

A PMLP BASED METHOD FOR CHAOTIC TIME SERIES PREDICTION

Hongying Yang, Hao Ye, Guizeng Wang, Maiying Zhong

Department of Automation, Tsinghua University, Beijing, 100084, P.R. China

E-mail: yanghy02@mails.tsinghua.edu.cn

Abstract: This paper proposes a new method for prediction of chaotic time series based on Parallel Multi-Layer Perceptron (PMLP) net and dynamics reconstruction technique. The PMLP contains a number of multi-layer perceptron (MLP) subnets connected in parallel. Each MLP subnet predicts the future data independently with a different embedding dimension. The PMLP determines the final predicted result according to the weighted average of all sub-outputs. Simulation results show the effectiveness of the method.
Copyright © 2005 IFAC

Keywords: chaos theory, time-series analysis, phase space, prediction, neural nets, parallel networks.

1. INTRODUCTION

The problem of nonlinear and chaotic time series prediction amounts to making predictions of future values of given chaotic data. It is one of the fundamental problems in science and engineering as well as in other disciplines. Since many signals generated from practical systems show chaotic behaviours, it has been widely applied in various areas, such as the heart beat research, the earthquake and the sunspot investigation, weather and power load forecasting and so on.

As an important way to study the characteristics of complicated systems, the interests in chaotic time series prediction have been increased over the past few years. There are many forecasting methods such as local-region method (Lv and Zhang, 2002a), neural network-based forecasting method and so on. Most methods are based on dynamics reconstruction to obtain information on the data characteristics. Among them, Camastra and Colla (1999) proposed a multi-layer perceptron (MLP) method for prediction using reconstruction techniques.

This paper presents a parallel multi-layer perceptron (PMLP) net for prediction of chaotic time series based on Camastra's approach (Camastra and Colla, 1999). The proposed method can reduce the error

accumulation effect and improve prediction stability of the MLP. In Section 2, the method of dynamics reconstruction to obtain the estimated value of correlation dimension is first introduced. Section 3 proposes a new method for chaotic time series prediction based on the PMLP net structure. Simulation results using the data of Chen's attractor are given in Section 4. In Section 5, some conclusions are drawn.

2. RECONSTRUCTION OF CHAOTIC TIME SERIES

This section will briefly review the idea of phase space reconstruction of chaotic time series (Camastra and Colla, 1999). The idea stems from the embedding theorem developed by Takens (1981) and Sauer, *et al.* (1991). The theorem regards a one-dimensional chaotic time series as the compressed information of higher dimension. Then, the time series $x(t), t = 1, 2, 3, \dots, N$ can be represented as a series of points $X(t)$ in a d -dimensional space

$$X(t) = (x(t), x(t-1), \dots, x(t-(d-2)), x(t-(d-1))) \quad (1)$$

where d is called *embedding dimension* of the system. According to *Takens' embedding theorem*

(Takens, 1981), in order to obtain an available reconstruction of the dynamics system, the embedding dimension d must satisfy.

$$d \geq 2D + 1, \quad (2)$$

where D is the dimension of the attractor, which is the key point to obtain correct embedding dimension. Once the time series is reconstructed and the minimum embedding dimension is estimated, the result will be used by PMLP.

Grassberger-Procaccia algorithm (Grassberger and Procaccia, 1983) is one of the popular methods for determining the so-called correlation dimension D_m , which is an estimation of the dimension of the attractor D . If the correlation integral $C_d(r)$ is defined as

$$C_d(r) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N H(r - |X_i - X_j|) \quad (3)$$

where H is the Heaviside function and d is the embedding dimension. It is shown (Lv, *et al.*, 2002b) that for r sufficiently small, and the number of observed values N sufficiently large

$$D_m = \lim_{r \rightarrow 0} \frac{\ln C_d(r)}{\ln(r)} \quad (4)$$

The algorithm plots $\ln C_d(r)$ against $\ln(r)$ curves through increasing the value of d until the slope of the curve's linear part is almost an invariable. Then, through regressing correlation dimension based on least squares (Zhao, *et al.*, 1999), the correlation dimension estimation D_m can be attained.

3. CHAOTIC TIME SERIES PREDICTION

How to predict chaotic time data with higher accuracy is the key of this paper. In practice, since most of the chaotic data is mixed with noise, it is rather difficult to get the exact estimation value of the minimum embedding dimension. Further, since the single MLP method (Camastra and Colla, 1999) uses the estimation value of the minimum embedding dimension as the number of inputs, it often gets inaccurate results due to the inaccurate estimation of the embedding dimension. So in this paper, a PMLP system consisting of multiple MLP is proposed to enhance system performance and reduce system error.

3.1 The Structure of PMLP.

The PMLP is composed of multiple MLP nets connected in parallel for predicting the time series whose structure is shown in Figure 1. The PMLP

contains n MLP subnets, which are denoted by $sub-MLP_i$, $i = 1, 2, \dots, n$, respectively. Each $sub-MLP$ net produces the predicted estimation independently at the same instant $t+1$ and the output of $sub-MLP_i$ is denoted as y_i . With proper combination of y_i , the PMLP outputs the final predicted value y .

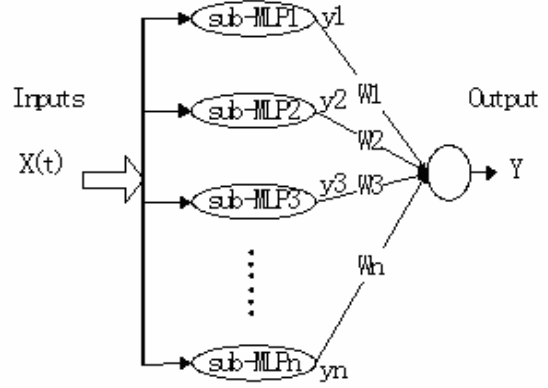


Fig. 1. The structure of PMLP.

3.2 Predict with a Single MLP.

In Figure 1, each multi-layer perceptron $sub-MLP_i$ has the same structure as that used in Camastra's method (Camastra and Colla, 1999), which takes a standard feedforward topology with only one hidden layer and full connection between adjacent, and uses mean square error (MSE) as the cost function. The numbers of hidden nodes are selected by optimization based on simple exhaustive search. The $sub-MLP_i$ is a multiple input single output (MISO) net with In_i inputs $x(t), x(t-1), \dots, x(t-m_i+1)$ and an output variable y_i . The equation can be expressed as

$$\overline{x_i}(t+1) = F[x(t), x(t-1), x(t-2), \dots, x(t-m_i+1)] \quad (5)$$

where the output $y_i = \overline{x_i}(t+1)$ is the one-step prediction of $x(t+1)$ at instant t . After the training of MLP by the historical data set, the one-step prediction value $\overline{x_i}(t+1)$ can be achieved.

3.3 The Input Nodes of Each Subnet.

In Camastra's method (Camastra and Colla, 1999), the estimated minimum embedding dimension is used as the number of inputs of the single MLP. In this paper, the same estimation value is taken as the number of inputs of the central subnet, and the other subnets take different input numbers calculated based on that value.

In practice, the minimum embedding dimension is chosen according to the estimated correlation dimension value D_m . Once the correlation dimension D_m is regressed by least squares through Grassberger-Procaccia algorithm, the number of input nodes of *center subnet sub-MLP* $_{[n/2]}$ can be determined as

$$In_{[n/2]} = 2D_m + 1 \quad (6)$$

where $[x] = a$, $a \leq x < a + 1$ ($a \in Z, x \in R$). n is the total subnet number of PMLP. In_i is the number of input nodes of the subnet MLP_i . When $i = [n/2]$, the *sub-MLP* $_i$ net is called *center subnet*. Then the number of input nodes of each subnet can be set as

$$In_i = In_{[n/2]} + (i - [n/2]) \quad (7)$$

where $i = 1, 2, \dots, n$.

3.4 The Computation of Weighted Factor.

Since each MLP subnet has different influence to the whole system, the employment of *weighted factor* ω is necessary. In this paper, the *weighted value* of the each subnet is determined according to the minimum predicted *absolute percent error* (APE) of y_i at each instant. Then, through the multiple linear regression method, the weighted parameters of each subnet can be obtained.

The overall output of the PMLP net is the weighted combination of the each individual MLP subnet and the final predicted result can be represented by the following equation.

$$\bar{x}(t+1) = \sum_{i=1}^n \omega_i \times \bar{x}_i(t+1) \quad (8)$$

where $\bar{x}_i(t+1)$ is the output of i th subnet y_i , $\bar{x}(t+1)$ is the output of y .

4. EXPERIMENTAL RESULTS

In this section, the simulation results of the Chen's attractor data are given to test the performance of the proposed method.

4.1 Time Series Data.

In the following, the chaotic time series based on the Chen's attractor data is used to test the proposed

method. Equation (9) generates the Chen's time series data (Lv, *et al.*, 2002b).

$$\begin{cases} x' = a(y - x) \\ y' = (c - a)x - xz + cy \\ z' = xy - bz \end{cases} \quad (9)$$

where the parameters are set with the values $a=35$, $b=3$, $c=28$, and integral time step is 0.0001. The 4-order Runge-Kutta algorithm with the initial values $x(0) = 0$, $y(0) = 1$, $z(0) = 0$ generates the time series. Fig 2 shows the construction of Chen's attractor and Fig 3 shows 400 points of x-component normalized Chen's time series data and noisy chaotic time series used in the prediction respectively. The noisy time series are produced by Chen's time series mixed with Gauss white noise.

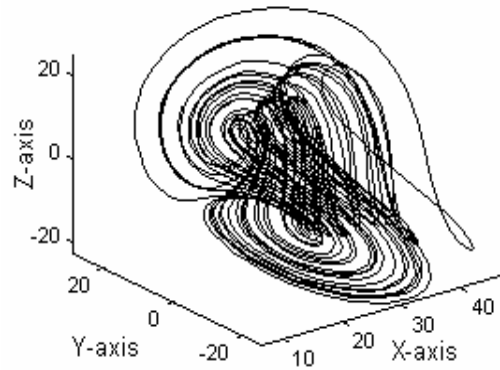


Fig. 2. Chen's attractor.

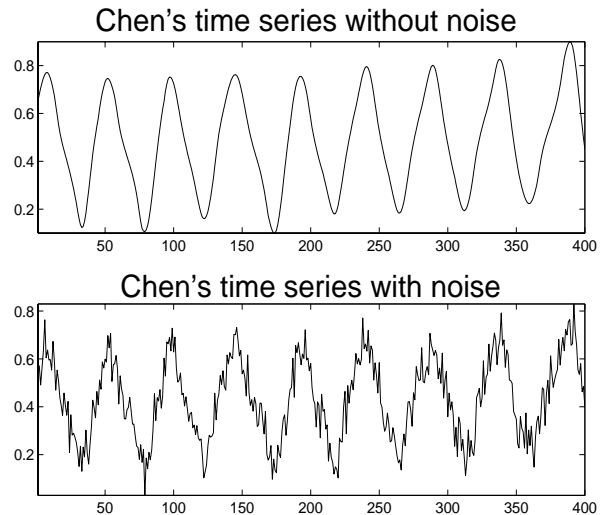


Fig. 3. Chen's time series data (x-component).

According to Grassberger-Procaccia method, the $\ln C_d(r)$ against $\ln(r)$ curves are plotted with the increase of the embedding dimension d (see Figure 4, Figure5).

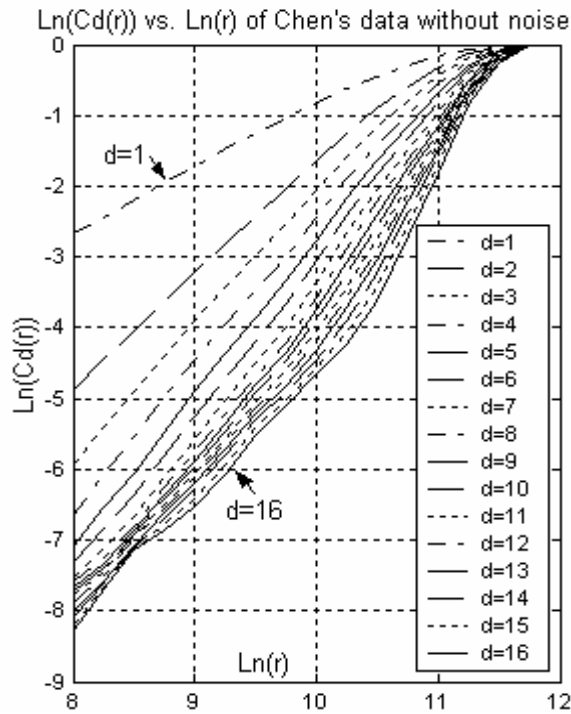


Fig. 4. Plot of $\ln C_d(r)$ against $\ln(r)$ of Chen's time series without noise.

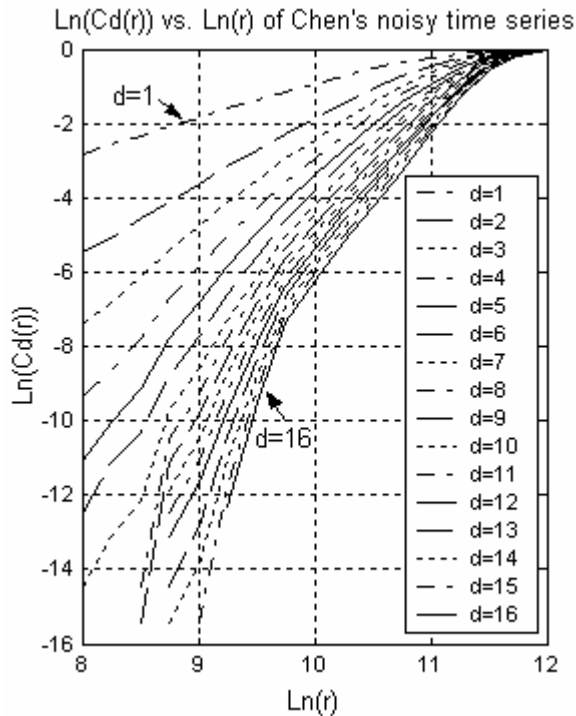


Fig. 5. Plot of $\ln C_d(r)$ against $\ln(r)$ of Chen's time series with noise.

From the figure 5, it can be seen that the curves is divergent and the slope of the Chen's noisy time series curve's linear part can not reach an invariable. So the accurate estimation of correlation dimension with noisy time series is rather difficult. The MLP doesn't fit this case.

Table 1 reports the computed result using the method of regression correlation dimension by least squares (Zhao, *et al.*, 1999). So the input nodes of the *center subnet* can be calculated.

Table 1: The Estimation of the Minimum Dimension

| Data set | Correlation dimension |
|--------------------------|-----------------------|
| Chen's time series | 2.4235 |
| Chen's noisy time series | 3.4715 |

4.2 Prediction Time Series with MLP and PMLP.

As a whole, 1,250 normalized Chen's time series data without noise were generated for the using of prediction. The data were divided into two training sets and one testing. The first 1000 samples are used for MLP net training. The next 100 samples are employed to decide the optimal value of weighted factor ω . The last 150 samples are used for test of the predicted accuracy of MLP and PMLP.

The tests were performed by a three-layer MLP and the PMLP net. The input nodes of MLP and the *center subnet* of PMLP are 6, which is the estimated value of the minimum embedding dimension. The net was tried on the test set. After the presentation of each training sample, the weights and biases are updated. Compared with MLP, a PMLP net was also tried. The parameters of PMLP are

$$\begin{aligned}
 I_{[n/2]} &= 6 & ; & & I_{in} &= [4, 5, 6, 7, 8] & ; \\
 \omega &= [0.25, 0.18, 0.40, 0.10, 0.07] & & & & & (10)
 \end{aligned}$$

Figure6 shows the one-step predicted results using the PMLP net. Figure 7 presents the APE errors of the two methods. The error curves indicate that the prediction stability of the PMLP system is better than the MLP net.

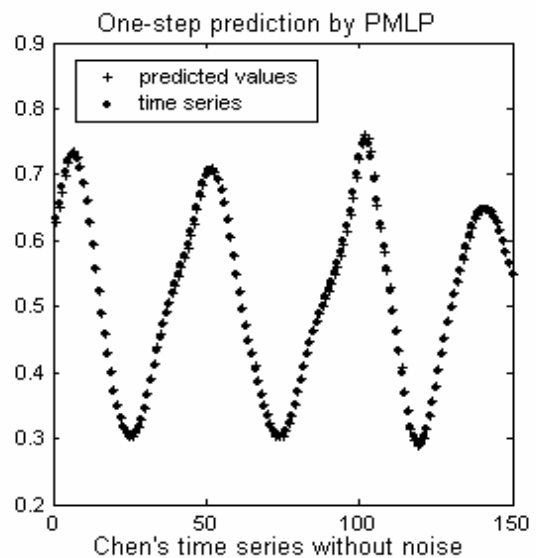


Fig. 6. One-step prediction by PMLP.

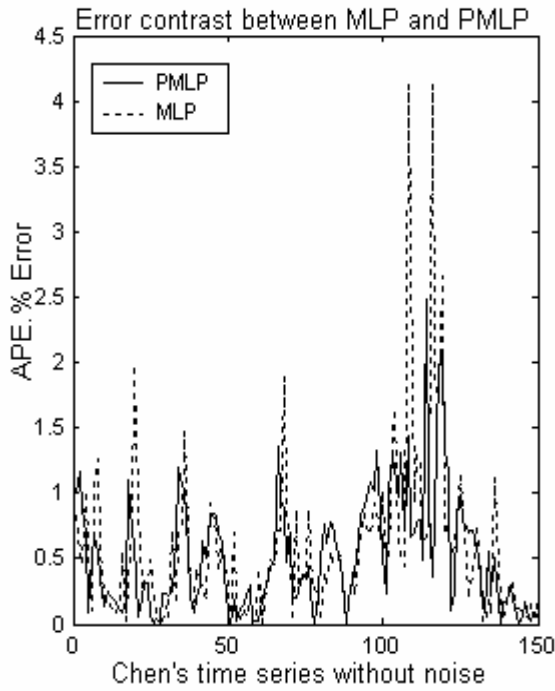


Fig. 7. Error contrast between MLP and PMLP.

Table 2 shows the percentage analysis according to APE error between MLP and PMLP. From the table, the predicted errors of the PMLP net are less than 2.5%. What's more, the ratio of the predicted error below 1.5% can reach 98%. But for the MLP, some of the predicted errors are larger than 2.5%. Also, the results are measured in terms of the max APE, *mean absolute percent error* (MAPE) and *root mean square* (RMS) reported in Table 3.

Table 2: The error percentage analysis of MLP and PMLP

| APE. % Error | MLP | PMLP |
|--------------|-----------|-----------|
| > 2.5 | 2.6667 % | 0.0000 % |
| 1.5~2.5 | 3.3333 % | 2.0000 % |
| < 1.5 | 94.0000 % | 98.0000 % |

Table 3: The error comparison between MLP and PMLP

| Method | MLP | PMLP |
|---------|-------------|-------------|
| MAX. % | 4.1240 | 2.4865 |
| MAPE. % | 0.5794 | 0.5174 |
| RMS | 7.4077e-005 | 4.6192e-005 |

4.3 Noisy Chaotic Time Series Prediction.

In this part, a 1,250 normalized Chen's time series data with Gauss white noise were generated. The data also were divided into two training sets and one testing as above. The MLP and the PMLP net were performed on normalized Chen's noisy data set. The

input nodes of MPL is 8 and the PMLP net parameters are

$$I_{[n/2](noise)} = 8 ; I_{in(noise)} = [6, 7, 8, 9, 10] ;$$

$$\omega_{(noise)} = [0.25, 0.14, 0.24, 0.20, 0.17] \quad (11)$$

Figure 8 shows the one-step predicted result of the noisy time series by the PMLP net. Figure 9 presents the noisy data's APE error contrast between MLP and PMLP.

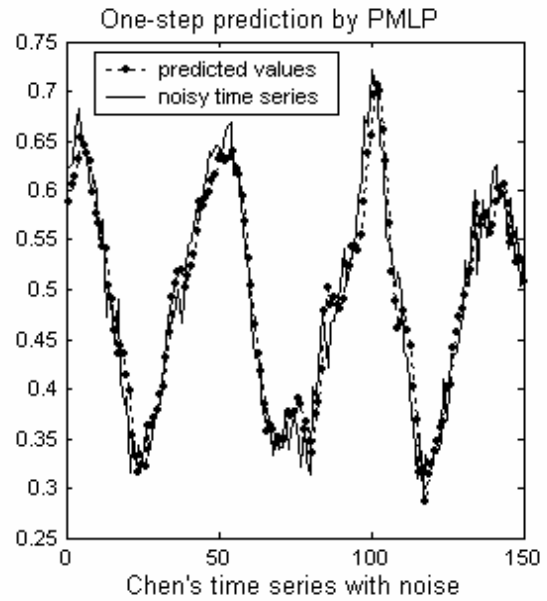


Fig. 8. One-step prediction by PMLP (with noisy data).

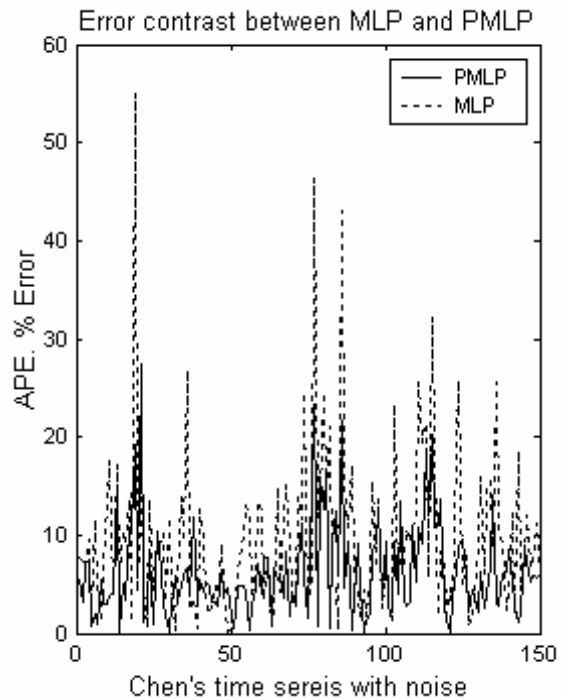


Fig. 9. Error contrast between MLP and PMLP (with noise data).

Also, Table 4 presents percentage analysis of the error of MLP and PMLP. Table 5 shows the predicted result measured according to the max APE, MAPE and RMS.

Table 2 and Table 4 indicate that the prediction performance of the PMLP net is obviously better than the MLP. And the max APE, MAPE and RMS of the PMLP net is also less than the MLP according to Table 3 and Table 5, especially for the predicted result of Chen's noisy time series.

Table 4: The error percentage analysis of MLP and PMLP (with noisy data)

| APE. % Error | MLP | PMLP |
|---------------------|------------|-------------|
| > 25 | 5.3333 % | 0.6667 % |
| 15~25 | 10.6667 % | 5.3333 % |
| < 15 | 84.0000 % | 94.0000 % |

Table 5: The error comparison between MLP and PMLP (with noisy data)

| Method | MLP | PMLP |
|----------------|------------|-------------|
| MAX. % | 55.0054 | 27.5234 |
| MAPE. % | 9.6015 | 6.2038 |
| RMS | 0.0165 | 0.0062 |

From the predictive result of above, the PMLP is obviously efficient than the MLP net for the case with noise present in the data. The coefficients in equation (11) shows the importance of the subnets other than the central subnet corresponding to the minimum embedding dimension (i.e. 8 in this example) and further demonstrates the necessity of introducing the multiple MLP nets.

5. CONCLUSIONS

This paper proposes to use the parallel multi-layer perceptron (PMLP) net to predict chaotic and noisy chaotic time series. The method is based on dynamics reconstruction technique. The PMLP contains multiple numbers of MLP nets connected by parallel. Each MLP subnet predicts the future value of the same instant independently with different embedding dimension. The PMLP determines the final value estimated by weighted average of every sub-output.

The experimental results show that compared with MLP, the PMLP can not only reduce the effect of error accumulation, but also heighten the prediction stability of the system. In particular, for the noisy chaotic time series the PMLP has a more available result than the MLP net. If properly applied, the

theory is an available tool for prediction and analysis of chaotic data.

ACKNOWLEDGEMENTS

The research is supported by the National Natural Science Fund of China and the 863 program of China.

REFERENCES

- Camastra, F. and A.M. Colla (1999). Neural Short-Term Prediction Based on Dynamics Reconstruction. *Neural Processing Letters*, Vol. 9, pp. 45-52
- Grassberger, P. and I. Procaccia (1983). Characterization of Strange Attractors. *Physics Review Letters*, Vol. 50, pp. 346-349.
- Lv J H and Zhang S C (2002a). Application of adding-weight one-rank local-region method in electric power short-term load forecast. *Control Theory and Applications*, Vol. 19, pp. 767-770. (In Chinese)
- Lv J H, Lu J A and Chen S C (2002b). *Chaotic Time Series and Its Application*. pp.34-36, Wuhan University Press, Wuhan. (In Chinese)
- Sauer, T., J.A. Yorke and M. Gasdagli (1991). Embedology. *J. Stat. Phys.*, Vol. 65, pp. 579-616.
- Takens, F. (1981). Detecting Strange Attractors in Turbulence. *Lecture Note in Mathematics*, Vol. 898, pp. 366-381.
- Zhao G B, Shi Y F, Duan W F and Yu H R (1999). Computing Fractal Dimension and The Kolmogorov Entropy from Chaotic times series. *Chinese Journal of Computational Physics*, Vol. 16, pp. 309-315. (In Chinese)