

网络流量预测方法

董梦丽, 杨 庚, 曹晓梅

(南京邮电大学计算机学院, 南京 210003)

摘 要: 分析网络流量的行为特性并建立模型进行预测, 对于网络管理以及安全预警具有重要意义。基于此, 针对网络异常处理滞后、网络服务质量差等问题, 研究多种经典流量预测方法, 从流量特性、建模复杂性、预测精度及应用场景等多角度进行分析比较。实验结果证明, 预测模型与具体场景密切相关, 实际操作时需根据流量特性及预测目标选择合适的模型。

关键词: 时间序列; 小波变换; 神经网络; 预测精度; 复杂度

Methods of Network Traffic Prediction

DONG Meng-li, YANG Geng, CAO Xiao-mei

(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

【Abstract】 Modeling and predicting of network traffic are important for internet management and security prediction. Considering about drizzle disposal and bad service quality, different kinds of classic traffic prediction methods are reviewed. Comparison about their suitable data, modeling complexity, prediction precision and suitable scene are presented. Experimental results prove that prediction model should correlate to scene tightly. It need to choose different models according to the traffic characters and prediction target in practices.

【Key words】 time series; wavelet transform; neural network; prediction precision; complexity

DOI: 10.3969/j.issn.1000-3428.2011.16.033

在通信网络技术发展的过程中, 针对网络流量的建模和预测研究一直备受关注。本文研究分析多种经典流量预测方法, 并分别对各个模型适合的流量特性、建模复杂性、预测精度以及适用场景进行比较。

1 短相关流量特性方法描述

短相关流量特性主要通过线性时间序列模型描述, 如 AR、MA、ARMA 等。这类模型是基于传统的时间序列预测技术, 又称为 Box-Jenkins 方法, 其实质是用一个含有足够多参数的模型去拟合一系列的数据集, 使其达到最佳拟合度。这类模型预测算法简单、精度高, 但前提是用于刻画平稳流量序列, 且只能描述短相关特性的流量。

由于线性时间序列模型的基本原型就是 AR 模型和 MA 模型, 因此在此仅介绍两者的组合模型——ARMA 模型的基本概念^[1]: 把具有如下结构的模型称为自回归滑动平均模型, 简记为 ARMA(p, q), 其中, p 为自回归模型 AR 模型的阶数, q 为滑动模型 MA 模型的阶数。

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ Ex_t \varepsilon_t = 0 \quad \forall s < t \end{cases}$$

易知, 当 $q=0$ 时, ARMA(p, q)退化为 AR(p)模型。当 $p=0$ 时, ARMA(p, q)退化为 MA(q)模型。

由 ARMA(p, q)建立模型的前提是假设时间序列是平稳时间序列, 或通过差分法以分离趋势成分、周期成分和随机成分来进行流量的分析和建模。如果引入差分使得序列平稳后再用 ARMA 建模, 就成了 ARIMA 模型。线性时间序列模型计算简单, 可以用于对精度要求不是很高但要求能量损耗小、延时短的场景, 如 WSN 或在线预测。这类模型只能用于刻画平稳流量序列, 在将非平稳的网络流量数据向平稳状态转换时, 会造成部分流量信息的损失。此外, 由于大规模网络本身是复杂非线性系统, 同时又受多种复杂外界因素的影响,

其宏观流量行为往往复杂多变, 这些特点都难以用单一的 ARMA(p, q)模型来描述。

2 长相关流量特性方法描述

随着网络的不断发展, 在真实环境中的网络流量呈现出相当明显的非线性、多尺度特性, 如自相似性、长相关性、多重分形性、突发性等特性。这就表明传统的网络流量模型, 如 Markov 模型、Poisson 模型、AR 模型、ARMA 模型等, 已经不能有效地刻画网络流量的新特性, 因此一些能够捕捉流量长相关、非线性等特性的流量预测模型应运而生, 如 FARIMA^[1]、灰色模型^[2]、TAR^[3]以及神经网络模型^[4]、小波模型^[5]、SVM^[6]等人工智能的模型。

2.1 FARIMA 模型

FARIMA(p, d, q)是模型 ARIMA(p, d, q)差分阶数 d 取分数的过程^[1], $0 < d < 0.5$ 的 FARIMA(p, d, q)是一个平稳的长相关过程, 可以用于描述网络流量的长相关特性。

当 p 和 q 不同时为 0 时, FARIMA(p, d, q)可以看作被一个分数差分噪声 FARIMA(0, d , 0)驱动的 ARMA(p, q)过程, 由此也可以看出此模型可以同时长相关和短相关建模, d 代表着长相关的强度, p, q 代表短相关的强度。然而, 大量研究表明, FARIMA(p, d, q)在同时处理长相关和短相关方面有优势, 但计算量比较大, 不能够进行实时预测, 且不能很好地捕获突发性^[7]。

2.2 小波变换

小波分析(Wavelet Analysis)是处理非平稳时间序列最有效的一种方法。对于网络流量当中的长相关性, 小波变换在

基金项目: 国家自然科学基金资助项目(60873231)

作者简介: 董梦丽(1988—), 女, 硕士, 主研方向: 无线传感器网络安全; 杨 庚, 教授、博士; 曹晓梅, 副教授、博士

收稿日期: 2011-03-11 **E-mail:** yangg@njupt.edu.cn

数学上具有其自身特有的优势。其含义为: 把称为基本小波的函数 $\Psi(t)$ 做位移 τ 后, 再在不同尺度 a 下与待分析信号 $x(t)$ 做内积:

$$WT_X(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \varphi\left(\frac{t-\tau}{a}\right) dt \quad a > 0$$

经过小波分解, 时序上比较复杂的网络流量可以被分解为近似不相关的时间序列, 而且小波分解对时间序列进行了平滑, 这样小波方法可以将复杂的网络流量转化为用传统的预测方法来对分解后的时间序列进行预测。因此, 近年来小波方法在网络流量建模以及预测中应用越来越广泛, 但小波变换分解层数需要根据实际情况人工衡量确定, 增加了小波运算的复杂度。

2.3 神经网络模型

神经网络模型通过大量的神经元的复杂连接, 以及通过学习、自组织和非线性动力学所形成的并行分布方式来处理难以语言化的模式信息, 当前人工智能方法被用来进行网络流量建模, 其中应用最多的就是人工神经网络, 它适合学习比较复杂的非线性关系。

神经元为其基本处理单元, 一般表现为一个多输入、单输出的非线性器件, 神经元的输出函数有多种, 一般用 S 型函数表达式来表现函数的非线性特征:

$$f(\mu_i) = \frac{1}{1 + \exp(-\mu_i/c)^2}$$

其中, c 为常数; μ_i 为神经元 i 的内部状态。

神经网络模型以其非常强的容错性、快速的并行计算以及强大的学习功能弥补了传统线性预测模型的不足, 不需要预测数据的前期知识, 具有较强的学习能力, 因而在流量预测中应用较广。但在实际预测中, 选择什么样的神经网络, 以及神经网络的层次结构、神经元数、训练方法等都还没有一个明确的理论依据, 一般凭经验用试凑法来决定, 算法复杂, 这也是神经网络实际应用发展中的一个瓶颈。

2.4 灰色模型

灰色预测是灰色系统理论的组成部分, 其中应用较为广泛的是传统灰色模型 GM(1,1)^[2], 建模简单。

灰色系统建模是从系统的一个或几个离散数列中找出系统的变化关系, 建立系统的连续变化模型, 在反映数据的趋势性变化上效果明显且所需样本数据少, 预测精度高, 但灰色系统本身没有很强的并行计算能力, 系统的轻微变化都将导致重新计算, 这对于灰色模型的应用而言是非常不利的^[2]。

3 组合预测模型

以往的这些平稳的或非平稳的流量预测模型, 虽然在一定程度上实现了对网络流量的拟合, 但这些单一模型都存在各自的局限性, 为了能更全面更有效地反映网络流量的真实性, 一些混合性模型被相继提出, 如灰色神经网络模型、小波神经网络模型、小波与时间序列模型结合的模式等。

3.1 灰色神经网络模型

神经网络能有效修正灰色预测模型达到提高预测精度的目的。文献[2]提出一种最优组合灰色神经网络预测模型, 主要思想为:

设 f_1 是灰色动态群预测值, f_2 是神经网络预测值, f_c 是最优组合预测值, 预测误差分别为: e_1 和 e_2 , 取 w_1 和 w_2 为相应的权系数, 且有 $w_1 + w_2 = 1$, $f_c = w_1 f_1 + w_2 f_2$, 则误差为 $e_c = w_1 e_1 + w_2 e_2$, 方差为:

$$v_{ar}(e_c) = w_1^2 v_{ar}(e_1) + w_2^2 v_{ar}(e_2) + 2w_1 w_2 \text{cov}(e_1, e_2)$$

灰色系统和神经网络都可以看作数值化、非数学模型的函数估计器。利用灰色方法求解, 所需的计算量小, 在少样本情况下可达到较高精度, 利用神经网络计算精度高, 而且误差可控, 故两者融合起来, 优势互补。实验结果表明^[5], 组合模型效果及预测精度均优于单一灰色预测模型或神经网络模型。

3.2 小波与时间序列模型相结合的模式

要改进网络流量模型的预测精度及运算复杂度就必须同时吸取平稳模型和非平稳模型的优点, 通过仔细的研究和分析, 一些小波时间序列模型被相继提出。基本思想是: 先用小波变换来分解流量, 一层层地将流量分解到不同频域上, 分解后的信号在频域上比原始信号单一, 且对信号作平滑处理, 因此分解后信号的平稳性比原始信号好得多, 分解后的流量很适合用平稳时间序列模型来精确建模。这种组合模型一方面克服了小波模型、FARIMA 模型等一般长相关模型计算量大的缺点, 保持了算法的简单性; 另一方面弥补了传统平稳线性模型不能进行非平稳预测的缺陷, 提高了预测精度。

3.3 小波神经网络模型

小波变换是处理非平稳时间序列最有效的一种方法; 而人工神经网络具有较强的非线性逼近功能和自学习、自适应特点。基于两者优点, 将小波变换与人工神经网络进行耦合这一新的组合模型被提出^[5]。

小波变换与神经网络的结合一般有 2 种方法: (1) 将小波放在神经网络之前, 提前用小波对网络流量进行除噪, 然后利用单纯的神经网络进行流量预测。(2) 使用小波代替传统神经网络的 Sigmoid 作为神经网络的隐节点激励函数, 构成一个前馈型神经网络。第 1 种方法简单易于实现, 但神经网络学习周期长, 易陷入局部最优, 而第 2 种方法利用小波的多尺度分析特性, 缩短了小波神经网络的学习时间, 使神经网络避免陷入局部最优, 从而提高了神经网络的预测性能。

4 各模型的综合比较

第 1 代流量预测模型具有模型计算简单、预测精度高的特点, 但前提是流量为平稳序列, 且只能刻画流量中的短相关特性, 具有一定的局限性, 但由于建模简单, 时间序列模型常被用于无线传感器网络环境或在线预测场景。

第 2 代流量预测模型能有效的刻画流量的长相关、自相似特性且预测精度都比较高, 在一些网络或交通流量的预测中起着重要的作用, 但都仍然有自己的不足, 如 FARIMA 不能够描述流量的非平稳特性; 神经网络模型所需训练样本多, 算法复杂; 灰色模型仅能适用原始序列按指数规律变化且变化速度不快的场合; 小波变换法增加了不稳定模型的计算复杂度等。概括起来第二代模型的突出特性就是高精度高复杂度, 对能量和预测时间都是一个挑战, 所以一般都是应用于有足够能量支持的网络环境当中。

第 3 代流量预测模型在各单一模型的基础上突破其局限性, 扬长避短, 能更加有效地反映网络流量的真实特性, 但有时复杂度会有所增加, 虽然组合模型比单一模型预测精度高, 但有些组合模型的方法并没有减少各分量的预测误差, 而是在重构中也将误差进行了相应的重构, 并没有降低预测误差的干扰。因此, 组合网络预测模型更需要仔细分析各子模型优缺点, 结合实际场景需求, 实现模型的最佳结合, 进而最大限度地达到预测效果。

各模型的综合比较情况如图 1 所示。

指标	ARMA	FARIMA	小波变换	神经网络	灰色模型	灰色神经网络模型	小波时间序列模型	小波神经网络模型
适合描述流量类型	短相关	短相关或长相关				短相关或长相关		
算法复杂度	简单	FARIMA、小波变换、神经网络都比较复杂，灰色模型相对简单且需要的样本数据少				复杂		
预测精度	用于预测短相关流量时精度较高	FARIMA 适合描述短相关或长相关流量，预测精度较高；小波变换最适合描述非平稳流量特性，预测精度较高；神经网络模型在描述非线性流量方面有独特优势，精度较高；灰色模型在反映数据的趋势性变化上效果明显，但并行计算能力差，预测精度较高				灰色神经网络模型：采用神经网络对灰色模型预测残差进行修正，达到少样本高精度的效果；小波时间序列模型：克服了传统模型的缺陷，精度进一步得到提高；小波神经网络模型：将小波处理的高效性与神经网络逼近非平稳时间序列的强功能完美结合，精度比各单一模型有所提高		
适用场景	在线预测或能量有限的场景	对预测精度有较高要求，同时能量以及预测时间没有特殊限制的场景				对预测精度有很高要求，但同时能量供给或时间要求都比较宽松的场景		

图 1 各模型比较结果

5 结束语

本文对现有的网络流量预测模型，分别在其适用场景、预测精度、建模复杂度等方面进行了详细的分析和比较。现有的流量预测方法随着网络规模的不断扩大而不断发展，流量预测的建模要建立在流量特性的基础之上，每种预测模型都有自己的适用场景，如 WSN 网络环境中 ARMA 仍然是比较成熟的预测模型，而普通以太网中却用到了小波神经网络等组合模型，这也就解释了第 1 代预测模型仍未被淘汰的原因。在今后的工作中，不仅要发掘已有模型的优缺点，更要在已有工作的基础上从阶数估计、参数确定等方面出发，设计更适合场景的流量预测模型。

参考文献

[1] 魏武雄. 时间序列分析——单变量多变量方法[M]. 北京: 中国人民大学出版社, 2005.

[2] Liu Yuan, Cao Jianhua. Network Traffic Prediction Based on Grey Neural Network Integrated Model[C]//Proc. of 2008 International

Conference on Computer Science and Software Engineering. [S. l.]: IEEE Press, 2008.

[3] Yin Hao, Lin Chuang, Berton S. Network Traffic Prediction Based on a New Time Series Model[J]. International Journal of Communication Systems, 2005, 18(8): 711-729.

[4] 蔡娜娜, 陈月辉, 李 伟. 基于神经网络的蛋白质三级结构预测[J]. 计算机工程, 2010, 36(9): 176-177.

[5] 雷 霆, 余镇危. 一种网络流量预测的小波神经网络模型[J]. 计算机应用, 2006, 26(3): 56-58.

[6] Feng Huifang, Shu Yantai, Wang Shuyi. SVM-based Models for Predicting WLAN Traffic[C]//Proc. of IEEE International Conference on Communications. [S. l.]: IEEE Press, 2006.

[7] Wang Feng, Li Dou, Zhao Yuping. Prediction of Self-similar Traffic and Its Application in Network Bandwidth Allocation[C]//Proc. of 2007 International Conference on Wireless Communications. [S. l.]: IEEE Press, 2007.

编辑 陈 文

(上接第 83 页)

5.2 实验

本文采用一个通信系统的真实软件进行实验。分 2 个测试组对该软件的部分功能进行背靠背测试：第 1 组使用随机测试(Random Testing, RT)策略,第 2 组使用改进的 ART 策略。RT 策略不考虑测试用例之间的关系，测试用例完全随机生成。

实验结果表明，改进的 ART 策略比 RT 策略可以更快、更多地触发软件失效。RT 策略共执行测试用例 106 个，触发了 8 个软件失效；改进的 ART 策略共执行测试用例 83 个，触发了 12 个软件失效。改进的 ART 策略在计算测试用例相似性时需要额外的开销，但总体效果要优于 RT 策略。

6 结束语

本文在现有 ART 的基础上，结合软件工程实践，依据软件失效所表征的现象和规律，提出一种基于测试用例相似性度量的软件自适应测试策略。该策略是对现有 ART 的扩展和补充，能在一定程度上弥补现有 ART 的局限，其测试效果优于随机测试。由于该策略不需要软件源代码，只注重软件的执行序列和失效的外部行为，因此比较适合于第三方软件测试。下一步的研究工作将对测试用例的特征给出更详细的描述和分类，探索在黑盒测试条件下测试用例与失效之间的潜

在关系，寻求更多的理论依据和实验证明，并应用于自动化测试，从而进一步推广 ART 的应用。

参考文献

[1] Chen T Y, Leung H, Mak I K. Adaptive Random Testing[C]//Proc. of the 9th Asian Computing Science. Chiang Mai, Thailand: [s. n.], 2004.

[2] Mayer J, Schneckenburger C. Adaptive Random Testing with Enlarged Input Domain[C]//Proc. of QSIK'06. Beijing, China: IEEE Computer Society, 2006.

[3] Ciupa I, Leitner A, Oriol M, et al. ARTOO: Adaptive Random Testing for Object-oriented Software[C]//Proc. of ICSE'08. Leipzig, Germany: ACM Press, 2008.

[4] 孙 林, 周兴社, 董云卫, 等. 改进的局部 ART 测试算法研究[J]. 计算机科学, 2008, 35(9): 192-194.

[5] 陈永祥. 自适应随机测试研究与应用[D]. 上海: 华东师范大学, 2009.

[6] 林 鹏, 包晓露, 王小娟, 等. 基于自适应遗传算法的测试用例生成工具[J]. 计算机工程, 2010, 36(11): 50-51.

编辑 金胡考