Mining Multivariate Temporal Patterns for Event Characterization and Prediction

G.Aswini[1], A.R.Ashok Kumar[2], D.Durai kumar[3]
[1]M.Tech Student, [2]Assistant Professor, [3]Associate Professor & Head,
Dept. of Information Technology,
Ganadipathy Tulsi's Jain Engineering College, Vellore, Tamilnadu, India

***Abstract* - Characteristic and prediction of the events are essential in many applications, such as forecasting economic growth, financial decision making etc. This can be done by processing the temporal patterns which are observed event data sequence often closely related to certain time-ordered structures. Among several existing method reconstructed phase space work well but only for univariate data sequence. So we propose a multivariate reconstructed phase space which is uses supervised clustering for characteristic and prediction of event from these dynamic data sequence. An optimization method is applied finally to estimate the parameters of the classifier that defines an optimal decision boundary in the Multivariate RPS.**

***Keywords* - Temporal Patterns, Reconstructed Phse Space, Uni-variate Data, Multivariate RPS**

1. INTRODUCTION

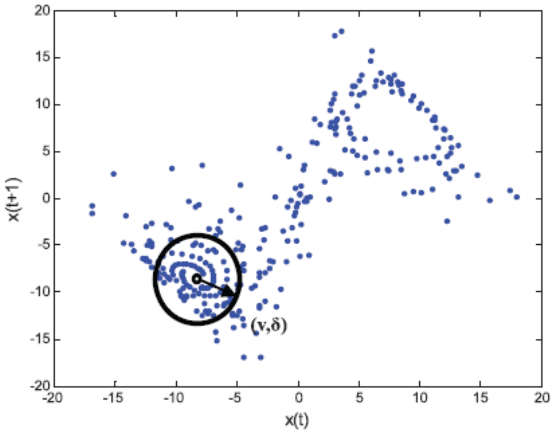
A temporal pattern is defined as a segment of signals that repeats frequently in the whole time-based signal sequence. The extracted patterns translate the characteristics of the original temporal sequence and can be used for pattern detection. For example, the temporal signal sequences could be the movements of head, hand, and body etc. noted down for a specific time interval. The patterns of the body movement represent the habit of a person. In many applications, the multivariate dynamic data system observed in is frequently complex and chaotic. Events are important occurrences or interpretation sparkly the internal states of the complex data system. For occurrence, in a water behavior plant, a point of reading of a chemical might indicate the not working of the plant.

Events, as an application specific subset in the target sequence, are frequently strictly related to convinced time-ordered structures, called temporal patterns. Discovering such temporal patterns that are characteristic and predictive of the events are critical in many applications, such as determining the timing of positions of certain financial securities, forecasting economic growth, detecting a medical anomaly condition, and interpreting of the dynamics in the underlying system. Specifically, the evolutional patterns of GDP are used to address the characteristics of the economic growth. Even though much concentration has been on the research of time-varying data streams categorization], time series similar, clustering growing stream data and symbolization, much less explore has addressed the problem of characterizing and predicting essential events and anomalies in the dynamic data system. In further words, we are not only paying attention in detecting usual temporal patterns in the data sequence, but more paying attention in identifying the patterns that are characteristic and predictive of essential events in the target data sequence, and exploring the fundamental relationships between the events and causing variables in the multivariate data sequence.

2. PROBLEM STATEMENT

RPS method works well in the uni-variate case, one of the limitations of this method is that it only considers a single data sequence, which contains both the temporal patterns and events of interest. It may result in poor performance. Among several existing methods for such a task, one of them is the reconstructed phase space (RPS) method. The fundamental concept of the RPS method is to embed the uni-variate data sequence into a multidimensional time-legged phase space with appropriate time delay and embedding dimension. It capable of representing temporal patterns of nonlinear dynamic sequence data that is typically chaotic and irregular. It has been applied in a variety of research fields, such as ECG signal pattern identification of the atrial electrophysiology and security index forecasting.

Figure



1. Example of a 2D temporal pattern cluster with center v and δ.

3. RELATED WORK

3.1 MULTIVARIATE EMBEDDING

This module embeds the multivariate sequence into the MRPS by estimating time delays and embedding dimensions for each and every single sequence. The mutual information function was found to be effective in estimating the time delay. The dimension for each variable is determined using a false nearest-neighbors technique. By applying the mutual in sequence function which provides quantitative quantify of spatial patterns in phase space.

The mutual information is calculated by,

$$M(\mathbf{x}_{ji}, \mathbf{x}_{j,i+\tau_j}) = \sum_{n,m} p_{nm}(\tau_j) \ln\left(\frac{p_{nm}(\tau_j)}{p_n p_m}\right),$$

A difference measure is defined by

$$r_i = \sqrt{\frac{\left\|\mathbf{x}_{nj}^{Q_j+1} - \mathbf{x}_{mj}^{Q_j+1}\right\|^2 - \left\|\mathbf{x}_{nj}^{Q_j} - \mathbf{x}_{mj}^{Q_j}\right\|^2}{\left\|\mathbf{x}_{nj}^{Q_j} - \mathbf{x}_{mj}^{Q_j}\right\|^2}},$$

3.2 RECONSTRUCTED PHASE SPACE WITH TRANSFORMATION

Different starting values or local trends can cause partition of structural parallel temporal patterns into different regions of phase space. To address this problem, To consider applying a linear transform on phase space embeddings. As a result, the new phase space gives a detrained demonstration of the data sequence. Though the Euclidean distance in the new phase space, can better measure the structural parallel between temporal patterns than in a unadventurous phase space. Regard as two sample data sequences with time delay,

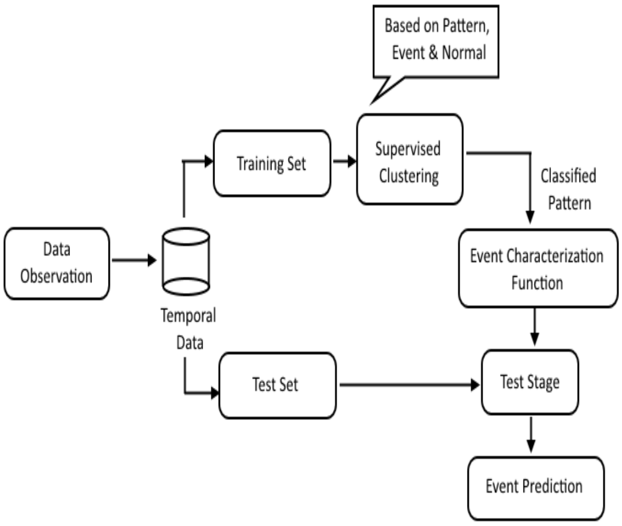
$$\mathbf{x} = \{x_{t-(Q-1)\tau}, x_{t-(Q-2)\tau}, \ldots, x_t\},$$
$$\mathbf{y} = \{y_{t-(Q-1)\tau}, y_{t-(Q-2)\tau}, \ldots, y_t\},$$

The correspondence of temporal structures stuck between two sequences can be calculated by,

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{Q} \left(x_{t-(Q-i)\tau} - y_{t-(Q-i)\tau} - d_0\right)^2,$$

3.3 GAUSSIAN MIXTURE MODEL CLASSIFICATION

Figure



Architecture for MRPS-GMM framework.

As a data sequence can be measured as a mixture of three classes: normal, pattern, and event state of variables, which represent frequent states. The data sequences, are separated by these three states. Local temporal structures and arithmetic property of illumination variables can be viewed as two sets of facial appearance in classifying patterns in complex dynamic data sequence. The Gaussian mixture log-likelihood score predictable based on these score they are confidential.

Gaussian mixture model defined as,

$$P(\mathbf{x}) = \sum_{i=1}^{M} P(\omega_i)\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

The observed data is splitter in temporal data. The temporal data is set of data's with in the time series. In this diagram testing set consisting of position of inputs, implementation conditions and predictable outputs. The training set is applied to the supervised clustering algorithm based on pattern, event, normal. Supervised clustering is one kind of algorithm which is used to grouping the similarity classes or data's.

3.4 TESTING STAGE

In the testing stage, apply the predictive pattern classifier obtained in the training stage to predict the events in the target sequence. At each time t, based on the classification decision in, a forecast will be made whether or not an event will occur. In testing stage the training set will compared with test set. Based on that the testing stage will be done. In this stage it is used to shows the whether it is correct data sequence or not.

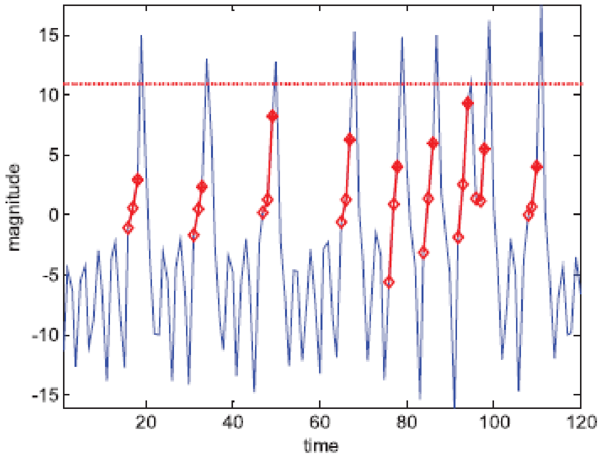
3.5 PREDICTION STAGE

This module makes use of the constructed event identification function and extracts the events from the given temporal patterns and forms a dynamic data model from which it can predict the occurrence of a particular event. The event is selected by the client itself which will be unknown to the user before the construction of event identification function; this function gives a list of event from which user can select a particular event based on their will. The user itself is used to select by him ownself.

4. EXPERIMENTAL RESULTS

We evaluate the performance of proposed MRPS method by comparing to four baseline algorithms. The chaotic time series data sets as illustrated in Fig. 3 are evaluated in this section. The accuracy measure and true positive rate measure are defined as follows: true positive rate TP/(TP+FN), accuracy (TP+FN)/(TP+FN+TN+FP), where TP, TN, FP, and FN correspond to the number of true positive, true negative, false positive, and false negative predictions, correspondingly.

Figure



Comparison of Prediction Results

The resulting temporal patterns associated with defined event function are plotted in Fig. 3. The patterns are different in structure and starting values compared with rather identical patterns found in [7]. This result indicates that our new approach is not only capable of identifying similar patterns, but also patterns that are rather different in structure and initial starting values.

Table

Method	Lorenz Map		Rossler Map	
	True Positive Rate	Accuracy	True Positive Rate	Accuracy
MRPS-GMM	98.94%	99.11%	94.44%	98.69%
TDNN-Rprop	93.62%	98.44%	83.33%	98.06%
TDNN-LM	96.23%	98.63%	84.21%	98.23%
BDT-Ada	91.49%	98.13%	73.47%	98.27%
BDT-Logit	92.55%	98.27%	71.59%	98.35%

5. CONCLUSION

The MRPS-GMM method for identify temporal patterns projecting of events in a multivariate data system. The proposed technique make obvious a number of reward over the existing methods. We introduced a latest MRPS. This embedding creates a new feature space combining all the creature embedding of each unpredictable progression. This algorithm in adding together provides a discriminative module that uses the Gaussian mixture model to achieve temporal patterns based on the following likelihood. We also established the effectiveness of our method by applying it to three experiments and showed considerable improvements in predictive pattern detection compared with baseline methods. A semi-supervised classification come within reach of to the pattern identification problem shows advantages in excess of existing methods and offers better statistical understanding between patterns and events.

REFERENCES

[1] Daniel B. Neill·Gregory F. Cooper proposed," A multivariate Bayesian scan statistic for early event detection and characterization," Physical Rev. E, vol. 64, article 036209, 2001.

[2] Jiang,Qin proposed,"Method for characterization, detection and prediction for target events" Physical Rev. E, vol. 64, article 036209, 2001.

[3] Koga and R. Aerosp proposed," Single-event effect characterization " IEEE Trans. Neural Networks, vol. 8, no. 4, pp. 863-873, July 1997.

[4] Povininelli, R.J proposed," A new temporal pattern identification method for Event characterization and prediction of complex time series events " World Bank Economic Rev., vol. 14, no. 2, pp. 221-250, 2000.

[5] Richard J proposed," Identifying Temporal Patterns for Characterization and Prediction of Financial Time Series Events" Physical Rev. vol. 45, no. 6 pp. 3403- 3411, 1992.

[6] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, "Feature-Based Classification of Time-Series Data," Int'l J. Computer Research, pp. 49-61, 2001.

[7] J. Yang and J. Leskovec, "Patterns of Temporal Variation in Online Media," Proc. Fourth ACM WSDM Int'l Conf. Web Search and Data Mining, pp. 177-186, 2011.