

基于小波分解的网络流量时间序列建模与预测

张 晗, 王 霞

(吉林大学 计算机科学与技术学院, 长春 130012)

摘 要: 提出一种基于小波分解的网络流量时间序列的分析和预测方法。将非平稳的网络流量时间序列通过小波分解成为多个平稳分量,采用自回归滑动平均方法分别对各平稳分量进行建模,将所有分量的模型进行组合,得到原始非平稳网络流量时间序列的预测模型。在仿真实验中,利用网络流量文库的时间序列数据建立了预测模型,并对其进行独立测试检验。仿真结果表明,本预测方法提高了网络流量时间序列的预测准确率,是一种有效、稳健的网络流量预测方法。

关键词: 网络流量; 小波分解; 时间序列; 预测

中图分类号: TP181 **文献标志码:** A **文章编号:** 1001-3695(2012)08-3134-03

doi:10.3969/j.issn.1001-3695.2012.08.087

Modeling and forecasting for network traffic based on wavelet decomposition

ZHANG Han, WANG Xia

(College of Computer Science & Technology, Jilin University, Changchun 130012, China)

Abstract: This paper proposed a network traffic forecasting methods based on wavelet decomposition and time series analysis method. Firstly, the method decomposed the network traffic time series in multiple stationary components by wavelet decomposition, then used the autoregressive moving average method to model the each stationary component separately. Finally combined all the components of the model to get the forecasting model of the original non-stationary network traffic time series. It carried out the simulation experiment on time series data of the network library. The simulation results show that, the proposed method improves the network traffic time series forecasting accuracy rate, and it is an efficient, robust network traffic forecasting method.

Key words: network traffic; wavelet decomposition; time series; forecasting

网络流量一般按一定的时间间隔进行采样,是一种典型的时间序列数据,因此时间序列分析方法在网络流量的预测中得到了广泛的应用,可以较好地揭示网络流量内在统计特征和变化规律^[1]。在实际应用过程中,时间序列分析通常假设网络流量变化是一种平稳时间序列数据,对非平稳网络流量时间的分析不多^[2]。然而网络系统十分复杂,网络流量大多数都具有非平稳特点,因此探讨非平稳时间序列分析对网络流量的分析和预测具有重要的意义^[3]。

在非平稳网络流量时间序列的预测应用中,通常先将其转换成平稳时间序列,然后用平稳时间序列预测方法进行分析和预测,如采用线性自回归滑动平均模型 (autoregressive moving average, ARMA)、带控项的自回归模型 (controlled autoregressive, CAR)、差分自回归滑动平均模型 (autoregressive integrating moving average, ARIMA)^[4-6]。这些传统时间序列分析方法均属于线性模型,适合于季节性、周期性等特征平稳的时间序列预测,对于一些非平稳、非线性的网络流时间序列,由于它们不能全面反映时间序列的复杂变化特征,预测准确率比较低^[7]。

小波分解由 Meyer 和 Mallat 提出,能够将时间序列数据分解成为比较简单的时间序列^[8]。为了提高网络流量的预测准确率,本文提出一种基于小波分解的网络流量时间序列的分析和预测方法。将非平稳的网络流量时间序列通过小波分解成为多个平稳分量;然后采用自回归滑动平均方法分别对各平稳

分量进行建模;最后将所有分量的模型进行组合,得到原始非平稳网络流量时间序列的预测模型。

1 网络流量的建模方法

1.1 时间序列的小波分解

通过小波分解的多分辨率功能对时间序列进行分解,从而将原时间序列分解成为低频和高频系数两部分。通过重构运算从低频系数中可以得到原时间序列的低频分量,从高频系数中可以得到原时间序列的高频分量^[9]。具体通过 Mallat 算法来实现时间序列的分解,分解关系如下:

$$\begin{cases} a_{j+1} = h_0 * a_j \\ d_{j+1} = h_1 * a_j \end{cases} \quad j=0,1,\dots \quad (1)$$

其中: h_0 表示低通分解滤波器; h_1 表示高通分解滤波器;* 表示卷积运算符; a_j 表示低频系数; d_j 表示高频系数。

从式(1)可知,当 $j=0$ 时, a_0 表示原始时间序列,原始时间序列通过 h_0 和 h_1 经过多次分解后,可以得到原始时间序列低频和高频系数。

1.2 对分解后时间序列的重构

时间序列经过小波分解后,得到低频和高频系数,然后将其进行重构运算,得到时间序列的低频和高频分量,小波重构关系如下:

收稿日期: 2011-11-24; 修回日期: 2011-12-26

作者简介: 张晗(1990-),男,山东栖霞人,主要研究方向为计算机科学与技术应用(zhanghanj190@163.com);王霞(1980-),女,山东栖霞人,高级工程师,博士,主要研究方向为计算机科学与技术应用。

$$\begin{cases} A_j = g_0 * a_j \\ D_j = g_1 * d_j \end{cases} \quad j=0,1,\dots \quad (2)$$

其中: g_0 表示低通重构滤波器; g_1 表示高通重构滤波器; A_j 表示低频分量; D_j 表示高频分量。

这样原始时间序列 S 与低频分量 A_j 和低频分量 D_j 的关系为

$$S = A_1 + D_1 \quad (3)$$

1.3 时间序列预测模型的建模步骤

a) 设收集到一个时间序列数据 S 。

b) 对其非平稳性进行判断。如果是非平稳时间, 采用小波分解对原始时间序列进行分解, 得到时间序列的低频和高频系数。

c) 采用重构运算算法对分解后的时间序列的低频和高频系数进行重构, 得到低频分量 A_1 和高频分量 D_1 。

d) 对 A_1 和 D_1 重新进行平稳性检测。若某一分量符合平稳时间序列特征, 则不需要进行分解; 否则需要对时间序列分量继续进行分解和重构, 直到其转换为平稳时间序列。

e) 如果时间序列分量的自相关函数呈现拖尾特征, 偏相关函数在时间延迟第 p 步后截尾, 则采用自回归 (autoregressive, AR) 模型对该时间序列分量进行建模预测。

f) 如果时间序列分量的偏相关函数呈现拖尾特征, 自相关函数在时间延迟第 p 步后截尾, 则采用滑动平均 (moving average, MA) 模型对该时间序列分量进行建模预测。

g) 如果时间序列既不适合采用 AR 模型, 也不适合采用 MA 模型预测, 那么就采用 ARMA (m, n) 模型进行建模预测。模型最佳阶数通过 BIC (Bayes information criterion) 或 AIC (Akaike information criterion) 信息准则来确定, 模型参数通过最小二乘法来估计。

2 仿真实验

2.1 数据来源

实验数据来源于网络流量文库 (<http://newsfeed.ntcu.net/~news/2006/>)^[10], 收集了主节点路由器 Incoming articles 从 2011 年 7 月 1 日到 11 月 10 日的网络每小时访问流量, 获得了 500 个数据, 组成一个网络流量时间序列 $\{x(t), t=1, 2, \dots, 500\}$ 。其中前 450 个数据作为训练样本进行建模, 并作为历史资料进行拟合, 后 50 个数据作为测试样本, 对模型性能进行预测检验。具体数据如图 1 所示。

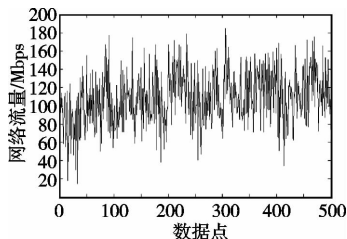


图1 采集的网络流量时间序列

2.2 数据分析

对网络流量时间序列的自相关函数和偏相关函数进行分析, 判断原始网络流量时间序列的平稳性, 采用的分析软件为 DPS 9.0, 得到的结果如图 2 和 3 所示。从图 2 和 3 可以判断, 原始网络流量时间序列的自相关函数和偏相关函数均没有明显衰减的趋势, 呈振荡波动。因此, 原始网络流量时间序列不符合平稳时间序列的特征, 采用 ARMA 模型直接对其进行建

模得到的预测结果不可靠。

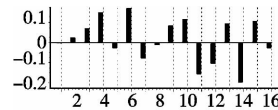


图2 网络流量时间序列的自相关函数

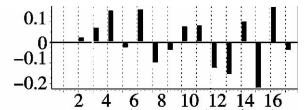


图3 网络流量时间序列的偏相关函数

同时从网络流量的变化曲线可以得知, 该网络流量数据序列不具备一定的季节性和周期性特征, 因此也不能直接采用 ARIMA、CAR 非平稳时间序列预测模型对其进行建模和预测。

2.3 网络流量的小波分解

针对网络流量的非平稳时间序列特性, 本文采用小波将其分解成平稳时间序列, 然后采用 ARMA 网络流量的平稳时间序列进行建模和分析。采用小波分析中的 db4 小波函数对网络流量时间序列进行分解, 再分别经过低通重构滤波器和高通重构滤波器进行重构运算, 网络流量时间被分解成为低频分量和高频分量, 具体如图 4 和 5 所示。

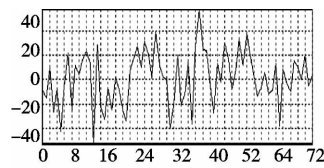


图4 网络流量的低频分量

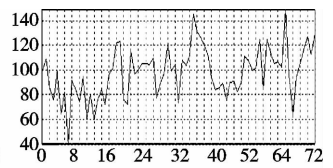


图5 网络流量的高频分量

2.4 网络流量时间的建模

对网络流量数据的低频分量 A_1 进行平稳性检测和建模, 低频分量的自相关函数和偏相关函数分别如图 6 和 7 所示。

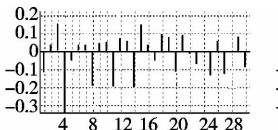


图6 网络流量低频分量的自相关函数

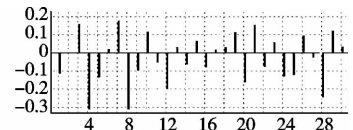


图7 网络流量低频分量的偏相关函数

从图 6 可知, 在时间延迟第 4 步, 低频分量 A_1 的自相关函数开始迅速衰减; 同时从图 7 可知, 网络流量的偏相关函数则出现拖尾现象, 这表明低频分量 A_1 此时成为一种平稳时间序列, 可采用 ARMA 模型对其进行建模与分析。根据网络流量低频分量的自相关函数和偏相关函数可知, 采用 MA 模型对其进行预测, 模型阶数为 4, 采用最小二乘法对 MA 模型的参数进行辨识, 通过 DPS 6.5 来实现, 得到的结果为

$$x_t = \varepsilon_t + 0.15\varepsilon_{t-1} - 0.774\varepsilon_{t-2} + 0.89\varepsilon_{t-3} - 0.165\varepsilon_{t-4} \quad (4)$$

其中: ε_t 表示均值为 0 的白噪声序列; x_t 表示网络流量时间序列的低频部分。

对网络流量数据的高频分量 D_1 进行平稳性检测和建模, 高频分量的自相关函数和偏相关函数分别如图 8 和 9 所示。

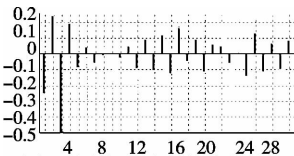


图8 网络流量高频分量的自相关函数

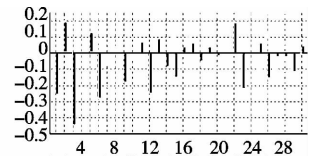


图9 网络流量高频分量的偏相关函数

从图 8 可知, 在时间延迟第 3 步, 高频分量 D_1 的自相关函数开始迅速衰减, 同时其偏相关函数出现拖尾现象, 这些特性表明, 此时高频分量 D_1 成为一种平稳时间序列, 可采用 ARMA 模型对其进行建模与分析。根据网络流量高频分量的自相关函数和偏相关函数可知, 采用 AR 模型对其进行预测, 模型的阶数为 3, 采用最小二乘法对 AR 模型的参数进行辨识, 通过 DPS 6.5 来实现, 得到的结果为

$$x_t = \varepsilon_t + 2.44\varepsilon_{t-1} - 0.157\varepsilon_{t-2} - 1.31\varepsilon_{t-3} \quad (5)$$

其中: ε_t 表示均值为0的白噪声序列; x_t 表示网络流量时间序列的高频部分。

根据时间序列预测模型的组合公式 $S = A_1 + D_1$ 将低频分量的网络流量值和高频分量的网络流量值进行组合,得到网络流量时间序列的预测模型。

2.5 结果与分析

用上述模型对网络流量训练样本进行拟合,并对测试样本进行预测检验,得到的结果如图10和图11所示。从图10的拟合结果可以看出,该模型对训练样本的拟合程度相当高,说明该模型具有较好的拟合效果,可以对测试样本进行检测;从图11的预测结果可知,预测的准确率相当高,预测效果比较令人满意。同时与没有进行小波分解的ARMA模型进行对比,对比结果表明,基于小波分解的网络流量模型预测准确率高于对比模型。

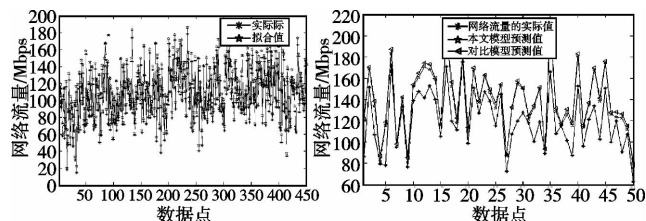


图10 网络流量时间序列的拟合结果

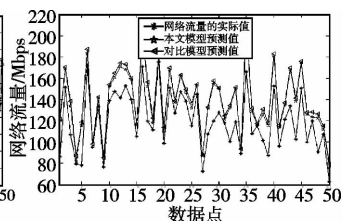


图11 网络流量时间序列的预测结果

小波分解是一种非线性分析方法,十分适合对非平稳数据进行处理。从对网络流量的预测结果可以看出,对于非平稳的网络流量时间序列,通过小波分解,将非平稳的网络流量时间序列分解成单一、平稳时间序列分量,然后采用平稳时间分析方法对平稳时间序列进行建模,有利于网络流量时间序列建模,进一步提高了网络流量的预测准确率。

在网络流量实际应用中,导致其非平稳的因素相当多,如周期、趋势因素等,采用一层小波进行分解,很难得到比较平稳的时间序列分量,需要进行多层分解,但是非平稳时间序列的最大分解层次受时间序列长度的影响相当大。当长度比较短时,最大分解层次满足不了分离各分量的影响要求,这就需要采用二进小波变换进行处理,这是在进行网络流量预测实践中

应该注意的问题。

3 结束语

网络流量预测是网络管理的基础,由于受经济、网络用户、节假日和周末等多种因素的影响,具有非平稳性,传统时间序列分析方法难以获得高准确率的预测结果。为此,本文提出一种基于小波分解的网络流量时间序列建模方法。采用具体网络流量数据进行仿真实验的结果表明,与传统时间序列预测方法相比,基于小波分解的时间序列预测方法提高了网络流量预测准确率,预测效果令人满意,是一种稳健、有效的网络流量预测方法。

参考文献:

- [1] JUVA I, VATON S, VIRTAMO J. Quick traffic matrix estimation based on link count covariances [C]//Proc of IEEE International Conference on Communications. 2006:603-608.
- [2] 高波,张钦宇,梁永生. 基于EMD及ARMA的自相似网络流量预测[J]. 通信学报, 2011, 32(4): 47-56.
- [3] NICOL D M, YAN Guan-hua. High-performance simulation of low-resolution network flows[J]. Simulation, 2006, 82(1): 21-42.
- [4] 程光,龚俭. 大规模网络流量宏观行为周期性分析研究[J]. 小型微型计算机系统, 2003, 24(6): 992-994.
- [5] CHEN Z, DELLS A, WEI P. A pragmatic methodology for testing intrusion prevention systems[J]. Computer Journal, 2009, 52(4): 429-460.
- [6] 姜明,吴春明,胡大民. 网络流量预测中的时间序列模型比较研究[J]. 电子学报, 2009, 37(11): 2353-2359.
- [7] 饶云华,曹阳,杨艳. 自相似网络通信量的多尺度预测研究[J]. 计算机工程与应用, 2005, 41(28): 26-28.
- [8] PHILIPPE A. Bayesian analysis of autoregressive moving average processes with unknown orders[J]. Computational Statistics & Data Analysis, 2006, 51(3): 1904-1923.
- [9] 李捷,侯秀红,韩志杰. 基于卡尔曼滤波和小波的网络流量预测算法研究[J]. 电子与信息学报, 2007, 29(3): 725-725.
- [10] 孙知信,张玉峰. 基于多维支持向量机的P2P网络流量识别模型[J]. 吉林大学学报:工学版, 2010, 40(5): 1298-1303.