Brain Activity Observed Special Cluster Notion For Identifying Disorders

[I]Saranya.T, [II]ShanthaKumari.S

Abstract

The function of brain can be studied by using fMRI. The huge volume of data which is complex in nature is stored as a dataset. This complex dataset requires efficient data mining techniques to study the pattern of brain. In proposed method, the complex interactions patterns are studied by using a clustering technique. In this proposed system each subject is modelled as a multivariate time series in which single dimensions represents different fMRI anatomical regions. An efficient algorithm for partitioning the cluster an interactions K-means (IKM) algorithm is used. Furthermore enhance accuracy of tumour using hierarchical clustering.

Keywords

Clustering, K-means clustering, multivariate time series, tumor detection, interaction among brain regions

I. Introduction

The brain is the part of the central nervous system located in the skull. It controls the mental processes and physical actions of human being. The brain along with the spinal cord and networks of nervous controls information flow throughout the body, voluntary actions such as walking, reading, talking and involuntary reactions such as breathing and digestion. Biomarker is a biological marker refers to a broad sub category of medical signs. It can be used as an indicator of particular disease state or some other physiological state of an organism. Schizophrenia is a psychotic disorder marked by several impaired thinking, emotions and behaviors. fMRI is functional Magnetic Resonance Imaging is a technique that directly measures the blood flow in the brain, thereby providing information on brain activity. functional MRI allows taking a patient through a set of tasks or stimulus like finger tapping, audio, visual and smell. Then see which part of the brain is activated. fMRI also called BOLD imaging.

It is a Magnetic Resonance Imaging based neuroimaging technique which makes it possible to detect the brain areas which are involved in task, a process or an emotion. BOLD effect is the detection of brain areas which are used during a condition is based on the Blood Oxygenation Level Dependent (BOLD) effect. When neurons are activated, the resulting increase need for oxygen is over compensated by large increase in perfusion. fMRI analysis is taken by the variation of the BOLD signal induced by the paradigm is very low (< 2%) and cannot be detected visually. Therefore, advanced statistical methods must be used to identify the voxels in which the signal varies according to the paradigm. In a first step, the images are pre-processed slice-timing correction, realignment of fMRI series to correct patient movements, registration with an anatomical scan, spatial normalization to a brain atlas, segmentation, and spatial filtering. In a second step, the preprocessed images are statistically analyzed, with a model describing the experiment the general linear model. Thresholded fMRI activation maps can be overlaid in color on a high resolution anatomical MR image or displayed on a 3D reconstruction of the brain. Brain Magix performs a user-friendly analysis of clinical fMRI images.

A voxel is a unit of graphic information that defines a point in 3-dimensionalspace. Since a pixel (picture element) defines a point in 2-dimensional space with its x and y coordinates. 3-D space is defined in terms of its position, color and density. Cluster is a number of similar things that occur together. Clustering is a process of partitioning a set of data or object in a set of meaningful subclasses called clusters.

K-means clustering in a data mining is a machine learning algorithm used to cluster observations into groups of related observations without any prior knowledge of those relationships. It is commonly used in medical imaging, biometrics and related field. EEG is a technology provides useful tools for researching the brain and helping patients with brain disorders. An electroencephalogram (EEG) records brain waves, which are produced by electrical activity in the brain. It is obtained by positioning electrodes on the head and amplifying the waves with an electroencephalograph. EEGs are valuable in diagnosing brain diseases such as epilepsy and tumors.

II. System Overview

A. Model Finding

Before addressing the problem to find the clusters the set of models MC can be computed from the set of objects OC which are associated to a cluster C. Models involving a large part or even all dimensions are not generalizable and hard to interpret. To determine the really relevant dimensions, the greedy stepwise algorithm for model finding in combination with the Bayesian Information Criterion (BIC) 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, BIC which determines a balance between goodness of fit and complexity of the model and is defined by:

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

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

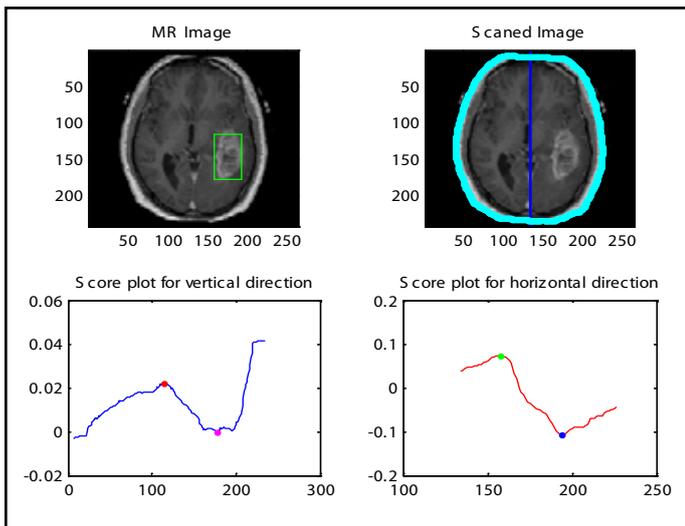Fig. 1: Resulting image of fMRI scanning experiment.

Fig. 2: Tumor Position is detected by bounding box.

B. Interaction K-Means Clustering

Analogously to K-means, the first step of IKM is the initialization. For IKM it is favorable that the initial clusters are balanced in size to avoid over fitting. Then partitions the data set into K equally sized random clusters and finds a set of models for each cluster. After initialization, IKM iteratively performs the following two steps. In the assignment step, each object is assigned to the cluster with respect to which the error is minimal. 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. In different iterations of IKM, BIC may select different numbers of explanatory variables to be included in the cluster models.
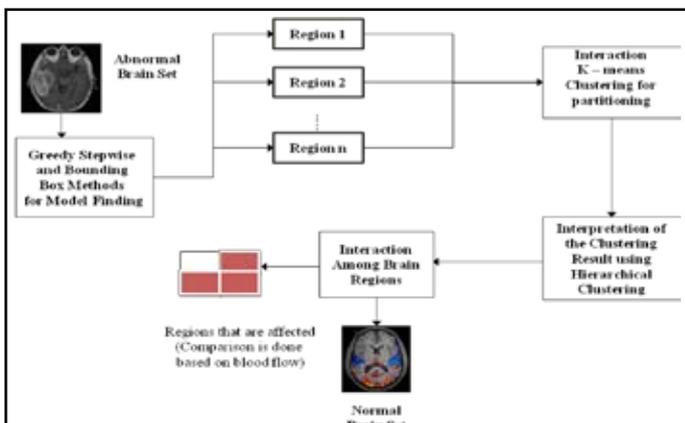

Fig. 3: System Architecture

C. Interpretation of the Clustering Result

Considering a pair of clusters, then generate the models of each individual cluster of from the training data. Compute the error of the test objects with respect to all models and sum up all errors. To obtain a ranking of the models regarding their ability to discriminate among the clusters consider errors with respect to the correct cluster of the test object with a positive sign (these errors should be small) and errors with respect to the other cluster with a negative sign, respectively. Finally, sort all models ascending according to the error. The top ranked models best discriminate among the clusters. The clustering result together with the information about which models best discriminate among clusters is a good basis for user interaction. Expert users can easily select the most

relevant dimensions of the multivariate time series based on this information. Also, experts can easily verify their hypotheses on which dimensions, in the neuroscience application corresponding to anatomical brain regions are most relevant. After selecting the relevant regions, IKM can be run again.
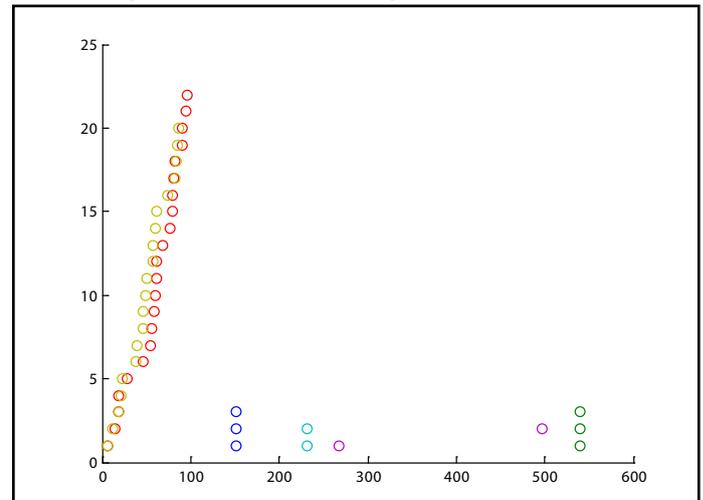

Fig. 4: Hierarchical clusters are displayed

D. Interaction among Brain Regions

1. Functional Magnetic Resonance Imaging

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. This approach is based on a set of time-series. Basically voxel time series from the images are used. However for neighboring voxels signal activity is very similar. Moreover, medical experts often desire to obtain results at the level of anatomical regions which facilitates interpretation.

2. Somatoform Pain Disorder

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 subjects underwent alternating blocks of pain and non painful stimulation while in the scanner. After pre processing segment the data of each subject into 90 anatomical regions of interest (ROIs).

3. Schizophrenia

Schizophrenia is characterized by the impaired interaction between distributed brain regions particularly the striatum. Increased dopamine activity in the striatum is essential for schizophrenia and antidopaminergic treatment the main therapy of the disorder. Intrinsic brain networks are characterized by synchronous brain activity at rest.

5. Feature Selection for Interaction-Based Clustering

Feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. Redundant features are those which provide no more information than the currently selected features, and irrelevant features provide no

useful information in any context. Feature selection techniques are a subset of the more general field of feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. Feature selection techniques are often used in domains where there are many features and comparatively few samples. The archetypal case is the use of feature selection in analyzing DNA microarrays, where there are many thousands of features, and a few tens to hundreds of samples. Feature selection techniques provide three main benefits when constructing predictive models:
Improved model interpretability
Shorter training times

III. Algorithms

A. Greedy Stepwise Algorithm
A greedy stepwise algorithm for model finding in combination with the Bayesian Information Criterion (BIC) 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 BIC which determines a balance between goodness of fit and complexity of the model and is defined by

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

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.

B. K-Means Clustering
K-means clustering is a method of vector quantization originally from signal processing that is popular for cluster analysis in data mining. Kmeans clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

Algorithm:
```
algorithm IKM (data set DS, integer K):
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∈CEO,C
//update
for each C ∈C do
MC := findModel(C);
if improvement of objective function bestClustering
:=C;
```

```
end while
end for
return bestClustering;
```

C. Interpretation of the Clustering Result
```
algorithm dimensionRanking
(Cluster Ci, Cluster Cj): ranking
error in models := new ARRAY [d];
//leave-one-out-validation
for each O ∈OCi ∪OCj do
test := O
OCi :=OCi \test; OCj :=OCj \test;
findModel(Ci); findModel(Cj);
for each cluster ∈{Ci,C j} do
if O.cid = cluster.id then
sign := 1; else sign :=−1;
end if
for i := 1...d do
error in models[d]+ = sign∗ cluster.models[d].
calculateErrorFrom(O. getTimeSeries(d));
end for
end for
end for
sort(error in models);
return error in models;
```

D. Hierarchical Clustering
In data mining, hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering:

1. Agglomerative
This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

2. Divisive
This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

IV. Conclusion
In this system the Interaction K-means was combined with the hierarchical clustering to produce good results, Also the images were robust against noise. The multivariate time series were introduced to detect the region of the brain in an efficient way. The interactions between the clusters were determined and studied to get the exact region that is defected.
The hierarchical clustering method which works well in the combination with interaction K means clustering. The clustering is done by splitting the brain image into various regions. These regions are ranked in a way with the defected region in the highest priority. Likewise the clustering is followed for the normal brain image too. Now this is compared with the defected brain image. On fMRI data the studies on Somatoform Pain Disorder and Schizophrenia and detects very interesting and meaningful interaction patterns.
The Feature Selection Method is used to find out the region that is affected in the highest way. This algorithm improves the model interpretability and enhances the generalization by reducing over fitting. Also the time is shortened by dividing the brain regions

into various segments and finally selecting the particular region that is defected.

Future work can be extended by adding a efficient way to calculate the interaction among the neurons of the brain by concentrating on time consumption.

V. AcknowledgEment

References

[1] Boris Milovic1, Milan Milovic2, (2012), 'Prediction and decision making in health care using Data Mining' Vol. 1, No.12

[2] DSudipta Roy. M, Samir Bandyopadhyay. K, (June 2012), 'Detection and quantification of brain tumor from MRI of brain and its symmetric analysis,' Volume 2 No. 6.

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

[4] Ge. X, Smyth. P, (2000), 'Deformable markov model templates for time-series pattern matching,' in KDD, pp. 81–90.

[5] Jessica Damoisaux. S, Katherine Prater. E, Br Miller. L, Michael Greicius. D, (2011), 'Functional connectivity tracks clinicaldeterioration in Alzheimer's disease'.

[6] Lin. J, Vlachos. M, Keogh E. J, Gunopulos. D, Liu.J.W, Yu. S. J, and Le. J. J, (2005), 'A map-based iterative clustering algorithm augmented by nearest neighbors search for time-series data streams,' in PAKDD, pp. 333–342.

[7] Li. C, Khan. L, and Prabhakaran. B, (2007), 'Feature selection for classification of variable length multiattribute motions,' in Multimedia Data Mining and Knowledge Discovery.

[8] Oates. T, Firoiu. L, and Cohen. P. R, (1999), 'Clustering time series with hidden markov models and dynamic time warping,' in In Proceedings of the IJCAI-99 Workshop on Neural, Symbolic and Reinforcement Learning Methods for Sequence Learning, pp. 17–21.

[9] Oates. T, (1999), 'Identifying distinctive subsequences in multivariate time series by clustering,' in Proc. ACM SIGKDD, pp. 322–326.

[10] Owsley. L, Atlas. L, and Bernard. G, (1997), 'Automatic clustering of vector time-series for manufacturing machine monitoring,' Acoustics, Speech, and Signal Processing, IEEE International Conference on, vol. 4, p. 3393.

Author's Profile and Image

Completed M.E Software Engineering, Assistant Professor in SMIT

Doing my M.E Computer Science and Engineering Final year