Local Prediction Model of Chaos Network Traffic Based on Orthogonal Polynomials

Lei Ting
College of Science, Beijing Forestry University, Beijing 100083, China

Xie HuiYang*
College of Science, Beijing Forestry University, Beijing 100083, China

Feng ShuangLei
China Electric Power Research Institute, Beijing 100192, China

Wang HongQing
College of Science, Beijing Forestry University, Beijing 100083, China

*Corresponding author: Xie HuiYang (email: yangrtjw@126.com)

Abstract
When the local prediction model of chaos traffic sequence is at high order, its normal equations usually produce ill-conditioned phenomenon which reduces performance of the model and even cause serious distortion. Combined with phase space reconstruction theory, this paper aims to propose a high-order local prediction model to eliminate ill-conditioning and reduce calculation workload based on orthogonal polynomials, and solves the conflict between prediction accuracy and computation complexity to a certain extent, the usefulness of robustness to the number of neighbor points in phase space is also considered. Chaos characteristics of the time series from actual network traffic were analyzed by Lyapunov exponent. On this basis, the evolution trend of the time series was predicted by local prediction model, the prediction results of new model and traditional model are also compared, and the effectiveness of new model is supported by these results.
1. Introduction

Many empirical studies confirmed that traffic in Internet which is a complex system has abundant chaos behavior characteristics [1-5]. Therefore, prediction method based on chaos theory is more suitable to study problems concerning network traffic prediction [6-11].

The research on forecasting method of chaotic time series is the important part of the nonlinear chaotic dynamic systems. In this field, many prediction algorithms such as local prediction model based on phase space reconstruction can be found in literature [10-15]. Among them, local prediction model attracts increasing attentions for its good approximation function, which often uses polynomial to construct the approximation function [10-15]. Generally speaking, high-order model has better prediction effect compared to low-order one. However, high-order model has heavy calculation workload and its normal equations are often morbid, with poor robustness. The polynomial orders that could be approximated are strictly limited by round-off error. It will generate unstable numbers to any algorithms and often produce meaningless results [14-16].

According to the problems of traditional local prediction model, the aim of this paper is to provide a high-order local model for chaos network traffic prediction based on orthogonal polynomials. Due to the good mathematical characteristics of orthogonal polynomials, this model could improve prediction accuracy, and reduce computation complexity, eliminate ill-conditioning of normal equations, therefore it get better prediction results than traditional methods, and it is robust to the number of neighbor points in phase space.

2. Theory of phase space reconstruction

To establish a prediction model of chaos time sequence, phase space reconstruction of the time sequence is the primary task, which is to “restore” the geometric structure of the phase space of complex system. Its theoretical basis is the Takens theorem.

Takens theorem [17]: Set a time sequence \( \{x(k)\}_{k=1,2,\ldots} \). Given delay time \( \tau \) and embedding dimension \( m \), some smooth mapping \( F: \mathbb{R}^n \to \mathbb{R}^m \) could be found on the attractor, which meets \( Y(t+1) = F(Y(t)) \), where \( Y(t) \) is a \( m \)-dimensional vector.

\[
Y(t) = (x(t), x(t-\tau), \ldots, x(t-(m-1)\tau),
\]

The Takens theorem reveals that orbit in the reconstructed phase space is equivalent to the original chaos system under diffeomorphism significance. Hence, the reconstructed phase space could maintain same geometric characteristics and information with the original dynamic system. This enables us not only to comprehend correlation of variables of the chaos system, but also study the dynamic behaviors of the system based on partial information.

3. Establishment of orthogonal local prediction model

Here, the local prediction model of chaos time sequence was established based on theory of phase space reconstruction.

Suppose \( \{x(k)\}_{k=1,2,\ldots} \) is a known time sequence. Its future evolution \( x(n+1), x(n+2), \ldots \) is predicted with the established local model.

Firstly, make phase space reconstruction of this time sequence. According to the Takens theorem, \( N = n-(m-1)\tau \) phase points of this time sequence in the \( m \)-dimensional reconstructed phase space could be gained by choosing appropriate delay time \( \tau \) and embedding dimension of phase space (m): \( Y(1) = (x(1), x(1+\tau), \ldots, x(1+(m-1)\tau)) \), \( Y(2) = (x(2), x(2+\tau), \ldots, x(2+(m-1)\tau)) \), \ldots, \( Y(N) = (x(N), x(N+\tau), \ldots, x(n)) \).

Later, find out the mapping \( F \) in the Takens theorem. Point \( Y(N+1) \) in the reconstructed phase space could be predicted from \( Y(N+1) = F(Y(N)) \) by using historical data. In this
way, the predicted value of time sequence \(x(n+1)\) could be gained, because it is the last component of phase point \(Y(N+1)\). Repeat and iterate continuously, and \(x(n+1), x(n+2), \ldots\) could be predicted. This is the ideal of modeling based on phase space reconstruction.

Therefore, the key of time sequence prediction is to get a good approximation of mapping \(F\). Since chaos behavior only occurs under nonlinearity of \(F\), it is necessary to find a nonlinear expression of \(F\). High-order polynomial fitting is the first choice, because any continuous phenomena, we can make full use of good mathematical characteristics of orthogonal polynomials to establish the high-order local prediction model of chaos traffic time sequence.

For the given time sequence \(\{x(k), k=1,2,\ldots\}\), polynomial sequence \(\{U_m(x), m=1,2,\ldots,q\}\) is chosen, where \(U_m(x)\) is the \(m\)-order polynomial about \(x\). The inner product is defined and shall meet:

\[
\{U_m, U_j\} = \sum_{i=1}^{q} w(k) U_m(x(k)) U_j(x(k)) = \begin{cases} A_{mj}, & m=j, A_{mj} > 0 \\ 0, & m \neq j \end{cases}
\]

Under this circumstance, \(\{U_m(x), m=1,2,\ldots,q\}\) is obviously the orthogonal polynomial sequence of \(w(k)\) weighted time sequence \(\{x(k), k=1,2,\ldots\}\).

Model (2) could be rewritten as:

\[
x(k+1) = \beta_0 + \beta_1 U_1[x(k)] + \beta_2 U_2[x(k)] + \cdots + \beta_q U_q[x(k)],
\]

\(i=1,2,\ldots,M\)

Comparison between model (3) and (2) reveals that \(b_i\) and \(\beta_i\) \((i=1,2,\ldots,q)\) have a simple function relationship. Therefore, \(b_0,b_1,\ldots,b_q\) could be calculated if \(\beta_0,\beta_1,\ldots,\beta_q\) is known.

View

\[
x(k+1) = \beta_0 + \beta_1 U_1[x(k)] + \beta_2 U_2[x(k)] + \cdots + \beta_q U_q[x(k)]
\]

as the multivariate function about \(\beta_0,\beta_1,\ldots,\beta_q\) and calculate them by weighted least square method. The square sum of model prediction error is:

\[
e(\beta_0, \beta_1, \ldots, \beta_q) = \sum_{i=1}^{q} w(k) \left[ \sum_{j=0}^{q} \beta_j U_j[x(k)] - \sum_{j=0}^{q} \beta_j U_j[x(k)] \right]^2
\]

Where \(w(k)\) is the weight of neighboring phase point \(Y(k)\). According to the principle of least square method, coefficients \(\beta_0, \beta_1, \ldots, \beta_q\) shall be able to get minimum \(e(\beta_0, \beta_1, \ldots, \beta_q)\). Hence, calculate partial derivatives about \(\beta_0, \beta_1, \ldots, \beta_q\) and make them all equal to zero:

\[
\frac{\partial e}{\partial \beta_j} = 2 \sum_{i=1}^{q} w(k) \left[ \sum_{j=0}^{q} \beta_j U_j[x(k)] \right] U_j[x(k)] = 0
\]

\(i=0,1,\ldots,q\)

Then,

\[
\sum_{n=0}^{M} \sum_{i=1}^{q} w(k) U_n[x(k)] U_i[x(k)] = \sum_{i=1}^{q} w(k) x(k+1) U_i[x(k)]
\]

\(i=0,1,\ldots,q\)
Information technologies

This is the normal equations of \( q+1 \) unknowns and \( q+1 \) equations. According to the definition of inner product, normal equations could be simplified as:

\[
\sum_{m=0}^{q} \beta_m \{ U_m, U_j \} = \{ x, U_j \}
\]

Let

\[
U = \begin{bmatrix}
\{ U_0, U_0 \} & \{ U_0, U_1 \} & \cdots & \{ U_0, U_p \} \\
\{ U_1, U_0 \} & \{ U_1, U_1 \} & \cdots & \{ U_1, U_p \} \\
\vdots & \vdots & \ddots & \vdots \\
\{ U_p, U_0 \} & \{ U_p, U_1 \} & \cdots & \{ U_p, U_p \}
\end{bmatrix},
\]

\[
C = \begin{bmatrix}
\{ x, U_0 \} \\
\{ x, U_1 \} \\
\vdots \\
\{ x, U_q \}
\end{bmatrix}, \quad \beta = [\beta_0, \beta_1, \cdots, \beta_q]^T
\]

Then, get the matrix form of normal equations:

\[
U\beta = C
\]

(4)

Under reversible \( U \), the unique solution

\[
\beta = [\beta_0, \beta_1, \cdots, \beta_q]^T = U^{-1}C
\]

to the equations could be gained. Then, model coefficients \( b_0, b_1, \cdots, b_q \) could be calculated.

If \( U_m(x) \) is the common polynomial, elements in matrix \( U \) may not be zero. At this moment, the normal equations are morbid and have heavy calculation workload. Since \( U_m(x) \) is an orthogonal polynomial, it can be known from orthogonal nature that elements in the matrix \( U \) except for diagonal are all zero, thus making \( U \) a diagonal matrix. Therefore, the equations (4) will be simplified as:

\[
\begin{bmatrix}
\{ U_0, U_0 \} & 0 & \cdots & 0 \\
0 & \{ U_1, U_1 \} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \{ U_q, U_q \}
\end{bmatrix}\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_q
\end{bmatrix} = \begin{bmatrix}
\{ x, U_0 \} \\
\{ x, U_1 \} \\
\vdots \\
\{ x, U_q \}
\end{bmatrix}
\]

The calculation workload is reduced significantly and ill-conditioning of the normal equations as well as coefficient correlation is eliminated.

However, there still have two problems that have to be solved: how to construct weighted orthogonal polynomial series \( \{ U_m(x(k)), m = 1, 2, \cdots, q \} \) and how to determine weight function \( w(k) \) of neighbor points \( Y(k) \).

With respect to construction of polynomial sequence \( \{ U_m(x(k)), m = 1, 2, \cdots, q \} \), based on the idea of Schmidt orthogonalization, let \( U_0[x(k)] = 1 \),

\[
U_r[x(k)] = (x(k) - \rho_r) U_0[x(k)],
\]

\[
U_{s+1}[x(k)] = [x(k) - \rho_{s+1}] U_s[x(k)] - \lambda_s U_{s+1}[x(k)]
\]

Where

\[
\rho_{s+1} = \frac{\sum_{i=1}^{M} w(i) x(i) U_s^2[x(i)]}{\sum_{i=1}^{M} w(i) U_s^2[x(i)]},
\]

\[
\lambda_s = \frac{\sum_{i=1}^{M} w(i) U_s^2[x(i)]}{\sum_{i=1}^{M} w(i) X^2_{s-1}[x(i)]}
\]

Therefore, it is easy to confirm that the gained \( \{ U_m(x(k)), m = 1, 2, \cdots, q \} \) meets orthogonality conditions.

Finally, weights \( w(k) \) of neighboring phase points \( Y(k) \) are determined. Prediction accuracy of model depends more on neighbor points closer to the central point, because they carry more information about prediction center. The model shall consider this important factor. Hence, weight function of neighbor points \( Y(k) \) shall be constructed based on distance:

\[
w(k) = e^{-\frac{1}{2} \sum_{i=1}^{M} \rho_i^{-1}}
\]

Where, \( R_i \) denotes distance between \( Y(N) \) and \( Y(k) \). This weight function depicts different effects of different neighbor points on dynamic behaviors at the prediction center. Neighbor points closer to the prediction center influence the prediction center more significantly.

Now, the high-order local prediction model of network traffic sequence is established.

4. Numerical simulation

In this part, chaos behavior characteristics of actual traffic sequence is analyzed. On this basis, validity of the established model is verified through numerical simulation.

Numerical simulation data comes from visit traffic of INNP server in the USENET system from January 1st, 2006 to March 31st (the URL is http://newsfeed.ntcu.net/~news/2006/). Sampling time granularity is hour and 24h per day. A total of traffic records of 2,160h were collected. We established the high-order orthogonal local model with the first 2,135 data and rest 25 data were used to testing prediction effect of the model. The time sequence composed by the first 2,135 data is \( \{ x(t), t = 1, 2, \cdots, 2135 \} \) (Fig.1). Now, let’s discuss the chaos behavior characteristics of this sequence.

Chaos appears an irregular movement and is a random-like behavior that doesn’t need to add any random factors in the determinant complex
system. Chaos degree of a system can be depicted by the maximum Lyapunov exponent $\lambda$ which is a quantity that characterizes the rate of separation of infinitesimally close trajectories for a dynamical system. Quantitatively, two trajectories in phase space with initial separation $d(0)$ diverge at a rate given by $|d(t)| = e^{\lambda t} |d(0)|$.

A positive $\lambda$ is taken as an indication that the time series is chaotic, and the larger the $\lambda$ is, the higher the chaos degree will be [14-5].

In this paper, whether this traffic sequence has chaos characteristics was determined from qualitative analysis based on phase diagram of attractor and images of Poincare section [14-16]. Attractor is a kind of abstract mathematical model. It is the set of infinite points in phase space. Here, we displayed attractors of actual traffic sequence in a three-dimensional phase diagram (Fig.2). Attractor orbits of flow sequence gather together in the three-dimensional reconstructed phase space and has evident characteristics of rough and irregular geometric structure as well as concentration. This is the intuitive representation that network has chaos behavior characteristics.

The Poincare section of network traffic on two-dimensional plane is shown in Fig.3. Poincare section has fractal geometric structure, that is, similar overall and partial geometric structure. This confirms that this traffic sequence may have chaos characteristics from another perspective.

In the following text, Lyapunov exponent $\lambda$ of the traffic sequence is calculated. Hence, phase space of the sequence is reconstructed firstly. It can be seen from the Taken theorem that the phase space reconstruction needs two important parameters: embedding dimension $m$ and delay time $\tau$. Here, $m$ could be determined through GP algorithm [18]. As shown in Fig.4, the horizontal axis is embedding dimension and the vertical axis is mutual information, (0) is taken as an indication that diverge at a rate $\tau = 14$.

A total of 5 neighboring phase points in the reconstructed phase space which are the closest to the prediction center were chosen as fitting reference points. The corresponding orthogonal local model when $q = 2$ and 3 was established. To compare prediction effect, the traffic sequence was also predicted by using the traditional local linear model.

To evaluate prediction effect of the model, relative error of prediction points was defined:

$$E = \frac{|x(t) - \hat{x}(t)|}{|x(t)|}$$

The regularization mean square error (RMSE) defined to describe the overall prediction performance of the model is:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} (x(t) - \hat{x}(t))^2}{\sum_{t=1}^{N} (x(t))^2}}$$

Where $\hat{x}(t)$ and $x(t)$ are mean value and predicted value of the sequence.

Prediction effect is presented in Fig.7. Obviously, prediction results of models are close to the actual value of traffic. Proportions of relative error of models in prediction points within a certain interval and mean prediction relative error of the models are listed in Table 1. Compared to traditional model, the orthogonal local model shows higher proportion of prediction points whose relative error falls within the interval of small value. This implies that model improvement indeed increases prediction performance of the model.

Mean prediction square errors of models are exhibited in Table 2. Compared to traditional model, second-order and third-order orthogonal models show significantly higher prediction accuracy. Although the third-order orthogonal model has little higher prediction accuracy, no essential difference of prediction effect has been observed between them. This indicates that we could achieve good approximation effect of traffic sequence by using the third-order orthogonal local model.

Information technologies
Information technologies

Figure 1. Network traffic sequence

Figure 2. Phase diagram of traffic sequence attractor in three-dimensional reconstructed space

Figure 3. Poincaré section of network traffic

Figure 4. Determining embedding dimension of the reconstructed phase space

Figure 5. Determining delay time through mutual information method

Figure 6. Calculate Lyapunove index of traffic sequence

Figure 7. Prediction effect of models and comparison

Table 1. Relative error intervals of prediction models

<table>
<thead>
<tr>
<th>Error interval</th>
<th>Prediction points in different error intervals</th>
<th>Traditional model</th>
<th>Second-order orthogonal model</th>
<th>Third-order orthogonal model</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.01</td>
<td></td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>[0.01,0.02)</td>
<td></td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>[0.02,0.03)</td>
<td></td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>[0.03,0.04)</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>≥0.04</td>
<td></td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Mean relative error</td>
<td></td>
<td>0.02168</td>
<td>0.01749</td>
<td>0.01648</td>
</tr>
</tbody>
</table>
Robustness of prediction models is discussed in the following text based on amount of neighbor points (M), an important parameter in modeling. For a robust local model, its prediction performance shall be not very sensitive to M. Fig.8 shows variation of RMSE of prediction models when M increases from 30 to 55 at a growth rate of 5. When M decreases, RMSE of all prediction models increases. This is because when neighbor points reduce, information of dynamic behaviors at prediction center contained in them attenuates. Such change is more obvious in traditional model than rest two orthogonal models. In other words, as M decreases, RMSE of the traditional model increases more quickly. When M decreases to 40, its RMSE has increased greatly. On the contrary, RMSE of the second-order and third-order orthogonal models increases slowly. Hence, we can draw to the conclusion that orthogonal model has higher robustness compared to the traditional model. This also inspires us that model prediction accuracy shall be improved by reducing prediction error according to the basic modeling principle and method rather than simply increasing neighbor points.

Acknowledgements
This work was supported by national natural science foundation of China (Grant No. 61370193)

References
Information technologies


