

基于混沌理论的煤矿监控网络流量时间序列的预测

邵小强, 马宪民

(西安科技大学 电气与控制工程学院, 陕西 西安 710054)

摘 要: 针对目前煤矿监控网络的流量增大趋势, 为了改进和提高网络的 QoS 质量, 提出了基于混沌时间序列预测网络流量的方法. 从相空间重构, 用互信息量法和虚假临近点法确定了延迟时间和嵌入维数, 用小数据量法求解了最大 Lyapunov 指数, 由此证明了网络流量时间序列的混沌特性, 并且建立相应模型, 成功地对其做出了预测. 仿真结果表明, 该方法具有较高的准确度.

关键词: 网络流量; 混沌; 相空间重构; 最大 Lyapunov 指数

中图分类号: TD65

文献标识码: A

文章编号: 1000-7180(2011)09-0005-03

Based on Chaos Theory to Predict the Series Time of Mine Monitoring and Control Network Traffic

SHAO Xiao-qiang, MA Xian-min

(College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: Coal mine for the current trend of increasing network traffic monitoring, in order to improve and enhance the quality of the network QoS, we propose a chaotic time series prediction based on the network traffic. From the phase space reconstruction, with the mutual information and false near the point of law determined by the delay time and embedding dimension, with a small amount of data the maximum Lyapunov index method, which proved that the chaotic time series of network traffic characteristics, and the establishment of The corresponding model, successfully made its forecast. Simulation results show that the method has high accuracy.

Key words: chaos; network traffic; phase space reconstruction; Lyapunov index

1 引言

目前国内外传统的预测多采用以统计学为理论基础的网路业务模型, 如自回归模型 (Auto Regressive, AR)、滑动平均模型 (Moving Average, MA)、自回归滑动平均模型 (Autoregressive moving average model, ARMA). 但是这些模型多少存在一定的局限性, 其主要用于线型稳定过程的建模问题. ARIMA^[1] 和 ARFIMA 主要用于解决线性不稳定过程的建模问题. 而网络流量是个非线性系统, 采用传统方法能力有限. 并且随着网络规模的快速膨胀, 网络流量表现出的特征也日益复杂化, 其表现出来

的强非线性和长期相关性 (自相似性) 使得传统预测技术越来越力不从心. 针对这一问题提出基于混沌的监控网络流量的时间预测方法可以改进传统方法的不足.

2 监控网络流量混沌特征分析

2.1 网络流量的混沌特性

网络流量是一个复杂的非线性过程, 混沌本身是一个非线性系统, 并且网络流量的自相似性与混沌具有紧密的联系^[2], 而混沌系统的一个特性就是对初始条件的敏感性, 初始条件的微小变化都会引起系统的相轨迹按指数速率变化, 某些特征量具有

收稿日期: 2011-01-12; 修回日期: 2011-02-22

基金项目: 陕西省教育厅自然科学基金专项 (2010JK663); 西安科技大学培育基金项目 (200835)

相同的值,而煤矿监控系统网络中就可以借鉴小变化特征对其重要监控对象进行分析和预测,通过计算网络流量的关联维数和 Lyapunov 指数,确定其网络特征存在混沌。

确定一个系统是否存在混沌,根据 Lyapunov 指数法^[3]进行分析,如在相空间中初始条件不同的两条相轨迹随时间按指数率吸引或分离的程度,进行比较和判断。

故利用混沌时间序列来对网络流量进行短期预测是可行的。

2.2 相空间重构及要求

时间序列为一维向量,而混沌运动至少要在三维的自治动力系统中才能出现,所以通过时间序列分析运动系统状态的演化,必须对时间序列进行相空间重构。从几何观点,混沌时间序列是高维相空间混沌运动的轨迹在一维空间上的投影,在该投影的过程中,本来不相邻的两点投影到一维空间时成为相邻的两点,就是混沌时间序列呈现出无规律的原因所在。重构相空间,就是从混沌时间序列中恢复混沌运动的轨迹,随着嵌入维数的增大,混沌运动的轨迹将会逐渐打开,混沌运动的轨迹就会得到恢复。

根据状态空间的重构理论,假设已经观测到的时间序列为:

$$\{x(t)\}, \quad t = 1, 2, \dots, N \quad (1)$$

其中,嵌入维数记为 m ,时间延迟记为 τ ,这样就可以得到重构的相空间为: $\{x(t)\} = (x(t), x(t+\tau), \dots, x(t+(m-1)\tau), X(t) \in R^m$ 。

根据 Takens 嵌入定理有下式:

$$X(t+T) = f(X(t)) \quad (2)$$

由重构相空间可以看出,坐标延迟相空间重构有两个关键参数:嵌入维数 m 和时间延迟 τ 。Takens 嵌入定理^[4]表明一维时间序列在无限长且无噪声的情况下,延迟时间取任意值时都能重构原系统的相空间,但实际情况中,时间序列是有限长的且不可避免的被噪声污染。因此,要对实测时间序列进行相空间重构,必先确定 m 和 τ 等参数。网络重构中这两个参数不能选的太大,这样收敛速度慢;参数太小细节反映不明显。

3 监控网络混沌模型建立和参数确定

3.1 时间延迟 τ 的确定

确定时间延迟 τ 的思想是:使 x_n 和 x_{n+1} 在某种程度上相互独立但又不完全无关,以使它们能在重构的相空间中作为独立的坐标处理。考虑到混沌系

统是一个非线性系统,所以文中主要介绍互信息量法^[5]来确定时间延迟 τ 。

假设两离散信息系统 $\{a_1, a_2, \dots, a_n\}$ 和 $\{b_1, b_2, \dots, b_n\}$ 构成的系统为 A 和 B ,则其信息熵为:

$$H(A) = - \sum_{i=1}^n P_a(a_i) \log_2 P_a(a_i) \quad (3)$$

$$H(B) = - \sum_{j=1}^n P_b(b_j) \log_2 P_b(b_j)$$

其中, $P_a(a_i)$ 和 $P_b(b_j)$ 分别为系统 A 和 B 中事件 a_i 和 b_j 的概率。

在 A 给定的情况下,得到的关于系统 B 的信息,称为 A 和 B 的互信息,即:

$$I(B, A) = H(B) - H(B | A) \quad (4)$$

因此有:

$$I(B, A) = \sum_i \sum_j P_{\phi}(a_i, b_j) \log_2 \left[\frac{P_{\phi}(a_i, b_j)}{P_a(a_i) P_b(b_j)} \right] \quad (5)$$

其中, $P_{\phi}(a_i, b_j)$ 为事件 a_i 和 b_j 事件的联合概率分布。定义:

$$[a, b] = [x(t), x(t) + \tau] \quad (6)$$

即 a 代表时间序列 $x(t)$, b 代表延迟时间为 τ 的时间序列 $x(t+\tau)$,则 $I(B, A)$ 是关于时间延迟 τ 的函数,记作 $I(\tau)$,它的大小代表了在已知系统 A 即 $x(t)$ 的情况下,系统 B 即 $x(t+\tau)$ 的确定性大小。如果 $I(\tau) = 0$,则表示 $x(t+\tau)$ 完全不可预测,或者 $x(t)$ 和 $x(t+\tau)$ 完全不相干; $I(\tau)$ 取得极小值,则表示 $x(t)$ 和 $x(t+\tau)$ 是最大可能的不相干。重构时使用 $I(\tau)$ 的第一个极小值作为最有延迟时间。

3.2 嵌入维数的确定

关联维数的计算对于嵌入维数 m 的确定有着重要的意义,因为 Takens 定理给出了相空间重构的充分条件: $m \geq 2d + 1$,其中 d 是混沌吸引子所在的相空间的拓扑维数。下面采用虚假最邻近点法来确定嵌入维数 m 。

d 维相空间中,每一个相点矢量 $X(i) = \{x(i), x(i+\tau), \dots, x(i+(d-1)\tau)\}$,都有一个某距离内的最邻近点 $X^{NN}(i)$,其距离为:

$$R_d(i) = \| X(i) - X^{NN}(i) \| \quad (7)$$

当相空间维数从 d 维增加到 $d+1$ 维时,这两个相点之间的距离就会发生变化,变为:

$$R_{d+1}^2(i) = R_d^2(i) + \| x(i+\tau d) - x^{NN}(i+\tau d) \|^2 \quad (8)$$

如果 $R_{d+1}^2(i)$ 比 $R_d(i)$ 大的多,说明这是由于高维混沌吸引子中两个不相邻的点投影到低维轨道上

时变成相邻的两点造成的,所以满足这样条件的点,就称之为虚假临近点. 令

$$a_1(i, d) = \frac{\|x(i + \tau d) - x^{NN}(i + \tau d)\|}{R_d(i)} \quad (9)$$

如果 $a_1(i, d) > R_r$, 则 $X^{NN}(i)$ 是 $X(i)$ 的虚假最邻近点, 阈值 R_r 可在 $[10, 50]$ 选取.

如果数据长度有限且含有噪声, 则可以这样判断: 如果

$$\frac{R_{d+1}(i)}{R_A} \geq 2 \quad (10)$$

则 $X^{NN}(i)$ 是 $X(i)$ 的虚假最临近点, 其中

$$R_A = \frac{1}{N} \sum_{i=1}^n [x(i) - \bar{x}], \quad \bar{x} = \frac{1}{N} \sum_{i=1}^n x(i) \quad (11)$$

对于实测时间序列, 从嵌入维数的最小起始值开始, 计算虚假最邻近的比例, 然后增加 d , 直到虚假最近邻点的比例小于 5% 或者虚假最邻近点不再随着 d 的增加而减少时, 可以认为混沌吸引子已完全打开, 此时 d 的即为嵌入维数, 即可以确定嵌入维数 m .

3.3 确定最大 Lyapunov 指数

由文献[6]可知 Lyapunov 指数体现了系统轨道演化过程的特性, 定量该系统对于初始条件依赖的敏感性, 其各指数参数可以较好的判断系统是否会出现混沌现象.

对于一维系统, 如果指数 $\lambda < 0$, 表明系统处于大尺度周期状态; $\lambda = 0$, 表明系统处于混沌临界状态; $\lambda > 0$, 表明系统处于混沌状态. 对于多维系统, 只要有一个指数为正值, 就表明系统处于混沌状态, 所以只需要求解多维系统的最大 Lyapunov 指数即可, 根据煤矿监控网络特征采用了小数据量法^[7].

相空间重构后, 假设轨道上第 j 点的最邻近点为 $Y_{j'}(t_0)$, 即满足如下关系:

$$d_j(0) = \min \|Y_j(t_0) - Y_{j'}(t_0)\|, \quad |j - j'| > p \quad (12)$$

其中, p 为时间序列的平均周期, 可通过能量谱平均功率的倒数估计出来, 所以最大 Lyapunov 指数就可以通过基本轨道上每个点的最近临近点的平均发散速率估计出来. Sato 等人推导如下:

$$\lambda_1(i, k) = \frac{1}{k\Delta t} \frac{1}{M-K} \sum_{j=1}^{M-K} \ln \frac{d_j(i+k)}{d_j(i)} \quad (13)$$

其中, k 是常数, Δt 为样本周期, $d_j(i)$ 是基本轨道上第 j 对最邻近点对经过 i 各离散时间步长后的距离, 最大 Lyapunov 指数表征的是初始轨道的指数的发

散程度, 所以有:

$$d_j(i) = C_j e^{\lambda_1(i\Delta t)}, \quad C_j = d_j(0) \quad (14)$$

将式(15)两边取对数, 得到:

$$\ln d_j(i) = \ln C_j + \lambda_1(i\Delta t) \quad (15)$$

$$j = 1, 2, \dots, M$$

可见, 最大 Lyapunov 指数 λ_1 相当于式(15)的斜率, 通过最小二乘法逼近直线得到:

$$y(i) = \frac{1}{\Delta t} < \ln d_j(i) > \quad (16)$$

其中, $< \ln d_j(i) >$ 表示所有关于 j 的平均值.

4 预测仿真及结果分析

4.1 网络流量统计格式及时间序列混沌性的判定

数据采用神木柠条塔矿网络监控中心的 ROUTER, 其下挂华为 2126S 的 SWITCH, 网络拓扑为在 2126S 上配置 VLAN, 将日交换数据, 利用 NETFLOW 记录了一周(6 天) 每 15 分钟记录一次主节点路由器交换的数据量, 总共 576 个数据向量, 形成了一个网络流量时间序列 $\{x(0), x(1), \dots, x(t)\}$, $t = 0, 1, 2, \dots, 576$. 其中使用 NETFLOW 记录流量转换数据的格式如表 1 所示.

表 1 流量数据格式表

srcIP	dstIP	port	srcPort	dstPort
192.168.0.2	192.168.43.18	10	1655	1598
192.168.22.11	192.168.13.24	11	2345	2201
192.168.54.10	192.168.10.14	12	553	545

使用 bin_to_text.sh 对格式中对应字段数值进行求和, 即可以得到计时时间段中的总流量.

用互信息量法求系统的延迟时间, 系统取得第一个极小值时 $\tau = 5$, 所以该系统的延迟时间为 5. 用虚假临近点法求得嵌入维数 $m = 6$.

根据小数据量法, 计算得到最大 Lyapunov 指数 $\lambda = 0.018402 > 0$, 由此可以证明网络流量时间序列满足混沌特性.

4.2 结果分析

根据混沌时间序列做出的实测与预测时间序列结果, 如表 2 所示.

表 2 误差分析

实测数据	预测数据	误差率(%)
1.745	1.739	0.343 839
1.607	1.596	0.684 505
2.032	1.997	1.722 440
1.582	1.591	0.568 900
2.122	2.119	0.141 376

(下转第 12 页)

和指数增益控制功能,且具有低功耗、低失真度和低输入等效噪声的特点.同时通过采用高效增益控制电路,实现系统功耗在压缩状态下的进一步减小.

参考文献:

- [1] Jose Silva — Martinez, Sergio Solis — Bustos, Martin Schellenberg. A CMOS hearing aid device[J]. Analog Integrated Circuit and Signal, 1999(21):163—172.
- [2] Cristiano Azzolini. A 1—V CMOS audio amplifier for low cost hearing aids[C]//Proc. of 15th, IEEE. Int. Conf. On Electronics, Circuit and Systems. University of Malta, 2008
- [3] Alex Lollo, Giacomo Bollati. A class—G headphone amplifier in 65nm CMOS technology[J]. IEEE Journal of Solid—State Circuit, 2010, SC—45(12):2530—2542.
- [4] Ilker E Ocak, Reha Kepenek. A high performance delta sigma readout circuitry for ug resolution micro accelerometers[J]. Analog Integrated Circuit and Signal processing, 2010(64):137—145.
- [5] Jorg Hauptmann, Franz Dielacher, Reinhard Steiner. A low—noise amplifier with automatic gain control and anti—clipping control in CMOS technology[J]. IEEE Journal of Solid—State Circuit, 1992, SC—27(7): 974—981.
- [6] Sunyoung Kim, Jae—Youl Lee, Seong—Jun Song, et al. An energy—efficient analog front—end circuit for a sub—1V digital hearing aid[J]. IEEE Journal of Solid—State Circuit, 2006, SC—41(4):876—882.
- [7] Sarpeshkar R, Baker M, Salthouse C, et al. An analog bionic ear processor with zero—crossing detection[C]//Proc. ISSCC. San Francisco, CA, 2005:78—79.
- [8] Panagiotopoulos D A, Newcomb R W, Singh S K. A current—mode exponential amplifier[J]. IEEE Trans. Circuit Sys, II, Analog Digit Signal, 2000, 47(6):548—551.
- [9] Xiaodan Zou, Yong Lian. A 1 V 450nW fully integrated programmable biomedical sensor 2009, interface chip [J]. IEEE Journal of Solid—State Circuit, 2009, SC—44(4): 1067—1077.
- [10] Yazicigolu R F, Merken P, Puers R, et al. A 60uW 60nV/sqrHz readout front—end for portable bio—potential acquisition systems[J]. IEEE J. Solid—State Circuit, 2007, 42(15):1100—1110.

作者简介:

李凡阳 男, (1983—), 博士研究生. 研究方向为模拟集成电路设计.

(上接第7页)

从仿真结果和数据分析看到,预测结果较好的反映实际测试结果情况,误差较小,但第三组数据与实际值偏差较大,经分析误差可能有以下原因:(1)没有考虑多负荷增减异变的影响;(2)与采样样本数据有关.针对以上问题可以加大样本数据收集,再对不同时间段训练数据样本建立相应优化模型,提高预测的准确度.

5 结束语

文中证明了网络流量时间序列的混沌特性,建立了预测模型.通过仿真实验,验证了该方法有较高的准确性.改进了煤矿监控网络流量预测的传统方法的不足,为今后提高和管理网络传输的 QoS 提供了可借鉴的理论依据.

参考文献:

- [1] 孙明谦,姚淑萍,胡昌振.服务器负载的小波—神经网络—ARMA 预测[J]. 计算机工程与应用, 2007, 43(10): 154—169.
- [2] 宣蕾,卢锡城,于瑞厚,等.网络威胁时序的自相似性分析[J]. 通信学报, 2008, 29(4):45—50.
- [3] 向昌盛,周子英.混沌时间序列的支持向量机预测[J]. 统计与决策, 2010 (1):7—10.
- [4] 王永生,孙瑾,王昌金,等.变参数混沌时间序列的神经网络预测研究[J]. 物理学报, 2008, 57(10):6120—6030.
- [5] 朱志伟,孟义朝,方捻,等.基于小波变换的光混沌信号消噪与 Lyapunov 指数计算[J]. 光子学报, 2008, 37(10):2103—2107.
- [6] 黄文标,施式亮.工业事故时间序列混沌特性判定及其实证研究[J]. 西安科技大学学报, 2010, 30(1):34—38.
- [7] Christiansen M L, Collins J J, De Luca C J. A practical method for calculating largest Lyapunov exponents form small data sets[J]. Physica D, 1993(65):117—134.

作者简介:

邵小强 男, (1976—), 博士研究生, 讲师. 研究方向为电气及网络控制.