ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

Interaction K-means Clustering for Finding Similar Interaction Patterns among Brain Regions

S. Bencyangel[1], J. Dhivya Merlin[2]

P.G. Scholar, Department of CSE, Adhiyamaan College of Engineering, Hosur, India[1]

Assistant Professor, Department of CSE, Adhiyamaan College of Engineering, Hosur, India[2]

Abstract: Brain is an important part, which controls all the functions of the human body. Brain activity is the only resource to understand whether any brain disorders exist or not .Functional Magnetic Resonance Imaging (FMRI) paves the way to study about the brain functions. But the information content from that is large in volume and complex and so the data requires some effective and efficient data mining techniques. To understand the complex interaction patterns among brain regions novel clustering technique is proposed.The objective is to assign objects exhibiting a similar intrinsic interaction pattern to common cluster. Based on this novel cluster notion, interaction K-means (IKM) is proposed, 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, fMRI data, Interaction patterns, Multivariate time series.

I.INTRODUCTION

Human brain activity is very complex and not fully understood.Abnormal brain activity is the only resource to understand psychiatric disorders. Functional magnetic resonance imaging opens up the opportunity to study human brain activity in a noninvasive way. The basic signal of fMRI depends on the blood-oxygen-level-dependent effect, which allows indirectly imaging brain activity by changes in the blood flow related to the energy consumption of brain cells.

fMRI data are time series of 3-dimensional volume images of the brain and are massive in volume with more than hundred thousands of voxels and hundreds of time points.These kind of data represent complex brain activity and also the information content is expected to be highly complex. Recently, resting-state FMRI has been considered to a great extent. In resting state fMRI, subjects are instructed to just close their eyes and made to relax while they are in the scanner. To understand thecomplex brain activity, it is essential to understand the complex interplay among brain regions during task and at rest.

Inspired by this idea, we propose a novel technique for mining the different interaction patterns in healthy and diseased subjects by clustering.The core of our method is a novel cluster notion: A cluster is defined as a set of subjects sharing a similar interaction pattern among their brain regions. A cluster analysis of motion stream data potentially identifies clusters with 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.
We model each subject as a data object which is represented by a multivariate time series. Each of the dimensions is a time series corresponding to the fMRI signal of a specific anatomical brain region. Our approach Interaction K-means (IKM) clusters the data and discovers the relevant cluster specific interaction patterns. The algorithm IKM is a generaltechnique for clustering multivariate time series and not limited to FMRI data.

II.RELATED WORK

M.D. Fox and M. E. Raichle, 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.

Chuanjun Li, Latifur Khan, and Balakrishnan Prabhakaran 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 there is use of 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 length for even

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

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 Support Vector Machines (SVM). 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.

H.-P. Kriegel, P. Krger, A. Pryakhin, M. Renz, and A. Zherdin,Approximate clustering of time series using the compact model based descriptions: A time series represents a collection of values obtained from sequential measurements over time. The purpose of time-series data mining is to obtain the meaningful knowledge from the unlabeled set of data. Clustering time series data has a drawback that length of time series has a negative influence on run time. Approximate clustering used in the existing system suffers from low accuracy.In this work,a method is proposed for compression of time series based on mathematical models and every time series is represented by a set of some specific reference time series.So the cost depends only on the number of reference time series rather than the length of time series.Thus using small number of reference time series provides good accuracy by reducing the storage cost and runtime of clustering algorithm.

T. W. Liao, Clustering of time series data a survey: Mainly the survey is based on three key components of time series clustering algorithm, the similarity/dissimilarity measure, and the evaluation criterion. The author has observed the goal of clustering is to identify structures 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**.**

M. Vlachos, J. Lin, E. Keogh, and D. Gunopulos, "A wavelet-based anytime algorithm for k-means clustering of time series: Most of the data mining and machine learning algorithms do not work well for clustering of time series data, due to its high dimensionality, very high feature correlation and large amount of noise. These challenges are addressed by a novel anytime algorithm, version of K-means clustering algorithm of time series. An initial clustering is done with quick and dirty data. This is used to initialize clustering at a slightly finer level of approximation. The process is repeated until the clustering results stabilize or until the approximation is raw data.

D. Arthur, B. Manthey, and H. Rglin, Smoothed analysis of the k-means method: This paper introduces the smoothed running time of the k-means method. 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.

E. J. Keogh, S. Lonardi, and C. A. Ratanamahatana, "Towards parameter-free data mining: Most of the data mining algorithms require the input parameters. The problem of working with the algorithm having parameters is that first, incorrect settings may cause the algorithm to fail. Second, it may overestimate the significance of reported pattern. Data mining algorithm should have few parameters as possible because a parameter free algorithm may limit our ability to impose our expectation. In this work, we describe about parameter free data mining algorithm, which have showed good results in bio informatics and computational theory. This approach is very helpful for anomaly detection, classification and clustering with time series, text, and video datasets.

III.PROPOSED SYSTEM

In this section, we describe the algorithm interaction K-means (IKM), an iterative algorithm which minimizes the clustering objective function.

Algorithm IKM

Similar to K-means, we run IKM several times with different random initializations. 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.

The first step of IKM is the initialization. For initialization, we randomly partition datasets DS into K clusters. After initialization, IKM iteratively performs two steps until convergence. The two steps are Assignment and Update. In the Assignment step, each object O is assigned to the cluster with the possibility of minimum error occurrence. After Assignment, in the Update step, the models of all clusters are reformulated. As an iterative partitioning clustering algorithm, IKM follows a similar algorithmic paradigm as K-means. However some of the differences exists. our cluster notion requires a similarity measure which is always evaluated between an object and a cluster but not between two objects.

In contrast to K-means or K-medoid algorithms, we cannot state that a data object is the representative of a cluster.The cluster representative in IKM is a set of models describing a characteristic pattern of

Algorithm
IKM (data set DS, integer K):
Clustering C
Clustering best Clustering;
//initialization
forinit := 1 . . . maxInit**do**

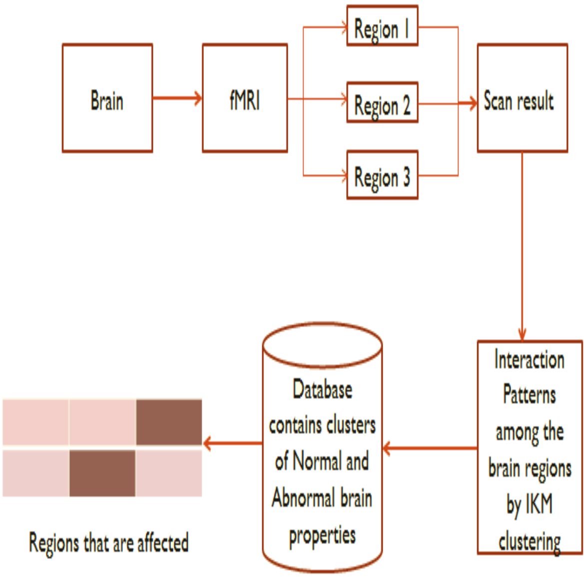
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

C :=randomInit(DS, K);
for each C ∈Cdo
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 ∈Cdo
Mc:=findModel(C);
ifimprovement of objective function
bestClustering := C;
end while
end for
returnbestClustering;

As for ordinary K-means, the runtime of IKM scales linearly with the number of objects n, since the complexity of assignment step is linear in n and usually only a few iterations are performed. Clearly, the update step is the most computationally expensive step, since model finding involves matrix inversion with complexity of O(d2) combined with the greedy stepwise algorithm with complexity of O(d3).

System architecture



The architecture describes how to find out whether the brain is affected or not. In the initial step brain will get scanned in the scanner machine. Functional magnetic resonance imaging, or fMRI, is a technique for measuring brain activity. It works by detecting the changes in blood oxygenation and When a brain area is more active it takes more oxygen and to meet this increased demand blood flow increases to the active area. This scan result is then use for finding the interaction patterns among the brain. The interaction k-means algorithm is used for finding the interaction patterns and initially a cluster containing the similar interaction patterns is formed. This cluster will be then compared with the database. The database contains two types of clusters-one is normal brain cluster and another one is diseased brain cluster.The cluster is finally compared with the database cluster. After comparison, the result will show us which part of brain is affected or not.

IV.RESULTS AND CONCLUSION

Our experimental evaluation demonstrates that the interaction based cluster notion is a valuable complement to existing methods for clustering multivariate time series. IKM achieves good results on synthetic data and on real world data from various domains, but especially excellent results on EEG and fMRI data. Our algorithm is scalable and robust against noise. Moreover, the interaction patterns detected by IKM are easy to interpret and can be visualized. Nonlinear models show their superiority in the corresponding real world data. In ongoing and future work, we plan to extend our ideas to differential equations. We want to consider different models for different regions of the time series. We intend to work on methods for suitable initialization of IKM, since existing strategies for K-means cannot be straightforwardly transferred to IKM because of the special cluster notion. We are also investigating in feature selection for interaction-based clustering.

REFERENCES

[1]M. D. Fox and M. E. Raichle, “Spontaneous fluctuations in brainactivity observed with functional magnetic resonance imaging,”Nat. Rev. Neurosci., vol. 8, no. 9, pp. 700–711, 2007.
[2]C. Li, L. Khan, and B. Prabhakaran, “Feature selection for classificationof variable length multiattribute motions,” in MultimediaData Mining and Knowledge Discovery, V. A. Petrushin and L. Khan,Eds. London, U.K.: Springer, 2007.
[3]H.-P. Kriegel, P. Krger, A. Pryakhin, M. Renz, and A. Zherdin, Approximate clustering of time series using compact modelbased descriptions, in Proc. DASFAA, New Delhi, India, 2008, pp. 364379.
[4]T. W. Liao, “Clustering of time series data – A survey,” Pattern Recognition., vol. 38, no. 11, pp. 1857–1874, 2005.
[5]D. Arthur, B. Manthey, and H. Rglin, Smoothed analysis of the k-means method, J. ACM, vol. 58, no. 5, pp. 19:119:31, Oct. 2011.
[6]C. Bhm, L. Ler, C. Plant, and A. Zherdin, Model-based classification of data with time series-valued attributes, in Proc. BTW, 2009, pp. 287296.
[7] 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.
[8]S. Smith et al., Correspondence of the brains functional architecture during activation and rest, PNAS, vol. 106, no. 31, p. 13040, 2009.
[9]M. Vlachos, J. Lin, E. Keogh, and D. Gunopulos, “A wavelet-based anytime algorithm for k-means clustering of time series,” in Proc. Workshop Clustering High Dimensionality Data Its Applications, 2003,pp. 23–30.