

文章编号: 2095-4980(2019)01-0131-05

小波系数感知的网络流量预测机制

林志达, 吕华辉

(广东省电力设计研究院, 广东 广州 510623)

摘 要: 对网络流量的精确预测, 可以准确把握网络运行趋势, 及时防范网络故障。针对长期网络流量预测准确度低, 收敛速度慢的问题, 提出一种小波系数感知的网络流量预测(WCNTP)机制。借助重标极差(R/S)序列分析法初步评估网络流量在大时间尺度上的统计特性; 利用离散小波变换将非平稳的网络流量分解为多个相对平稳的流量序列; 利用分数自回归求和滑动(FARIMA)模型对网络流量进行预测。结果表明, 所提机制在长期网络流量预测过程中, 具有较高的准确度且收敛速度快, 能够精确评估网络性能, 在保证网络平稳运行的同时, 提高网络服务质量。

关键词: 网络流量预测; R/S 序列分析; 离散小波变换; 分数自回归求和滑动模型; Hurst 参数; 突发特性

中图分类号: TN915.06; TP393.04 **文献标志码:** A **doi:** 10.11805/TKYDA201901.0131

Wavelet coefficient aware network traffic prediction

LIN Zhida, LYU Huahui

(Guangdong Electric Power Design Institute, Guangzhou Guangdong 510623, China)

Abstract: Precise prediction of network traffic makes great contributions to grasping the network running trends and avoiding network failure. Aiming at the problem of low accuracy and slow convergence in the long-term network traffic prediction, a Wavelet Coefficient-aware Network Traffic Prediction (WCNTP) mechanism is proposed. By using Rescaled Range(R/S) sequence analysis, the statistical characteristics of network traffic on the large time scale are evaluated. Then the non-stationary network traffic is decomposed into a number of relatively stable network traffic sequences by discrete wavelet transform. Finally, the network traffic is predicted by using the Fractional Auto-Regressive Integration Moving Average(FARIMA) model. Results show that, the proposed mechanism has high accuracy and fast convergence speed in the process of long-term network traffic prediction, by which the network performance can be evaluated accurately, thereby improving the network service quality and ensuring the smooth operation of the network.

Keywords: network traffic prediction; R/S sequence analysis; discrete wavelet transform; Fractional Auto-Regressive Integration Moving Average(FARIMA) model; Hurst parameters; burst characteristics

近年来, 互联网技术高速发展, 人们对网络服务质量的要求日趋严格^[1], 但网络数据量爆炸式增长, 大规模网络中, 大数据流量的管理不佳以及恶意数据流攻击的防范不足所导致的网络拥塞、崩溃等故障频繁发生, 给网络服务质量的控制和管理带来巨大挑战, 对网络安全及性能造成极大威胁, 严重影响用户体验^[2]。通过建立合理的网络流量模型, 深入具体分析, 实现网络流量及网络运行情况精准预测, 对维护网络安全, 提高网络性能, 提升用户服务质量至关重要^[3], 设计合理精确的网络流量预测机制, 加强网络流量管理刻不容缓。

网络流量预测是指收集网络流量的历史数据, 并根据所收集的数据, 建立合理的网络流量模型, 通过对网络历史流量数据的分析, 实现对未来网络流量的预测, 提前获知网络运行趋势, 发现网络潜在危机, 从而及时采取防范措施^[4]。网络流量预测机制中, 网络流量模型的建立严重影响流量预测机制的准确性^[5], 一个恰当的网络流量模型, 可以准确分析网络流量数据的统计特性, 精确估计网络性能, 准确预测网络流量趋势^[6]。

收稿日期: 2017-10-31; 修回日期: 2017-12-06

基金项目: 国家高技术研究发展计划(“863”计划)基金资助项目(2012AA050801); 国家电网公司科技基金资助项目(5455H7150035)

目前,研究者已经对网络流量预测机制进行了深入研究,建立了许多经典的网络流量模型。根据流量模型的不同,网络流量预测机制可分为两种:线性预测机制和非线性预测机制。对于线性流量预测机制,在假设网络中流量数据服从泊松分布的基础上,文献[7]提出的基于自回归滑动平均(ARMA)模型的流量预测机制,具有较低的算法复杂度,易于实现,但以上机制由于对网络的自相似及异构性等特征描述不准确,导致预测精确度不高。为此,文献[8]在 ARMA 机制的基础上,提出基于自回归求和滑动平均(ARIMA)模型的流量预测机制,借助多次差分使网络流量序列更加平稳,更准确地分析网络流量数据的统计特征,提高了预测精确度,在短期流量预测中表现出较好的准确度性能,但进行长期流量预测时效果不佳。对于非线性流量预测机制,比较经典的是基于神经网络^[9-11]的流量预测机制,利用神经网络的自主学习功能,对流量序列加以训练,实现高精度的流量预测,但此类机制中,由于神经网络使用的经验风险最小化原则,学习过程中过分依赖经验,会出现过学习,收敛速度慢等问题,严重制约了其发展。

针对上述问题,本文提出一种小波系数感知的网络流量预测(WCNTP)机制,根据网络流量的自相似特性,借助 R/S 序列分析法初步评估网络流量在大时间尺度上的统计特性,针对网络流量序列的非平稳和突发特性,利用离散小波变换将非平稳的网络流量分解为多个相对平稳的流量序列,进而利用分数自回归求和滑动(FARIMA)模型对网络流量进行预测。所提机制同时考虑了网络流量的长相关和短相关特性,可以高精度地对网络流量进行长期预测,且具有较好的收敛性,对网络性能及服务质量的提升具有深远重大的现实意义。

1 网络流量特征分析及分解

1.1 R/S 时间序列分析

网络流量所呈现出的自相似特性表明网络流量是可预测的^[12]。由于衡量网络业务流的自相似特性的唯一参数是 Hurst 参数,因此,需要合理的 Hurst 参数评估模型对其进行评估。由于 R/S 分析法能够很好地刻画时间序列的分行特征和长期记忆过程,因此,本文使用 R/S 分析法对网络流量的统计特性进行评估。

将某一节点 θ 的已生存时间划分为 n 个单位时间长度,由于流量的突发特性,在不同单位时间到达节点 θ 的数据包数量差异非常大,用 $E(X) = \frac{1}{n} \sum_{i=1}^n X_i$ 表示在第 i 个单位时间内到达节点 θ 的数据包数量, $E(X) = \frac{1}{n} \sum_{i=1}^n X_i$

序列的均值以及序列与均值的差值如式(1)~(2)所示,其中 $E(X) = \frac{1}{n} \sum_{i=1}^n X_i$ 的均值为 0。

$$E(X) = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

$$\Delta X(i) = X_i - E(X) \quad (2)$$

式中: X_i 表示真实网络流量值; n 为网络流量数据采样的次数; $\Delta X(i)$ 的累加值能够表示不同时间尺度下流量的波动情况,因此能够在一定程度上体现网络流量的分型特性,如式(3)所示。通过计算 $\Delta X(i)$ 的累加值,可得到 R/S 分析法中核心参数 R_n 的计算公式,如式(4)所示。

$$\Delta_j = \sum_{i=1}^j \Delta X(i), j=1,2,\dots,n \quad (3)$$

$$R_n = \max(0, \Delta_1, \Delta_2, \dots, \Delta_n) - \min(0, \Delta_1, \Delta_2, \dots, \Delta_n) \quad (4)$$

为了便于分析和计算,需要将 R_n 进行归一化处理,即 R_n / S_n , 因此需要先计算 X_i 序列的方差,如式(5)所示。根据网络流量的自相似特性, R_n / S_n 与所划分的时间段数量 n 符合幂律特性 $R_n / S_n \propto n^H$, 两边取对数,可以得到式(6),其中 C 为常数。需要注意, n 值不能太小,否则所评估得到的 Hurst 参数值偶然性较大,使得其与实际流量的 Hurst 参数值误差较大。

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n E[(X_i - \bar{X})^2] \quad (5)$$

$$\log(R_n / S_n) = H \log(n) + C \quad (6)$$

将式(6)做适当变换,得到最终 H 参数的预测公式,如式(7)所示。

$$H = \frac{\log(R_n / S_n) - C}{\log(n)} \quad (7)$$

通过上述方式可对网络业务流的统计特性有一个初步评估,接下来需要进一步建立模型对网络流量进行预测。

1.2 网络流量离散小波变换

由于网络流量序列具有突发特性和非平稳特性，网络流量在时间序列上呈现出极度不稳定的形态^[13]，因此首先利用离散小波变换将原始非平稳流量分解为多个相对平稳的流量序列。通过变换的信号是原始信号的另一种表现方式，信号的内容不会随着变换而改变，即在变换过程中不会出现信号的衰减。通过离散小波变换把原始的网络流量信号分解为两个信号，即基于选定的尺度函数 ϕ_0 和小波函数 ψ_0 ，进而构建带通小波函数基 $\psi_{j,k}$ 和低通尺度函数基 $\phi_{j,k}$ ，这样原始信号可以通过小波基的平移和收缩来表示，即近似信号可以通过一系列小波系数来表示。小波函数基与小波函数以及尺度函数基与尺度函数的关系如式(8)所示。通过上述过程，可以得到离散小波变换的近似系数与细节系数，如式(9)所示，其中 $f(t)$ 为时间序列过程， M 为分解层数。

$$\begin{cases} \phi_{j,k}(t) = 2^{-\frac{j}{2}} \phi_0(2^{-j}t - k) \\ \psi_{j,k}(t) = 2^{-\frac{j}{2}} \psi_0(2^{-j}t - k) \end{cases} \quad (8)$$

$$\begin{cases} W_\phi(j,k) = \frac{1}{\sqrt{M}} \sum_t f(t) \phi_{j,k}(t) \\ W_\psi(j,k) = \frac{1}{\sqrt{M}} \sum_t f(t) \psi_{j,k}(t) \end{cases} \quad (9)$$

2 基于 FARIMA 的网络流量预测

本文前面已将网络流量进行了小波分解，通过分解的近似部分与细节部分需要建立模型对其进行预测。由于实际的网络流中存在自相似特性，而 FARIMA 模型可以同时描述网络流量的长相关特性与短相关特性，因此本文运用 FARIMA 模型对分解后的网络流量进行预测。刻画 FARIMA 模型的 3 个最重要的参数为 p, d, q ，其中 p 表示自回归项， q 表示移动平均项， d 表示差分阶数。通常 p 和 q 用来描述网络流量的短相关特性， d 用来描述网络流量的长相关特性。FARIMA 模型由 ARMA 模型演变而来，本文通过将 FARIMA 过程划分为“分数差分”和“ARMA 过程”两部分，以过程化的顺序执行方式预测网络流量。FARIMA 过程如式(10)所示，其中 $d \in (-0.5, 0.5)$ ， d 的取值是 FARIMA 模型与传统的 ARMA 模型的最大区别。 B 为滞后算子，其中 $-1 < B < 1$ ， ε_t 为白噪声序列，其均值为 0，方差为 σ^2 。 $\Phi(B^{-1})$ 为自回归多项式，其阶数为 p ， $\Theta(B^{-1})$ 为滑动平均多项式，其阶数为 q ，式(11)和式(12)分别为 $\Phi(B^{-1})$ 和 $\Theta(B^{-1})$ 的演化方程。

$$\Phi(B^{-1})(1 - B^{-1})^d X_t = \Theta(B^{-1})\varepsilon_t \quad (10)$$

$$\Phi(B^{-1}) = 1 - \phi_1 B^{-1} - \phi_2 B^{-2} - \dots - \phi_p B^{-p} \quad (11)$$

$$\Theta(B^{-1}) = 1 - \theta_1 B^{-1} - \theta_2 B^{-2} - \dots - \theta_q B^{-q} \quad (12)$$

分数差分过程是预测的关键，因此，首先将式(10)简化，如式(13)所示，其中 $\hat{X}_t$ 和 Δ^d 的表达式如式(14)和式(15)所示。

$$\hat{X}_t = \Delta^d X_t \quad (13)$$

$$\hat{X}_t = \Phi^{-1}(B^{-1})\Theta(B^{-1})\varepsilon_t \quad (14)$$

$$\Delta^d = (1 - B^{-1})^d \quad (15)$$

式中 Δ^d 为分数差分算子，可对其进行二项展开，表示为

$$\Delta^d = (1 - B^{-1})^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B^{-1})^k \quad (16)$$

则 $\hat{X}_t$ 可进一步表示为式(17)所示，其中 π_j 为滤波器的权值系数，其表达式如式(18)所示。

$$\hat{X}_t = \sum_{k=0}^{\infty} \binom{d}{k} (-B^{-1})^k X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad (17)$$

$$\pi_j = \frac{(-1)^j \Gamma(1+d)}{\Gamma(1+j) \Gamma(1+d-j)} \quad (18)$$

根据上述建模分析过程，即可对经过小波变换的网络流量进行预测，可得预测后的近似部分与细节部分，然后通过小波重构即可获得初始业务流的预测值。

3 数值结果分析

本文所采用的实验分析数据来源于 <http://newsfeed.ntcu.net/~news/2006/>, <http://newsfeed.ntcu.net/~news/2012/>, <http://newsfeed.ntcu.net/~news/2013/> 等多个流量文件, 使用多个流量文件的目的是为了防止流量文件过于单一, 单一流量文件容易出现过拟合等问题, 导致结果缺乏依据。流量文件主要收集了主节点路由器在 5 天之内不同时间段的访问流量, 如图 1 所示。通过 R/S 时间序列分析, 图 1 中时间流量序列的自相似指数为 0.906 5, 表明网络流量具有很高的自相似性。此外, 从图中可知网络的最大流量和最小流量之间存在较大差距, 表明网络流量的非平稳以及高突发性。

3.1 小波变换

本文使用小波变换将收集得到的网络流量分解成近似分量和细节分量两部分, 如图 2(a)、图 2(b)所示。近似分量反映了网络流量的变化趋势, 而细节分量表征噪声特性。

3.2 WCNTD 机制预测

为了验证本文提出的 WCNTD 机制的有效性, 本文将其与 ARIMA 预测机制进行比较。为了更好地评价两种机制的预测效果, 引入评价指标: 均方误差 (Mean Square Error, MSE), 归一化平均绝对误差 (Normalized Mean Absolute Error, NMAE), 其表达式如式(19)、式(20)所示。

$$MSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (X_i - \hat{X}_i)^2} \quad (19)$$

$$NMAE = \frac{\sum_{i=1}^n |X_i - \hat{X}_i|}{\sum_{i=1}^n X_i} \quad (20)$$

式中 $\hat{X}_i$ 表示预测网络流量值。显然, MSE 与 NMAE 越小, 预测误差越小, 预测性能越佳。

本文利用所提出的 WCNTD 机制对持续时间为 120 h 的网络流量数据进行预测, 采样间隔为 1 h, 预测效果如图 3 所示。图 3 中的 real traffic 数据同样来自于图 1 中的数据, 为了防止过拟合问题, 本文增大训练数据量, 使得训练数据量占总数据量的比例能够保持在一定程度。同时, 本文对数据中存在的纯净数据进行了清洗。因此, 参数估计与性能预测时不存在过拟合问题。可以看出, 通过 WCNTD 机制预测的网络流量更加趋近于真实网络流量。ARIMA 预测机制不能同时预测网络流量的长相关特性与短相关特性, 降低了预测的准确度。此外, ARIMA 预测机制并没有对原始网络流量进行初步处理, 进一步放大了 ARIMA 模型的弊端。而本文所提机制对网络流量进行小波变换, 将真实网络流量分解成近似部分与细节部分, 根据 FARIMA 模型可以同时描述网络流量的长相关特性与短相关特性, 完成对网络流量预测, 有效提高了预测准确度。

为了更加直观地反映 WCNTD 机制的特性, 对 WCNTD 和 ARIMA 的预测结果进行比较, 如图 4 和表 1 所示。图 4 横坐标为误差值区间, 纵坐标为统计比例。从图 4 可知, WCNTD 的误差区间为 (-25.1, 25.6), ARIMA 的误差区间为 (-36.9, 37.9), 而且对于误差区间在 (-10, 10) 之间的统计, WCNTD 明显高于 ARIMA, 提高了近 10%。从表 1 可以看出, 相比于 ARIMA, 本文提出的 WCNTD 机制的均方

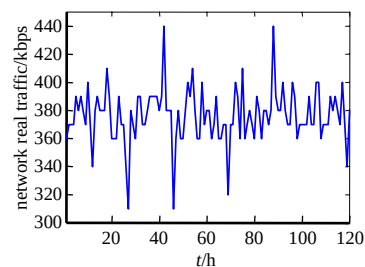
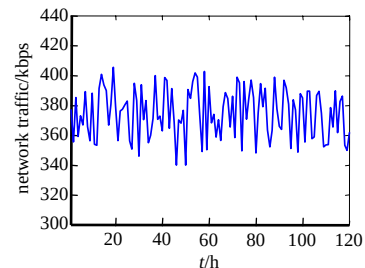
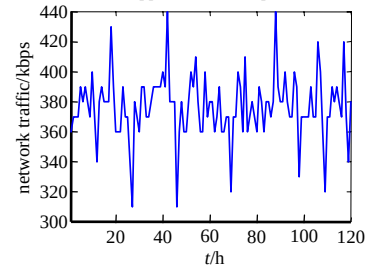


Fig.1 Network traffic
图 1 网络流量图



(a) approximate component



(b) detail component

Fig.2 Approximate component (a) and detail component (b)
图 2 近似分量(a)和细节分量(b)

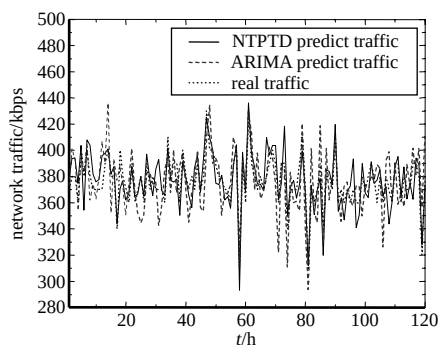


Fig.3 Comparison of WCNTD and ARIMA prediction results
图 3 WCNTD 和 ARIMA 预测结果对比

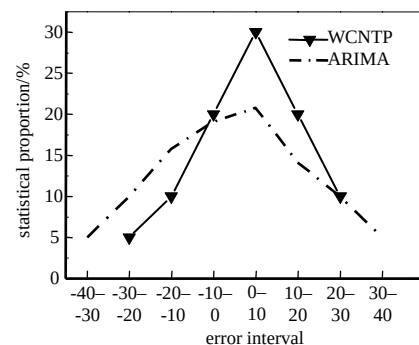


Fig.4 WCNTD and ARIMA error statistics
图 4 WCNTD 和 ARIMA 误差统计图

误差和归一化平均绝对误差均减少了 35%以上,表明小波分解和 FARIMA 模型在流量预测过程中的作用十分明显。

4 结论

针对长期网络流量预测准确度低,收敛速度慢的问题,本文提出一种小波系数感知的网络流量预测机制。该机制借助 R/S 序列分析法初步评估网络流量在大时间尺度上的统计特性,然后,利用离散小波变换将非平稳的网络流量分解为多个相对平稳的流量序列,最终,利用 FARIMA 模型对网络流量进行预测,极大地提高了长期网络流量预测的准确度,对维持网络平稳运行,提高网络服务质量意义深远。

参考文献:

- [1] WU D P,ZHANG H P,WANG H G,et al. Quality-of-protection-driven data forwarding for intermittently connected wireless networks[J]. IEEE Wireless Communication, 2015,22(4):66-73.
- [2] WU D P,ZHANG P N,WANG H G,et al. Node service ability aware packet forwarding mechanism in intermittently connected wireless networks[J]. IEEE Transactions on Wireless Communications, 2016,15(12):8169-8181.
- [3] QU H,MA W T,ZHAO J H,et al. Prediction method for network traffic based on maximum correntropy criterion[J]. China Communications, 2013,10(1):134-145.
- [4] ALARCON-AQUINO V,BARRIA J V. Multiresolution FIR neural-network-based learning algorithm applied to network traffic prediction[J]. IEEE Transactions on Systems, 2006,36(2):208-220.
- [5] MORALES F,RUIZ M,GIFRE L,et al. Virtual network topology adaptability based on data analytics for traffic prediction[J]. IEEE Journal of Optical Communications and Networking, 2017,9(1):35-45.
- [6] ABADI A,RAJABIOUN T,IOANNOU P A. Traffic flow prediction for road transportation networks with limited traffic data[J]. IEEE Transactions on Intelligent Transportation Systems, 2015,16(2):653-662.
- [7] 邹伯贤,刘强. 基于 ARMA 模型的网络流量预测[J]. 计算机研究与发展, 2002,39(12):1645-1652. (ZOU Boxian,LIU Qiang. Network traffic prediction based on ARMA model[J]. Computer Research and Development, 2002,39(12):1645-1652.)
- [8] 金旗,裴昌幸,朱畅华. ARIMA 模型法分析网络流量[J]. 西安电子科技大学学报, 2003,5(1):6-10. (JIN Qi,PEI Changxing,ZHU Changhua. Use ARIMA model to analyze network traffic[J]. Journal of Xidian University, 2003,5(1):6-10.)
- [9] 周贞贞,孙桦. 基于自适应结构的径向基神经网络模式识别[J]. 太赫兹科学与电子信息学报, 2013,11(3):474-479. (ZHOU Zhenzhen,SUN Hua. Pattern recognition using Radial Basis Function Neural network based on the adaptive structure[J]. Journal of Terahertz Science and Electronic Information Technology, 2013,11(3):474-479.)
- [10] CHAN K Y,DILLON T S. On-road sensor configuration design for traffic flow prediction using fuzzy neural networks and Taguchi method[J]. IEEE Transactions on Instrumentation and Measurement, 2013,62(1):50-59.
- [11] HALL J,MARS P. Limitations of artificial neural networks for traffic prediction in broadband networks[J]. IET Proceedings Communications, 2000,147(2):114-118.
- [12] WANG Y B,SCHUPPEN J H,VRANCKEN J. Prediction of traffic flow at the boundary of a motorway network[J]. IEEE Transactions on Intelligent Transportation Systems, 2014,15(1):214-227.
- [13] YU Y H,SONG M N,FU Y,et al. Traffic prediction in 3G mobile networks based on multifractal exploration[J]. Tsinghua Science and Technology, 2013,18(4):398-405.

作者简介:



林志达(1983-),男,广州市人,本科,高级工程师,主要研究方向为信息系统运行和安全管理.email:linzhida11@126.com.

吕华辉(1968-),男,广州市人,本科,高级工程师,主要研究方向为网络安全.

表 1 预测误差对比表

mechanism	prediction error	
	WCNTF	ARIMA
MSE	1.038	1.643 0
NMAE	0.023	0.038 8