



# **Survey on Feature Subset Selection Algorithm in Brain Interaction Patterns**

K.Vidhyadevia<sup>1</sup>, M.Beema Mehraj<sup>2</sup>, Dr.K.P.Kaliyamurthie<sup>3</sup>

Dept of CSE, Bharath University, 173, Agaram Road, Selaiyur, Chennai, India<sup>1,2,3</sup>.

**ABSTRACT:** Functional magnetic resonance imaging (FMRI) patterns provides the prospective to study brain function in a non-invasive way. The FMRI data are time series of 3-dimensional volume images of the brain. The data is traditionally analyzed within a mass-univariate framework essentially relying on classical inferential statistics. Handling of feature selection and clustering is a complicated process in interaction patterns of brain datasets. To understand the complex interaction patterns among brain regions system proposes a novel clustering technique. System models each subject as multivariate time series, where the single dimensions characterize the FMRI signal at different anatomical regions. In this survey paper, compares various research parameters for detection of clusters of objects with similar interaction patterns classification and clustering techniques used in it. The study papers was effective to understand the techniques and gives idea to propose select the key features in the preprocessed dataset based on the threshold values and Dimension Ranking algorithm was used to select the best cluster for assuring best result.

**KEYWORDS:** Data mining, Clustering, Interaction K-means Algorithm, Feature Subset Selection, Dimension Ranking, FMRI.

## **I. INTRODUCTION**

Feature selection for interaction based clustering in interaction patterns among brain images by using feature subset selection algorithm. Feature selection for interaction based clustering in interaction patterns among brain images by using feature subset selection algorithm to provide efficient selection process. The project is to analyze the brain interaction patterns for identifying the state of the brain. Functional magnetic resonance imaging (FMRI) opens up the opportunity to study human brain function in a noninvasive 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. Recently, resting-state FMRI has attracted considerable attention in the neuroscience community. 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. FMRI produce images at higher resolution than other scanning techniques. 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.

Recent findings suggest a modular organization of the brain into different functional modules. In tracking cognitive processes with functional MRI mental chronometry it says that Functional magnetic resonance imaging (FMRI) is used widely to determine the spatial layout of brain activation associated with specific cognitive tasks at a spatial scale of millimeters. Recent methodological improvements have made it possible to determine the latency and temporal structure of the activation at a temporal scale of few hundreds of milliseconds. Despite the sluggishness of the hemodynamic response, FMRI can detect a cascade of neural activations - the signature of a sequence of cognitive processes. Decomposing the processing into stages is greatly aided by measuring intermediate responses. By combining event-related FMRI and behavioral measurement in experiment and analysis, trial-by-trial temporal links can be established between cognition and its neural substrate. 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. Inspired by this idea, a novel technique is proposed for mining the different interaction patterns in healthy and diseased subjects by clustering. At the core of method is a novel cluster notion: A cluster is defined as a set of subjects sharing a similar interaction pattern



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

among their brain regions. A cluster analysis of motion stream data potentially identifies clusters with similar movements, usually performed by different persons. Time series data are frequently large and may contain outliers.

In addition, time series are a special type of data set where elements have a temporal ordering. Therefore clustering of such data stream is an important issue in the data mining process. Numerous techniques and clustering algorithms have been proposed earlier to assist clustering of time series data streams. The clustering algorithms and its effectiveness on various applications are compared to develop a new method to solve the existing problem. After standard pre-processing including parcellation into anatomical regions, each subject is modeled as a data object which is represented by a multivariate time series. The nature of the considered images and the objective of the segmentation being multiple, there is no unique technique for image segmentation and segmenting an image into meaningful regions remains a real challenge. The algorithm is used to evaluate the values within a regional time and group them together based on the merging criteria, resulting in a smaller list and more number of information is collected and compared with the database and finally abnormality can be detected.

The layout of the paper is as follows. In section II, address the above mentioned techniques and also give a brief on the literature being reviewed for the same. Section III, presents a comparative study of the various research works explored in the previous section. Section IV, describes about future work. Section V gives the conclusion in and lastly provides references.

## II. RELATED WORK

In this paper [1] IKM achieves good results on synthetic data and on real world data. It is scalable and robust against noise. Ranking Algorithm improves the efficiency of clustering result there is no separate algorithm for feature selection process Complexity is high. Auto class technique is not applicable if the number of time points varies among the objects, a case frequently occurring in FMRI data. There is no separate mechanism for feature selection

In this paper [2] The k-means method is in fact a perfect candidate for smoothed analysis: it is extremely widely used, it runs very fast in practice, and yet the worst-case running time is exponential. It did not make a huge effort to optimize the exponents as the arguments are intricate enough even without trying to optimize constants. The smoothed analyses so far are unsatisfactory as the bounds are still super-polynomial in the number  $n$  of data points

In this paper [3] Classification decisions are supported by class-specific interaction patterns within the time series of a data object. It would be also interesting to design model-based classifiers for FMRI data and combined FMRI-EEG data sets which are very challenging because of the large number of time series in FMRI data kind of object representation is very natural and straightforward in many applications, not much research on data mining methods for objects of this particular type.

In this paper [4] This leads to a new view about the neural basis of human perceptual decision-making processes. It is also tempting to speculate that the general principles derived from the studies of simple perceptual decision processes reviewed here extend to other settings. Seen in this light, it is not surprising that motor structures seem to have a role in decision formation. It is not yet clear how these structures contribute to decisions that are not linked to particular actions.

In [5] author has monitored the goal of clustering is to identify structure in an unlabeled data set by objectively organizing data into homogeneous groups where the within-group-object similarity is minimized and the between-group-object dissimilarity is maximized. None of these in the paper which included in this survey handle multivariate time series data with different length for each variable. Clustering seasonality patterns of retail data, Cluster analysis of country's energy consumption, discovering consumer power consumption patterns for the segmentation of markets

In this paper [6] The great benefit is that the size of our approximation depends only on the number of coefficients of the model (i.e. the number of reference time series). The distance computation requires rather high runtimes and, if the time series are indexed by a standard spatial indexing method such as the R-Tree [1] or one of its variants, this index will perform rather badly due to the well-known curse of dimensionality. The key issue for approximate clustering is of course to generate accurate results.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

In this paper [7] By applying Support Vector Machines (SVM) to the feature vectors, we can efficiently classify and recognize real world multi-attribute motion data using only a single motion pattern in the database to recognize similar motions allows for less variations in similar motion real time recognition of individual isolated motions accurately and efficiently.

In [8] author has monitored the shape deformation patterns in Anatomical structures show up evidently. Shrinkage of the hippocampus and Cortical and sub-cortical gray matter along the directions. Increasing adas-cog and cdr.sb (measures of increasing cognitive degeneration) and decreasing AVLT and Logical scores (measures of audio verbal learning and logical memory).

In this paper [9] Provide a systematic approach to learn the nature of such time warps while simultaneously allowing for the variations in descriptors for actions. Activity recognition is not very robust to intra- and inter-personal changes of the same action, and are extremely sensitive to warping of the temporal axis due to variations in speed profile activity recognition are not very robust to intra- and inter-personal changes.

In this paper [10] This should provide enough information to allow anyone to reproduce Auto Class, or to use the same evaluation functions in other contexts where these models might be relevant. We will only deal with real and discrete attributes the task of supervised classification: A related problem is unsupervised classification, where preparation cases are also unlabeled

### III. COMPARATIVE STUDY

In this section analyzed the various research works on several parameters and presented their comparison in the table below.

TABLE 1. COMPARISON OF VARIOUS RESEARCH WORKS

S. No	Title	Author	Issues	Tool/ Technology used	Methods and algorithm	Advantages and Disadvantages
1	Mining Interaction Patterns among Brain Regions by Clustering	Claudia Plant, Andrew Zherdin, Christian Sorg, Anke Meyer-Baese, and Afra M. Wohlschläger	There is no separate mechanism for feature selection	Java	Interaction K-means Algorithm	Advantages IKM achieves good results on synthetic data and on real world data. Disadvantages Auto class technique is not applicable if the number of time points varies among the objects, a case frequently occurring in FMRI data.
2.	Smoothed Analysis of the k-Means Method	David Arthur, Bodo Manthey, Heiko Rogli	The smoothed analyses so far are unsatisfactory as the bounds are still super-polynomial in the number n of data points.	Java	K-Means Method	Advantages The k-means method is in fact a perfect candidate for smoothed analysis: it is extremely widely used, it runs very fast in practice, and yet the worst-case running time is exponential. Disadvantages It did not make a huge effort to optimize the exponents as the arguments are intricate Enough even without



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

						trying to optimize constants.
3	Model-based Classification of Data with Time Series-valued Attributes	Christian Bohm, Leonhard Laer, Claudia Plant, Andrew Zherdin	Kind of object representation is very natural and straightforward in many applications, there has not been much research on data mining methods for objects of this special type.	Java	Model-based classifier, iterative stepwise algorithm	<p>Advantages Classification decisions are supported by class-specific interaction patterns within the time series of a data object.</p> <p>Disadvantages It would be also interesting to design model-based classifiers for FMRI data and combined FMRI-EEG data sets which are very challenging because of the large number of time series in FMRI data.</p>
4	The neural systems that mediate human perceptual decision making	Hauke R. Heekeren, Sean Marrett and Leslie G. Ungerleider	Seen in this light, it is not surprising that motor structures seem to have a role in decision formation. It is not yet clear how these structures contribute to decisions that are not linked to particular actions	Java	machine-learning algorithms	<p>Advantages This leads to a new view about the neural basis of human perceptual decision-making processes.</p> <p>Disadvantages It is also tempting to speculate that the general principles derived from the studies of simple perceptual decision processes reviewed here extend to other settings.</p>
5	Clustering of time series data—a survey	T. Warren Liao	Clustering seasonality patterns of retail data, Cluster analysis of country's energy consumption, Discovering consumer power consumption patterns for the segmentation of markets	Java	Relocation clustering, Relocation clustering, k-Means and fuzzy c-means	<p>Advantages The goal of clustering is to identify structure in an unlabeled data set by objectively organizing data into homogeneous groups where the within-group-object similarity is minimized and the between-group-object dissimilarity is Maximized.</p> <p>Disadvantages None of these in the paper which included in this survey handle multivariate</p>

## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

						time series data with different length for each variable.
6	Approximate Clustering of Time Series Using Compact Model-based Descriptions	Hans-Peter Kriegel, Peer Kroger, Alexey Pryakhin, Matthias Renz, and Andrew Zherdin	The key issue for approximate clustering is of course to generate accurate results	Java	Mathematical Model, Efficient Approximate Clustering	<p>Advantages The great benefit is that the size of our approximation depends only on the number of Coefficients of the model (i.e. the number of reference time series).</p> <p>Disadvantages The distance computation requires rather high runtimes and, if the time series are indexed by a standard spatial indexing method such as the R-Tree [1] or one of its variants, this index will perform rather bad due to the well-known curse of dimensionality.</p>
7	Feature Selection for Classification of Variable length Multi-attribute Motions	Chuanjun Li, Latifur Khan, and Balakrishnan Prabhakaran	Real time recognition of individual isolated motions accurately and efficiently.	Java	Support Vector Machines (SVM)	<p>Advantages By applying Support Vector Machines (SVM) to the feature vectors, we can efficiently classify and recognize real world multi-attribute motion data.</p> <p>Disadvantages Using only a single motion pattern in the database to recognize similar motions allows for less variation in similar motion.</p>
8	Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures	Nikhil Singh, P. Thomas Fletcher, J. Samuel Preston, Linh Ha, Richard King, J. Stephen Marron, Michael Wiener, and Sarang Joshi	Increasing adas-cog and cdr.sb (measures of increasing cognitive degeneration) and decreasing AVLT and Logical scores (measures of audio verbal learning and logical memory).	Java	Partial Least Squares (PLS) method	<p>Advantages The shape deformation patterns in Anatomical structures show up evidently shrinkage of the hippocampus and Cortical and sub-cortical gray matter along the directions.</p> <p>Disadvantages Shrinkage of the hippocampus and Cortical and sub-cortical gray matter along the</p>



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

						directions.
9	The Function Space of an Activity	Ashok Veeraraghavan, Rama Chellappa, Amit K. Roy-Chowdhury	Activity recognition are not very robust to intra- and inter-personal changes of the same activity, and are extremely sensitive to warping of the temporal axis due to variations in speed profile	Java	Nominal activity trajectory and a function space	Advantages Provide a systematic approach to learn the nature of such time warps while simultaneously allowing for the variations in descriptors for actions. Disadvantages We will only deal with real and discrete attributes
10	Bayesian Classification Theory	Robin Hanson, John Stutz, Peter Cheeseman	The task of supervised classification: A related problem is unsupervised classification, where training cases are also unlabeled.	Java	Bayesian approach	Advantages This should provide enough information to allow anyone to reproduce Auto Class, or to use the same evaluation functions in other contexts Disadvantages We will only deal with real and discrete attributes

### IV. FUTURE WORK

The case study was very useful to understand the techniques. It is well understood that how the techniques are used to feature Subset Selection Algorithm for Handling Brain Interaction Patterns. Our future work is to design an efficient technique for image segmentation and segmenting an image into meaningful regions in brain image processing.

### V. CONCLUSION

The research, a new approach for the feature selection before clustering mechanism has been presented. Feature Subset Selection method used to choose a subset of the original features to be used for the subsequent processes. In this algorithm, initially collect the preprocessed dataset then find out the ranges of the respected dataset. After that, select the key features in the preprocessed dataset based on the threshold values. Interaction K-means (IKM), a partitioning clustering algorithm used to detect clusters of objects with similar interaction patterns classification and clustering. Interaction K-means (IKM), a partitioning clustering algorithm suitable to detect clusters of objects with similar interaction patterns classification and clustering. Finally, Best Ranking algorithm used to select the best cluster for assuring best result.

### REFERENCES

- [1] D. Arthur, B. Manthey, and H. Röglin, "Smoothed analysis of the k-means method," J. ACM, vol. 58, no. 5, pp. 19:1–19:31, Oct. 2011.  
[2] Nikhil Singh, P. Thomas Fletcher, J. Samuel Preston, Linh Ha, Richard King, J. Stephen Marron<sup>2</sup>, Michael Wiener<sup>3</sup>, and Sarang Joshi : Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures, Springer, MICCAI 2010.





# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

- [3] C. Böhm, L. Läer, C. Plant, and A. Zherdin, "Model-based classification of data with time series-valued attributes," in Proc. BTW, 2009, pp. 287–296.
- [4] H. Heekeren, S. Marrett, and L. Ungerleider, "The neural systems that mediate human perceptual decision making," Nat. Rev. Neurosci., vol. 9, no. 6, pp. 467–79, Jun. 2008.
- [5] H.-P. Kriegel, P. Kröger, A. Pryakhin, M. Renz, and A. Zherdin, "Approximate clustering of time series using compact modelbased descriptions," in Proc. DASFAA, New Delhi, India, 2008, pp. 364–379.
- [6] 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.
- [7] 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
- [8] A. Veeraraghavan and A. K. R. Chowdhury, "The function space of an activity," in Proc. CVPR, New York, NY, USA, 2006, pp. 959–968.
- [9] T. W. Liao, "Clustering of time series data – A survey," Pattern Recognit., vol. 38, no. 11, pp. 1857–1874, 2005.
- [10] 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.

## BIOGRAPHY

**K. Vidhyadevi** received the B.E degree in Computer Science and Engineering from Arunai Engineering College, Tirunavanamalai in 1998. She Worked as Lecturer for 3 Years at Polytechnic College and Worked as MIS Co-Ordinator for 2 Years at Govt. SSA Chennai. Pursing M.Tech Computer Science and Engineering from Bharath University. Participated in workshops on Android PHP .Net C# programming language.

**M. Beema Mehraj** received M.Tech in Bharath university Now She is working as an Assistant professor at Bharath University for a years.

**Dr. K.P. Kaliyamurthi** received Ph.D and M.Tech from the Bharath University He had 23 years experience in teaching .Now he is working as a HOD in the Department of CSE at Bharath University. He has published more than 40 research papers in international journals and national journals.