Journal of Southeast University (English Edition) Vol. 21, No. 2, pp. 229 - 232 June 2005 ISSN 1003—7985

Local polynomial prediction method of multivariate chaotic time series and its application

Fang Fen Wang Haiyan

(College of Economics and Management, Southeast University, Nanjing 210096, China)

Abstract: To improve the prediction accuracy of chaotic time series, a new method formed on the basis of local polynomial prediction is proposed. The multivariate phase space reconstruction theory is utilized to reconstruct the phase space firstly, and on its basis, a polynomial function is applied to construct the prediction model, then the parameters of the model according to the data matrix built with the embedding dimensions are estimated and a one-step prediction value is calculated. An estimate and one-step prediction value is calculated. Finally, the mean squared root statistics are used to estimate the prediction effect. The simulation results obtained by the Lorenz system and the prediction results of the Shanghai composite index show that the local polynomial prediction errors of the multivariate chaotic time series are small and its prediction accuracy is much higher than that of the univariate chaotic time series.

Key words: chaotic time series; phase space reconstruction; local polynomial prediction; stock market

In practical problems, it is hard to build up exact analytic models for complex systems (such as the stock market) because their constructions are very intricate and the information available is incomplete and inaccurate. Complex systems are usually analyzed by time series observed or measured from the systems. In the conventional prediction studies, most of the methods are based on the univariate chaotic time series. However, the multivariate time series can always be observed or measured from the complex system. The prediction errors with multivariate chaotic time series are much smaller than those with univariate chaotic time series when using the local mean prediction, local linear prediction and BP neural networks prediction[1]. In this paper we extend the local polynomial prediction method[2,3] from the univariate chaotic time series to the multivariate chaotic time series.

1 Phase Space Reconstruction

Suppose that we have observed an M-dimensional multivariate chaotic time series $\{\boldsymbol{x}_n\}_{n=1}^N = \{(x_{1,n}, x_{2,n}, \ldots, x_{M,n})\}_{n=1}^N$. As in the case of a univariate time series (where $M=1$), we make the time delay reconstruction:

$$\begin{aligned} \boldsymbol{V}_n = \{&x_{1,n}, x_{1,n-\tau_1}, \ldots, x_{1,(m_1-1)\tau_1}; x_{2,n}, x_{2,n-\tau_2}, \ldots, x_{2,(m_2-1)\tau_2}; \ldots; \\ &x_{M,n}, x_{M,n-\tau_M}, \ldots, x_{M,(m_M-1)\tau_M}\} \qquad n = J_0, J_0+1, \ldots, N; J_0 = \max_{1\leqslant i\leqslant M}(m_i-1)\tau_i + 1 \end{aligned} \tag{1}$$

where τ_i, m_i ($i = 1, 2, \ldots, M$) are the time delays and the embedding dimensions, respectively[4,5]. Following Takens's delay embedding theorem[6], when $m = \sum_{i=1}^{M} m_i$ or each m_i is large enough, there exists an M continued function f_i: $\mathbf{R}^m \rightarrow \mathbf{R}$, such that

$$x_{i,n+1} = f_i(\boldsymbol{V}_n) \qquad i = 1, 2, \ldots, M \tag{2}$$

Thus the evolution from $\boldsymbol{V}_n$ to $x_{i,n+1}$ reflects the motion of the original unknown dynamics. This means that the geometrical characteristics of the attractor in the reconstructed space are equivalent to the original state space. So any differential or topological invariant quantities computed for the reconstructed attractor are identical to those in the original state space.

We find the time delays τ_i with the mutual average information method[6,7] separately for each univariate time series $\{x_{i,n}\}_{n=1}^N$, $i = 1, 2, \ldots, M$. We use the same method for choosing the embedding dimensions m_i as that of Ref. [5].

Received 2004-11-11.

Biographies: Fang Fen (1977—), female, graduate; Wang Haiyan (corresponding author), male, doctor, associate professor, hywang@ seu. edu. cn.

2 Local Polynomial Prediction of Multivariate Chaotic Time Series

Suppose that the state vector at time T is $\boldsymbol{V}_T$. Time p later than T on attractor is approximated by the function

$$x_{1,T+p} \cong f_1(\boldsymbol{V}_T) \tag{3}$$

where $x_{1,T+p}$ is determined by the d-th order polynomial $f_i(\boldsymbol{V}_T)$

$$\begin{aligned} x_{1,T+p} \cong f_i(\boldsymbol{V}_T) &= f_0 + \sum_{k_1=0}^{m_1-1} f_{1k_1}^{(1)} x_{1,T-k_1\tau_1} + \dots + \sum_{k_d=k_{d-1},\dots,k_2=k_1,k_1=0}^{m_1-1} f_{dk_1k_2\dots k_d}^{(1)} x_{1,T-k_1\tau_1} \dots x_{1,T-k_d\tau_1} + \\ &\sum_{k_1=0}^{m_2-1} f_{1k_1}^{(2)} x_{2,T-k_1\tau_2} + \dots + \sum_{k_d=k_{d-1},\dots,k_2=k_1,k_1=0}^{m_2-1} f_{dk_1k_2\dots k_d}^{(2)} x_{2,T-k_1\tau_2} \dots x_{M,T-k_d\tau_2} + \dots + \\ &\sum_{k_1=0}^{m_M-1} f_{1k_1}^{(M)} x_{M,T-k_1\tau_M} + \dots + \sum_{k_d=k_{d-1},\dots k_2=k_1,k_1=0}^{m_M-1} f_{dk_1k_2\dots k_d}^{(M)} x_{M,T-k_1\tau_M} \dots x_{M,T-k_d\tau_M} \end{aligned} \tag{4}$$

In the local prediction method, the change of $\boldsymbol{V}_T$ with time on the attractor is assumed to be the same as those of nearby points, $\boldsymbol{V}_{T_h}(h=1,2,\dots,k)$. Using k of $\boldsymbol{V}_{T_h}$ and $\boldsymbol{V}_{T_h+p}$, for which the values are already known, the coefficient of $\boldsymbol{f}$ is determined by the solution of the following equation.

$$\boldsymbol{v} \cong \boldsymbol{A}\boldsymbol{f} \tag{5}$$

where

$$\boldsymbol{v} = \{x_{T_1+p}, x_{T_2+p}, \dots, x_{T_k+p}\}^{\mathrm{T}} \tag{6}$$

$$\boldsymbol{f} = \{f_0, f_{10}^{(1)}, \dots, f_{1(m_1-1)}^{(1)}, \dots, f_{10}^{(M)}, \dots, f_{1(m_M-1)}^{(M)}, \dots, f_{d(m_M-1)\dots(m_M-1)}^{(M)}\}^{\mathrm{T}} \tag{7}$$

and $\boldsymbol{A}$ is the $k \times s$ ($s = \sum_{i=1}^{M} C_{m_i+d}^{d} - M + 1$) matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & x_{1,T_1} & \dots & x_{1,T_1-(m_1-1)\tau_1} & x_{1,T_1}^2 & \dots & x_{1,T_1}^d & \dots & x_{M,T_1} & \dots & x_{M,T_1-(m_M-1)\tau_M}^d \\ 1 & x_{1,T_2} & \dots & x_{1,T_2-(m_1-1)\tau_1} & x_{1,T_2}^2 & \dots & x_{1,T_2}^d & \dots & x_{M,T_2} & \dots & x_{M,T_2-(m_M-1)\tau_M}^d \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 1 & x_{1,T_k} & \dots & x_{1,T_k-(m_1-1)\tau_1} & x_{1,T_k}^2 & \dots & x_{1,T_k}^d & \dots & x_{M,T_k} & \dots & x_{M,T_k-(m_M-1)\tau_M}^d \end{bmatrix} \tag{8}$$

In order to obtain a stable solution, the number of rows in the matrix $\boldsymbol{A}$ must satisfy the given relation:

$$k \geqslant s \tag{9}$$

In this paper the polynomial order d is taken as $d=2$, and p is taken as $p=1$ in Eq. (4). Furthermore, in order to evaluate the prediction result, we use the following indices, namely the root mean squared prediction error (RMSE) and normalized mean squared prediction error (NMSE) described in Ref. [1].

3 Simulation

Consider the Lorenz system $\frac{\mathrm{d}x_1}{\mathrm{d}t} = \sigma(x_2 - x_1)$, $\frac{\mathrm{d}x_2}{\mathrm{d}t} = x_1(R - x_3) - x_2$, $\frac{\mathrm{d}x_3}{\mathrm{d}t} = x_1x_2 - bx_3$, where $\sigma = 10$, $R = 28$ and $b = 8/3$. The initial points are $x_{1,0} = 15.34$, $x_{2,0} = 13.68$ and $x_{3,0} = 37.91$ and the step length of the integral is $h = 0.04$. We use the fourth-order Runge-Kute integral method and obtain two simulated time series of x_1 and x_2 individually with 1.21×10^4 data points. To reduce the influence of transition, we abandon first 10^4 data points and only keep the last 2 100 data points. The former 2 000 data points are used to train samples and the latter 100 data points are used as a prediction data set.

For the univariate time series $\{x_{1,n}\}_{n=1}^{2\,000}$, we obtain $\tau = 4$ and $m = 3$ [4]. We reconstruct a phase space with $(x_{1,n}, x_{1,n-4}, x_{1,n-8})$. Then we get the specific one-step prediction equation:

$$\begin{aligned} x_{1,T+1} \cong f_1(\boldsymbol{V}_T) &= f_0 + f_{10}^{(1)} x_{1,T} + f_{11}^{(1)} x_{1,T-4} + f_{12}^{(1)} x_{1,T-8} + f_{200}^{(1)} x_{1,T}^2 + f_{201}^{(1)} x_{1,T}x_{1,T-4} + \\ &f_{202}^{(1)} x_{1,T}x_{1,T-8} + f_{211}^{(1)} x_{1,T-4}^2 + f_{212}^{(1)} x_{1,T-4}x_{1,T-8} + f_{222}^{(2)} x_{1,T-8}^2 \end{aligned} \tag{10}$$

For the multivariate time series $\{(x_{1,n}, x_{2,n})\}_{n=1}^{2\,000}$, we obtain $\tau_1 = \tau_2 = 4$, $m_1 = 2$, $m_2 = 1$ [4]. We reconstruct a phase space with $(x_{1,n}, x_{1,n-4}, x_{2,n})$. Then we get the specific one-step prediction equation:

$$\begin{aligned} x_{1,T+1} \cong f_1(\boldsymbol{V}_T) &= f_0 + f_{10}^{(1)} x_{1,T} + f_{11}^{(1)} x_{1,T-4} + f_{200}^{(1)} x_{1,T}^2 + f_{201}^{(1)} x_{1,T}x_{1,T-4} + \\ &f_{211}^{(1)} x_{1,T-4}^2 + f_{10}^{(2)} x_{2,T} + f_{200}^{(2)} x_{2,T}^2 \end{aligned} \tag{11}$$

The number of the nearest neighbors is 12 (i. e. $k = 12$). The prediction errors are shown in Tab. 1.

Tab. 1 One-step prediction errors of the last 100 data points

Method	Time series	k	$\varepsilon_{\text{RMSE}}$	$\varepsilon_{\text{NMSE}}$
Local polynomial Prediction	Univariate	12	7.1×10^{-3}	$8.031\,4 \times 10^{-7}$
	Multivariate	12	1.2×10^{-3}	$2.294\,1 \times 10^{-8}$

From Tab. 1, we can see that the predicted results with the multivariate local polynomial prediction method are in extremely good agreement with the real time series, and the results are better than those with the univariate local polynomial prediction method.

4 Application

As is well known, data of the stock market, e. g. , stock prices, often shows greatly complicated behavior, therefore it is very difficult to predict its movement accurately. In order to set up a good prediction model about such financial indices, to seek a suitable variable affecting price index is important. In Ref. [8], the authors have proven the chaos of Shanghai securities market, so we can apply the chaotic prediction method given in this paper to the stock market. In this paper, as well as the composite index, the Chang index is considered, since it is one of the key factors influencing the dealer's mind.

We selected the Shanghai composite index and the Chang index as two variants and we used the multivariate local polynomial prediction method to predict the Shanghai composite index. We selected 1 069 data points from the Shanghai composite index and the Chang index individually from December 16, 1996 to May 28, 2001, and we denote the Shanghai composite index (SCI) by P and the Change index by ΔP. From this we have the two variants time series $\{P_n\}_{n=1}^{1\,069}$ and $\{\Delta P_n\}_{n=1}^{1\,069}$ (where $\Delta P_n = P_n - P_{n-1}$).

In order to reduce noise, we use the linear logarithm diminish (LLD) method to $\{P_n\}_{n=1}^{1\,069}$ and get the time series $\{x_{1,n}\}_{n=1}^{1\,069}$, where $x_{1,n} = \ln P_n - (a + bn)$. We denote $\{\Delta P_n\}_{n=1}^{1\,069}$ as $\{x_{2,n}\}_{n=1}^{1\,069}$. In order to compare the prediction result, we reconstructed the attractor in the following two ways. The first way is to reconstruct the attractor with the univariant local polynomial prediction method. By the mutual average information method and false nearest neighbor method[6], we get $\tau = 35$ and $m = 12$. So the reconstructed phase space is $\boldsymbol{V}_n = \{x_{1,n}, x_{1,n-\tau}, x_{1,n-2\tau}, \dots, x_{1,n-11\tau}\}$.

The second way is to reconstruct the attractor with the multivariate local polynomial prediction method. From the mutual average information method and the method in Ref. [5], we get $\tau_1 = 35$, $\tau_2 = 5$, $m_1 = 5$, $m_2 = 1$. So the reconstructed phase space is $\boldsymbol{V}_n = \{x_{1,n}, x_{1,n-\tau_1}, x_{1,n-2\tau_1}, x_{1,n-3\tau_1}, x_{1,n-4\tau_1}, x_{2,n}\}$.

For the univariate time series $\{x_{1,n}\}_{n=1}^{1\,069}$ or the two variant time series $\{x_{1,n}, x_{2,n}\}_{n=1}^{1\,069}$, the former 1 000 data points are used as training samples and the latter 69 data points are used as the prediction data set. The number of the nearest neighbors is 100 (i. e. $k = 100$). The prediction errors are shown in Tab. 2.

Tab. 2 One-step prediction errors of the last 69 data points

Method	Time series	k	$\varepsilon_{\text{RMSE}}$	$\varepsilon_{\text{NMSE}}$
Local polynomial Prediction	Univariate	100	2.621 9	4.91×10^{-2}
	Multivariate	100	7.63×10^{-2}	8.4×10^{-3}

From Tab. 2, we also can see that the predicted results with multivariate local polynomial prediction method are in extremely good agreement with the real time series' as shown in Fig. 1, and the results are better than the ones from the univariate local polynomial prediction method. So it can be said that the prediction with the multivariate local polynomial prediction method is successful.

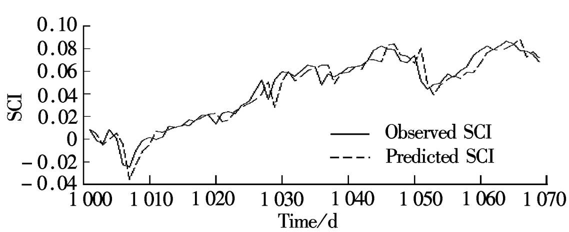

Fig. 1 Observed and predicted SCI of the last 69 data points

5 Conclusion

In this paper, we present a local polynomial prediction method of multivariate chaotic time series. All errors show that this prediction method is better than the univariate one. Therefore, in practical applications, when multivariate time series data are available, we can use the proposed method to obtain better predictions.

References

[1] Wang Haiyan, Zhu Mei. A prediction comparison between univariate and multivariate chaotic time series [J]. *Journal of Southeast University* (*English Edition*), 2003, **19**(4): 414 − 417.

[2] Farmer J D, Sidorowich J J. Exploiting chaos to predict the future and reduce noise [A]. In: Lee Y C, ed. *Evolution*, *Learning and Cognition*[C]. River Edge: World Scientific, 1987. 277 − 330.

[3] Kasim K, Levent S, Orhan S. Nonlinear time series prediction of O_3 concentration in Istanbul [J]. *Atmospheric Environment*, 2000, **34**(8): 1267 − 1271.

[4] Wang Haiyan, Sheng Zhaohan, Zhang Jin. Phase space reconstruction of complex system based on multivariate time series [J]. *Journal of Southeast University* (*Natural Science Edition*), 2003, **33**(1): 115 − 118. (in Chinese)

[5] Cao Liangyue, Mee A, Judd K. Dynamics from multivariate time series [J]. *Physica D*, 1998, **121**(1, 2): 75 − 88.

[6] Kantz H, Schreiber T. *Nonlinear time series analysis* [M]. Cambridge: Cambridge University Press, 1997. 127 − 133.

[7] Fraser A M, Swinney H L. Independent coordinates for strange attractors from mutual information [J]. *Phys Rev A*, 1986, **33**(2): 1134 − 1140.

[8] Gao Hongbing, Pan Jin, Cheng Hongmin. Judgment of chaos in our securities market [J]. *System Engineering*, 2001, **18**(6): 28 − 32. (in Chinese)

多变量混沌时间序列局部多项式预测方法及应用

方　芬　　王海燕

(东南大学经济管理学院, 南京 210096)

摘要: 为了改善混沌时间序列的预测精度, 提出了一种新的多变量混沌时间序列的局部多项式预测方法. 它首先利用多变量时间序列的相空间重构理论重构相空间, 并据此利用多项式函数构造预测模型, 该模型根据嵌入维数构造数据矩阵, 进行模型的参数估计和计算一步预测值, 最后根据平均根统计量推断预测效果. Lorenz 系统的模拟仿真和上海综合股价指数的局部预测结果表明: 用多变量混沌时间序列局部多项式预测法进行预测的误差小, 且比单变量混沌时间序列局部多项式预测法的预测精度高.

关键词: 混沌时间序列; 相空间重构; 局部多项式预测; 证券市场

中图分类号: O175; O241