Vol 18 No 6, June 2009
1674-1056/2009/18(06)/2194-06
Chinese Physics B
© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

Small-time scale network traffic prediction based on a local support vector machine regression model*

Meng Qing-Fang(孟庆芳)[a)b)], Chen Yue-Hui(陈月辉)[a)], and Peng Yu-Hua(彭玉华)[b)]

[a)] *School of Information Science and Engineering, University of Jinan, Jinan 250022, China*

[b)] *School of Information Science and Engineering, Shandong University, Jinan 250100, China*

(Received 30 October 2008; revised manuscript received 30 November 2008)

In this paper we apply the nonlinear time series analysis method to small-time scale traffic measurement data. The prediction-based method is used to determine the embedding dimension of the traffic data. Based on the reconstructed phase space, the local support vector machine prediction method is used to predict the traffic measurement data, and the BIC-based neighbouring point selection method is used to choose the number of the nearest neighbouring points for the local support vector machine regression model. The experimental results show that the local support vector machine prediction method whose neighbouring points are optimized can effectively predict the small-time scale traffic measurement data and can reproduce the statistical features of real traffic measurements.

Keywords: network traffic, small-time scale, nonlinear time series analysis, support vector machine regression model

PACC: 0545

1. Introduction

Network traffic analysis and modeling play a major role in characterizing network performance, so they have aroused the interest of many researchers. Models that accurately capture the salient characteristics of the traffic are useful for analyzing and simulating network dynamics, and they facilitate further our understanding of network dynamics, so they have a fundamental significance for designing networks and solving engineering problems, e.g., the traffic balance scheme, router, switcher designing, the manage devices, and relevant support software development.

Complexity is a key issue in network geometry and information traffic. Evidence of traffic complexity appears in many forms, such as the long-range correlations and self-similarity found in the statistical analysis of traffic measurements. There is also strong evidence of these phenomena on several different time scales. The complexity revealed from the traffic measurements leads to the fact that the network traffic cannot be analyzed in the framework of available traffic models.[1−3] Alternative reliable traffic models and tools for quality assessment and control should be developed.[4,5]

Recently, communication and network technologies have been developed rapidly, which causes the traffic characteristics to change greatly. The research emphasis on the network traffic analysis and modeling has switched from the large-time scale to the small-time scale. It is shown that the traffic characteristics of the small-time scale are different from those of the large-time scale.[6,7] So the large-time scale network traffic models are unsuitable for small-time scale network traffic.

Nonlinear time series analysis, which has been extensively studied in the past two decades,[8−24] has provided a new idea and method to deal with complex phenomena and irregular time series. There is growing evidence that the nonlinear deterministic component exists in the network traffic time series.[25,26] In the present paper we apply systematically the nonlinear time series analysis method to the traffic measurement data. The experimental results show that the nonlinear time series analysis method can be successfully used for further understanding the main features of traffic data.

*Project supported by the National Natural Science Foundation of China (Grant No 60573065), the Natural Science Foundation of Shandong Province, China (Grant No Y2007G33), and the Key Subject Research Foundation of Shandong Province, China (Grant No XTD0708).

2. Nonlinear analysis of small-time scale network traffic data

The reconstructed phase space, which is equivalent to the original phase space of a system from a scale time series, is the basis of nonlinear time series analysis.[8] Based on Whitney's topological embedding theorem, Takens[9] has proved the delay embedding theorem which offers the possibility of reconstructing m-dimensional dynamics from a scale time series in the absence of noise. However, the delay embedding theorem does not directly show how to choose its parameters-embedding dimension and delay time. So the key to performing the phase space reconstruction from a real time series is to determine its optimal parameters. Suppose that an observed scalar time series is $\{x_i\}$, $i = 1, 2, \ldots N$. According to Takens' delay embedding theorem, the embedding of the strange attractors can be obtained by reconstructing the delay vectors $X(n) = [x_n, x_{n-\tau}, \ldots, x_{n-(m-1)\tau}]^T$, $n = (m-1)\tau+1, (m-1)\tau+2, \ldots, N$, where m is the embedding dimension and τ is the delay time.

There have been many discussions on how to determine the embedding dimension from a scalar time series.[10−16] There are four methods that are usually used. They are the computing invariant method,[10,11] the singular value decomposition method,[12] and the false neighbour method,[13,14] which have some disadvantages, such as inapplicability in a short time series, sensitivity to noise, time-consuming in computation, subjective criterion, etc., and the method based on nonlinear prediction,[15,16] which has some advantages, such as applicability in a short time series, stable to noise, computationally efficient, and no parameter introduced purposely. The last method is based on the system predictability. Increasing the embedding dimension of the nonlinear prediction model parameterized by the embedding dimension, one notes that the optimal embedding dimension occurs when the prediction error stops drastically declining. Reference [15] is based on a local linear prediction model and reference [16] is based on a global polynomial prediction model, which optimize both parameters: the embedding dimension and the nonlinear degree, at the same time. In the present paper, according to a local nonlinear prediction model we determine the embedding dimension from the traffic measurement data from the following equation:

$$C(m) = \ln \varepsilon(m)^2 + 2m/N, \tag{1}$$

where

$$\varepsilon(m)^2 = \frac{\sum_{n=m+1}^{N} [\hat{x}_n(m) - x_n]^2}{\sum_{n=m+1}^{N} [x_n - \bar{x}]^2}. \tag{2}$$

Here x_n $(n = 1, 2, \ldots, N)$ are the actual values, $\hat{x}_n$ $(n = 1, 2, \ldots, N)$ are the prediction values, and $\varepsilon(m)^2$ is the normalized variance of the prediction error.

Predicting the time evolution of a system from observations is one of the most important classical problems in time series analysis. With the development of chaos theory and research on its applications, various nonlinear time series prediction methods have been proposed to predict irregular time series,[17−23] which can be grouped into two major classes: global and local. The global methods use the whole time series to model the true function of the whole attractor. It is clear that methods of this kind have the disadvantage that if new information is taken into account then all the parameters involved in the models may change, and that a long parameter estimation time may be required. At the same time, the true function of the whole attractor is very difficult to model. The local methods[17−23] overcome this drawback by building a model only on the local attractor and utilizing only part of the past information. Farmer and Sidorowich have proved that the prediction performance from the local methods is more superior than that obtained from the global methods with the same embedding dimension.[17] So the local prediction methods have received considerable attention and become one of the most popular methods to predict nonlinear time series.

The first step of the local prediction is to find the neighbouring points of the current state $X(n)$ in the reconstructed phase space. According to the reconstructed trajectory, we first compute the distances between the current delay vectors $X(n)$ and its $n-1$ preceding delay vectors $X(i)$ $(i = 1, 2, \ldots, n-1)$ with an imposed metric,

$$d(i) = \|X(i) - X(n)\|_2, \tag{3}$$

and then find the k nearest neighbouring points $X(n_i)$ $(i = 1, 2, \ldots, k)$ of the current state $X(n)$. The local linear prediction model in an m-dimensional delay embedding space is an m-order autoregressive model, and the prediction model is expressed as a linear superposition of the m elements of delay vector, i.e.

$$\hat{x}_{n+T} = a_n X(n) + b_n. \tag{4}$$

The coefficients a_n and b_n are determined by using the least squared criterion to minimize

$$\min \sum_{i=1}^{k} |x_{n_i+T} - a_n X(n_i) - b_n|^2. \tag{5}$$

In the local linear prediction method, linear mapping is used to approximate the evolvement function of the current state $X(n)$ and the future value $\hat{x}_{n+T}$, and also the neighbouring points $X(n_i)$ $(i = 1, 2, \ldots, k)$ of $X(n)$ to approximate this linear mapping.

3. Local prediction and the nearest neighbouring points

Research on the local methods is continuing. But most researchers have proposed that the local nonlinear model has a similar performance to that of the local linear model and the linear model is usually used in the local prediction method.[17] However, in recent years researchers have tried to use the nonlinear model to replace the linear model in the local prediction method. For example, in Refs.[22] and [23] the support vector machine regression model was used to replace the linear model and the local support vector machine prediction method was suggested where the neighbouring points were used to train this local support vector machine regression model. Namely, we train the support vector machine, where the input is $X(n_i)$ $(i = 1, 2, \ldots, k)$ and the output is x_{n_i+T} $(i = 1, 2, \ldots, k)$, and obtain the corresponding support vector and weighting coefficient, then exploit this local support vector to compute the prediction value $\hat{x}_{n+T}$ of the future value.

The number of nearest neighbouring points is an important parameter for the local prediction method, which has important influence on the prediction accuracy and computation complexity of the local model. To our knowledge, under the local linear constraint the local nonlinear model may have a similar performance to that of the local linear model. In the present paper, we suggest that in order to take the nonlinear fitting advantage of the nonlinear model in local prediction and to improve the prediction accuracy of the local nonlinear model, we should choose the appropriate neighbouring points for the local prediction method. The local nonlinear model is not restricted by the local linear constraint. So the number of neighbouring points of the local nonlinear model can be larger than that of the local linear model. But if the number of neighbouring points are too large, the computation complexity will be increased and the model will be overfitting. So we should appropriately choose the number of neighbouring points for the local nonlinear prediction model.

In the present paper, we use a local support vector machine regression model to predict the traffic data, and we adopt the neighbouring point selection method,[24] which is based on an information criterion, to choose the number of nearest neighbouring points for the local support vector machine prediction model, which can be formulated as follows:

$$N(k) = \ln \varepsilon(k)^2 + (k+1)\frac{\ln N}{N}, \tag{6}$$

where $\varepsilon(k)^2$ is the normalized variance of the prediction error.

4. Experimental results and analysis

In this paper we use the TCP traffic data which is issued by the Lawrence Berkeley Laboratory. These traffic data contain a one-hour wide-area traffic between the Digital Equipment Corporation and the rest of the world. The data package used in this paper is DEC-Pkt1 in millisecond time precision.[27] The traffic data aggregated with time bin 0.1 s, that is, the arrived package numbers distributed within the 0.1 s time interval, are shown in Fig.1.

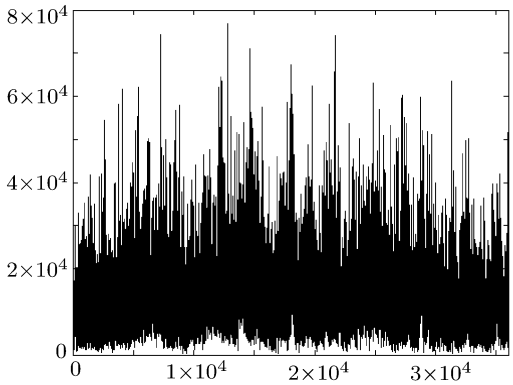


Fig.1. Traffic measurement data aggregated within 0.1 s.

Generally we may consider the traffic measurements as a sum of a regular process and a stochastic part, related to the high-frequency noise. The elimination of the noisy part may simplify the analysed time series. In order to separate the regular component of the dynamical process from the stochastic noise component, we apply the wavelet soft threshold noise reduction method to the traffic measurement data. The difference between the original time series and the filtered signal corresponds to the noisy component. Figure 2 presents part of the original traffic series, the corresponding filtered signal and the noisy component. Then the filtered traffic measurement data are normalized to the interval [–1, 1].

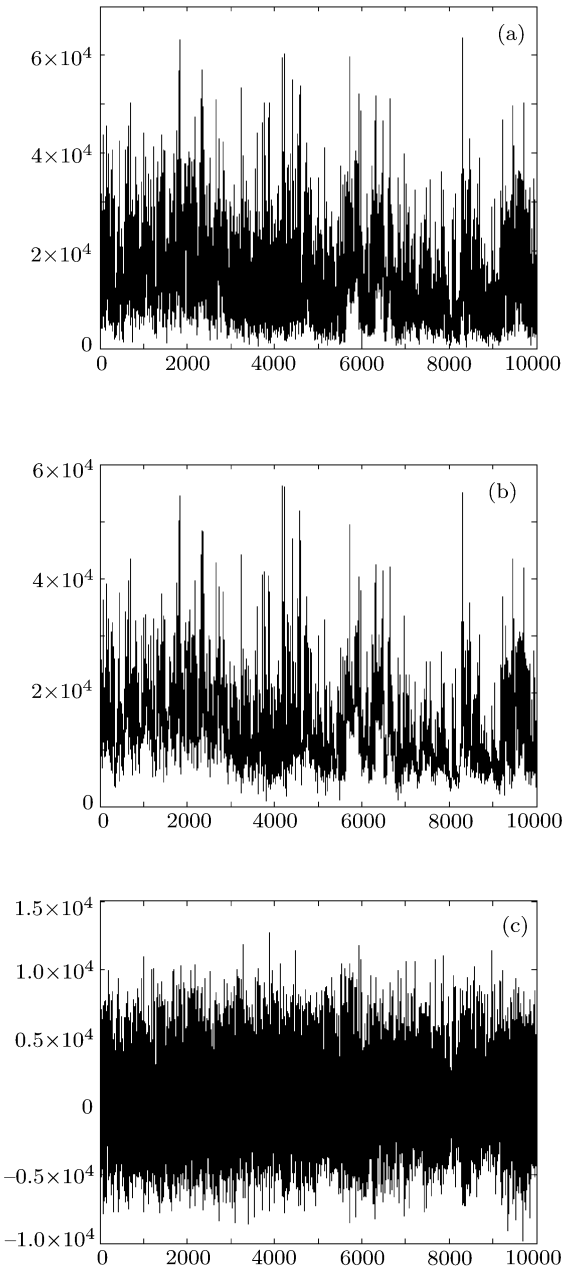


Fig.2. Wavelet soft threshold noise reduction results of traffic measurement data, showing original data (a), filtered data (b), and noisy component (c).

We use the method based on nonlinear prediction[16] to determine the embedding dimension of the traffic data. In this paper, according to the prediction error of the local support vector machine regression model the embedding dimension is determined. The results are shown in Fig.3. One can see that when $m = 5$, $C(m)$ takes a minimum value, therefore, the embedding dimension for this traffic time series should be 5. The delay time takes a value of 1 for the traffic data.

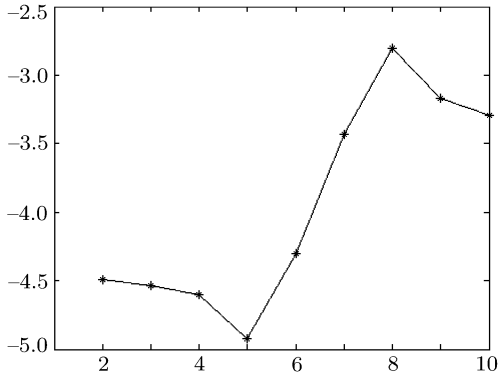


Fig.3. $C(m) - m$ for the traffic measurement data.

We use the BIC-based neighbouring point selection method[24] to choose the number of nearest neighbouring points for the local support vector machine prediction model. The results are shown in Fig.4. One can see that when $k = 16$, $N(k)$ takes a minimum value, therefore, the number of nearest neighbouring points for the local support vector machine prediction model should be 16.

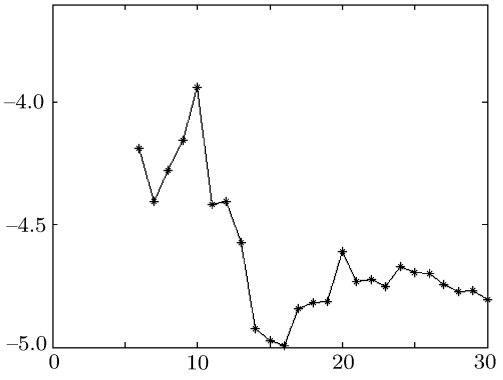


Fig.4. $N(k) - k$ for the traffic measurements data.

We use the local support vector machine prediction method to predict the traffic measurement data,

where the number of nearest neighbouring points, k, is 16. The length of this traffic measurement data is 36000. The first 33000 data points are used as the training set, and the last 3000 data points are used as the test set. The normalized mean squared error is used to evaluate the prediction accuracy,

$$\mathrm{NMSE} = \frac{(1/L)\sum_{i=1}^{L}|x_{N+i} - \hat{x}_{N+i}|^2}{(1/N)\sum_{i=1}^{N}[x_i - \bar{x}]^2}, \qquad (7)$$

where $L = 3000$. The radial basis function kernel is chosen as the kernel function of the support vector machine regression model, and the parameters are $C = 10^6$, $\varepsilon = 0.00001$, and $\sigma = 10$. The prediction results are shown in Fig.5. The statistical distribution of the traffic measurements and the prediction values are shown in Fig.6. The statistical distribution of the prediction errors is shown in Fig.7, where the abscissa represents the value of the prediction errors and the vertical axis refers to the probability of the prediction errors.

From Fig.5, it can be clearly seen that the local support vector machine prediction method, whose neighbouring points are optimized, can effectively predict the traffic measurement data, the prediction accuracy is good, and the normalized mean squared error NMSE is as small as 1.0135×10^{-2}.

From Fig.6, it can be clearly seen that the time series generated by using the local support vector machine regression model has a very similar probability distribution to that of the actual traffic measurement time series. So the local support vector machine regression model can reproduce quite well the statistical distribution of the real traffic measurement data.

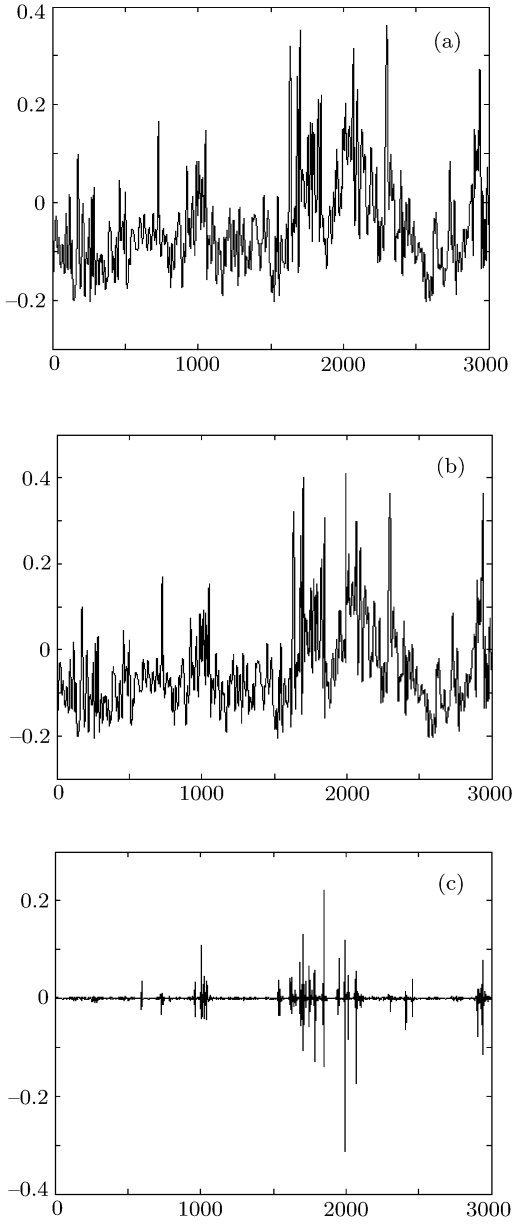


Fig.5. Prediction results of the local support vector machine prediction method for the traffic measurement data, where k=16, panel (a) shows actual values, panel (b) indicates prediction values, and panel (c) displays prediction errors with NMSE=1.0135×10^{-2}.

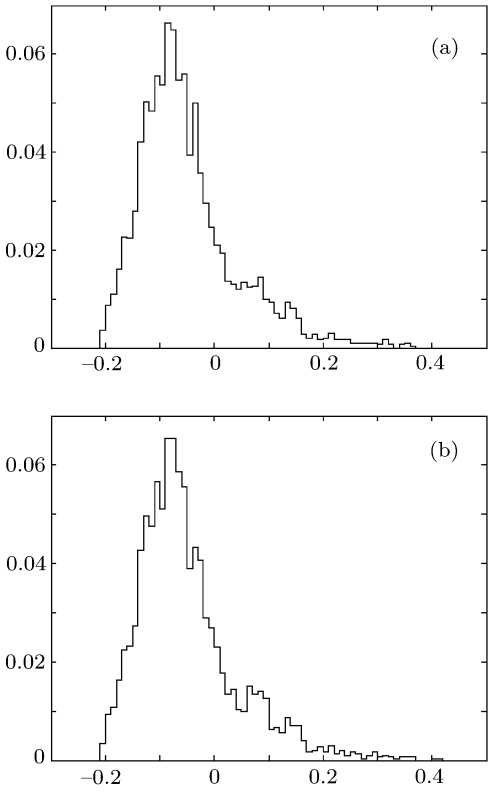


Fig.6. Statistical distributions of traffic measurement data (a) and generated time series (b).

From Fig.7, it can be clearly seen that for the traffic data, the prediction errors obtained by using

the local support vector machine prediction method mainly concentrate in the vicinity of zero, and, in general, the probabilities of the prediction errors with a large absolute value are very small.

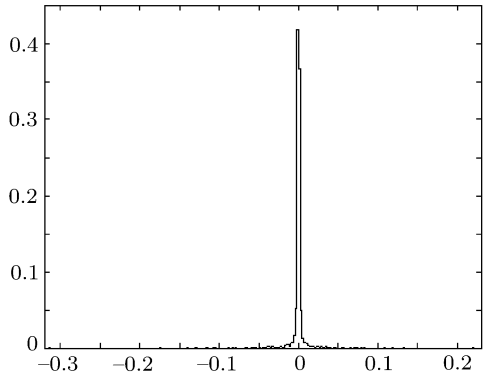


Fig.7. Statistical distribution of prediction errors.

5. Conclusion

We apply systematically the nonlinear time series analysis method to the traffic measurement data. The experimental results show that the nonlinear time series analysis method can be successfully used for further understanding the main features of the traffic data. The prediction-based method is used to determine the embedding dimension of the traffic data. Based on the reconstruction of phase space, the local support vector machine prediction method is used to predict the traffic measurement data, and the BIC-based neighbouring point selection method is used to choose the number of the nearest neighbouring points for the local support vector machine prediction model. The experimental results show that the local support vector machine prediction method whose neighbouring points are optimized can effectively predict the traffic measurement data, the prediction accuracy is good, the normalized mean squared error NMSE is very small, the prediction error mainly concentrates in the vicinity of zero, and the time series generated by the local support vector machine regression model has a very similar probability distribution to that of the actual traffic measurement time series.

References

[1] Orosz G, Krauskopf B and Wilson R E 2005 *Physica* D **211** 277
[2] Orosz G, Wilson R E and Krauskopf B 2004 *Phys. Rev.* E **70** 026207
[3] Gasser I, Sirito G and Werner B 2004 *Physica* D **197** 222
[4] Doulamis A D, Doulamis N D and Kollias S D 2003 *IEEE Trans. on Neur. Netw.* **14** 150
[5] Xie Y B, Wang W X and Wang B H 2007 *Phys. Rev.* E **75** 026111
[6] Zhang Z L, Ribeiro V J, Mooon S and Diot C 2003 *IEEE Infocom.* **3** 1826
[7] Uglig S 2004 *ACM SIGCOMM Computer Communications Review* **34**
[8] Kantz H and Schreiber T 2003 *Nonlinear Time Series Analysis* (Second Edition) (Cambridge: Cambridge University Press)
[9] Takens F 1981 *Dynamical Systems and Turbulence* [C] (Berlin: Springer) p366
[10] Xiao F H, Yan G R and HanY H 2005 *Acta Phys. Sin.* **54** 550 (in Chinese)
[11] Grassberger P and Procaccia I 1983 *Phys. Rev. Lett.* **50** 346
[12] Broomhead D S and King G P 1986 *Physica* D **20** 217
[13] Kennel M, Brown R and Abarbanel H D I 1992 *Phys. Rev.* A **45** 3403
[14] Cao L 1997 *Physica* D **110** 43
[15] Sugihara G and May R M 1990 *Nature* **344** 734
[16] Meng Q F, Peng Y H and Xue P J 2007 *Chin. Phys.* **16** 1252
[17] Farmer J D and Sidorowich J J 1987 *Phys. Rev. Lett.* **59** 845
[18] Giona M, Lentini F and Cimagalli V 1991 *Phys. Rev.* A **44** 3496
[19] Kugiumtzis D, Lingjarde O C and Christophersen N 1998 *Physica* D **112** 344
[20] Ragwitz M and Kantz H 2002 *Phys. Rev.* E **65** 056201
[21] Meng Q and Peng Y 2007 *Phys. Lett.* A **370** 465
[22] Li H C and Zhang J S 2005 *Chin. Phys. Lett.* **22** 2776
[23] Zhang J S, Dang J L and Li H C 2007 *Acta Phys. Sin.* **56** 67 (in Chinese)
[24] Meng Q F, Peng Y H, Qu H J and Han M 2008 *Acta Phys. Sin.* **57** 1423 (in Chinese)
[25] Shang P, Li X and Kamae S 2005 *Chaos, Solitons and Fractals* **25** 121
[26] Shang P, Li X and Kamae S 2006 *Phys. Lett.* A **357** 314
[27] Internet traffic archive. http://ita.ee.lbl.gov/