Optik 135 (2017) 98–103

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Time series prediction using dynamic Bayesian network

Qinkun Xiao*, Chu Chaoqin, Zhao Li

Department of Electrical and Inforamtion Engineering, Xi'an Technological University, Xi'an, Shaanxi, 710032, PR China

ARTICLE INFO

Article history:
Received 2 November 2016
Accepted 25 January 2017

Keyword:
Time series prediction
Echo state network
KFM
DBN

ABSTRACT

Time series prediction is a challenging research topic, especially for multi-step-ahead prediction. In this paper, a novel multi-step-ahead time series prediction model is proposed based on combination of the Kalman filtering model (KFM) and the echo neural networks (ESN). Recently, the studies demonstrate the ESN model is a promising strategy for multi-step-ahead time series prediction, at the same time, the KFM is a recursion-based sequence information processing approach, which has been used effectively for prediction, filtering and smooth of time series data. In this paper, we consider to use the recursion-based KFM to enhance performance of the ESN-based direct prediction model. A novel graph model named the E-KFM that generated from combination of the ESN and the KFM is developed to predict multi-step-ahead time series data. The simulation and comparison results show that the proposed model is more effectiveness and robustness.

1. Introduction

Recently, time series forecasting is concerned by many researches, most of time prediction models focus on one-step-ahead prediction, the multi-step-ahead prediction is still a challenge topic. We can divide the multi-step-ahead time series forecasting into two kinds of ways [1–3]: direct-based and iterate-based approaches. The direct-based approaches always build prediction models to estimate multi-step-ahead values directly. The iterate-based approaches often first constructs one-step-ahead forecasting framework, and the predictions are usually used as known information to estimate the next ones. In iterate-based approaches, the prediction accuracy is great influenced by cumulative errors, on the other hand, for direct-based methods, time cost of methods always is an important element to be considered.

In split of aforementioned two main trend approaches, other time series forecasting approaches are also concerned, for example, such as, multi-input several multi-outputs (MISMO) forecasting framework [6,7], the multi-input multi-output (MIMO) approach [5], and DirRec method [4], and so on. The MIMO and the MISMO strategy usually concern to get higher prediction accuracy, however, it is often with the higher time cost. The MISMO approaches divide original forecasting problem into subtasks, using optimal solution or cross-validation, to estimate system outputs [8], on the one hand, the forecasting performance is usually not bad, however, the time cost of computation still need to be improved.

In this article, a new graph-based approach is proposed for improving performance of time series forecasting, and the algorithm is based on combination of the echo state network (ESN) [9,10] and Kalman filtering frame. The prediction system is described by the dynamic Bayesian network (DBN), the DBN can present random sequence signals entirely. Our motives can be described as follow.

* Corresponding author.
E-mail addresses: xiaoqinkun10000@163.com, xiaoqinkun@tsinghua.edu.cn (Q. Xiao).
URL: http://mailto:xiaoqinkun10000@163.com (Q. Xiao).

http://dx.doi.org/10.1016/j.ijleo.2017.01.073
0030-4026/

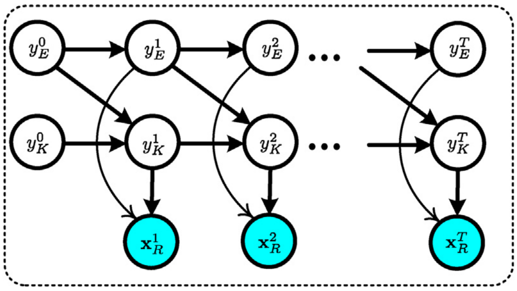

Fig. 1. DBN presentation for E-KFM.

Firstly, in last years, many scholars have proved that the ESN is promising time series prediction system. In many fields, the ESN has been used to predict sequence information [9,10], hence, in this paper, the ESN model is selected as one part of our forecasting frameworks.

Secondly, in split of the ESN method has good prediction performance, only short-time sequence information is used to estimate the coming information. Such as, for the ESN real-time forecasting [9], only previous part of short-time data is utilized as input signals to predict data, if all previous sequence data is used to estimate the coming data, the forecasting performance would be further improved.

In short, in this paper, the short-time forecasting model would be combined with the long-time sequence estimation approach together. The sequence information processing methods usually can use sequence signals entirely and effectively. For example, the Kalman filtering model [11,12] is an effective algorithm to handle sequence signals data, such as, sequence data smoothing, filtering and prediction [13–15]. Further, for random signals representation and processing, the dynamic Bayesian network (DBN) is a good selection. According to [13], the DBN-based sequence processing system is with more flexible representation, on the other hand, the model can work with other models easily. Such, in our paper, a combination frame is proposed, the frame put the ESN-based forecasting model and the KFM together, and the combination system is described using DBN graph model too, it is called the E-KFM time series prediction model.

The rest parts are organized as follows. Section 2 describes the detail steps of the E-KFM for multi-step-ahead prediction. Simulation results are analyzed in Section 3, and conclusions are drawn in Section 4.

2. Prediction model

To build a multi-input and one-output nonlinear system model: $f:\mathbb{R}^D \to \mathbb{R}$, which is used for h-step-ahead prediction, and it is written as:$\hat{y}(n+h)=f(\mathbf{x}_R(n))$. The $x_R(n)=[x(n), \cdots, x(n-(D-1)\tau)]^T$, where the τ is time delay coefficient, the precondition of time series data reconstruction is $D \geq 2d_1+1$, where the d_1 is correlation dimension. Based on the DBN graph model theory and phase space reconstruction theory [16], the DBN graph model of time series prediction is built in Fig. 1, which is a combination of the ESN-based prediction and the KFM model, it is called the E-KFM model.

The ESN is a recurrent-based calculation neural network, it is made up of input layer of signals, a recurrent-based reservoir and signals output layer. Let $\mathbf{x}_R^t$ and y_E^t be the input and output signals in time t, respectively. The echo state e^t inside the reservoir and output y_E^t are generated as follows [10]: $\mathbf{e}^t = f(\mathbf{W}_{in}\mathbf{x}_R^t + \mathbf{W}_z\mathbf{e}^{t-1})$, and $y^t = e^t W_{\text{out}}$, where $f(.)$ denotes the reservoir activation function, and W_{in},W_e and W_{out} are weight value matrixes are corresponding to input layer, echo layer and output layer, respectively. In the E-KFM, if let the $x_R(n)$ be input of the ESN, and $y_E(n+h)$ be the h-step-ahead prediction data of the ESN, we have: $\hat{y}_E(n+h)=f_{ESN}(\mathbf{x}_R(n))$, where the $f_{ESN}(.)$ is a function denoted the operation of ESN-based prediction.

Similar to ESN-based prediction, let $x_R(n)$ be the measurements of the KFM, h-step-ahead prediction data $y_K(n+h)$ be the true states of the KFM, we have: $\hat{y}_K(n+h)=f_{KFM}(\mathbf{x}_R(n))$. For simple, let $\mathbf{x}_R^t$, y_K^t and y_E^t be the value of $x_R(n)$, $y_K(n+h)$ and $y_E(n+h)$ at time t, respectively. Given the observation sequence $\mathbf{x}_R^{1:t}$, the hidden sequence $y_K^{1:t}$ is obtained using the DBN probability inference, the obtained $y_K^{1:t}$ would be the higher accuracy prediction. The detail is discussed as follows.

Based on the DBN graph in Fig. 1, considering the E-KFM model as a physical system, we have:

$$\begin{cases} y_K^{t+1} = \mathbf{F}_t\mathbf{Y}_{K+E}^t + \mathbf{w}_t \\ \mathbf{x}_R^t = \mathbf{H}_t\mathbf{Y}_{K+E}^t + \mathbf{v}_t \end{cases} \tag{1}$$

where $\mathbf{Y}_{K+E}^t = (y_K^t, y_E^t)^T$ and $y_E^t = f_{ESN}(\mathbf{x}_R^t)$, and the F_t, H_t, w_t and v_t are state transition matrix, observation matrix, system noise and measurement noise, respectively. Both of the w_t and v_t are Gaussian distributions: $\mathbf{w}_t = \mathbb{N}(0, \mathbf{Q}_t)$, $\mathbf{v}_t = \mathbb{N}(0, \mathbf{R}_t)$, where the Q and R are system noise matrix and measurement noise matrix, respectively. Suppose all continue variables in the E-KFM are Gaussian distributions, let prior state of distribution of E-KFM be:$p(y_K^t) = \mathbb{N}(\mu_K, \Sigma_K)(y_K^t)$, and $p(y_E^t) = \mathbb{N}(\mu_E, \Sigma_E)(y_E^t)$. Let the state condition distribution of the E-KFM be $p(y_K^{t+1}|\mathbf{Y}_{K+E}^t) = \mathbb{N}(\mathbf{F}\mathbf{Y}_{K+E}^t, \mathbf{Q}_t)(\mathbf{Y}_{K+E}^t)$, and the measurement distribution of the E-KFM be $p(\mathbf{x}_R^t|\mathbf{Y}_{K+E}^t) = \mathbb{N}(\mathbf{H}\mathbf{Y}_{K+E}^t, \mathbf{R}_t)(\mathbf{x}_R^t)$. The DBN parameters are calculated using the Maximum Likehood [11] (ML) algorithm.

Based on above assumptions, prediction data sequence $y_K^{1:t}$ is calculated using DBN inference based on recursion strategy, the steps are described as follows.

(1) *One-step-ahead data pre-estimation and update*. Firstly, in E-KFM, to calculate $p(y_K^1)$ according to Bayesian rule [11]:

$$\begin{aligned} P(y_K^1) &= \iint_{y_K,y_E} P(y_K^1|\mathbf{Y}_{K+E}^0)P(\mathbf{Y}_{K+E}^0)dy_K dy_E \\ &= \iint_{y_K,y_E} P(y_K^1|y_E^0, y_K^0)P(y_E^0, y_K^0)dy_K dy_E \end{aligned} \tag{2}$$

Due to variable y_E^0 and y_K^0 are independent of each other, hence, $P(y_E^0, y_K^0) = P(y_E^0)P(y_K^0)$. The next, to update y_K^1 based on measurements $\mathbf{x}_R^1$, using Bayesian rule [11]:

$$\begin{aligned} P(y_K^1, y_E^1|\mathbf{x}_R^1) &= \alpha P(\mathbf{x}_R^1|y_K^1, y_E^1)P(y_K^1, y_E^1) \\ &= \alpha P(\mathbf{x}_R^1|y_K^1)P(\mathbf{x}_R^1|y_E^1)P(y_E^1)P(y_K^1) \\ &= \alpha P(\mathbf{x}_R^1|y_K^1)P(\mathbf{x}_R^1|y_E^1)P(y_E^1) \\ &\times \iint_{y_K,y_E} P(y_K^1|y_E^0, y_K^0)P(y_E^0, y_K^0)dy_K dy_E \end{aligned} \tag{3}$$

(2) *Recursion-based filtering calculation*. In E-KFM, we use measurement data sequence $\mathbf{x}_R^{1:t}$ to update the y_K^t, the all known data is used entirely, According to the probability graph model and Markov theory [11], the recursion-based filtering calculation is:

$$\begin{aligned} P(y_K^t, y_E^t|\mathbf{x}_R^{1:t}) &= P(y_K^t, y_E^t|\mathbf{x}_R^t, \mathbf{x}_R^{1:t-1}) \\ &= P(y_K^t, y_E^t|\mathbf{x}_R^t)P(y_K^t, y_E^t|\mathbf{x}_R^{1:t-1}) \\ &= \alpha P(\mathbf{x}_R^t|y_K^t, y_E^t)P(y_K^t, y_E^t) \cdot P(y_K^t, y_E^t|\mathbf{x}_R^{1:t-1}) \\ &= \alpha P(\mathbf{x}_R^t|y_K^t, y_E^t)P(y_K^t, y_E^t) \\ &\times \iint_{y_K,y_E} P(y_K^t, y_E^t|y_K^{t-1}, y_E^{t-1})P(y_K^{t-1}, y_E^{t-1}|\mathbf{x}_R^{1:t-1}) \\ &= \alpha P(\mathbf{x}_R^t|y_K^t)P(\mathbf{x}_R^t|y_E^t)P(y_K^t)P(y_E^t) \\ &\times \iint_{y_K,y_E} P(y_K^t|y_K^{t-1}, y_E^{t-1})P(y_E^t|y_K^{t-1}, y_E^{t-1})P(y_K^{t-1}, y_E^{t-1}|\mathbf{x}_R^{1:t-1}) \\ &= \alpha P(\mathbf{x}_R^t|y_K^t)P(\mathbf{x}_R^t|y_E^t)P(y_K^t)P(y_E^t) \\ &\times \iint_{y_K,y_E} P(y_K^t|y_K^{t-1})P(y_E^t|y_K^{t-1})P(y_K^{t-1}, y_E^{t-1}|\mathbf{x}_R^{1:t-1}) \end{aligned} \tag{4}$$

Because in E-KFM, according to Kalman filtering, we have:

$$\begin{aligned} P(y_E^t|\mathbf{x}_R^{1:t}) &= \alpha P(\mathbf{x}_R^t|y_E^t)P(y_E^t) \\ &\times \int_{y_E} P(y_E^t|y_E^{t-1})P(y_E^{t-1}|\mathbf{x}_E^{1:t-1}) \end{aligned} \tag{5}$$

Based on Eqs. (4) and (5), we have:

$$\begin{aligned} P(y_K^t|\mathbf{x}_R^{1:t}) &= \frac{P(y_K^t, y_E^t|\mathbf{x}_R^{1:t})}{P(y_E^t|\mathbf{x}_R^{1:t})} \\ &= \frac{P(\mathbf{x}_R^t|y_K^t)P(y_K^t)\iint_{y_K,y_E} P(y_K^t|y_K^{t-1})P(y_K^{t-1}, y_E^{t-1}|\mathbf{x}_R^{1:t-1})}{\int_{y_E} P(y_E^{t-1}|\mathbf{x}_R^{1:t-1})} \end{aligned} \tag{6}$$

Table 1
Performance comparison of multi-step-ahead prediction for *Lorenz* data.

Algorithm	$E^{h=1}_{NRMSE}$	$E^{h=6}_{NRMSE}$	$E^{h=12}_{NRMSE}$	$E^{h=18}_{NRMSE}$
E-KFM	0.0602	0.0605	0.0638	0.0762
LV-SVM [18]	0.053	0.1330	0.1617	0.1858
ESN [9]	0.0854	0.1063	0.1576	0.1755
RBF [12]	0.2097	0.1848	0.1876	0.2308

Lastly, the value of h-step-ahead prediction is computed as:

$$\hat{y}^t_K = \mathbf{E}[y^t_K|\mathbf{x}^{1:t}_R] = \int_{y_K} y^t_K P(y^t_K|\mathbf{x}^{1:t}_R)dy^t_K \tag{7}$$

The total prediction algorithm is described as:
Algorithm: Time series prediction based on E-KFM
Input: training data $\{x_t\}_{t=1:t_{trn}}$, checking data $\{x_t\}_{t=1:t_{chk}}$
Output: h-step-ahead prediction $\hat{\boldsymbol{y}}_{pre}(n+h)$
1.Based on Takens theorem, to compute phase space reconstruction parameters: D, τ;
2.The data are reconstructed based on the parameters:
$\{\mathbf{x}^t_R\}_{t=1:T}, \mathbf{x}^t_R = [x(n), \cdots, x(n-(D-1)\tau)]^T$ and $\{y^t\}_{t=1:T}, y^t = x(n+h)$;
3. To build the E-KFM model, based on training data $\{\mathbf{x}^{1:t_{trn}}_R, y^{1:t_{trn}}\}$ to obtain optimal parameters: $\{\mu_E$, μ_K, Σ_E, Σ_K, F, H, Q, $R\}$ based on ML method;
4. To get initial system prediction value: $\mathbf{Y}^0_{K+E} = [y^0_K, y^0_E]^T$;
5. For $t = 1:T$
6. Based the E-KFM to calculate $P(y^t_K)$ based on Eq. (2);
7. To compute the $P(y^t_K, y^t_E|\mathbf{x}^{1:t}_R)$ according to Eq. (4);
8. To calculate $P(y^t_E|\mathbf{x}^{1:t}_R)$ according to Eq. (5);
9. Calculating the $P(y^t_K|\mathbf{x}^{1:t}_R)$ according to Eq. (6);
10. To compute the $y^t_K - E[y^t_K|\mathbf{x}^{1:t}_R]$ according to Eq. (7);
11. End
12. To output h-step-ahead prediction $\hat{\boldsymbol{y}}_{pre}(n+h) = \{y^{1:T}_K\}$.

3. Simulations

We use one example to evaluate the performance of the proposed model. Lorenz attractor is linked with the most famous 3D self-control non-linear dynamical system, can be used for modeling of turbulent flow in fluid, the dynamical equation is written as [17]:

$$\begin{cases} \dfrac{dx(t)}{dt} = -\sigma x(t) + \sigma y(t) \\ \dfrac{dy(t)}{dt} = -x(t)z(t) + rx(t) - y(t) \\ \dfrac{dz(t)}{dt} = x(t)y(t) - bz(t) \end{cases} \tag{8}$$

when σ, r, b = (10, 28, 8/3), the system is chaotic state. To set system initial parameters are $x_0 = 1$, $y_0 = 2$, $z_0 = 3$. Based on fourth order Runge-Kutta method, we can get the numerical solution of x(t) in Eq. (8). According to Takens theorem [19], the optimal embedding dimension and embedding delay of x(t) are calculated. Hence, we use $x_R = [x(t-11), \cdots, x(t-1), x(t)]$ as input data to predict output data: x(t + h). We extract 1000 pairs input-output data, the first 300 is the training data and the remaining is the checking data.

Based on the proposed E-KFM model, the prediction results ($h = 1$) are shown in Fig. 2, according to [9], the first 100 data is moved, from Fig. 2, we know there is little difference between target values and predicted values, and the errors range is also very small, the results show the proposed prediction model has good popularized value.

For fair comparison, the same environment is used to assess the performance of the proposed model with of the other methods, including the ESN [9], the RBF [12], and the LS-SVM [18]. The normalized root mean square error (NRMSE) is used to access the performance of prediction.

$$E_{RMSE} = \sqrt{\frac{\sum_{j=1}^{N} \left(y(j) - \hat{y}(j)\right)^2}{\sigma^2 N}} \tag{9}$$

where the σ is the standard deviation (STD) of chaotic time series data. The comparison results are shown in Table 1. The kernel function of LS-SVM is Gaussian RBF kernel, and the parameters are selected based on training data according to

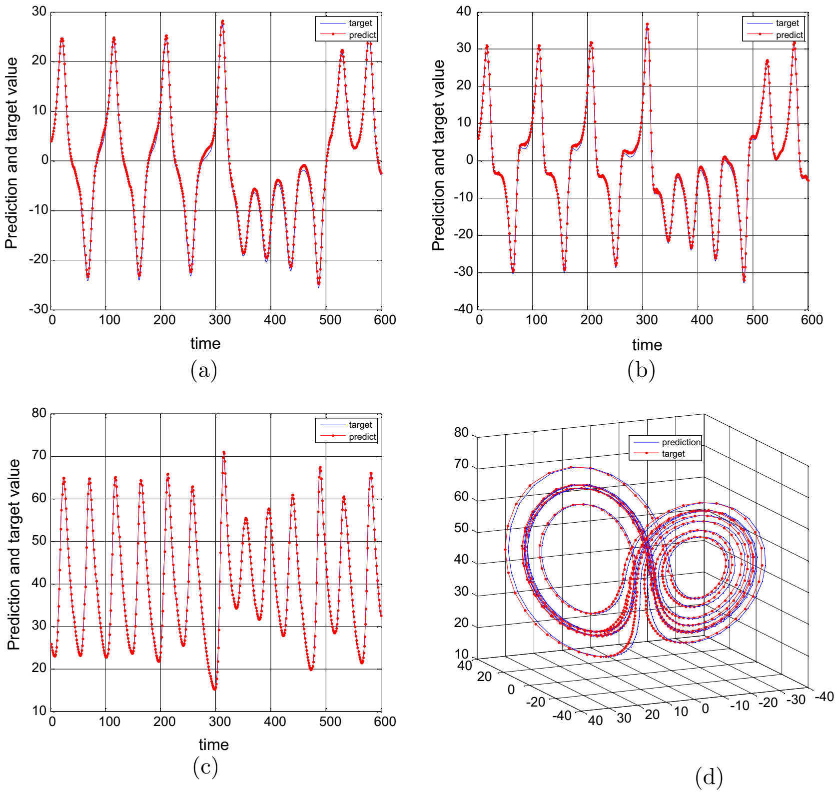

Fig. 2. *Lorenz* time series prediction results. (a) The prediction and target of $x(t)$. (b) The prediction and target of $y(t)$. (c) The prediction and target of $z(t)$. (d) The prediction and target data in 3D space.

leave-one-out method or cross-validation method [16]. Based on simple search and 5 fold cross-validation method, optimal parameter is: $\gamma = 8500, \sigma^2 = 3$. The number of hidden layer in RBF is 5, the excitation function is $g(x) = \exp\left(-\frac{||x-c_i||_2^2}{\sigma_i^2}\right)$, where the c_i is the central vector of i-th node, the σ_i is the base width. The excitation function of ESN is Sigmoid function $f(x) = \frac{1-e^x}{1+e^x}$, the scope of reserve pooling is 40, the spectral radius of internal connection weights is 0.8, The input unit scale is 0.2, sparse degree is 0.03.

From the comparison results of prediction in Table 1, the E-KFM prediction model is superior to other models, it also indicate that the proposed model has better performance and robustness.

4. Conclusions

In this paper, a multi-step-ahead time series prediction models that combine ESN with the KFM is proposed. Our contribution can be described as:

(1) We propose a novel graph-based time series prediction model named the E-KFM, the model combines the neural network and Bayesian inference together effectively, and uses recursion-based method to predict multi-step-ahead time series data.
(2) Based on some theories, such as Bayesian rules, dynamic Bayesian networks and phase space reconstruction, the probability-based recursion calculation structure is presented to obtain the higher accuracy of multi-step-ahead prediction. Experimental results show that the E-KFM model has better performance than some existing algorithms.

Acknowledgment

This research is supported by the project (61271362) of the National Natural Science Foundation of China.

References

[1] A.F. Atiya, S.M. El-Shoura, S.I. Shaheen, M.S. El-Sherif, A comparison between neural-network forecasting techniques-case study: river flow forecasting, IEEE Trans. Neural Netw. 10 (1999) 402–409.

[2] G. Chevillon, Direct multi-step estimation and forecasting, J. Econ. Surv. 21 (2007) 746–785.
[3] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, A. Lendasse, Methodology for long-term prediction of time series, Neurocomputing 70 (2007) 2861–2869.
[4] A. Sorjamaa, A. Lendasse, Time series prediction using DirRec strategy, in: European Symposium on Artificial Neural Networks Bruges, (Belgium) 26–28 April, 2006.
[5] G. Bontempi, Long term time series prediction with multi-input multi-output local learning, in: European Symposium on Time Series Prediction, Helsinki Finland, 2008, pp. 145–154.
[6] S. Ben Taieb, G. Bontempi, A. Sorjamaa, A. Lendasse, Long-term prediction of time series by combining direct and MIMO strategies, in: IEEE International Joint Conference on Neural Networks, Atlanta, U.S.A., 2009, pp. 3054–3061.
[7] S. Ben Taieb, G. Bontempi, A.F. Atiya, A. Sorjamaa, A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition, Expert Syst. Appl. (2012).
[8] S. Ben Taieb, A. Sorjamaa, G. Bontempi, Multiple-output modeling for multi-step-ahead time series forecasting, Neurocomputing 73 (2010) 1950–1957.
[9] S.P. Wen, Z.G. Zeng, T.W. Huang, Exponential stability analysis of memristor-based recurrent neural networks with time-varying delays, Neurocomputing 97 (2012) 233–240.
[10] Shi ZW, HanM, Support vector echo-state machine for chaotic time-series prediction, IEEE Trans. Neural Netw. 18 (2) (2007) 359–372.
[11] J.L. Gross, J. Yellen, Graph Theory and Its Applications, 2nd edn., Chapman and Hall/CRC, Boca Raton, 2011.
[12] Simon Haykin, Neural Networks and Learning Machines, 3rd edn., Prentice Hall, 2009.
[13] Qinkun Xiao, Li Xing, Gao Song, Time series prediction using optimal theorem and dynamic Bayesian network, Optik 127 (2016) 11063–11069.
[14] Xuedong Wu, Zhiyu Zhu, Shaosheng Fan, A novel online adaptive time series prediction model with input and output uncertainties, Optik 126 (June(11–12)) (2015) 1213–1220.
[15] H.-J. Yang, X. Hu, Wavelet neural network with improved genetic algorithm for traffic flow time series prediction, Optik 127 (October (19)) (2016) 8103–8110.
[16] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.-S. Young (Eds.), Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol 898, Springer-Verlag, 1981, pp. 366–381.
[17] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (1963) 130–141.
[18] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore, 2002.
[19] M.B. Kennel, R. Brown, HDI Abarbanel determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A 45 (6) (1992) 3403–3411.