7-1-2012

Multivariate Temporal Patterns Detection Using Reconstructed Phase Space and Gaussian Mixture Model in Dynamic Data System

Xin Feng
Marquette University

Wenjing Zhang
Marquette University, wenjing.zhang@marquette.edu

Presented version. Presented at the *10th World Congress on Intelligent Control and Automation*, 2012.

Predictive Temporal Patterns Detection in Multivariate Dynamic Data System

Wenjing Zhang and Xin Feng

Department of Electrical Engineering and Computer Engineering

Marquette University

Milwaukee, Wisconsin 53201, U.S.A

July 2012

Presentation Outline

- Problem statement
- Multivariate temporal pattern detection
- Experimental results
- Conclusion and future work
- References

Presentation Outline

MARQUETTE UNIVERSITY
Be The Difference.

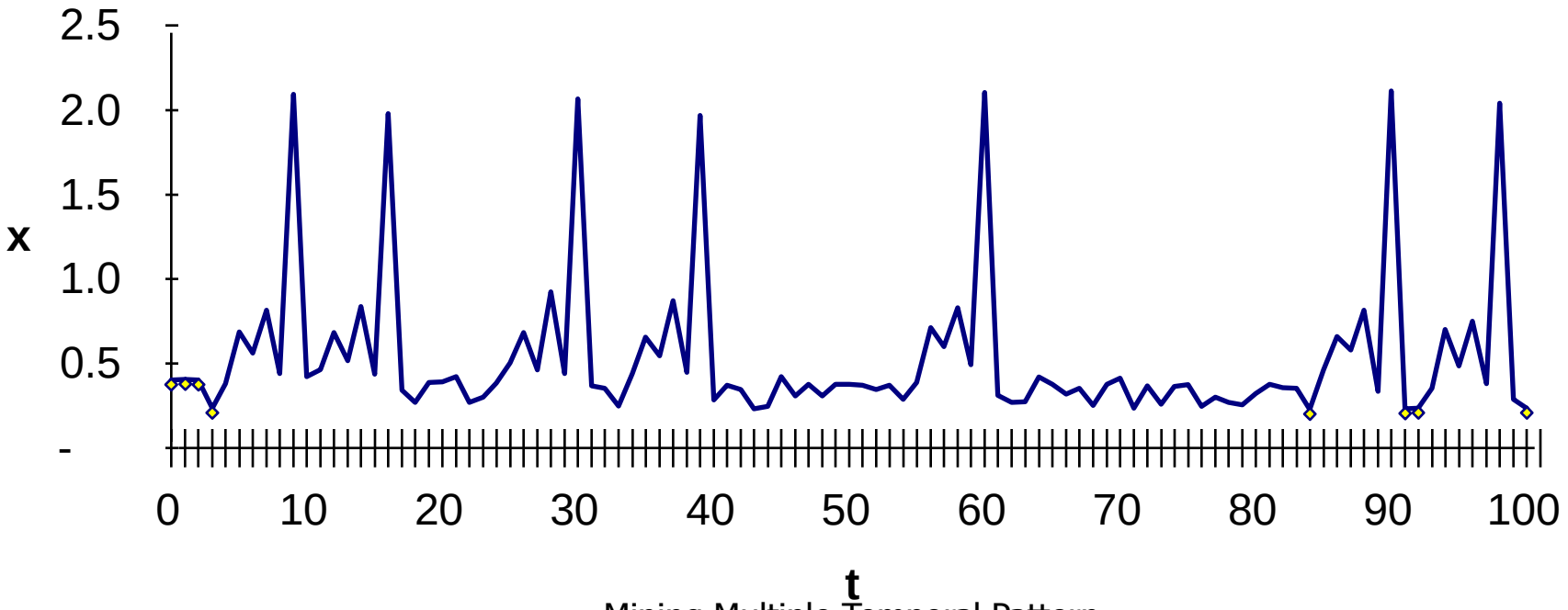
- Problem statement
  - Early work
  - A new proposed MRPS method
  - Experimental results
  - Conclusion and future work
  - References

Problem Statement

- A Dynamic Data System (DDS) is observed by time ordered data:

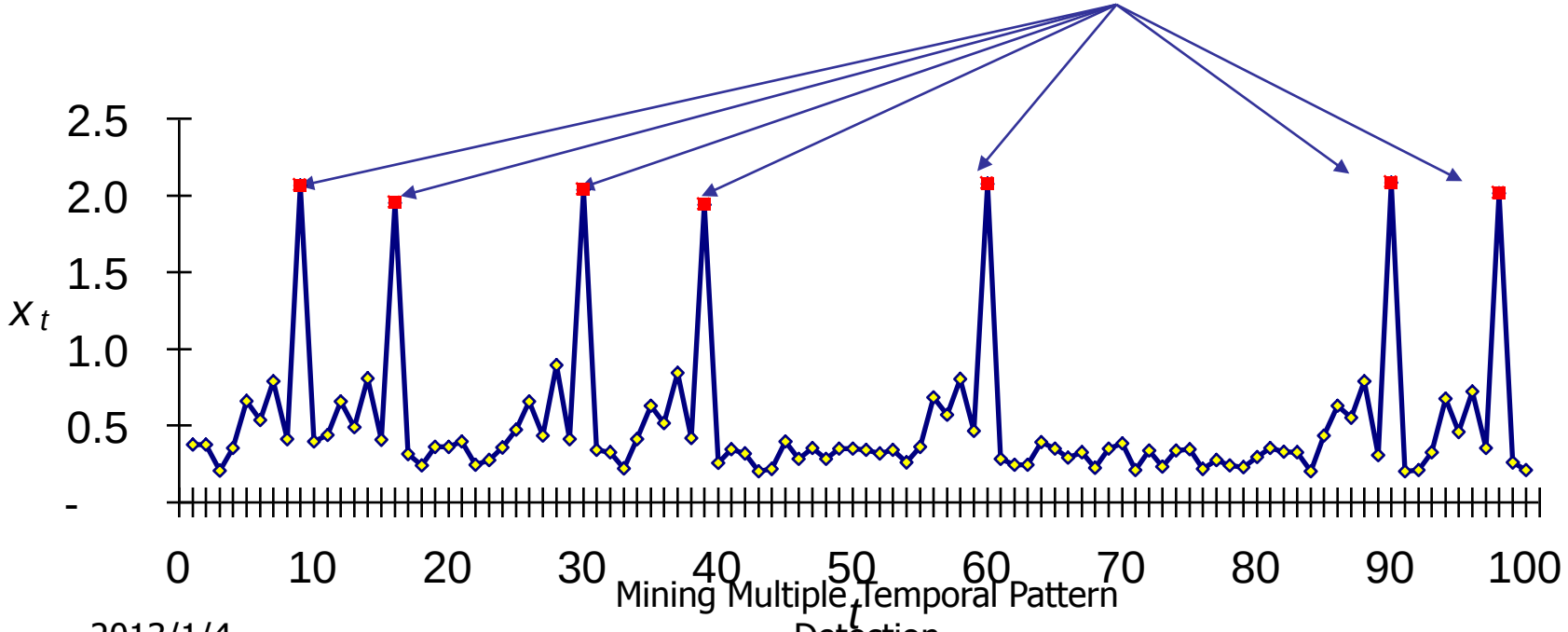
$$X(t,m) = \{x_t(m) \mid t = 1,2,3,\ldots,n;\ m=1..,K\}$$

- ***Multi-dimensional*** in most cases



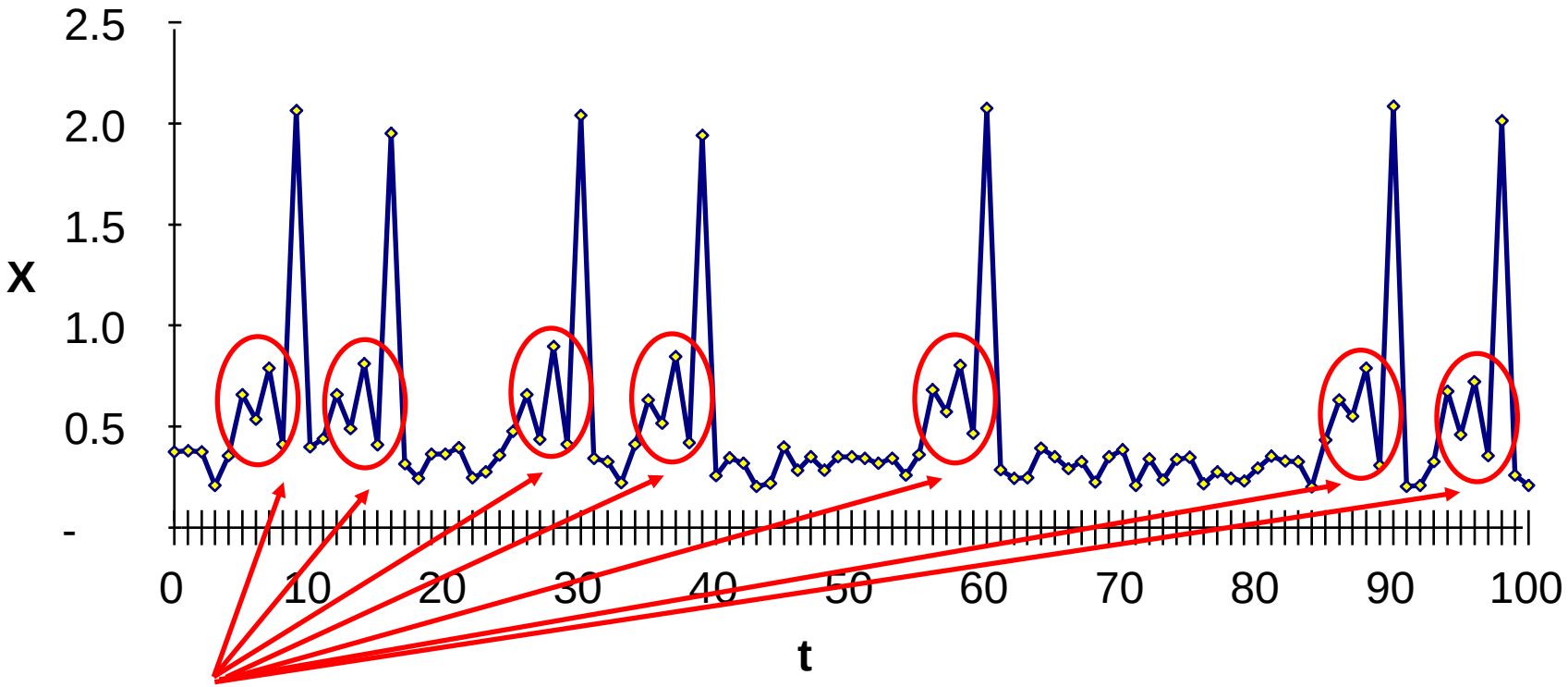
Problem Statement

- Non-stationary, non-periodic
- Chaotic deterministic time series whose attractors are non-stationary
- May contain “Events of Interest (events)”



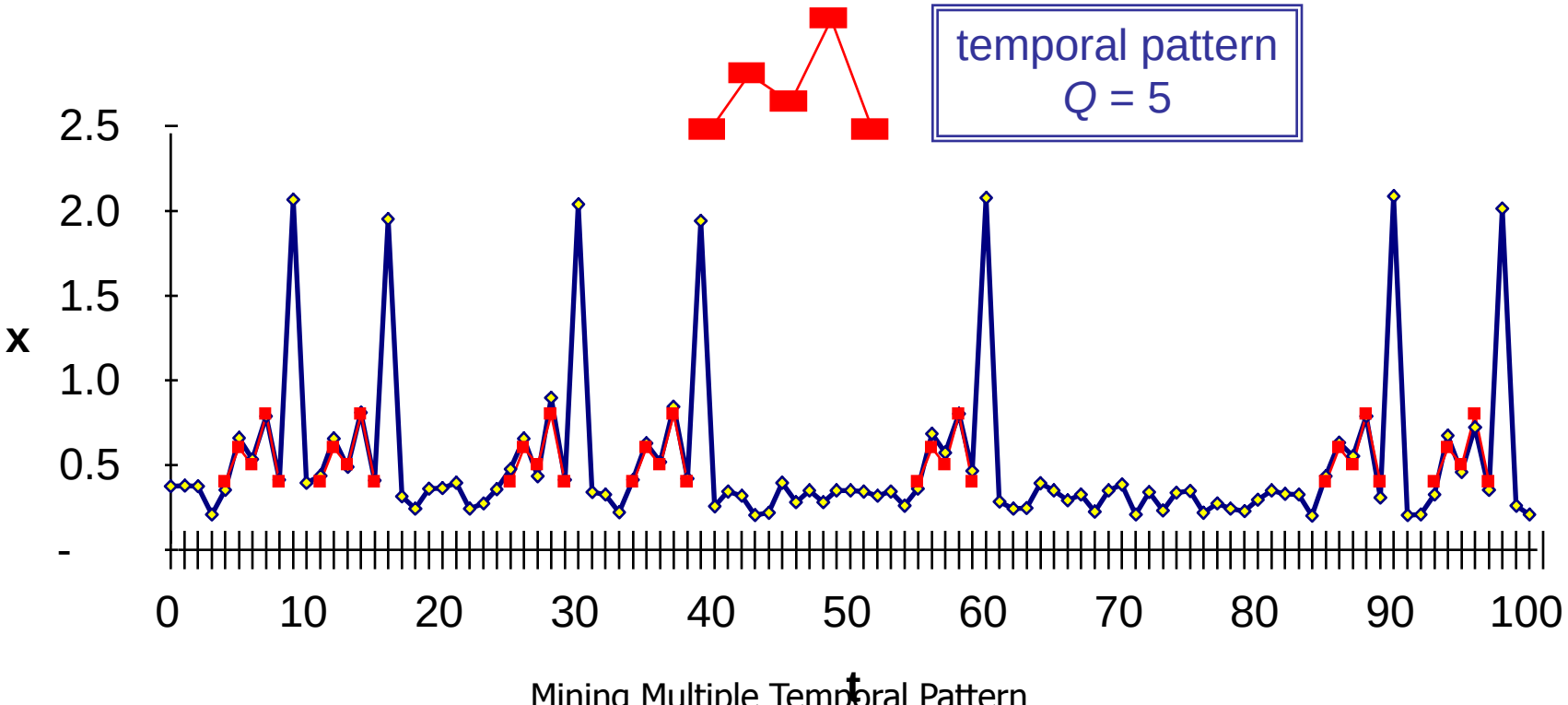
Problem Statement

- May also contain "temporal patterns"
- Can we use such temporal patterns to detect events?



Problem Statement: Temporal Patterns

- Find temporal patterns in a "Phase Space"
- $p \in P \subseteq \Re^Q$, a vector of length Q



Presentation Outline

- Problem statement
- Early work
- A new proposed MRPS method
- Experimental results
- Conclusion and future work
- References

Early Work

MARQUETTE UNIVERSITY
Be The Difference.

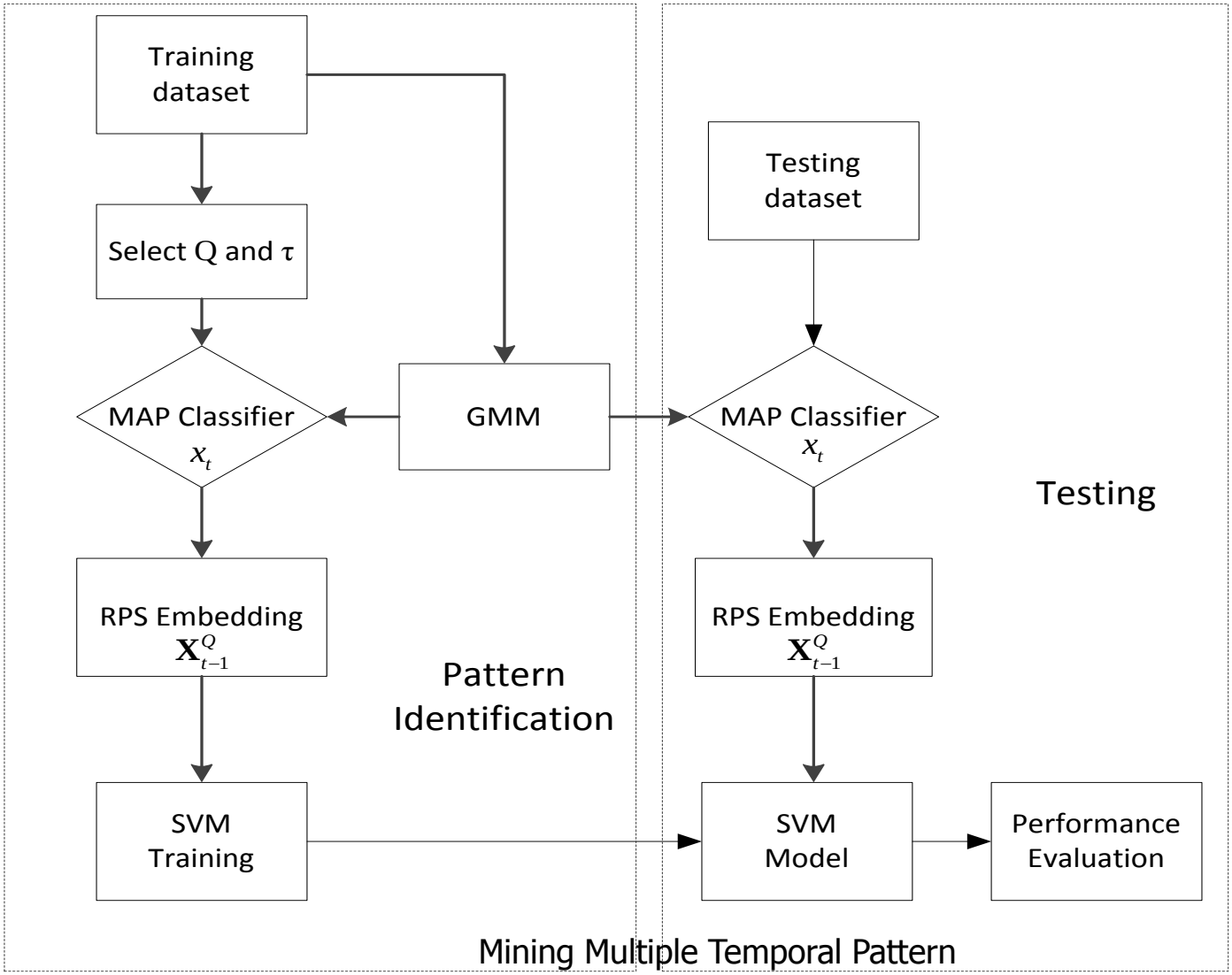
- TSDM (Povinelli, Feng and Huang 2002, 2005)
  - A systematic data mining approach
  - Phase space embedding in RPS
  - Search of a single temporal patterns in phase space (optimization)
  - Demonstrated effective for a variety of applications

Early Work

MARQUETTE UNIVERSITY
Be The Difference.

- Time Delay Embedding: transformation of time series into a Q-dimensional reconstructed phase space (RPS) $\mathbf{R}^Q$ with time delay τ.
- Takens Theorem: if Q is large enough, the phase space is homeomorphic to the state space that generated sequence data.
- Provides a sufficient condition for phase space to be embedding of the state space from which time series (1) was sampled is that Q is greater than twice the dimension of the original state space.

The GMM-SVM method



Early Work

MARQUETTE UNIVERSITY
Be The Difference.

- The GMM-SVM method
- Provides an unique method for detecting temporal patterns that are predictive of future events in analyzing chaotic and dynamic system data.
- Successfully been applied to a variety of applications.
- Only for the uni-variate system.
- Continued work to apply it to the multivariate system.

Presentation Outline

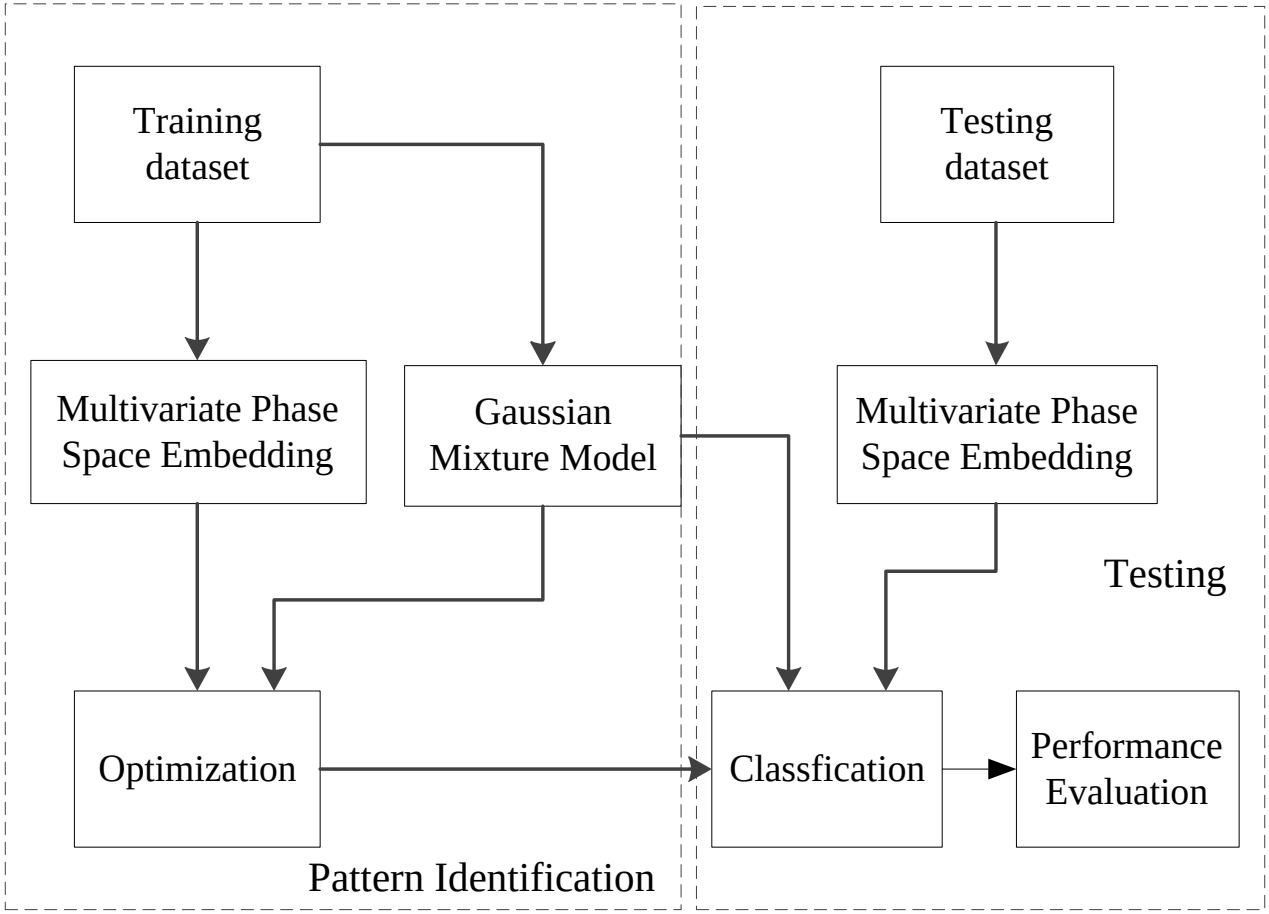
- Problem statement
- Early work
- A new proposed MRPS method
- Experimental results
- Conclusion and future work
- References

Proposed Multiple Reconstructed Phase Space (MRPS) Method

MARQUETTE UNIVERSITY
Be The Difference.

- Search for optimal temporal patterns that are characteristic and predictive of events of interest in Multivariate Dynamic Data System
- Estimate the embedding dimension Q and time-delay τ for each data sequence and define event function, g for specific problem
- Preliminary clustering by Gaussian Mixtures
- Embed the time sequence into a phase space
- Find the "optimal" decision boundary within Phase Space by objective function optimization

Overview of MRPS method



Multivariate Phase Space Embedding

MARQUETTE UNIVERSITY
Be The Difference.

- Estimate embedding dimension Q_j and time delay t_j for each data sequence $X_j, j = 1, 2, ¼ m$

- Construct Multivariate Phase Space by combining embeddings of each data sequence into a product space $\boldsymbol{X}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i}, ..., \mathbf{x}_{ji}, \mathrm{K}, \mathbf{x}_{mi}, \mathbf{x}_{e,i})$

- The dimension of the multivariate embedding is the sum of each embedding dimension

Multivariate Phase Space Embedding(Cont.)

MARQUETTE UNIVERSITY
Be The Difference.

Observation matrix can be represented by

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \\ X_e \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mN} \\ x_{e1} & x_{e2} & \cdots & x_{eN} \end{bmatrix}_{(m+1)\times N}$$

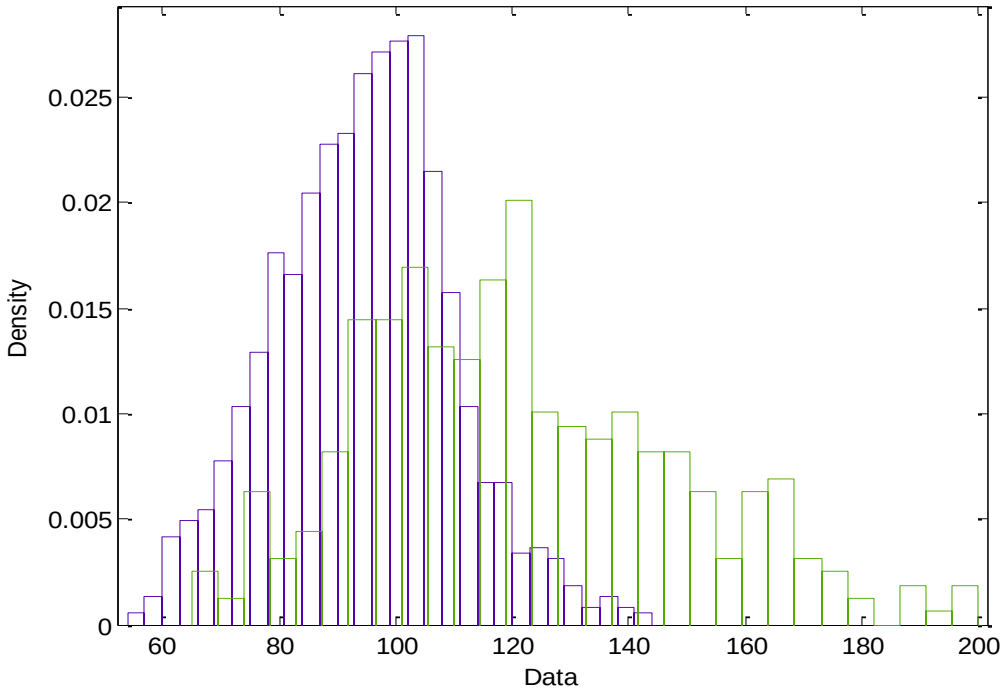
The multivariate phase space embedding can then be constructed for each time i as

$$\boldsymbol{X}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i}, \ldots, \mathbf{x}_{ji}, \cdots, \mathbf{x}_{mi}, \mathbf{x}_{e,i})$$

where $j = 1, 2, \ldots, m+1$ and $\mathbf{x}_{ji} = [x_{ji} \quad x_{j,i-t_j} \quad \cdots \quad x_{j,i-(Q_j-1)t_j}]$ represents the phase space embedding for jth variable x_j with the time delay t_j and dimension Q_j. The dimension Q of the multivariate embedding is the sum of each embedding dimension Q_j, $Q = \sum_j Q_j$.

The Gaussian Mixture Model (GMM)

- Preliminary labeling by event function $g(x_t)$
  - $g(x_t)=+1, \{x_{t-Q\tau}, x_{t-Q\tau-1}, \ldots, x_t\}$ are potential pattern data points

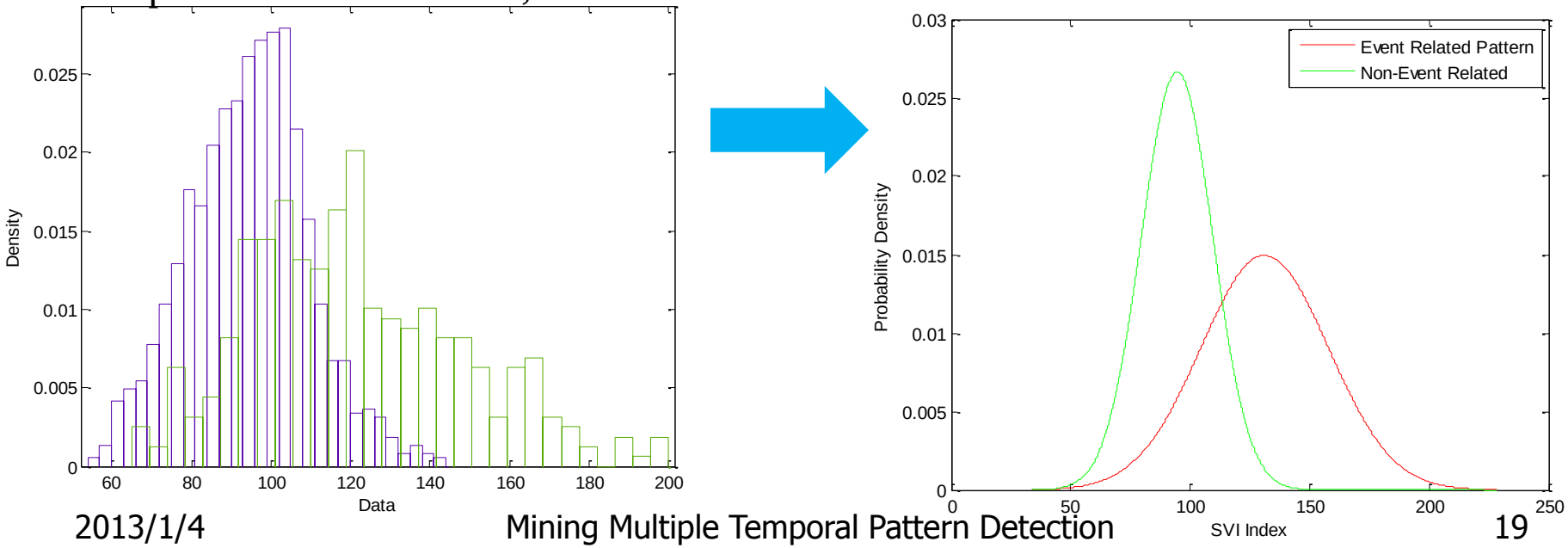


Gaussian Mixture Modeling

- Expectation-Maximization

$$\hat{p}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n}\hat{p}\left(\omega_i \middle| x_k, \hat{\theta}\right) \qquad \hat{\mu}_i = \frac{\sum_{k=1}^{n} p\left(\omega_i \middle| x_k, \hat{\theta}\right) x_k}{\sum_{k=1}^{n} p\left(\omega_i \middle| x_k, \hat{\theta}\right)} \qquad \hat{\Sigma}_i = \frac{\sum_{k=1}^{n} P\left(\omega_i \middle| x_k\right)\left(x_k - \hat{\mu}_i\right)\left(x_k - \hat{\mu}_i\right)^T}{\sum_{k=1}^{n} P\left(\omega_i \middle| x_k\right)}$$

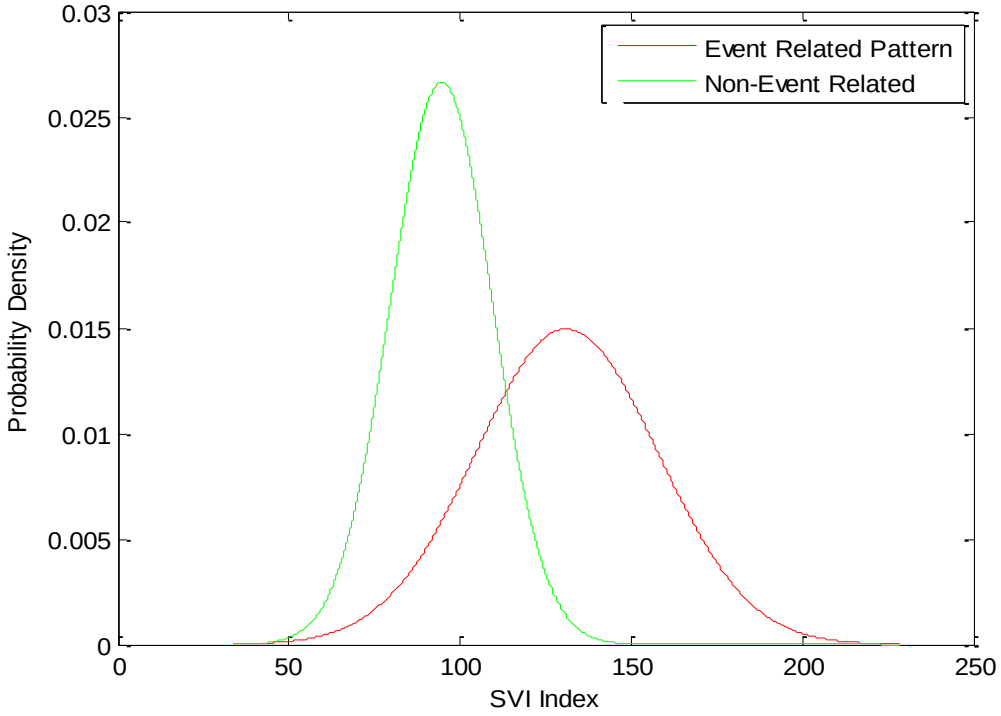
mixture components with each component distributed as $\mathrm{N}(x \mid \mu_i, \Sigma_i)$ with mean μ_i and covariance matrix $\sum_i$, and $p(\omega_i)$ is the marginal distribution for ith component of the mixtures, with constraints .



Gaussian Mixture Modeling

MARQUETTE UNIVERSITY
Be The Difference.

- Discriminative Log-odds Ratio
  - A first-stage clustering by posterior estimation
  - Construct a basis function by $j(\mathbf{x}_t) = \log \frac{P(w_p \mid \mathbf{x}_t)}{P(w_n \mid \mathbf{x}_t)}$



Optimization for Pattern Detection

MARQUETTE UNIVERSITY
Be The Difference.

- Goal: Search for an optimal decision boundary that classify event related patterns
- Labeled by event function, e.g. g=±1
- Embed the filtered data sequence into a phase space
- Solve the following optimization problem:

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

where $$f(\mathbf{x}) = \sum_{i=1}^{N} a_i \exp(\|f(\mathbf{x}) - f(\mathbf{x}_i)\|^2 / s^2) + b_1 j(\mathbf{x}) + b_0$$

- The new vector **x** in phase space will be classified as a member of the pattern class if $f(\mathbf{x}) > 0$

Presentation Outline

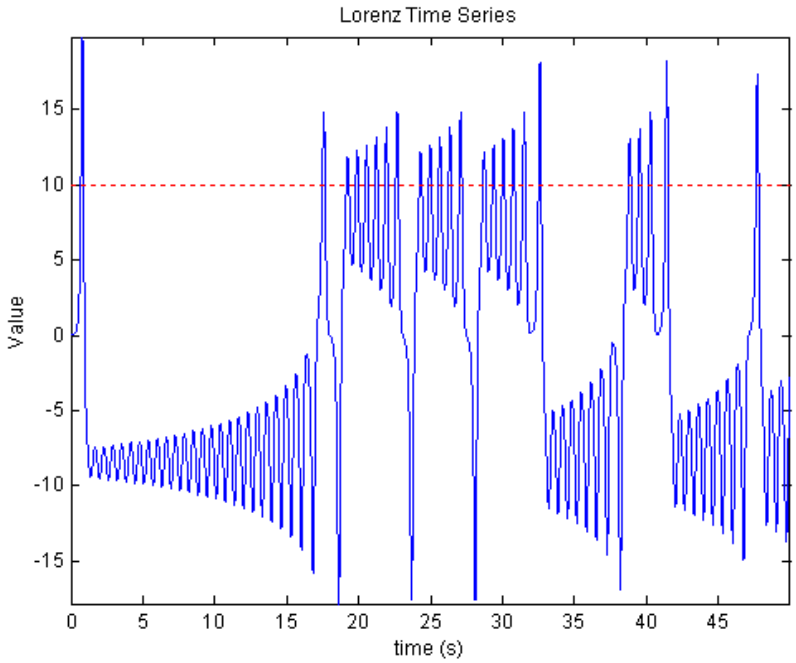
MARQUETTE UNIVERSITY
Be The Difference.

- Problem statement
- Previous work
- A new proposed MRPS method
- **Experimental results**
- Conclusion and future work
- References

Experimental results
Lorenz chaotic series

$$\begin{cases} \dfrac{dx}{dt} = \sigma(y - x) \\ \dfrac{dy}{dt} = rx - y - xz \quad where\, \sigma = 10.0, r = 28.0\, and\, b = \dfrac{8}{3} \\ \dfrac{dz}{dt} = xy - bz \end{cases}$$

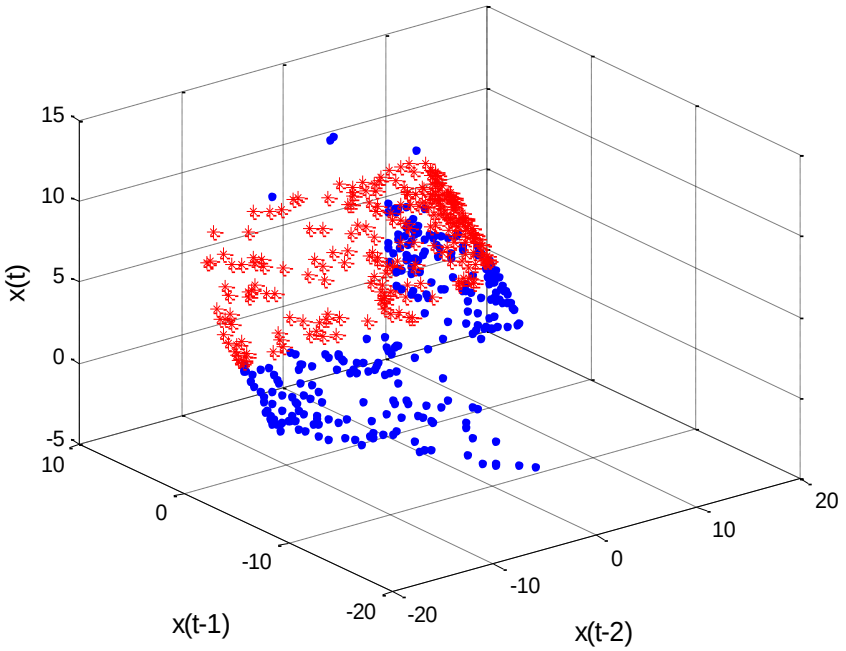
- t=0.2s, Q=3
- Event X(t+1)>10.0
- 500 Second
  - 2500 Data Points



Lorenz Chaotic Series

MARQUETTE UNIVERSITY
Be The Difference.

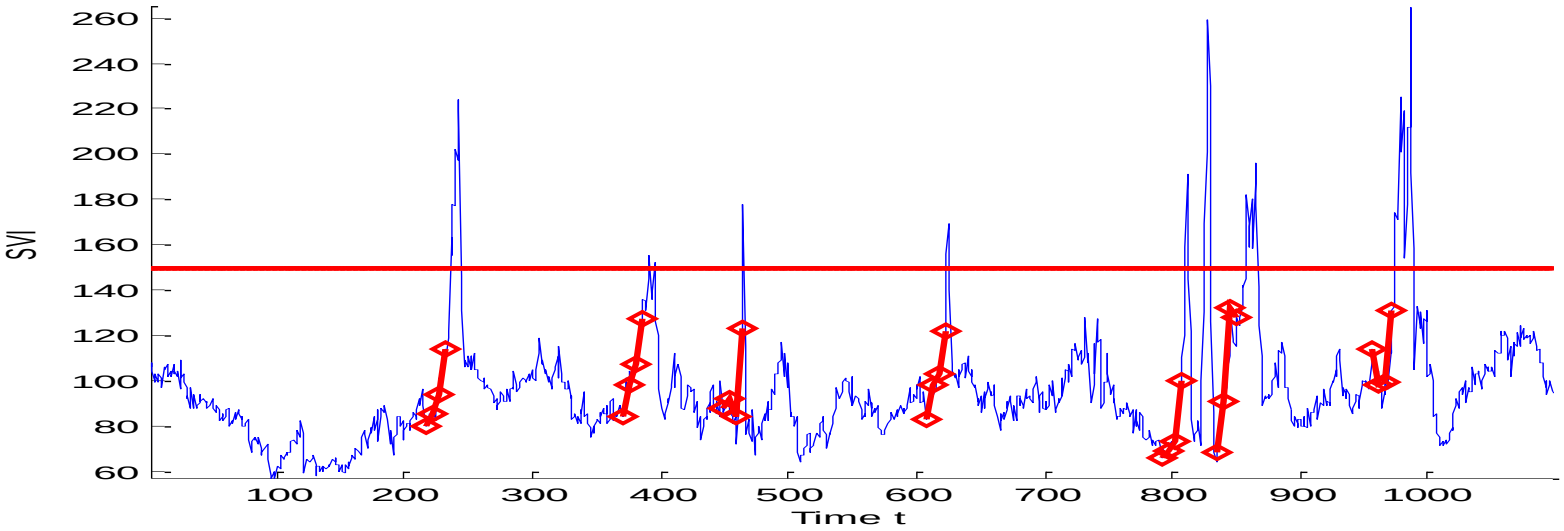
- Pattern points in Phase space are labeled with the definition of event function
  - Red points are related to events
  - Blue points are non-event related



Mining Multiple Temporal Pattern Detection

Sludge Volume Index (SVI) Series

- Sludge bulking is one of the primary causes of water treatment plant failure as the bulking conditions result in exceeding discharge limitations.
- According to U.S. government regulation, a breakout of 150 will result in a fine.
  - g(x_t)=+1 if $\max\{x_{t+1},\ldots,x_{t+3}\}-150.0>0$



Dynamic Temporal Patterns of SVI

Lorenz Chaotic (SVI) Series

MARQUETTE UNIVERSITY
Be The Difference.

A Comparison of Prediction Results of Methods

	Training Set		Test Set	
Method	True Positive Rate	Accuracy	True Positive Rate	Accuracy
MRPS	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%

Sludge Volume Index (SVI) Series

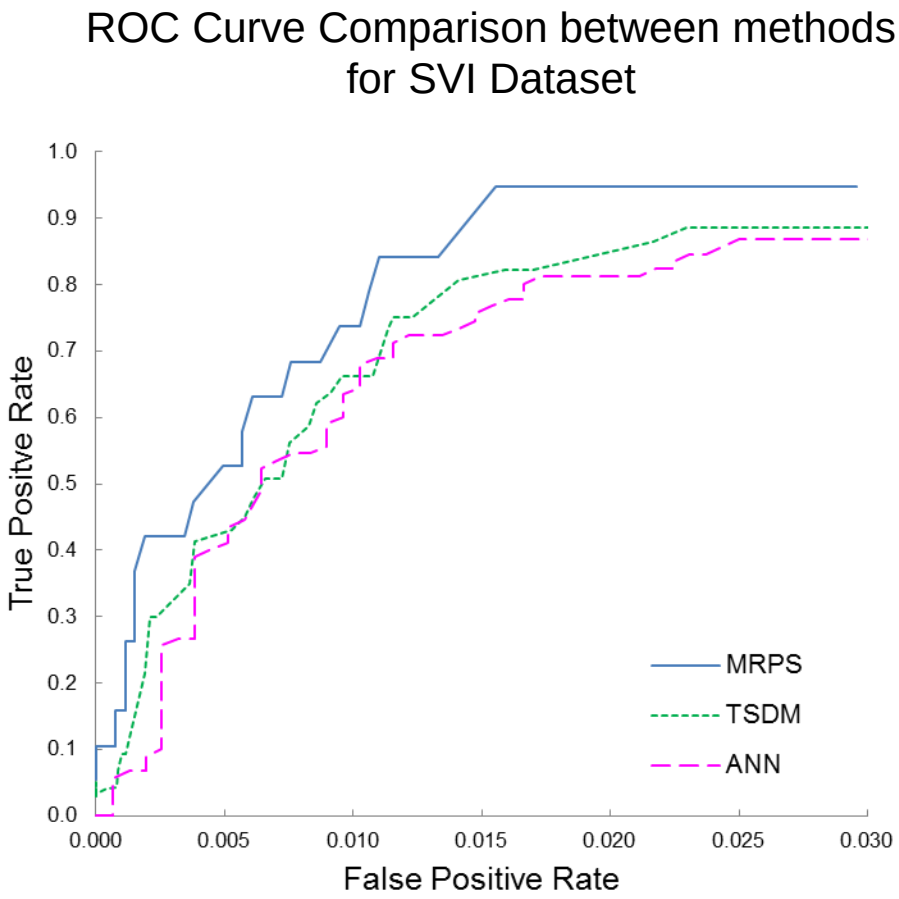
MARQUETTE UNIVERSITY
Be The Difference.

A Comparison of Prediction Results of Methods

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

ROC Curve Comparison

MARQUETTE UNIVERSITY
Be The Difference.



Conclusions & Future Work

MARQUETTE UNIVERSITY
Be The Difference.

- A new MRPS algorithm based on the Multivariate Reconstructed Phase Space is proposed, for detecting dynamic temporal patterns.
- Provides a discriminative approach that utilizes both by Gaussian mixture model and optimization technique to classify temporal patterns that are statistically correlated with events in a dynamic data system.
- Continued publication and more applications

References

MARQUETTE UNIVERSITY
Be The Difference.

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

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

[6] 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.

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

[8] J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining Knowl. Disc.*, vol. 2, no. 2, pp. 1–47, 1998

[9] S.M. Weiss and N. Indurkhya, Predictive Data Mining: A Practical Guide. San Francisco: Morgan Kaufmann, 1998.

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

MARQUETTE UNIVERSITY

Be The Difference.

xin.feng@mu.edu

THANK YOU!