1-1-2012

A New Ensemble Learning Method for Temporal Pattern Identification

Wenjing Zhang
Marquette University

Xin Feng
Marquette University

Published version. *Procedia Computer Science,* Vol. 12 (2012): 102-109. DOI. © Elsevier. Used with permission.

Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 12 (2012) 102 – 109
Procedia Computer Science

Complex Adaptive Systems, Publication 2
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2012- Washington D.C.

A New Ensemble Learning Method for Temporal Pattern Identification

Wenjing Zhang[a], Xin Feng[a]*

[a]*Department of Electrical and Computer Engineering*
Marquette University, Milwaukee, Wisconsin 53201-1881, USA

Abstract

In this paper we present a method for identification of temporal patterns predictive of significant events in a dynamic data system. A new hybrid model using Reconstructed Phase Space (MRPS) and Hidden Markov Model (HMM) is applied to identify temporal patterns. This method constructs phase space embedding by using individual embedding of each variable sequences. We also employ Hidden Markov Models (HMM) to the multivariate sequence data to categorize multi-dimensional data into three states, e.g. normal, patterns and events. A support vector machine optimization method is used to search an optimal classifier to identify temporal patterns that are predictive of future events. We performed two experimental applications using chaotic time series and natural gas usage series related to the natural gas usage forecasting problem. Experiments show that the new method significantly outperforms the original RPS framework and neural network method.

Keywords: Temporal Pattern; Reconstructed Phase Space; Hidden Markov Model; Optimization; Dynamic Data System

1. Introduction

Consider the multivariate data sequences $\boldsymbol{X}(t)$ representing the measurements of a dynamic data system:

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

where t is the time index, N is the total number of observation, $\{x_{e,t}, t = 1, 2, \ldots N\}$ is called event sequence containing events of interest, $\{x_{it}, t = 1, 2, \ldots N\}$, $i = 1, 2, \ldots m$, is called multivariable sequence. In many applications, the underlying multivariate dynamic system observed by (1) is often complex and chaotic. Events in data systems are often closely related to time ordered structures, called temporal patterns. Discovering temporal patterns that are characteristic and predictive of to the events are important in many applications including in financial applications to determine the timing of positions of securities [2], forecasting economic growth and outlook, medical anomaly detection, interpretation of underlying system dynamics [9], [20]. In [22], evolutional patterns of GDP are used to address the characteristics of the economic growth.

In the area of data mining and pattern analysis, we are not only interested in detecting the events in the event data sequence $x_e(t)$, but also in exploring the causal relationship with the multivariate variable sequence. For instance, it can represent causing variables to the events in the sequence $x_{e,t}$.

Many existing methods are based on univariate data sequence. Among them is the Time Series Data Mining (TSDM) method [1] [3], which was used to identify patterns by embedding the univariate sequence data into the Reconstructed Phase Space (RPS) by estimating the optimized time delay τ and embedding dimension Q [11] [21]. Since the RPS approach is capable of representing temporal patterns of nonlinear dynamic sequence data that is typically chaotic and irregular [12] [13], it has been widely applied in a variety of research fields, such as ECG signal pattern identification of the atrial electrophysiology [8], wilding droplet identification [3] and security index forecasting [1]. Since the dynamic data sequences discussed here are highly contaminated by noise, the univariate based RPS embedding may result in poor performance due to lack of explanatory variables and significant noise in the event data sequence. Another drawback of the existing RPS framework is that similar temporal patterns may not appear in the same region of the feature space in the RPS due to some local trending and shifting effects, therefore resulting in missing temporal

1877-0509 © 2012 Published by Elsevier B.V. Selection and/or peer-review under responsibility of Missouri University of Science and Technology.
 doi:10.1016/j.procs.2012.09.038

Existing multivariate approaches include Discrete Fourier Transform (DFT) [4] and Discrete Wavelet Transform (DWT) [5][10], autoregressive modeling [19], nonlinear classification using neural networks, decision trees as well as clustering algorithms [10]. Since frequency domain approaches are based on spectral patterns of the feature space, data sequences with different nonlinear dynamic patterns but similar power spectrum may not be distinguished. Methods using Artificial Neural Network (ANN) suffer significantly from low generalizability and explainability. In [7], multiple patterns are used to detect effective patterns by predictability measure validation; however this approach is still based on one single variable.

In this paper, we propose a new method using multivariate phase space embedding to solve the problem of multivariate predictive pattern detection. The new method introduces an augmented Multivariate Reconstructed Phase Space (MRPS) which incorporates multivariate sequence. We embed the multivariate sequence into the MRPS by estimating multiple time delays and embedding dimensions for each and every single sequence. In addition, we address the problem of local trends and the shift effect. Experimental results show that the new method significantly outperforms the existing univariate RPS approach as well as other methods such as the neural network method.

We will first review the existing RPS approach in Section 2. Then, we will discuss in details the MRPS approach in Sections 3, 4, followed by experiment in section 5 and conclusion in 6.

2. Related Work

2.1. Reconstructed Phase Space and RPS Embedding

A reconstructed phase space is a Q-dimensional metric space into which a data sequence is unfolded [11]. In [12], Takens showed that if Q is large enough, the phase space is homeomorphic to the state space that generated the data sequence. Takens' Theorem guarantees that the reconstructed dynamics are topologically identical to the true dynamics of the system. This provides the theoretical justification for reconstructing state spaces using time-delay embedding.

Given a univariate data sequence $\{x_t, t = 1,2,\ldots N\}$, a state space topologically equivalent to the original state space can be reconstructed by a process called time-delay embedding [11]. The vectors, e.g. data points, $\mathbf{x}_t, t = 1,2,\ldots N$ in R^Q are represented by:

$$\begin{bmatrix} \mathbf{x}_{1+(Q-1)\tau} \\ \mathbf{x}_{2+(Q-1)\tau} \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} x_{1+(Q-1)\tau} & \cdots & x_{1+\tau} & x_1 \\ x_{2+(Q-1)\tau} & \cdots & x_{2+\tau} & x_{1+\tau} \\ \vdots & & \ddots & \\ x_N & \cdots & x_{N-(Q-2)\tau} & x_{N-(Q-1)\tau} \end{bmatrix}_{N-(Q-1)\tau \times 1} \tag{2}$$

where τ is the time-delay and Q is the dimension of embedding vector.

2.2. Temporal Pattern Identification in the RPS

Povinelli and Feng [3], applied an optimized genetic algorithm to search the phase space for optimal heterogeneous clusters of temporal patterns that are predictive of specific events.

Since genetic search algorithm is probabilistic and computational complex, the genetic search optimization approach may experience a high level of computing load and inconsistent results. The assumption of patterns are in well-defined clusters may become insufficient if the patterns in phase space are sparse and forming irregular non-spherical shapes. In [1], A Gaussian shaped fuzzy membership function was applied to temporal pattern clusters and objective functions. The center and radius of the pattern clusters are then calculated by the fuzzy membership function instead of forming a rigid region with fixed center and radius.

$$\max_{\mathbf{v},d}\{f(\mathbf{v},d)\} = \max_{\mathbf{v},d}\{\textstyle\sum_N \mu(\mathbf{x}_i)g(\mathbf{x}_i)\} \tag{3}$$

where $g(\mathbf{x}_i)$ is the event function defined specifically to the problem, and $\mu(\mathbf{x}_i)$ is the membership function for each embedding $\mathbf{x}_i$

$$\mu(\mathbf{x}_i) = \exp(\|v - x_i\|^2 / 2\sigma^2) \tag{4}$$

Although several attempts have been made to explore the clustering interpretation of temporal patterns and events, there is relatively less literature applying a discriminative approach in the framework of reconstructed phase space. In [7], the predictability measure was introduced and logistic regression is applied to find multiple temporal pattern clusters and evaluate the validity of these patterns. The estimate of probability that event will occur is as follows:

$$p(z_1, z_2, \cdots z_n) = \frac{e^{g(z_1, z_2, \cdots, z_n)}}{1 + e^{g(z_1, z_2, \cdots z_n)}} \tag{5}$$

where the temporal pattern indicator z_{it} is defined by:

$$z_{it} = \begin{cases} 1 & if\ \mathbf{x}_i\ is\ present\ at\ t \\ 0 & if\ \mathbf{x}_i\ is\ absent\ at\ t \end{cases}$$

3. Event function and Multivariate Reconstructed Phase Space (MRPS)

3.1. Definition of Event function in Multivariate Sequences

Consider the multivariate sequence in (1). the *j*th variable data sequence denoted as $X_j = \{x_{ji}, i = 1, 2, \ldots N\}$. An $(m+1)$-dimensional vector $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{mi}, x_{e,i})^T$ is observed at each time *i* as a sample instance in the feature space. The pattern vectors that are predictive of future events are characterized by a defined event function. A general form of the event function can be defined as follows:

$$g(\mathbf{x}_i) = \mathrm{sgn}\{\max\{x_{e,i+1}, \ldots, x_{e,i+k}\} - c\} \tag{6}$$

where *k* is the time-step ahead , c is the threshold that defines an event of interests. The sequence $\{x_{e,i}, i = 1, \ldots, N\}$ denotes the target sequence with events that is of interest. *k* is a predefined constant which is the maximum time window to make a forecast. Therefore, we can separate multivariate vector into event, pattern and normal state by the definition in (6). At training stage, the feature vector $\mathbf{x}_i$ is labeled with $g(\mathbf{x}_i) = \{+1, -1\}$ by a *k*-step ahead forecasting event function each time *i.*

3.2. Multivariate Phase Space Embedding

In some applications, it is imperative to find variables in the multivariate dynamic system that causes outcomes of events. Consider a general multivariate data sequences in (1) with *m* causing variables, one event variable. Denote $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{ji}, \ldots, x_{mi}, x_{e,i})^T$ as the observation at time *i,* where $i = 1, 2, \ldots N$, $j = 1, 2, \ldots m$. By estimating the time delay τ_j and dimension Q_j for each variable x_j, $j = 1, 2, \ldots m$, the multivariate phase space embedding can be constructed as $\boldsymbol{X}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i}, \ldots, \mathbf{x}_{ji}, \ldots, \mathbf{x}_{mi}, \mathbf{x}_{e,i})$ at each time *i*, where $\mathbf{x}_{ji}$ represent the phase space embedding for *j*th variable x_j given in (2) with the time delay τ_j and dimension Q_j at time *i*. The dimension Q of the multivariate embedding is the sum of each embedding dimension Q_j, $Q = \sum_j Q_j$.

4. RPS-HMM Classification in Phase Space

In this section, we present our approach to the temporal pattern recognition in the MRPS and classify patterns that statistically significantly related to events. This is done in three stages. We first extend the method of single RPS by constructing a multivariate embedding combining each variable's embedding. The time delay and the dimension of each variable are estimated by mutual information and false nearest neighbor method. The multivariate embedding is then transformed by a differencing operation to reduce the effect of local trend drift. The second stage is to apply Hidden Markov models (HMM) to the dataset estimating the statistical distribution of three separated states: normal, pattern and events. In the final stage, a new objective function is constructed and optimization method is applied to search the optimal classifier that identifies the temporal patterns in the multivariate feature space.

4.1. Multivariate Embedding and Parameter Estimation

To estimate the embedding time delay τ_j of the *j*th variable x_j, $j = 1, 2, \ldots m$, we applied the first minimum of mutual information function which provides the quantitative characteristics of spatial patterns in phase space [11], [24]. The minimum of the mutual information function was found to be effective in estimating the time delay. Given a data sequence and time delay τ_j, the mutual information is computed by:

$$M(\mathbf{x}_{ij}, \mathbf{x}_{i-\tau_j,j}) = \sum_{n,m} p_{nm}(\tau_j) \ln \frac{p_{nm}(\tau_j)}{p_n p_m} \tag{7}$$

The dimension Q_j for each variable is determined using a false nearest-neighbors technique [13]. For each data point $\mathbf{x}_{nj}^{Q_j}$, the changing rate of distance is defined by:

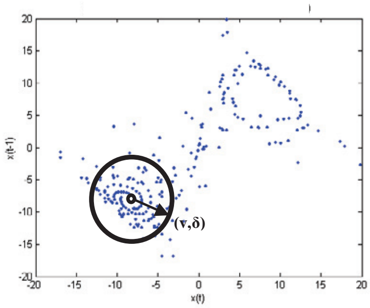

Fig. 1 Example of a 2-D temporal pattern cluster, with center v and radius σ

$$r_i = \sqrt{\frac{\left\|\mathbf{x}_{nj}^{Q_j+1} - \mathbf{x}_{mj}^{Q_j+1}\right\|^2 - \left\|\mathbf{x}_{nj}^{Q_j} - \mathbf{x}_{mj}^{Q_j}\right\|^2}{\left\|\mathbf{x}_{nj}^{Q_j} - \mathbf{x}_{mj}^{Q_j}\right\|^2}} \quad (8)$$

where $\left\|\mathbf{x}_{nj}^{Q_j} - \mathbf{x}_{mj}^{Q_j}\right\|$ is the Euclidean distance between its nearest neighbor

$$\mathbf{x}_{mj}^{Q_j} = \underset{\mathbf{x}_{mj}^{Q},\, n \neq m}{argmin} \left\|\mathbf{x}_{nj}^{Q_j} - \mathbf{x}_{mj}^{Q_j}\right\|$$

$\mathbf{x}_{nj}^{Q_j}$ is marked as having a false nearest neighbor, if r_i exceeds a given threshold ρ. The criterion for adequate embedding dimension Q_j is that the number of data points for which $r_i > \rho$ is zero in R^{Q_j}.

For the *j*th variable x_j, the time delay is τ_j and dimension is Q_j. The dimension Q of the MRPS embedding is the sum of each embedding dimension Q_j, $Q = \sum_j Q_j$.

4.2. Reconstructed Phase Space with Differencing

In some applications, the temporal structures of pattern may have some drifting effects that make the pattern have different initial conditions. In the RPS embedding, such drifting or trend effect causes the temporal pattern with same dynamics to be separately located in different regions in phase space. To address this problem, we consider a phase space embedding on differenced data series in that such transform usually result in a detrended representation of the data sequence. By the theorem of Filtered delay embedding prevalence in Sauer et al. [14], given a linear constant transformation, the resulting filtered delay mapping also gives an embedding of the underlying dynamic system. The transformed embedding also preserved the same local dynamics as the regular embedding of the original dataset. Moreover, the differenced embedding gives a better representation of the data as the similarity of local dynamics under such embedding does not depend on the initial starting point of the pattern. The Euclidean distance in the phase space now becomes a true mapping between the similarity of the temporal pattern and underlying dynamics.

Applying the difference transformation defined as $\nabla\mathbf{x}$ on the original embedding $\mathbf{x}$, we are able to obtain a new vector $f(\mathbf{x}) = \nabla\mathbf{x}$. The Euclidean distance in the new space f is equal to the similarity measure in (3). It can be noted that for a given Q, the difference operation ∇ is linear [14] which preserves the same dynamics of the new embedding.

4.3. Hidden Markov Model Scoring

As mentioned in section 3, the data sequences can be regarded as a mixture of three distinctive states: normal, pattern and event. In addition to the local temporal structures of predictive patterns, we apply HMM to estimate the underlying state transition within multivariate sequences. Although HMM is less effective in capturing the correlations between the observed variables separated by multiple time steps, this is well compensated by including a RPS model. Previous work under the RPS framework [1], [3], [7] employed the clustering method to identify temporal patterns. However, little discussion was made to apply discriminative approach characterizing patterns based on the statistical correlations of multivariate data sequence. The approach proposed here applies a multivariate Hidden Markov models to exploit the discriminative information that will be incorporated in the construction of objective function in the step of optimization.

HMM is widely used to explain the observation sequence $X = \mathbf{x}_1\mathbf{x}_2 \ldots \mathbf{x}_T$ in terms of underlying state sequence $Q = q_1q_2 \ldots q_T$. Under the assumption of three states, the HMM can be characterized by the following parameters.

a) $S = \{S_1, S_2, S_3\}$, set of states, where S_1, S_2, S_3 denotes normal, pattern and event state respectively. State at each time *t* is denoted by q_t.

b) $A = \{a_{ij}\}$, transition probability, a_{ij} denotes the transition probability from state S_i to S_j

$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i], \quad 1 \le i, j \le 3 \quad (9)$$

c) $B = \{b_j(\mathbf{x})\}$, observation probability distribution. $b_j(\mathbf{x})$ gives the probability of observing $\mathbf{x}$ in the state S_j at time *t*

$$b_j(\mathbf{x}_t) = P[\mathbf{x}_t \mid q_t = S_j], \quad 1 \le j \le 3. \quad (10)$$

d) $\pi = \{\pi_i\}$, the initial state distribution gives probability that current state is S_i at initial time

$$\pi_i = P[q_1 = S_i], \quad 1 \le i \le 3. \quad (11)$$

This parameter set can be written in a compact form

$$\theta = (A, B, \pi)$$

To estimate parameter set, maximum likelihood method is applied by using Baum-Welch (EM) [16] algorithm to maximize $P[X \mid \boldsymbol{\theta}]$. Instead of using maximum a posterior (MAP), we consider the log-odds $\varphi(\mathbf{x}_t)$ of pattern state and normal state of multivariate sequence at each time *t*.

$$\varphi(\mathbf{x}_t) = \log \frac{P(S_2 | X, \boldsymbol{\theta})}{P(S_1 | X, \boldsymbol{\theta})} \quad (12)$$

4.4. Optimization in Multivariate Feature Space

The traditional RPS method typically categorizes the embedded data points by clustering method with a defined event function to distinguish event and nonevent patterns. This approach would perform well for datasets with relatively low level of noise with respect to signals, but may not work well when the dynamics of the system is significantly affected by noise or trending effects. Under this circumstance, a well-defined cluster may not exist and similar patterns are separately located in several cluster regions. Considering the fact that RPS is well suited for capturing the long range time delayed structure, we categorize patterns based on temporal similarity in differenced data sequence and also the discriminant scores estimated by Hidden Markov Models (HMM) in the feature space.

For a given data sequence $X = \{x_i, i = 1, 2, \ldots N\}$, we obtain two sets of features $f(\mathbf{x}_i)$ and $\mathbf{x}_i$. The event function $g(\mathbf{x}_i) \in \{+1, -1\}$ defined in (10) is the indicator of each embedding. The problem of searching optimal classifier to detect temporal patterns can be formulated as the following optimization problem:

$$\min_{\boldsymbol{\beta}} \{L(g(\mathbf{x}), f(\mathbf{x}))\} = \min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \exp(-g(\mathbf{x}_i) f(\mathbf{x}_i)) \quad (13)$$

where $L(g, f(x))$ is the objective function defined as the weighted exponential sum of event function and classifier for each $\mathbf{x}_i$. The classifier is defined as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \exp\left(\frac{\| f(\mathbf{x}) - f(\mathbf{x}_i) \|^2}{\sigma^2} \right) + \beta_1 \varphi(\mathbf{x}) + \beta_0 \quad (14)$$

where $\varphi(\mathbf{x}) = \log(p(\omega_p | \mathbf{x}) / p(\omega_n | \mathbf{x}))$, $\boldsymbol{\beta} = (\alpha_1 \ldots \alpha_n, \beta_1, \beta_0)$

The first term in the classifier (18) denotes the similarity of the differenced data series represented by the kernel estimation in the phase space. The second term denotes the Hidden Markov log-likelihood score estimated from (16). This formulation considers both local temporal dynamics of the data sequence and the statistical interpretation given by Hidden Markov models. The parameters α_i and β_1 determines the constraints of the temporal dynamics and statistical correlations between features. In general, large weights in α_i indicates the events are more likely relevant to the local dynamics of the system. Whereas a large weight in β_1 would suggest the cause of events is more correlated with the feature variables than the temporal dynamics. The parameters α_i, β_1 and β_0 can be determined by applying Quasi-Newton method to minimize the objective function $L(g(\mathbf{x}), f(\mathbf{x}))$ defined in (13).

5. Experimental Results

We applied the new MRPS framework to two applications. The first is the simulated chaotic dynamic data sequence generated by Lorenz equation, with duration of 500-second and the sampling frequency of 125Hz. The second application is the detection of causing variables of the Sludge Volume Index (SVI) series which is an effective quality operation indicator in a typical water treatment plant. We compare the results by our method to the previous RPS approach as well as to the neural network method.

5.1. Analysis of Lorenz Chaotic Time Series

A Lorenz time series is generated by [11] with the initial values set to $x_0 = 0, y_0 = -0.01, z_0 = 0.01$, and parameters: $\sigma = 9, r = 25$, and $b = 3.3$. The time delay was estimated as $\tau = 0.2s$ and the embedding dimension *Q*=3. We use sequences of two variables $\mathbf{x}_t = (x_t, y_t)$ to forecast the event in the x_t sequence.

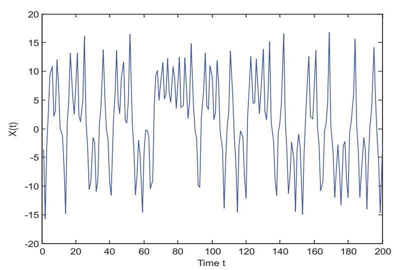

Fig.3. Sample Lorenz series x(t)

The event function is defined as

$$g(\mathbf{x}_t) = x_{t+1} - 11.0 > 0 \quad (15)$$

which means that in the following time instance, the magnitude of data point is greater than the threshold 11.0. The first 1500 data points are used for the training process and the other 700 data points are used for testing. We estimated the distribution of the normal, pattern and event data points using HMM. We embedded the multivariate sequence into MRPS and applied optimization method to classify patterns based on the event definition in (15).

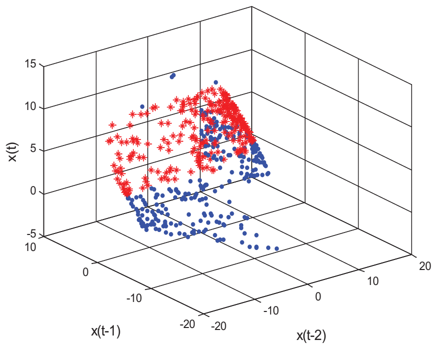

Fig. 4. Temporal Patterns of Lorenz Series

TABLE 1
TEST RESULTS OF HMM-RPS METHOD ON LORENZ SERIES

	Predicted as events	Predicted as nonevents
Actual events	TP=51	FN=10
Actual nonevents	FP =5	TN=435

TABLE 2
COMPARISON OF PREDICTION RESULTS

	Training Set		Test Set	
Method	True Positive Rate	Accuracy	True Positive Rate	Accuracy
RPS-HMM	87.45%	99.36%	91.07%	97.01%
TSDM	62.45%	99.82%	48.65%	96.52%
ANN	67.00%	89.35%	45.30%	84.59%

The comparative results of the new method and previous framework methods are presented in Table 2. It shows clearly that in the training phase both methods achieved 99% prediction accuracy, but our new method had higher true positive accuracy (TPR). In the testing phase, the results given by the new method is consistent with the training results, whereas the TSDM method had a poor recognition rate of events with a low positive accuracy.

5.2. Sludge Volume Index (SVI) Analysis

Sludge bulking is the primary causes of water treatment plant failure as the bulking conditions result in exceeding discharge limitations. There are many potential causing factors to sludge bulking, such as level of Dissolved Oxygen (DO). Efforts have been made to study the cause of this problem from the biological point of view, but failed to formulate a deterministic cause-effect relationship due to complexity. Also the modeling approaches have been applied including stochastic models and artificial neural systems [9].

The sludge volume index (SVI) used here is the primary indicator representing the bulking conditions. Fig. 5 shows the SVI daily sequence from 2002 to 2008. The SVI value above 150 indicates the occurrence sludge bulking events.

In this experiment, we consider two factors: the SVI and the DO index which is generally considered as one of the probable factors causing the bulking problem. We denote SVI and DO as $\mathbf{x}_t = (S_t, D_t)$. The events are defined as:

$$g(\mathbf{x}_t) = \max\{S_{t+1}, \ldots, S_{t+3}\} - 150.0 > 0 \tag{16}$$

The identified temporal patterns are presented in Fig.4

The temporal patterns are marked by squares lines with embedding dimension $Q = 4$. The last point of the pattern is the predicting point which indicates the sludge bulking is likely to occur with high probability. The results of proposed method and previous framework are presented in Table 3.

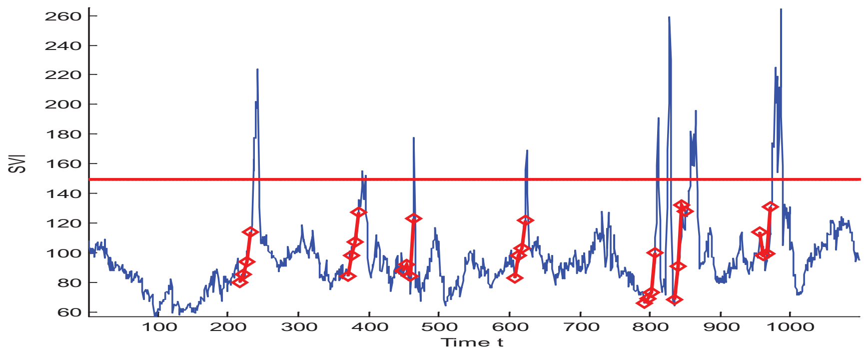

Fig.5. Temporal Patterns and Prediction of SVI

TABLE 3
CONFUSION MATRIX OF SVI USING NEW METHOD

	Predicted as events	Predicted as nonevents
Actual events	TP=29	FN=9
Actual nonevents	FP =6	TN=1605

TABLE 4
TESTING SET RESULTS COMPARISON OF SVI DATASET

Method	True Positive Rate	True Negative Rate	Accuracy
RPS-HMM	82.86%	97.33%	99.09%
TSDM	54.52%	95.45%	93.27%
ANN	51.37%	91.85%	90.56%

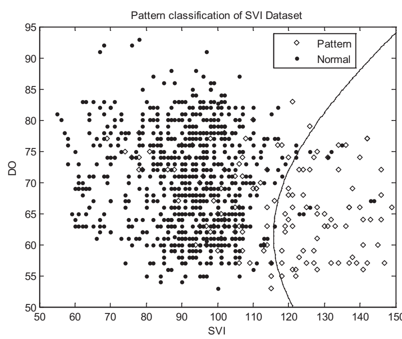

Fig.6. Temporal patterns of SVI index

In comparing the two methods, we can see that for this SVI prediction task, the RPS-HMM approach outperforms the TSDM approach by 28.32% difference in forecasting events in testing dataset. Fig. 6 shows the plot of pattern and normal points in the SVI-DO feature space. The results can be better understood by examining the temporal pattern plotted in Fig.5 .The patterns that related to the events are not simply consistent of their structures and shapes. Instead, the temporal patterns are time-evolving and not obvious to capture by traditional approaches. The results demonstrated that the new approach could be applied in a monitoring system for the sludge bulking conditions of water treatment plants to provide early alerts for the potential bulking problems.

The above results indicate that the proposed method outperforms the original RPS framework and the neural network approaches for identifying patterns in both high and low Signal-to-Noise ratios. From the results of SVI dataset in which the data sequence is noisy, we note that new approach achieves significantly better identification and prediction accuracy compared with TSDM and ANN methods.

6. Conclusion

The new MRPS framework presented here demonstrates several significant advantages over the existing methods. First, we introduced a new multivariate reconstructed phase space (MRPS). This embedding creates a new feature space combining all the individual embedding of each variable sequence. This algorithm also provides a discriminative method that utilizes the Hidden Markov model to score temporal patterns based on posterior likelihood. We also propose a new objective function and a new classifier to integrate the temporal patterns in MRPS and Bayesian discriminative scoring of the multivariate data sequences.

We also demonstrated the new MRPS framework by applying it to two experiments and showed significant improvements in event detection compared with the previous RPS method. Future work may include other mixture model, with more distributions applied for clustering.

References

[1] X. Feng and H. Huang, "A Fuzzy-Set-Based Reconstruction Phase Space Method for Identification of Temporal Patterns in Complex Time Series," *IEEE Trans. on Knowledge and Data Engineering*, vol.17, no. 5, pp. 601-613, 2005.

[2] Chiung-Hon Leon Lee, Alan Liu, Wen-Sung Chen, "Pattern discovery of fuzzy time series for financial prediction," *IEEE Trans. on Knowledge and Data Engineering*, vol. 18, no. 5, may 2006

[3] R.J. Povinelli and X. Feng, "A New Temporal Pattern Identification Method for Characterization and Prediction of Complex Time Series Events," *IEEE Trans. on Knowledge and Data Engineering*, vol.15, no. 2, pp.339-352, March/April 2003.

[4] C. Faloutsos, M. Ranganathan and Y. Manolopoulos, "Fast subsequence matching in time-series databases," *SIGMOD*, 1994, pp. 419-429

[5] K. Chan, A. Fu, "Efficient time series matching by wavelets", *In: ICDE* 1999, pp. 126–133

[6] E. Keogh, "A fast and robust method for pattern matching in time series databases", *Proc. 9th Int. Conf. on Tools with Artificial Intelligence, ICTAI, 1997*, pp. 578–584.

[7] X. Feng, O. Senyana, "Mining Multiple Temporal Patterns of complex dynamic data systems", *CIDM 2009*, pp. 411-417

[8] R. J. Povinelli, M. T. Johnson, A. C. Lindgren, and J. Ye, "Time series classification using Gaussian mixture models of reconstructed phase spaces", *IEEE Trans. on Knowledge and Data Engineering*, vol. 16, pp. 779 - 783, 2004

[9] Capodaglio, H. Jones, V. Novotny, X. Feng, "Sludge bulking analysis and forecasting: application of system identification and artificial neural computing technologies", *Water Research*, 1991; 25(10):1217–24.

[10] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, England: Cambridge University Press, 2000

[11] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 1997.

[12] F. Takens, "Detecting Strange Attractors in Turbulence", *Lecture Notes in Math*. Vol. 898, Springer, New York (1981).

[13] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev*. A 45, 1992, pp. 3403.

[14] T. Sauer, J.A. Yorke, and M. Casdagli, "Embedology," *J. Statistical Physics*, vol. 65, pp. 579-616, 1991.

[15] J. Iwanski and E. Bradley, "Recurrence Plot Analysis: To Embed or not to Embed?," *Chaos*, vol. 8, pp. 861-871, 1998.

[16] T.K. Moon, "The expectation-Maximization Algorithm," *IEEE Signal Processing Mag.*, vol. 13, pp. 47-59, 1996.

[17] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Society*, *Series B*, vol. 39,no. 1, 1977.

[18] J. Friedman, T. Hastie, R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *The Annals of Statistics*, vol. 28(2), pp. 337–407, 2000.

[19] G. Box and G. Jenkins, *Time series analysis: Forecasting and control*, San Francisco, CA: Holden-Day, 1976

[20] Y. Quilfen, A. Bentamy, P. Delecluse, K. Katsaros, and N. Grima, "Prediction of Sea Level Anomalies Using Ocean Circulation Model Forced by Scatterometer Wind and Validation Using TOPEX/Poseidon Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 4, July 2000

[21] D. Sciamarella and G. Mindlin, "Unveiling the topological structure of chaotic flows from data," *Phys. Rev. E*, vol. 64, pp.036 209:1-7,2001.

[22] L Pritchett, "Understanding Patterns of Economic Growth: Searching for Hills among Plateaus, Mountains, and Plains," *World Bank Economic Review*, vol. 14, issue 2, 2000

[23] W. Zhang, X. Feng, and N. Bansal, "Detecting temporal patterns using Reconstructed Phase Space and Support Vector Machine in the dynamic data system", *Proc. IEEE International Conference on Information and Automation, ICIA, 2011,* pp.209-214.

[24] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A33, 1986, pp. 1134-1146.