 2-3 seconds. We first applied standard pre-processing including realignment, normalization to a standard template and smoothing.

Our approach is based on a set of time-series. Basically we can use each voxel time series from the images. However, for neighbouring voxels signal activity is very similar. Moreover, medical experts often desire to obtain results at the level of anatomical regions which facilitates interpretation. Therefore, in Section 8.2, we use a brain atlas from [7] with a predefined mask of regions. As an alternative to anatomical regions, in Section 8.3 we use Independent Component Analysis (ICA). From the ICA result, we selected physiologically relevant components and rejected ICs reflecting motions-artefacts or noise.

A. Somatoform Pain Disorder

DS10 [6] has been obtained from a study on Somatoform Pain Disorder and consists of images of 13 subjects with pain disorder and 13 healthy controls. Somatoform Pain Disorder has severe impact on the quality of living of the affected persons since the main symptom is severe and prolonged pain for which there is no medical explanation. The causes of this psychiatric disorder are not fully understood but the hypothesis is that patients have altered mechanisms of observing and processing pain. Therefore, in our experiment, subjects underwent alternating blocks of pain- and non-painful stimulation while in the scanner. After pre-processing we segmented the data of each subject into 90 anatomical regions of interest [7] (ROIs). The task is to cluster persons based on the interaction patterns of the ROIs within the brain during the experiment. Each person is represented by a multivariate time series with 90 dimensions and 325 time points. There are four subjects with 216 time points only.

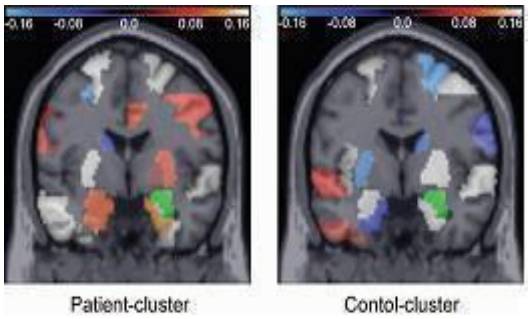



Figure.7. Graphic Representation Of The Models For The Right Amygdale (Green) In Patients And Controls Involving All 90 Aal-Rois. Red To White: Areas With A Signal-

Time Course With Positive Linear Coefficient In The Model, Blue: Negative Linear Coefficient, Respective

Our technique IKM does not require the multivariate time series subjected to a cluster analysis to be of equal length. The result of IKM is superior to the results of all comparison methods: One cluster is composed of nine subjects with somatoform pain disorder and four

healthy controls. The second cluster contains nine healthy controls and four subjects with somatoform pain disorder. Based on previous studies [6], [8], it is known that the right amygdale is strongly associated with somatoform pain disorder. The model of this region is the best separating model among the clusters. Figure 8 presents a visualization of the model of right amygdale.

These time series result in 26 multidimensional time series objects of healthy controls and patients. Causal influence from one area into another was modelled by Granger Causality between time series of brain network activity. Clustering based on nonlinear models reflecting Granger causality separated patients from controls with high cluster purity (84.6%) consistently for a model of the striatum. Each cluster consists of 13 persons. In total, only two persons (one control and one patient) have been incorrectly clustered. Changed influence on the striatum was found for several intrinsic brain networks, indicating an aberrant regulation of striate activity.

## VIII. CONCLUSION

In this paper, we define a cluster as a set of objects sharing a specific interaction pattern among the dimensions. Interaction K-means (IKM) simultaneously clusters the data and discovers the relevant cluster specific interaction patterns. The algorithm IKM is a general technique for clustering multivariate time series and not limited to fMRI data. In addition, we propose Essential cluster Identification algorithm for reduce the size and complexity of the clusters. Essential Cluster Model Grouping algorithm will make the models using the relation and dependencies of every cluster. For the mining the data we use DepMine algorithm and finds the dependencies between the every clusters model and mining the data, then find the result Our approach in the project to understand the complex interaction patterns among brain regions using essential clustering, fast mining and easy to identify Disorders.

## REFERENCES

[1]Claudia Plant, Andrew Zherdin, Christian Sorg, Anke Meyer-Baese, Afra M. Wohlschl¨ager“Mining Interaction Patterns among Brain Regions by Clustering”, 2013

[2] Eamonn Keogh,Shruti Kasetty University of California, Riverside” On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration”,2005.

[3] D. T. Larose, Data Mining Methods and Models. John Wiley & Sons, 2006.

[4]E. I. George, “The variable selection problem,” J. Amer. Statist. Assoc, vol. 95, pp. 1304–1308, 2000.

[5]J.B.MacQueen, “Some methods for classification and analysis of multivariate observations,” in Proc. of the fifth Berkeley Symposium on Mathematical Statistics and

Probability, L. M. L. Cam and J. Neyman, Eds., vol. 1. University of California Press, 1967, pp. 281–297.

[6]H. G¨ undel, M. Valet, C. Sorg, D. Huber, C. Zimmer, T. Sprenger, and T. T¨ olle, "Altered cerebral response to noxious heat stimulation in patients with somatoform pain disorder." Pain., vol. 137, pp. 413–421, Nov 2007.

[7]N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri singlesubject brain," NeuroImageVolume 15, pp. 273–289, January 2002.

[8]I. Strigo, A. Simmons, S. Matthews, A. Craig, and M. Paulus, "Association of major depressive disorder with altered functional brain response during anticipation and processing of heat pain." Arch Gen Psychiatry, vol. 65, no. 11, pp. 1275–84, Nov 2008.

[9]M. L. Kringelbach, "The human orbitofrontal cortex: linking reward to hedonic experience," Nature Reviews Neuroscience, vol. 6, pp. 691–702, 2005.

[10]Jessica Lin1, Michai Vlachos1, Eamonn Keogh1, Dimitrios Gunopulos1, Jianwei Liu2, Shoujian Yu2, and Jiajin Le2"A MPAA-Based Iterative Clustering Algorithm Augmented by Nearest Neighbors Search for Time-Series Data Streams"2002. specific interaction patterns. The algorithm IKM is a general technique for clustering multivariate time.Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. .