

混沌局域预测法在网络流量短期预测中的应用

欧阳旻^{1,2}, 张祖平¹

(1. 中南大学 信息科学与工程学院, 湖南 长沙 410083; 2. 湖南工业大学 计算机与通信学院, 湖南 株洲 412008)

摘要: 网络流量数据表现出一定的混沌特性, 而在传统的混沌预测算法中, 使用欧氏距离衡量相空间中相点间的相关性。由于欧氏距离的局限性, 在高维相空间中, 传统混沌预测算法的精度迅速下降。使用夹角余弦取代欧氏距离, 作为判别相点间相关性的标准; 并将相点视为向量, 以向量的模和夹角为优化目标, 进行预测参数识别。将以上算法应用于网络流量数据的短期预测, 结果表明, 在高维相空间中, 预测算法的精度得到了明显的提升。

关键词: 网络流量短期预测; 混沌; 局域法; 欧氏距离; 夹角余弦

中图分类号: TP393.1

文献标志码: A

文章编号: 1673-9833(2009)04-0049-04

Application of Chaotic Forecasting Algorithm to Network Traffic Forecasting

Ou Yangmin^{1,2}, Zhang Zuping¹

(1. School of Information Science and Engineering, Central South University, Changsha 410083, China;

2. School of computer and communication, Hunan University of Technology, Zhuzhou Hunan 412008, China)

Abstract: The network traffic data is chaotic, and in traditional chaotic forecasting algorithm, the Euclid distance is used to measure correlation between phase points in phase space. Because of limitation of Euclid distance, in high dimension phase space, the forecasting accuracy of the traditional chaotic forecasting algorithm decreases rapidly. The included angle cosine instead of Euclid distance is used as the standard to judge correlation between phase points. And regarding phase points as vectors and taking vectors module and angles as optimization objectives, identifies forecasting parameters. Applies the above method to forecast the network traffic data and the results indicate that the forecasting accuracy increases remarkably in high dimension phase space.

Keywords: network traffic short-term forecasting; chaos; local method; Euclid distance; the included angle cosine

0 引言

随着网络通信的迅猛发展, 对网络服务质量的要求越来越高, 如何减少网络拥塞、提高网络的服务性能, 使得网络流量建模与预测在自适应拥塞控制、带宽分配和网络管理中显得至关重要。国内外研究者提出了许多网络流量预测算法, 如神经网络法^[1-3], 小波分解法^[4-5], 支持向量机^[6-8]等, 这些算法为网络流量建模与预测做出了积极的贡献。混沌科学的发展, 为

网络流量预测提供了新的思路。混沌是一种在自然界和人类社会中广泛存在的非线性运动形式, 近年来的研究显示, 网络流量时间序列表现出一定的混沌特性^[9-10]。

混沌预测的理论基础是 Packard 和 Takens 提出的相空间重构理论^[11-12]。在此基础上, 混沌预测算法主要包括全域法^[13-14]和局域法^[15-16], 其中局域法因其较小的计算量和较高的预测精度而得到了广泛的应用。在

收稿日期: 2009-05-06

作者简介: 欧阳旻(1975-), 男, 湖南湘潭人, 湖南工业大学讲师, 中南大学硕士研究生, 主要研究方向为计算机网络及非线性建模, E-mail: zzyomin@vip.163.com;

张祖平(1962-), 男, 湖南长沙人, 中南大学教授, 博士, 主要研究方向为计算机网络。

局域法中,使用欧氏距离作为衡量相点间相关性的量度,但据研究,在高维相空间中,欧氏距离难以准确反映相点间的相关性^[17]。本文提出了一种基于夹角余弦的混沌局域加权线性算法,该算法以夹角余弦代替欧氏距离作为判别相点间相关性的准则,并将相点间相关性大小通过加权的方式作用于预测模型;在此基础上,将相点视为向量,在预测参数辨识过程中以向量的模和夹角取代欧式距离作为优化目标,提出了新的预测参数识别算法,从而克服了基于欧式距离的局域线性回归预测算法存在的缺点。将以上算法应用于某大学网络数据流量的预测中,结果表明,在高维嵌入空间中,采用本方法的预测精度较传统算法的预测精度有较大提升。

1 传统局域法的局限性

1.1 相空间重构理论

任何混沌时间序列预测的基础都是相空间重构,对混沌系统为:

$$Y_{t+1} = F(Y_t), \quad (1)$$

式中: $Y_t \in \mathbf{R}^n$, 为状态变量;

$F: \mathbf{R}^n \rightarrow \mathbf{R}^n$, 为1个光滑连续函数。

对式(1),能够观察到的往往是1个单维的混沌时间序列 $\{x(t), t=1, 2, \dots, N\}$ 。由相空间重构理论可得该序列在 m 维相空间中的相点为:

$$\mathbf{X}(t) = \{x(t), x(t+\tau), \dots, x[t+(m-1)\tau]\}, \quad (2)$$

$$(t=1, 2, \dots, L),$$

式中: m 为嵌入维;

τ 为时间延迟, $L=N-(m-1)\tau$ 。

根据 Takens 定理,当 τ 选择恰当,且 $m \geq 2n+1$ 时,存在确定性映射 $F^m: \mathbf{R}^m \rightarrow \mathbf{R}^m$, 使得:

$$\mathbf{X}(t+1) = F^m(\mathbf{X}(t)), \quad (3)$$

式(3)即为重构系统,与式(1)具有相同的动力学特性。

1.2 欧氏距离的局限性

多重非线性因素的综合作用,决定着网络流量的演化规律。在重构相空间的基础上,局域法基于相似性原理,寻找与基准相点类似的相点作为参考邻域,并通过预测参数识别,提取参考邻域中各相点在相空间中演化的动力学信息,然后将其作用于基准相点,以实现短期预测。在混沌局域预测中,参考邻域与基准相点的相关性越高,预测精度越有保证。这是由于非线性因素及变化实质上决定了相点性态及其演化规律,因此相关性越高的相点所对应的影响因素更加相似,以此作为参考可以实现更加精确的预测。而相关性高的相点在时间轴上具有相似的形状,当嵌入维较小时,通过欧式距离选择的参考邻域可以反映这种相关性,但当嵌入维逐渐增大时,其局限性则开始逐步显现。

以下例子可以说明这个问题,设相点(无量纲):

$$\begin{cases} \mathbf{X}_0 = (20, 22, 40, 45, 60, 80), \\ \mathbf{X}_1 = (30, 35, 55, 60, 70, 90), \\ \mathbf{X}_2 = (40, 38, 37, 55, 55, 90), \end{cases}$$

为6维空间的3个相点,以 \mathbf{X}_0 为基准相点,如图1所示,图中纵轴表示相点大小,横轴表示相点中元素的排列顺序。计算 $\mathbf{X}_1, \mathbf{X}_2$ 与 \mathbf{X}_0 的欧式距离:

$$D(\mathbf{X}_0, \mathbf{X}_1) = 30.3, D(\mathbf{X}_0, \mathbf{X}_2) = 29.8,$$

由于 $D(\mathbf{X}_0, \mathbf{X}_2) < D(\mathbf{X}_0, \mathbf{X}_1)$, 因此根据欧式距离选择 \mathbf{X}_2 作为参考相点,而实际上参考 \mathbf{X}_1 的演化规律可以得到更高的预测精度。

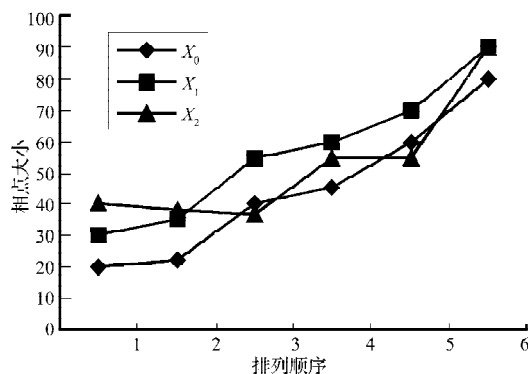


图1 1个计算相点间相关性的范例

Fig. 1 A relevance example of computing phase points

本文将相点视为向量,采用夹角余弦作为判别相点间相关性的标准,两相点夹角余弦值越大,夹角越小,表明相关性越高。计算 $\mathbf{X}_1, \mathbf{X}_2$ 与 \mathbf{X}_0 间夹角余弦值可得:

$$C(\mathbf{X}_0, \mathbf{X}_1) = 0.9947, C(\mathbf{X}_0, \mathbf{X}_2) = 0.9805,$$

由于 $C(\mathbf{X}_0, \mathbf{X}_1) > C(\mathbf{X}_0, \mathbf{X}_2)$, 选择 \mathbf{X}_1 作为参考相点。上例表明,使用夹角余弦作为选择参考临域的准则,在高嵌入维空间中,克服了欧式距离的局限,所选相点与基准点相关性更强,提高了预测精度。

2 基于夹角余弦的混沌局域预测算法

向量 α, β 间夹角余弦的定义为:

$$\cos(\alpha, \beta) = \frac{\alpha \cdot \beta}{|\alpha| \times |\beta|}. \quad (4)$$

对于重构相空间 $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L\}$, 以相点 \mathbf{X}_L 为基准相点,计算其它相点与 \mathbf{X}_L 的夹角余弦值 c ,选取其中最大的 m 个值所对应的相点作为参考临域 $\mathbf{X}_{Li}(i=1, \dots, m)$ 。设 c_m 为 c 中的最大值,根据夹角余弦值为临域点加权值 P :

$$P_i = \frac{\exp(c_i - c_m)}{\sum_{i=1}^m \exp(c_i - c_m)}. \quad (5)$$

根据参考临域 \mathbf{X}_{Li} , 对预测参数 a, b 进行辨识。这里将相点视为向量,以向量的模和夹角作为优化目

标, 要求拟和相点模尽量逼近目标相点模, 并同时使拟和相点与目标相点夹角最小。即对:

$$f(a, b) = \sum_{i=1}^p P_i * (|X_{Li+h}| - |ae + bX_{Li}|)^2, \quad (6)$$

$$g(a, b) = \sum_{i=1}^p P_i * \cos(ae + bX_{Li}, X_{Li+h}), \quad (7)$$

使 $f(a, b)$ 取得最小值的同时使 $g(a, b)$ 取得最大值, h 为预测步长。

由夹角余弦定义展开式 (7) 得:

$$g(a, b) = \sum_{i=1}^m P_i * \frac{K_1 a + K_2 b}{K_3 * \sqrt{ma^2 + K_4 ab + K_5 b^2}}, \quad (8)$$

式中: $K_1 = \sum_{j=1}^m x_{i+h}^j$;

$$K_2 = \sum_{j=1}^m x_i^j x_{i+h}^j;$$

$$K_3 = \sqrt{\sum_{j=1}^m x_{i+h}^{j2}};$$

$$K_4 = \sum_{j=1}^m 2x_i^j;$$

$$K_5 = \sum_{j=1}^m x_i^{j2}.$$

对式 (8) 分别求 a, b 的偏导, 并消去二次项得:

$$a = \frac{\left| \sum_{i=1}^m P_i (2K_1 K_5 - K_2 K_4) \right|}{\left| \sum_{i=1}^m P_i (K_1 K_4 - 2m K_2) \right|} * b, \quad (9)$$

令 $K = \frac{\left| \sum_{i=1}^m P_i (2K_1 K_5 - K_2 K_4) \right|}{\left| \sum_{i=1}^m P_i (K_1 K_4 - 2m K_2) \right|}$, 将 $a = Kb$ 代入式 (6) 并对

b 计算得:

$$b = \frac{\sum_{i=1}^m P_i \sqrt{\sum_{j=1}^m x_{i+h}^{j2} * \sum_{j=1}^m (K + x_i^j)^2}}{\sum_{i=1}^m P_i \sum_{j=1}^m (K + x_i^j)^2}, \quad (10)$$

$$a = K * \frac{\sum_{i=1}^m P_i \sqrt{\sum_{j=1}^m x_{i+h}^{j2} * \sum_{j=1}^m (K + x_i^j)^2}}{\sum_{i=1}^m P_i \sum_{j=1}^m (K + x_i^j)^2}. \quad (11)$$

将 a, b 值代入式 (12), 可得 $x(t)$ 序列的 h 步预测值 $\hat{X}_{L+h}(m)$

$$\hat{X}_{L+h} = ae + bX_L, \quad e = [1, 1, \dots, 1]_m^T. \quad (12)$$

3 算例分析和结论

计算机网络由众多计算机用户组成, 而网络流量时间序列则反映了众多用户使用网络的规律。单个用

户使用网络的随机性较强, 但众多网络用户的总体网络流量时间序列则呈现一定的规律性 (见图 1)。本文采用某大学的网络流量数据进行算例分析, 每 15 min 采样一次数据, 总时间长度为 30 d, 数据总计为 2 880 个。源数据如图 2 所示, 图中纵坐标为连续 6 d 的网络流量数据, 横坐标是采样点顺序。在图 2 中, 可以看到明显的 6 个周期, 由于测量时间从每日零时开始, 观察图 2 可以发现, 每日零时的网络流量为全天最低, 接近于 0, 在此之后, 网络流量迅速增大, 并多在中午 12 时前后出现一天中的最大值。

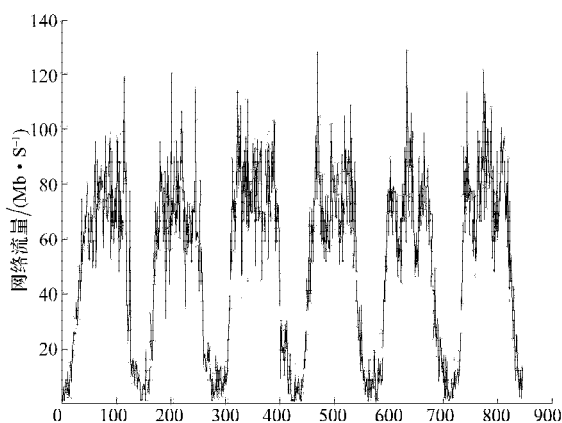


图 2 连续 6 天的网络流量数据

Fig. 2 The network traffic data of continuous six days

图 2 中数据呈现出一定的规律性, 但网络流量数据的峰值部分有较大的起伏, 显然含有一定的噪声, 为准确预测带来了一定的难度。使用基于欧式距离的局域加权线性算法 (简称算法 1) 和基于夹角余弦的混沌局域加权线性算法 (简称算法 2), 分别预测历史数据下一日的共 96 点的网络流量数据。采用 GP (Grassberger-Procassia) 算法和互信息量法分别计算嵌入维 m 和时间延迟 τ , 得 $m=7, \tau=11$, 预测误差统计表如表 1 所示。

表 1 算法 1 和算法 2 预测误差统计表

Tab. 1 The statistics of forecasting error of algorism 1 and 2

| 算法 | 误差 | | | 误差 < 3 % 的比例 |
|----|------|------|-------|--------------|
| | 平均 | 最大 | 最小 | |
| 1 | 5.13 | 1.53 | 18.31 | 22.03 |
| 2 | 3.86 | 0.62 | 11.03 | 45.03 |

本文时间序列的嵌入维为 7, 属于高嵌入维空间 (一般而言, 嵌入维超过 5, 使用传统预测算法的预测精度就会迅速下降, 因此嵌入维超过 5 即可认为是高嵌入维空间)。比较以上两组算法的预测误差统计, 可以发现算法 2 的各项统计指标都明显好于算法 1, 算例结果证明了算法 2 在高嵌入维空间中, 比算法 1 更能准确预测。考虑到原信号中存在一定的噪声, 以上预测结果是令人满意的, 这也证明基于夹角余弦的混沌

局域预测法是适合网络流量数据短期预测的。噪声的存在抑制了预测精度的进一步提高,因此如果结合有效的去噪算法,预测精度将可以进一步提升。

4 结语

网络流量数据的运行规律由多种非线性因素影响和驱动,其动力结构中表现出一定的混沌特性。在网络流量的混沌局域预测中,通常使用欧式距离选取参考临域,并将其作为优化目标识别预测参数,导致在高维相空间中预测精度迅速下降。针对这一问题,本文首先阐述了以上现象出现的原因,并提出使用夹角余弦取代欧式距离。通过图 1 所示范例,说明了在高维嵌入空间中,相比于欧式距离,夹角余弦选择的参考相点与基准相点间的相关性更密切。然后在此基础上,本文将相点视为向量,使用向量的模和夹角取代欧式距离进行线性预测模型的参数辨识,进一步提高了预测精度。在某大学网络流量算例中,本文算法的有效性得到了验证。

参考文献:

- [1] Tarraf A A, Habib I W, Saadawi T N. A Novel Neural Network Traffic Enforcement Mechanism for ATM Networks[J]. IEEE Journal on Selected Areas on Communications, 1994, 12(6): 1088-1096.
- [2] Alarcon-Aquino V, Barria J V. Multiresolution FIR Neural-Network-Based Learning Algorithm Applied to Network Traffic Prediction[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and reviews, 2006, 36(2): 208-220.
- [3] 余 健, 郭 平. 优化 Elman 神经网络用于网络流量预测[J]. 计算机工程与设计, 2008, 29(17): 4531-4534.
Yu Jian, Guo Ping. Application of Optimized Elman Neural Network to Network Traffic Prediction[J]. Computer Engineering and Design, 2008, 29(17): 4531-4534.
- [4] Riedi R H, Crouse M S, Ribeiro V J. A Multi-Fractal Wavelet Model with Application to Network Traffic[J]. IEEE Transactions on Information Theory, 1999, 45(3): 992-1018.
- [5] 胡 俊, 胡玉清, 肖中卿. 基于小波变换的网络流量预测模型[J]. 计算机工程, 2008, 34(19): 112-115.
Hu jun, Hu Yuiqing, Xiao Zhongqing. The Network Traffic Forecasting Based on Wavelet Transformation[J]. Computer Engineering, 2008, 34(19): 112-115.
- [6] Liu Zunxiong, Zhang Deyun, Hui Chuanliao. Multi-Scale Combination Prediction Model with Least Square Support Vector Machine for Network Traffic[J]. Advances in Neural Networks, 2005, 34(8): 385-390.
- [7] Jin Xuexiang, Zhang Yi, Yao Danya. Simultaneously Prediction of Network Traffic Flow Based on PCA-SVR[J]. Advanced in Neural Networks, 2007, 44(1): 1022-1031.
- [8] 杨 光, 张国梅, 刘星宇. 基于小波核 LS-SVM 的网络流量预测[J]. 微机发展, 2005, 15(12): 125-128.
Yang Guang, Zhang Guomei, Liu Xingyu. Network Traffic Forecasting Based on Wavelet LS-SVM[J]. Computer Development, 2005, 15(12): 125-128.
- [9] 陆锦军, 王执铨. 基于混沌特性的网络流量预测[J]. 南京航空航天大学学报, 2006, 38(2): 217-221.
Lu Jinjun, Wang Zhiquan. Network Traffic Forecasting Based on Chaotic Feature[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2006, 38(2): 217-221.
- [10] 李建华, 刘润洁, 申金媛, 等. 现代网络流量的混沌奇异吸引子[J]. 解放军理工大学学报: 自然科学版, 2008, 9(5): 427-430.
Li Jianhua, Liu Runjie, Shen Jinyuan, et al. The Strange Chaotic Attractors of Modern Network[J]. Journal of PLA University of Science and Technology: Natural Science Edition, 2008, 9(5): 427-430.
- [11] Packard N H, Crutchfield J P, Farmer J D, et al. Geometry from a Time Series[J]. Physical Review Letters, 1980, 45(6): 712-716.
- [12] Takens F. Detecting Strange Attractors in Fluid Turbulence [M]. Berlin: Springer-Verlag, 1981.
- [13] 简相超, 郑君里. 一种正交多项式混沌全局建模方法[J]. 电子学报, 2002, 30(1): 76-78.
Jian Xiangchao, Zheng Junli. A Chaotic Global Modeling Method Based on Orthogonal Polynomials[J]. Acta Electronica Sinica, 2002, 30(1): 76-78.
- [14] 张 森, 肖先赐. 混沌时间序列全局预测新方法——连分式法 [J]. 物理学报, 2005, 54(11): 5062-5068.
Zhang Sen, Xiao Xianci. A New Method of Global Prediction for Chaotic Time Series Based on Continued Fractions[J]. Acta Physica Sinica, 2005, 54(11): 5062-5068.
- [15] Kugiumtzis D, Lingjaerde C, Christophersen N. Regularized Local Linear Prediction of Chaotic Time Series[J]. Physica D, 1998, 112(3-4): 334-360.
- [16] 岳毅宏, 韩文秀, 王 健. 基于加权平均一阶发散度的混沌序列预测法[J]. 系统工程与电子技术, 2004, 26(5): 602-604.
Yue Yihong, Han Wenxiu, Wang Jian. Forecasting Method of Chaotic Sequence Based on Weighted Average First-Order Divergence Degree[J]. Systems Engineering and Electronics, 2004, 26(5): 602-604.
- [17] 蒋传文, 候志俭, 张勇传. 基于关联度的高嵌入维混沌预测方法研究[J]. 系统工程与电子技术, 2002, 24(12): 65-66.
Jiang Chuanwen, Hou Zhijian, Zhang Yongchuan. One High Embedded Dimensions Chaotic Forecasting Method Based on Degree of Incidence[J]. Systems Engineering and Electronics, 2002, 24(12): 65-66.

(责任编辑: 罗立宇)