Asian Journal of Engineering and Technology Innovation 02 (05) 2014 (33-37)

Asian Journal of Engineering and Technology Innovation

ISSN : 2347-7385

REVIEW ARTICLE

**Received on: 01-12-2014**
**Accepted on: 10-12-2014**
**Published on: 22-12-2014**

**Prof. Deepak Dharra, Abhishek Iyer, Nitika Gulhane, Sultana Parveen Khan, Vishal Dond**
Computer Engineering Department, Indira college of engineering and Management Pune University, Maharashtra, India
deepakdharrao@gmail.com
abhsheyy@gmail.com
nitikagulhane@gmail.com
sksultanakhan@gmail.com
vishal.dond@gmail.com

**QR Code for Mobile users**

Conflict of Interest: None Declared

# Event Characterization and Prediction Based on Temporal Patterns Using Mrps in Dynamic Data System

**Deepak Dharra, Abhishek Iyer, Nitika Gulhane, Sultana Parveen Khan, Vishal Dond**
Computer Engineering Department, Indira College of Engineering and Management Parandwadi, Pune.

**Abstract:**
Multivariate Reconstructed Phase space is a new method proposed for identifying and characterizing temporal patterns. Among the several existing methods this method is the most efficient and beneficial. This method along with Gaussian Mixture Model characterizes events in Dynamic data system. MRPS is an extension of original univariate phase space and can be applied in numerous research fields. As MRPS considers Multiple data sequences it has wide application in medical as well as in financial time systems. Nowadays , Pattern Identification of ECG is gaining importance since the classification algorithm in MRPS results accuracy in diagnostics. MRPS-GMM approach can also be applied in order to estimate the GDP(Gross Domestic Product ) of a nation's economy.

**Keywords** – Multivariate Reconstructed Phase space, temporal Patterns, Gaussian Mixture Model

**Cite this article as:**
Prof. Deepak Dharra, Abhishek Iyer,Nitika Gulhane, Sultana Parveen Khan,Vishal Dond
**Event Characterization and Prediction Based on Temporal Patterns Using Mrps in Dynamic Data System**,
Asian Journal of Engineering and Technology Innovation 02 (05); 2014; 33-37.

## 1. INTRODUCTION

Data mining deals with extraction of data from large databases. Temporal data mining focuses on data mining of large sequential data. Events are the occurrences in a complex data system which is related to temporal patterns. The frequently recurring signals in a time series data are called as temporal patterns. Basically, the pattern consists characteristics of the original data and can be used for data summarization and identification. Event characterization and pattern identification in Dynamic Data System is a new direction in the field of research.

The existing approach for detecting temporal patterns is Reconstructed Phase space which is a univariate approach. This approach discovers the correspondence between events and patterns in a same time sequence. Although RPS performs well with univariate case it has some limitations which can be overcome by the new proposed method. RPS considers a single data sequence, which consists of both temporal patterns and events .It often leads to poor performance due to lack of explanatory variables and noise in the event data sequence . RPS due to its shifting effects leads to missing of temporal patterns [7].

To overcome these drawbacks we propose a new MRPS (Multivariate Reconstructed Phase Space) which allows embedding of multiple time sequences simultaneously. MRPS can provide a more feasible and optimal solution. Using MRPS-GMM approach we can detect patterns in Multivariate environment and with GMM approach we can enhance the performance of a classifier. The proposed system uses a Gaussian Mixture Model which is a probabilistic model.GMM parameters can be estimated with the data obtained from training stage using the Iterative Expectation-Maximization (EM) Algorithm. The Expectation –Maximization algorithm is an iterative algorithm used to find the maximum likelihood of parameters in statistical models.

**This Paper is divided into 5 sections:** Section 1 includes introduction. The second section of literature survey describes the previous work in temporal data mining. Section 3 elaborates the system architecture and system features. The algorithms are briefly described in section 4 and section 5 includes conclusion.

## 2. LITERATURE SURVEY

In this section we discuss about the innovative methods and approaches for temporal pattern detection, characterization and prediction.

DevendraTayal [3] proposed various methodologies developed using Fuzzy time series .The time series analysis has its applications for engineering and economy problems. The proposed approach was compared with time variant and invariant fuzzy time series and gave better results. Although this method provided with numerous benefits, the time series had to be converted to periodic and stationary series in order to analyse it and should be examined in a phase space in order to get the requested pattern from it.

Jeremy Archuleta [1] presented a characterization of a Map Reduce-based data mining application on a general-purpose GPU (GPGPU).In order to guide the GPGPU programmer towards optimal performance within a broad design space the authors proposed eight performance characterizations of data mining application. The Experimental results[1] indicated that GPGPUs can provide better performance , but the "one –size-fits-all" approach is unsuitable for temporal data mining on graphics processors. The approach presented in this paper lacks some work for real time and data mining analysis.

Prof. RupeshMahajan [5] presented a paper on time series data. This paper introduces the work that link data mining with time series modeling. The proposed techniques like data reduction, transforms on time series and indexing method provides automatic time series identification and automatic outlier detection.

Kelvin Sim[2] contributed his work in temporal Data mining by identifying several financial problems, which require mining of actionable subspaces defined by objects and attributes over a sequence of time. The framework included in this paper is MASC algorithm. It includes three modules namely projection of Standard Relational Database, Subspace clustering, Database Binarization .Clustering aims at finding groups of similar objects and has the ability to suggest concrete and useful actions. This approach was robust to noise and could find efficient solutions to optimization problems. This approach had limitations like it require objects to have similar values across time, and it suffers from scalability.

Mohammed Abdul Khaleel [6] proposed a survey on data mining techniques. The main research in this paper focuses on finding data mining techniques that can be used to extract temporally frequent diseases. Various algorithms proposed in this paper are Time annotated sequences, time series data mining, fuzziness approach ,Multiscale matching .The work in this paper provides reusable methodology and extracts interesting TAS patterns. Although this had applications in medical data mining , the classical frequent pattern mining cannot utilize the time interval between events .

Mohamed F Ghalwash[4] in the year 2012 proposed the theory of Multivariate Shapelets detection. In this paper the authors have studied the problems using time series segments called shapelets.TheAlgorithms implemented in this approach are Multivariate Shapelet Detection Shapelet Pruning, UnivariateShapelets Detection ,Distance

Threshold method, Utility Score method ,Classification, Regressive Moving Average (ARMA) model. This approach is applicable to any domain and MSD allows early and patient-specific classification of multivariate time series. The most important drawback of this system was the results obtained using conventional classification methods on short time series is not accurate and the length of time series limits the applicability of the proposed approach to datasets with smaller number of data instances.

WenjingZhang [7] in the year 2014 proposed a new method in his paper that applies a Multivariate Reconstructed Phase Space. MRPS identifies multivariate temporal patterns that are predictive of events in dynamic data system. This method incorporates data categorization based on the definition of events. Multivariate Embedding and Parameter Estimation, Reconstructed Phase Space with Transformation,Gaussian Mixture Model Classification and optimization in Multivariate Feature Space. This method along with Gaussian Mixture Model and Bayesian approach can be applied on multiple data sequences. This new method is applied to overcome the drawbacks of Reconstructed Phase Space. RPS is based on univariate case but results in poor performance and loss of temporal patterns.

## 3. SYSTEM ARCHITECTURE

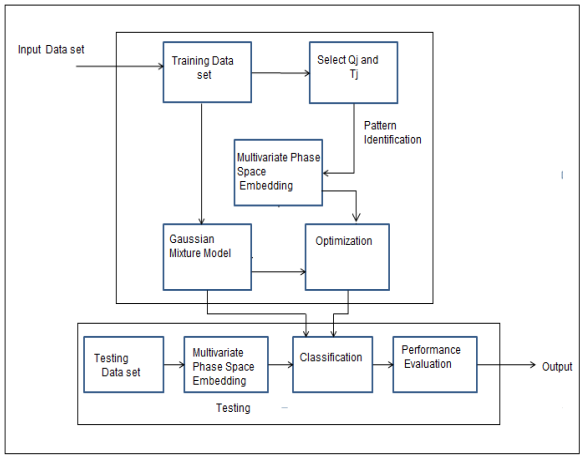


Fig 3.1 Block Diagram of MRPS-GMM Framework

### 3.1 SYSTEM FEATURES

Consider a multivariate data sequence X(t) in a dynamic data system,

$X(t) = [x_{1t}, x_{2t}, \ldots, x_{mt}, x_{et}]^T t = 1, 2, \ldots N$ .....[1]

Here t is the time index, $x_{et}$is the target sequence, i=1,2,....,m represents m multivariable sequences. The system is basically based on these multivariable sequence and the events occurring in these target sequence. Since the proposed MRPS-GMM algorithm can be applied in an ECG system, the input is a data set. The data sets are nothing but collection of data. The Lorenz Map, Sludge Volume Index (SVI), Rossler Map are the three data sets considered. If considering an example of ECG machine the input will be in the textual form and by applying MRPS-GMM algorithm in the Pattern Identification stage and then Testing it results a characterized output in tabular format. The similar MRPS algorithm can also be applied for predicting patterns in the field of economy. It can be applied in predicting the Gross Domestic Product. The stages of characterizing and predicting patterns using MRPS-GMM are explained below.

#### 3.1.1 Data Categorization and Preparation

The input Data set will be categorized in this module. This can be done by defining an event function. In dynamic Data system an Event is an occurrence that describes the changing state of internal system. Event function in terms of one-step prediction can be defined as:

$g(x_t) = x_{t+1}$

K-step forward event function can be defined as,

$g(x_t) = \max\{x_{t+1}, x_{t+2}, \ldots\ldots, x_{t+k}\}$

Event function characterizes embedding vectors in the RPS and discovers correspondence between patterns and events. This multivariate data sequences is then divided into training and testing data sets .The event function for target sequence is defined and the time delay of temporal patterns is obtained. The dimension of Phase space is also determined in this module.

#### 3.1.2 Training Stage

At this stage we construct a multivariate embedding. This can be achieved by combining all individual embedding of each sequence into an MRPS vector. The MRPS vector is given as:

$X_i = (x_{1i}, x_{2i}, \ldots\ldots, x_{ji}, \ldots, x_{mi}, x_{ei})$

This $X_i$ is then transformed into a topological equivalent phase space to obtain trend invariant embedding $\emptyset(X_t)$ .The task of training stage is construction of GMM from training data set. The Expectation Maximization algorithm is also performed at this stage .The EM algorithm performs two steps: Expectation step (E) and Maximization step (M).E step creates a function for the construction of log likelihood evaluated using the current estimate of the parameters and in M step it computes parameters maximizing the expected log-likelihood found in the E step. Training stage performs optimization to objective function and obtains minimizer and classifier for identifying temporal patterns.

### 3.1.3 Testing stage

The predictive pattern classifier obtained at the training stage is applied to testing stage and the events are predicted in the target sequence. A forecast will be made whether or not an event will occur.

### 3.1.4 Computation Complexity

Complexity of Gaussian Mixture Model is O(kn). For large data set limited memory BFSG method can be applied by storing only recent m step values. BFSG has complexity $O(n^2)$.

## 4. ALGORITHMS PROPOSED

At the training stage the multivariate data sequences are categorized into events, patterns, and normal state. The pattern vectors that are predictive of future patterns take positive values of event function and the negative values are taken by normal state vectors. These are then applied to the algorithms .

### 4.1 Multivariate Phase Space Embedding

In the embedding process the embedding dimension $Q_j$and the time delay $T_j$ is estimated using mutual information and the false nearest neighbor method. Multivariate Phase Space Embedding is the first stage in Pattern Identification.

### 4.2 Gaussian Mixture Model

This algorithm is applied on the data set estimating the statistical distribution of event state, pattern state and normal state. The Gaussian mixture model can be defined as:

$$\boldsymbol{P(x) = \sum_{i=1}^{M} P(\omega_i)\, N(x|\mu_i\, , \textstyle\sum_i)}$$

Where M is the number of mixtures,$P(\omega_i)$ is the marginal distribution of i th components of mixtures . The GMM parameters are evaluated using Expectation Maximization (EM ) algorithm. Gaussian Probability density function is given as:

$$\boldsymbol{P(x) = \frac{1}{\sqrt{(2\pi)^Q \quad |\textstyle\sum|}} exp\left(-\frac{1}{2}(x-\mu)\textstyle\sum^{-1}(x-\ \mu)^{\tau}\right)}$$

$|\sum|$ is determinant of the covariant matrix,

x and μ are row vectors.

This function estimates expected value and covariance.

### 4.3 Expectation Maximization Algorithm

EM algorithm is an iterative algorithm which is applied to statistical models. It estimates parameters for GMM.

## 5. Conclusion

MRPS has wide applications as it considers multivariate time sequences. It is used in ECG mechanisms in order to identify the varying temporal patterns .It can also be applied in financial time systems as well as in economic system like GDP. MRPS-GMM approach can obtain the Gross Domestic Product of a nation. The experimental evaluation [7] on three data sets concluded that MRPS is better than univariate phase. A new objective function is also defined.

## REFERENCES


1. Jeremy Archuleta ,Yong Cao, Wu-chunFeng , Tom Scogland. *Multi-Dimensional Characterization of temporal data mining on Graphics Processors.* Department of Computer Science ,Virginia Tech .2009
2. Kelvin Sim, ArdianKristantoPoernomo, VivekanandGopalkrishnan. *Mining Actionable Subspace Clustered in Sequential Data .* 2010.
3. DevendraTayal ,ShilpaSonawani, GunjanAnsari,Charu Gupta. *Fuzzy time series of Low Dimensional Numerical Data.* Jan-Feb 2012.
4. Mohamed F Ghalwash and ZoranObradovic. *Early Classification of Multivariate Temporal Observations by extraction of interpretable shapelets*. August 2012.
5. Prof. RupeshMahajan, Prof ArivanthamThangvelu ,ProfminalShahakar. *Data mining Techniques for identifying Temporal Patterns of Time series data.* Nov-Dec 2012 Department of IT, Pad. Dr DYPIET Pune-18.
6. Mohammed Abdul Khaleel ,Sateesh Kumar Pradhan, G.N Dash, F.A Mazarbhuiya. *A survey of Data mining Techniques on medical data for finding Temporally Frequent Diseases.*Dec 2013
7. WenjingZhang , Member IEEE, and XInFeng ,Senior member IEEE. *Event Characterization and Prediction Based on Temporal Patterns in Dynamic Data System.* Jan 2014.

8. Roshan Joy Martis, ChandanChakraborty, Ajoy .k .Ray .*A two segment mechanism for registration and classification of ECG Gaussian Mixture Model*. Feb 2009.
9. Richard J Povinelli and XinFeng*Data mining of Multiple non stationary time series*
10. Sachinsingh and NetajiGandhi.N*Pattern Analysis of different ECG signal using Pan – Tompkins algorithm* Department of physics ,Indian Institute of Technology Roorkee, India IJCSE
11. SrivatsanLaxman and P S Sastry*A survey of Temporal Data mining* Department of Electrical Engineering, Indian Institute of Science, Bangalore 560 012, India
12. Richard J. Povinelli, Senior Member, IEEE, and XinFeng, Senior Member, IEEE.*A new Temporal Pattern Identification method for Characterization and Prediction of Complex Time series Events.*IEEE Transactions on Knowledge and Data engineering, March/April 2003.