Effect of Dimension Reduction on Prediction Performance of Multivariate Nonlinear Time Series

Jun-Yong Jeong, Jun-Seong Kim, Chi-Hyuck Jun*
Department of Industrial and Management Engineering, Pohang University of Science and Technology,
Pohang, Korea

ABSTRACT

The dynamic system approach in time series has been used in many real problems. Based on Taken's embedding theorem, we can build the predictive function where input is the time delay coordinates vector which consists of the lagged values of the observed series and output is the future values of the observed series. Although the time delay coordinates vector from multivariate time series brings more information than the one from univariate time series, it can exhibit statistical redundancy which disturbs the performance of the prediction function. We apply dimension reduction techniques to solve this problem and analyze the effect of this approach for prediction. Our experiment uses delayed Lorenz series; least squares support vector regression approximates the predictive function. The result shows that linearly preserving projection improves the prediction performance.

Keywords: State space reconstruction, delay Lorenz series, Least Squares Support Vector Regression

* Corresponding Author, E-mail: chjun@postech.ac.kr

1. INTRODUCTION

A traditional time series approach such as ARIMA models interprets the underlying model of series with a linear equation. But, many time series can be generated from a nonlinear relationship. Thus in prediction of nonlinear time series, it is natural to apply the dynamical system approach which is intrinsically nonlinear. The dynamic system approach in time series is widely encountered in natural systems and social systems. It assumes that time series is observed from the state vector in the dynamical system. The system is known to be very sensitive to an initial condition. But in the long term, its behavior is constrained to a fractal finite region because of its invariant topological property. Since Lorenz series was discovered, this approach has been applied to many areas including meteorology (Harding *et al.*, 1990; Lorenz, 1995), medicine (Adeli *et al.*, 2008), economics

(Das and Das, 2007), signal processing (Kennel *et al.*, 1992), traffic flow (Shang *et al.*, 2005), climate (Dhanya and Kumar, 2011), biology (Mackey and Glass, 1977) and so on. For example, Lorenz series is the popular time series in a physics system, especially for explaining climate (Lorenz, 1963); it is the simplified version of the Navier Stokes equations and related to various systems like Raleigh-Bernard problem and other weather problems (Dudul, 2005).

Taken's embedding theorem (Takens, 1981) establishes the theoretical background for prediction in the dynamical system approach. By this theorem, one can build the nonlinear predictive function from a time delay coordinates vector to the future value of the observed series. Although the theorem is for univariate time series, expansion to multivariate time series is natural and showed the better prediction performance (Barnard *et al.*, 2001; Cao *et al.*, 1998).

For approximating the predictive function, several machine learning models are frequently used: Neural Network (Gholipour *et al*., 2006), Neuro-fuzzy Network (Singh and Borah, 2013), Support Vector Regression (Mukherjee *et al*., 1997), and Least Squares Support Vector Regression (LSSVR) (Mei-Ying and Xiao-Dong, 2004). The time delay coordinates vector from multivariate time series, however, can exhibit statistical redundancy which disturbs the performance of a machine learning model (Barnard *et al*., 2001; Han and Wang, 2009). Several studies apply dimension reduction techniques such as Independent Component Analysis (Barnard *et al*., 2001) and Principal Component Analysis (Han and Wang, 2009) to solve this problem. But there is no extensive comparison about the effect of dimension reduction.

Thus, the goal of this paper is to analyze the effect of dimension reduction on prediction of multivariate nonlinear time series; we select LSSVR to approximate the predictive function because of structural risk minimization and computational efficiency. To achieve this, the paper is organized as follows. In Section 2, basic theories about the dynamic system approach in time series are introduced. In Section 3, short descriptions about dimension reduction methods are shown. In Section 4, methodologies for our experiment are described and results from the experiment are given. A conclusion is given in Section 5.

2. DYNAMIC SYSTEM APPROACH IN TIME SERIES

2.1 State Space Reconstruction

A dynamic system consists of a state space $\boldsymbol{S}$, a set of time $\boldsymbol{T}$, and an evolution rule $\boldsymbol{F} : \boldsymbol{S} \times \boldsymbol{T} \to \boldsymbol{S}$. It is the mathematical concept in which the evolution rule describes how the state vector $s(t) \in \boldsymbol{S}$ evolves over time. From the state vector $\boldsymbol{s}(t)$, we have M dimensional observed series $\{x_1(k), x_2(k), \cdots, x_M(k)\}_{k=1}^{N}$ as follows:

$$x_i(k) = h_i(\boldsymbol{s}(kt_s)), \ i = 1, \cdots, M; \ k = 1, \cdots, N \quad (1)$$

where t_s is the sampling rate, $h_i(\cdot)$ is the observation function, and N is the length of the observed series

Our problem is to predict the future values of the observed series $\{x_1(k), x_2(k), \cdots, x_M(k)\}_{k=1}^{N}$ without the information of the original state space. For a univariate time series $\{x_M(k)\}_{k=1}^{N}$, we start by making the time delay coordinates vector that consists of the lagged values of the observed series:

$$\mathbf{X}_i(k) = \{x_i(k), x_i(k - \tau_i), \cdots, x_i(k - (m_i - 1)\tau_i)\}, \quad (2)$$
$$k = (m_i - 1)\tau_i + 1, \cdots, N$$

where τ_i is the time delay and m_i is the embedding dimension. By Taken's embedding theorem (Takens, 1981), the time delay coordinates vector can reconstruct a manifold topologically equivalent to the unknown original manifold in the state space. Under some regularity conditions, for almost τ_i and for some $m_i \geq 2[D] + 1$ (D is the dimension of the original manifold in $\boldsymbol{S}$), there exists a predictive function $f_i : R^{m_i} \to R^{m_i}$ such that

$$\mathbf{X}_i(k+1) = f_i(\mathbf{X}_i(k)) \quad (3)$$

Expansion of the theorem to multivariate time series $\{x_1(k), x_2(k), \cdots, x_M(k)\}_{k=1}^{N}$ is similar (Cao *et al*., 1998); we make the time delay coordinates vector from the multivariate time series as follows:

$$\mathbf{V}(k) = \begin{cases} x_1(k), x_1(k - \tau_1), \cdots, x_1(k - (m_1 - 1)\tau_1); \\ x_2(k), x_2(k - \tau_2), \cdots, x_2(k - (m_2 - 1)\tau_2); \\ \cdots\cdots; \\ x_M(k), x_M(k - \tau_M), \cdots, x_M(k - (m_M - 1)\tau_M); \end{cases} \quad (4)$$
$$k = J_0, J_0 + 1, \cdots, N; \ J_0 = \max_{1 \leq i \leq M} \{(m_i - 1)\tau_i + 1\}$$

where m_i is the embedding dimensions and τ_i is the time delay. If m_i or Σm_i is sufficiently large, there exists a predictive function $g : R^m \to R^m (m = \sum_{i=1}^{M} m_i)$ such that

$$\mathbf{V}(k+1) = g(\mathbf{V}(k)) \quad (5)$$

Equivalently, there exists a predictive function $g_i : R^m \to R$ such that

$$x_i(k+1) = g_i(\mathbf{V}(k)) \quad (6)$$

The state space reconstruction from multivariate series shows the better prediction performance than those from univariate series (Barnard *et al*., 2001; Cao *et al*., 1998).

2.2 Parameter Selection

As we build the time delay coordinates vector, essential task is selecting the time delay τ_i and embedding dimension m_i. The time delay τ_i is calculated separately for each univariate time series with mutual information method (Fraser and Swinney, 1986). Mutual information measures the dependency between $x_i(k)$ and $x_i(k + \tau)$ through a histogram. We select τ_i which shows the first minimum of the mutual information.

We apply False Nearest Neighbor (FNN) method, to compute the embedding dimension m_i. FNN for univariate time series arises from the topological equivalence between the state space and embedding space which is the space of the time delay coordinates vector (Kennel *et al*., 1992). For sufficiently large m_i, the nea-

rest neighbor of a point on the embedding space with dimension m_i is also close to the point on those with dimension $m_i + 1$. If the distance between these points becomes large on the higher dimension embedding space, the nearest neighbor is called false nearest neighbor; the time delay coordinates vector with embedding dimension m_i fails to preserve the topological property. FNN tries to find m_i that minimizes the ratio of false nearest neighbor. Multivariate version FNN is similar to the above except that it minimizes the average ratio of false nearest neighbor by increasing m_i by one in turn (Su, 2010). We start from $m_1 = m_2 = \cdots = m_M = 1$ and increase some m_i until this average goes to zero.

3. DIMENSION REDUCTION

3.1 Dimension Reduction Techniques

Suppose that a $(N - J_0 + 1) \times m$ matrix $\mathbf{V}$ represents a dataset. Assume that this dataset has an intrinsic dimension $d (d < m)$; the vectors $\mathbf{V}(k) \in R^m$, $k = J_0, J_0 + 1$, $\cdots, N$ are lying near the manifold that has dimension d and embedded in the m-dimensional space. Dimension reduction techniques find the mapping from the matrix $\mathbf{V}$ into a new $(N - J_0 + 1) \times d$ matrix $\mathbf{Z}$, while preserving some properties of the dataset (Van der Maaten, 2007). It is divided into four categories as shown in Table 1; linear techniques, global nonlinear techniques, local nonlinear techniques and variants of local nonlinear techniques. Table 1 shows the list of dimension reduction techniques which we applied in our experiment.

3.2 Linear Techniques

Linear techniques are based on a linear mapping. Principal component analysis (PCA) aims to construct the low-dimensional representation of the dataset that describes as much of its variance as possible (Hotelling, 1933). Multidimensional scaling (MDS) retains the pair-

Table 1. Dimension reduction techniques

Category	Technique	Property
Linear	PCA	Variance
	MDS	Pairwise distance
Global nonlinear	Isomap	Geodesic distance
Local nonlinear	Laplacian Eigenmaps	Distance to nearest neighbors
	LLE	Linear combination of nearest neighbors
	LTSA	Tangent space alignment
Variants of Local nonlinear	LLP	Same as Laplacian Eigenmaps
	NPE	Same as LLE
	LLTSA	Same as LTSA

wise distance as possible (Torgerson, 1952).

3.3 Global Nonlinear Techniques

Global nonlinear techniques construct the low-dimensional space which retains a global nonlinear property in the dataset. Isomap preserves the pairwise geodesic distances between the data points; the geodesic distance between two points is measured over the manifold (Tenenbaum *et al.*, 2000).

3.4 Local Nonlinear Techniques

Local nonlinear techniques aim to build the mapping which preserves a local nonlinear property. Laplacian Eigenmap retains the distances between a high-dimensional point and its nearest neighbors; in low-dimensional space, they are weighted by Gaussian kernel function (Belkin and Niyogi, 2002). Local Linear Embedding (LLE) expresses a high-dimensional point as the linear combination of its nearest neighbor (Roweis and Saul, 2000); it tries to preserve the weights in the linear combination. Local Tangent Space Analysis (LTSA) exploits the tangent space in the neighborhood of a high-dimensional point; it aligns these local tangent spaces in the low- dimensional space mapped from the original space (Zhang and Zha, 2004).

3.5 Variants of Local Nonlinear Techniques

Variants of local nonlinear techniques solve the same problem in local nonlinear techniques, but they are based on a linear transformation to solve an out-of-sample problem. Linearity Preserving Projection (LPP) finds the linear mapping that minimizes the cost function of Laplacian Eigenmaps (He and Niyogi, 2004). Neighborhood Preserving Embedding (NPE) is based on LLE (He *et al.*, 2005). Linear Local Tangent Space Alignment (LLTSA) constructs the linear transformation that minimizes the objective function in LTSA (Zhang *et al.*, 2007).

3.6 Correlation Dimension Estimation

Correlation dimension is applied to estimate the intrinsic dimension d. It is based on the fact that the number of points in a hypersphere with radius r is proportional to r^d (Grassberger and Procaccia, 2004).

4. EXPERIMENT WITH DELAYED LORENZ SERIES

To analyze the effect of dimension reduction on predicting multivariate nonlinear time series, we conducted an experiment with a generated series. Figure 1 shows a brief procedure. Details for the procedure are as follows.
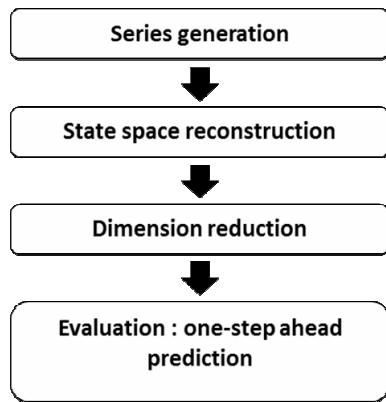
Figure 1. Procedure of the experiment.

4.1 Series Generation

We generated a length 4,000 delayed Lorenz series from the delayed differential equations in Eq. (7) (Zhi-Yong *et al.*, 2011). Parameters in Eq. (7) were chosen by $a = 16$, $b = 45.92$, $c = 4$, $e_1 = 1.2$, $e_2 = 0.75$, $e_3 = 1$ and $\gamma = 3$. The series started with $s_1(t) = s_2(t) = s_3(t) = 1$ for $t < \gamma$. The differential equations were solved by the explicit Runge-Kutta (2, 3) method where the step size of integral h is 0.01. The observed series were $x_i(k) = h_i(s(kt_s)) = s_i(kt_s)$, $i = 1, 2, 3$; $k = 1, \cdots, 4,000$ where the sampling rate t_s is 0.3. First 1,000 observations were burnt to reduce the effect of the initial points. From the remaining 3,000 observations, we considered the next 2,000 observations as a training set and the remainder 1,000 observations as a test set for an evaluation. Figure 2 shows the three series for the training and the test sets.

$$\begin{cases} \dfrac{ds_1}{dt} = -a\big(s_1(t) - s_2(t)\big) - e_1 s_3(t - \gamma) \\[2mm] \dfrac{ds_1}{dt} = (b - s_3)s_1 - s_2 - e_2 s_2(t - \gamma) \\[2mm] \dfrac{ds_1}{dt} = s_1(t)s_2(t) - c s_3(t - \tau) - e_3 s_3(t - \gamma) \end{cases} \quad (7)$$
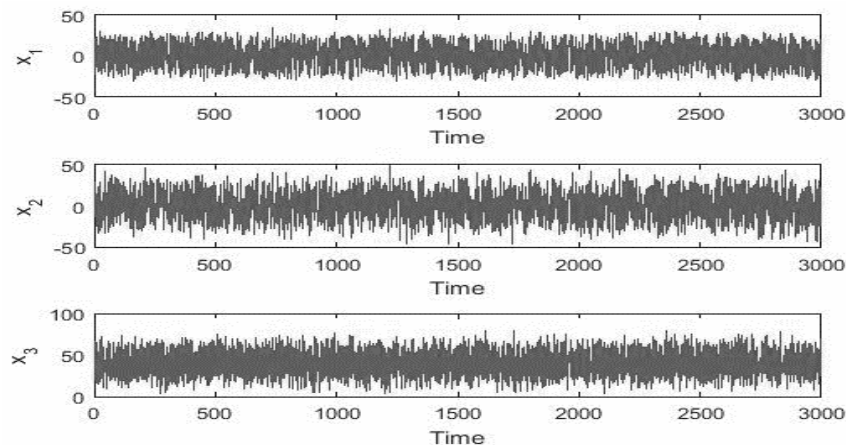
4.2 State Space Reconstruction

To build the time delay coordinates vector from multivariate time series, we used mutual information method and multivariate version FNN to estimate the time delay τ_i and embedding dimension m_i, respectively. For $\{x_1(k), x_2(k), x_3(k)\}_{k=1001}^{3000}$, we obtained the time delay $\tau_1 = \tau_2 = \tau_3 = 1$ and embedding dimension $m_1 = 6$, $m_2 = m_3 = 1$.

$$\mathbf{V}(k) = \begin{pmatrix} x_1(k), x_1(k-1), \cdots, x_1(k-5); \\ x_2(k); \\ x_3(k) \end{pmatrix}$$

For comparison purpose, we also built the time delay coordinates vector from univariate time series; the only difference from the above is that we used univariate version FNN.

$$\mathbf{X}_1(k) = \big(x_1(k), x_1(k-1), \cdots, x_1(k-4)\big)$$
$$\mathbf{X}_2(k) = \big(x_2(k), x_2(k-1), \cdots, x_2(k-5)\big)$$
$$\mathbf{X}_3(k) = \big(x_3(k), x_3(k-1), \cdots, x_3(k-4)\big)$$

4.3 Dimension Reduction

After the correlation dimension of $\mathbf{V}$ was estimated, we transformed $\mathbf{V}$ to $\mathbf{Z}$ through the dimension reduction techniques in Section 3. For local nonlinear techniques and variants of local nonlinear techniques, the number of nearest neighbor was set to 12 by trial and error.

$$\mathbf{V}(x) \in R^7 \to \mathbf{Z}(k) \in R^4$$

4.4 Evaluation: One-Step-Ahead Prediction

For each series, we learned the multivariate model with $\mathbf{Z}(k)$ for one-step ahead prediction by LSSVR.

Figure 2. Training and test sets.

Simplex methods with the initial points from simulated annealing optimized the parameters for LSSVR. The univariate model with $\mathbf{X}_i(k)$ and multivariate model with $\mathbf{V}(k)$ without dimension reduction for one-step ahead prediction were also learned similarly for model comparison. The performance of the model was estimated through Root Mean Square Error (RMSE) on the test set.

$$RMSM = \sqrt{\frac{\sum_{k=3001}^{4000}\left(\widehat{x_i}(k+1) - x_i(k+1)\right)^2}{1000}} \qquad (8)$$

4.5 Results

Table 2 summarizes the results of the experiment. The several models with the different input such as $\mathbf{X}_i(k)$, $\mathbf{V}(k)$, or $\mathbf{Z}(k)$ yielded the different results. The column named 'Univariate' represents the results from the univariate model with $\mathbf{X}_i(k)$ and the column named 'Without dimension reduction' represents the results from the multivariate model with $\mathbf{V}(k)$ without dimension reduction. Similarly, the column named 'PCA' represents the results from the multivariate model with $\mathbf{Z}(k)$ transformed from $\mathbf{V}(k)$ through PCA.

Analogous to the previous studies (Barnard *et al.*, 2001; Cao *et al.*, 1998), the multivariate model with $\mathbf{V}(k)$ without dimension reduction showed smaller RMSE than the univariate model with $\mathbf{X}_i(k)$. Especially, the decrease in RMSE between them was largest for $x_2(k)$ by $\frac{15.14-12.84}{15.14}\times100\% = 15.19\%$. For $x_1(k)$ and $x_3(k)$, RMSE decreased 9.82% and 1.84%, respectively.

Most of the multivariate models with $\mathbf{Z}(k)$, which is transformed from $\mathbf{V}(k)$ through dimension reduction, failed to improve the prediction performance of the multivariate model with $\mathbf{V}(k)$ without dimension reduction; among them, the models that use nonlinear techniques showed the worst prediction performance. LPP marginally reduced RMSE for all series; it reduced RMSE by less than 1%.

5. CONCLUSION

The effect of dimension reduction on predictability of multivariate nonlinear time series is analyzed in this paper. The time delay coordinates vector from multivariate time series can cause statistical redundancy which disturbs the ability of a machine learning model. Thus, we apply various dimension reduction techniques to solve it. From the experiment with delayed Lorenz series, variants of local nonlinear techniques could improve prediction performance of the multivariate model.

Our future work should be to extend the work to higher dimensional series in which dimension reduction could show a better result.

REFERENCES

Adeli, H., Ghosh-Dastidar, S., and Dadmehr, N. (2008), A spatio-temporal wavelet-chaos methodology for EEG-based diagnosis of Alzheimer's disease, *Neuroscience Letters*, **444**(2), 190-194.

Barnard, J. P., Aldrich, C., and Gerber, M. (2001), Embedding of multidimensional time-dependent observations, *Physical Review E*, **64**(4), 046201.

Belkin, M. and Niyogi, P. (2002), *Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering*, In: T. G. Dietterich, S. Becker, and Z. Ghahramani. (eds), *Advance in Neural Information Processing Systems*, MIT Press, 585-591.

Cao, L., Mees, A., and Judd, K. (1998), Dynamics from multivariate time series, *Physica D: Nonlinear Phenomena*, **121**(1), 75-88.

Chen, D. and Han, W. (2013), Prediction of multivariate chaotic time series via radial basis function neural network, *Complexity*, **18**(4), 55-66.

Das, A. and Das, P. (2007), Chaotic analysis of the foreign exchange rates, *Applied Mathematics and Compu-*

Table 2. Results of experiments

	Input										
		Multivariate									
	Univariate	Without dimension reduction	Linear		Global nonlinear	Local nonlinear			Variants of local nonlinear		
			PCA	MDS	Isomap	Laplacian Eigenmaps	LLE	LTSA	LPP	NPE	LLTSA
$x_1(k)$	10.99	9.91	11.14	11.14	12.12	11.61	12.59	10.9	**9.90**	10.96	10.22
$x_2(k)$	15.14	12.84	14.04	14.05	15.45	14.53	15.49	14.05	**12.78**	14.14	12.97
$x_3(k)$	10.83	10.63	10.9	10.9	12.01	11.38	11.78	11.05	**10.53**	11.34	10.61

tation, **185**(1), 388-396.

Dhanya, C. and Kumar, D. N. (2011), Multivariate nonlinear ensemble prediction of daily chaotic rainfall with climate inputs, *Journal of Hydrology*, **403**(3), 292-306.

Dudul, S. V. (2005), Prediction of a Lorenz chaotic attractor using two-layer perceptron neural network, *Applied Soft Computing*, **5**(4), 333-355.

Fraser, A. M. and Swinney, H. L. (1986), Independent coordinates for strange attractors from mutual information, *Physical Review A*, **33**(2), 1134.

Gholipour, A., Araabi, B. N., and Lucas, C. (2006), Predicting chaotic time series using neural and neuro-fuzzy models: a comparative study, *Neural Processing Letters*, **24**(3), 217-239.

Grassberger, P. and Procaccia, I. (2004), *Measuring the strangeness of strange attractors*, In: B. R. Hunt, J. A. Kennedy, T.-Y. Li and H. E. Nusse. (eds.), *The Theory of Chaotic Attractors*, Springer, 170-189.

Han, M. and Wang, Y. (2009), Analysis and modeling of multivariate chaotic time series based on neural network, *Expert Systems with Applications*, **36**(2), 1280-1290.

Harding, A. K., Shinbrot, T., and Cordes, J. M. (1990), A chaotic attractor in timing noise from the VELA pulsar?, *The Astrophysical Journal*, **353**, 588-596.

He, X. and Niyogi, P. (2004), *Locality preserving projections*, In: S. Thrun, L. K. Saul and B. Schölkopf. (eds), *In Advances in Neural Information Processing Systems 16*, The MIT Press, Cambridge, USA, MA, 153-160.

He, X., Cai, D., Yan, S., and Zhang, H.-J. (2005), *Neighborhood preserving embedding*, Proceedings of the Tenth IEEE International Conference on the Computer Vision (ICCV), **2**, 1208-1213.

Hotelling, H. (1933), Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, **24**(6), 417- 441.

Kennel, M. B., Brown, R., and Abarbanel, H. D. (1992), Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review A*, **45**(6), 3403-3411.

Lorenz, E. N. (1963), Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences*, **20**(2), 130-141.

Lorenz, E. N. (1995), *The essence of chaos*, University of Washington Press.

Mackey, M. C. and Glass, L. (1977), Oscillation and chaos in physiological control systems, *Science*, **197**(4300), 287-289.

Mei-Ying, Y. and Xiao-Dong, W. (2004), Chaotic time series prediction using least squares support vector machines, *Chinese Physics*, **13**(4), 454-458.

Monahan, A. H. (2000), Nonlinear principal component analysis by neural networks: Theory and application to the Lorenz system, *Journal of Climate*, **13** (4), 821-835.

Mukherjee, S., Osuna, E., and Girosi, F. (1997), Nonlinear prediction of chaotic time series using support vector machines, *Proceedings of the 1997 IEEE Workshop of the Neural Networks for Signal Processing*, 511-520.

Roweis, S. T. and Saul, L. K. (2000), Nonlinear dimensionality reduction by locally linear embedding, *Science*, **290**(5500), 2323-2326.

Shang, P., Li, X., and Kamae, S. (2005), Chaotic analysis of traffic time series, Chaos, *Solitons and Fractals*, **25**(1), 121-128.

Su, L.-y. (2010), Prediction of multivariate chaotic time series with local polynomial fitting, *Computers and Mathematics with Applications*, **59**(2), 737-744.

Suykens, J. A., De Brabanter, J., Lukas, L., and Vandewalle, J. (2002), Weighted least squares support vector machines: robustness and sparse approximation, *Neurocomputing*, **48**(1), 85-105.

Takens, F. (1981), Detecting strange attractors in turbulence. In: D. A. Rand and L. S. Young (eds.), *Dynamical Systems and Turbulence*, Warwick 1980, Springer, Berlin, German, 366-381.

Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000), A global geometric framework for nonlinear dimensionality reduction, *Science*, **290**(5500), 2319-2323.

Torgerson, W. S. (1952), Multidimensional scaling: I. Theory and method, *Psychometrika*, **17**(4), 401-419.

Van der Maaten, L. (2007), An introduction to dimensionality reduction using matlab, *Report*, **1201**(07-07), 62.

Vapnik, V. (2013), *The nature of statistical learning theory*, Springer Science and Business Media.

Zhang, T., Yang, J., Zhao, D., and Ge, X. (2007), Linear local tangent space alignment and application to face recognition, *Neurocomputing*, **70**(7), 1547-1553.

Zhang, Z.-Y. and Zha, H.-Y. (2004), Principal manifolds and nonlinear dimensionality reduction via tangent space alignment, *Journal of Shanghai University (English Edition)*, **8**(4), 406-424.

Zhi-Yong, Y., Guang, Y., and Cun-Bing, D. (2011), Time-delay feedback control in a delayed dynamical chaos system and its applications, *Chinese Physics B*, **20** (1), 010207.