

基于小波分解和支持向量机的网络流量组合预测

段 谟 意

(南京铁道职业技术学院 软件学院, 江苏 南京 210031)

摘 要: 研究网络流量预测问题, 网络流量具有突发性、周期性、非线性特点, 传统网络流量预测模型无法建立准确预测模型, 导致预测误差大, 预测精度低。为了提高网络流量的预测精度, 提出一种小波分解和支持向量机的网络流量预测模型。首先采用小波变换对网络流量进行分解, 把网络流量不同特性成分分离出来, 然后采用支持向量机对各分量进行预测, 最后采用小波变换对各分量预测结果进行重构, 得到网络流量的最终预测结果。仿真实验结果表明, 相对其它预测模型, 提高了网络流量的预测精度, 为网络流量预测优化提供了可靠依据。

关键词: 网络流量; 小波分解; 支持向量机; 粒子群算法

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2012)09-0193-04

Network Traffic Combination Forecasting by Wavelet Decomposition and Support Vector Machine

DUAN Mo-yi

(Nanjing Railway Vocational and Technical College, Nanjing 210031, China)

Abstract: Study of network traffic prediction, network traffic has a sudden, periodic, nonlinear characteristics, the traditional network traffic prediction model unable to establish accurate prediction model, lead to the prediction error, low accuracy of prediction. In order to improve the prediction accuracy of network traffic, put forward a kind of wavelet decomposition and support vector machine network traffic prediction model. Wavelet transform is adopted to network decomposition, the network traffic characteristics of components separated, then by using support vector machine to each component are predicted, the wavelet transform is adopted to predict the results of reconstruction of the components, network traffic has been the final forecasting result. The simulation results show that, compared with other prediction models, improve the prediction accuracy of network traffic, for the prediction of network traffic flow optimization provides reliable basis.

Key words: network traffic; wavelet decomposition; support vector machine; PSO

1 引言

随着网络通信技术发展, Internet 已经渗透到了社会的各个领域, 网络承载的应用服务也越来越多, 网络行为特征日益复杂, 这对网络管理、流量控制提出了更高的要求, 分析与预测网络流量具有越来越重要的现实意义^[1]。

早期网络流量预测模型主要为泊松分布和马尔可夫模型, 它们主要针对小规模网络, 随着网络规模增大, 这些方法已经不再适合网络流量的预测和分

析^[2]。Klivansky 等人根据网络流量自相似特性, 提出了多分形小波模型^[3]。随后有学者根据网络流量动态性采用时间序列模型来预测网络流量, 如 ARIMA、ARMA 等预测方法^[4]。这些传统预测模型只能处理平稳网络流量, 而现代大规模网络本身是一种非线性系统, 会随着时间、外部环境等不断地变化, 网络流量具有极其复杂的特性, 这些传统模型难以准确、全面刻画网络流量复杂的非线性关系^[5]。近些年, 网络流量分析和非线性预测受到了许多学者的关注, 引入小波理论、支持向量机理论和神经网络

理论等非线性预测方法,大量实践研究表明它们取得良好的预测效果^[6].然而网络流量受到上网人行为、经济、节假日、周末等因素影响,具有周期性、多尺度、非线性变化特点,极其复杂,传统单一预测模型不能够全面刻画网络流量变化特性,不可避免地会产生较大的预测误差,如何提高网络流量预测精度,成为网络流量预测中急待解决的难题^[7].

为了进一步提高网络流量预测精度,提出一种基于小波分解和支持向量机相结合的网络流量预测模型(WA-SVM).首先采用小波分析(WA)对网络流量进行分解,把网络流量不同特性成分分离出来,然后采用支持向量机(SVM)对各特性进行预测,最后采用 WA 对各特征预测结果进行重构,得到网络流量最终预测结果.

2 网络流量特性分析

由于网络异常、带宽拥塞、病毒入侵、用户上网行为、节假日等多种因素的影响,真实网络流量具有错综复杂的动态特性^[8].图 1 给出了实际网络流量数据,通过对图 1 的网络流量进行观察和分析,可以发现网络流量具有如下特性:

(1) 突发性.由于短时间内发送大量数据导致网络流量不同的时间尺度上具有突发性,其中图 1 种的尖峰表示网络流量的突发性.

(2) 周期性.从图 1 可知,网络流量在一定时间内呈现周期性变化,一般呈 24 小时为一个周期.

(3) 长相关性.在图 1 中,网络流量序列具有相同的统计特性,即长相关特性.

(4) 非线性.网络流量具有多变性,其变化过程与时间、用户数和网络拓扑结构均相关联,造成了网络流量的非线性.

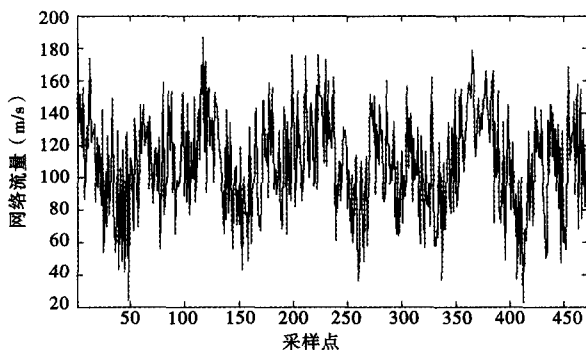


图 1 网络流量原始数据

综合上述分析可知,网络流量表现出明显多尺度,如自相似、长相关等特征,且周期性、混沌性等复

杂特性,要对网络流量进行准确预测,首先要解决的问题就是从实际网络流量中提取重要特性. WA 可以由粗到细地逐步观察信号,具有多分辨率特性,同时经过小波分解后,分解后的各分量彼此不相关,因此采用 WA 对网络流量进行分解,可以得到网络流量的不同特性成分,然后采用预测能力强的 SVM 对网络流量分量进行预测,最后重构成为网络流量的最终预测结果.

3 小波分析和支持向量机的网络流量建模

3.1 网络流量的小波分析

3.1.1 小波的定义

设 $\psi(t)$ 表示一平方可积函数,即有 $\psi(t) \in L^2(R)$,那么该函数的傅里叶变换 $\psi(\omega)$ 满足如下条件:

$$C_\psi = \int_R |\psi(\omega)|^2 / |\omega| d\omega < \infty \quad (1)$$

那么 $\psi(t)$ 就被称之为小波母函数.

对 $\psi(t)$ 进行伸缩和平移可以得到:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}|\psi[(t-\tau)/a]|} \alpha \quad (2)$$

其中, τ 表示平移因子, α 表示伸缩因子.

那么 $\psi_{a,\tau}(t)$ 被称之为 α, τ 的小波基函数^[9].

3.1.2 小波变换的 α Trous 算法

网络流量是一种离散时间序列数据,因此本文采用 α Trous 离散小波变换算法对其进行分解和重构.设获得网络流量时间序列为: $X(t), t=0, 1, \dots, N-1$,采用离散低通滤波器 h 可以得到网络流量分解尺度系数为:

$$c_{j+1}(t) = \sum_{l=-\infty}^{+\infty} h(l)c_j(t+2^j l) \quad (3)$$

根据 α Trous 离散小波变换性质,网络流量各尺度下的细节系数 d_j 可以表示为:

$$d_{j+1}(l) = c_j(t) - c_{j+1}(t) \quad (4)$$

那么网络流量时间序列的分辨率为 L 的小波变换为:

$$D = \{d_1, d_2, \dots, d_L, c_L\} \quad (5)$$

其中, $\{d_1, d_2, \dots, d_L\}$ 表示网络流量各尺度下的细节信, $\{c_L\}$ 表示网络流量近似信号.

网络流量时间序列 $X(t)$ 可以由细节信号和近似信号进行重构,即

$$X(t) = c_0(t) = c_L(t) + \sum_{j=1}^L d_j(t) \quad (6)$$

3.1.3 网络流量的小波分解步骤

设网络流量输入信号为 $X(t)$, $h(k)$ 表示小波变换低通滤波器, $g(k)$ 表示小波变换高通滤波器, 小波分解的级数为 L , 网络流量的小波分解步骤如下:

(1) 网络流量初始化.

$$c_0 = X(t), h_1(k) = h(k), g_1(k) = g(k), j = 1.$$

(2) 网络流量小波分解.

$$c_j[k] = c_{j-1}[k] * h_j[-k] \quad (7)$$

$$d_j[k] = d_{j-1}[k] * g_j[-k] \quad (8)$$

其中, $*$ 表示卷积运算.

(3) 对滤波进行插零处理.

$$h_{j+1}(k) = \begin{cases} h_j\left[\frac{k}{2}\right], & n \text{ 为偶数} \\ 0, & n \text{ 为奇数} \end{cases} \quad (9)$$

$$g_{j+1}(k) = \begin{cases} g_j\left[\frac{k}{2}\right], & n \text{ 为偶数} \\ 0, & n \text{ 为奇数} \end{cases} \quad (10)$$

(4) $j = j + 1$, 如果 $j < L$ 那么就返回到步骤(2)继续分解, 否则就结束分解.

综合所述, 网络流量的小波分解过程如图2所示. 通过对网络流量进行 α Trous 小波分解, 可以很好对网络流量不同成分进行分解, 为后继网络流量采用 SVM 进行建模提供基础.

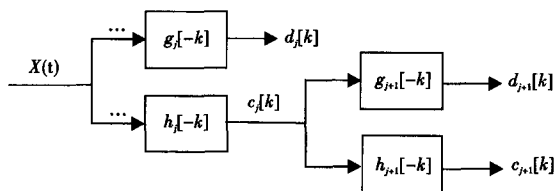


图2 网络流量的小波分解过程

3.2 网络流量的支持向量机建模

SVM 是一种基于结构风险最小化原则的新机器学习方法, 较好地解决了传统机器学习算法存在的问题, 非线性学习能力优异, 泛化能力强, 非常适合于具有时变性、非线性变化的复杂网络流量预测^[10].

网络流量经过小波分解, 原始网络流量序列中的低频信息被分离, 即获得了网络流量的周期性和长期趋势规律, 很好地消除原始网络流量中的突变部分; 高频分量很好地描述了网络流量的突变性、短相关性和随机性等特征, SVM 选择最适合的核函数, 然后对各个分量进行预测, 最后对各分量预测结果进行重构得到网络流量的最终预测结果.

设获得的网络流量分量为 $\{x_i\}$, m 表示网络流量的嵌入维数, 那么基于支持向量机的网络流量一

步预测模型为:

$$x_t = \text{SVM}(x_{t-1}, x_{t-2}, \dots, x_{t-m}) \quad (11)$$

其中, SVM 表示支持向量机模型.

3.3 网络流量的小波重构步骤

SVM 对网络流量进行预测后, 采用 WA 对网络流量进行重构, 小波重构实际上是小波分解逆过程, 即有

$$c_{j-1} = c_j \bar{h} + d_j \bar{g} \quad (j = 0, 1, 2, \dots, L) \quad (12)$$

其中, $\bar{h}$ 和 $\bar{g}$ 表示是 h 、 g 的对偶算子.

网络流量的小波重构过程如图3所示.

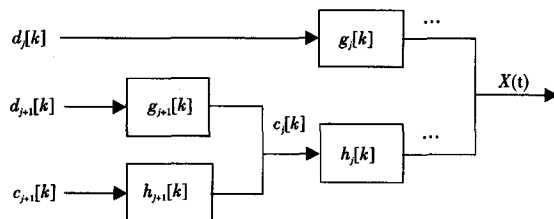


图3 网络流量的小波重构过程

3.4 网络流量建模预测流程

(1) 对网络流量数据进行收集, 得到原始网络流量时间序列 $X(t)$.

(2) 网络流量的小波分解. 设置小波分解尺度, 将原始网络流量分解成不同成分的网络流量分量.

(3) 网络流量分量的归一化处理. 由于网络流量的各分量变化范围大, 为了加快 SVM 学习速度和预测精度, 对其进行归一化处理, 具体为:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

式中, x_i 是原始网络流量, x'_i 归一化后的网络流量, $x_{\max}$, $x_{\min}$ 分别代表网络流量最大值和最小值.

(4) 模型初始化. 将数据分成为训练样本集和测试样本集, 采用 10 折交叉验证法确定每一个分量的嵌入维数, 然后根据嵌入维数确定 SVM 的输入向量和输出向量.

(5) SVM 核函数及参数确定. 对于网络流量的高频部分, 其近似于平稳, SVM 核函数选择线性核函数; 对于网络流量的低频部分, 其具有显著非平稳和非线性, SVM 核函数选择高斯核函数, 它们的参数均采用粒子群算法优化得到.

(6) SVM 训练和预测. 采用上述方法得到的参数对训练集进行学习, 建立预测模型, 然后对测试集进行预测.

(7) 对各个分量预测结果进行反归一化, 反归一化公式为:

$$x_i = x'_i \times (x_{\max} - x_{\min}) + x_{\min} \quad (14)$$

(8) 对归一化后网络流量分量采用小波变换技术进行重构,得到网络流量最终预测结果。

4 仿真实验

4.1 数据来源

对我院网站服务器的路由器网络流量进行实时采集,共获是 472 个数据点,前 372 个数据组成训练集进行建模,后 100 个数据作为测试集,对网络流量预测模型性能进行检验。仿真网络流量原始数据如图 1 所示。

4.2 网络流量的小波分解

首先对图 1 进行小波分析,得到的分析结果如图 4 所示。从图 4 网络流量频率可知,实际网络流量十分复杂,具有周期性、非线性、突变性,因此必须采用小波技术首先对其进行分解成各分量,取分解尺度为 3,得到分解结果如图 5 所示。

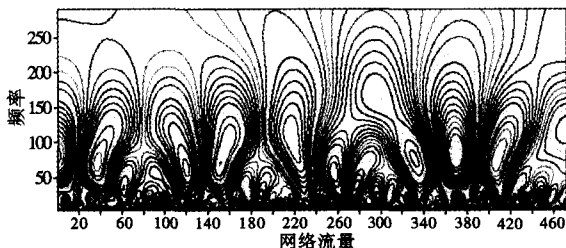


图 4 原始网络流量的小波分析

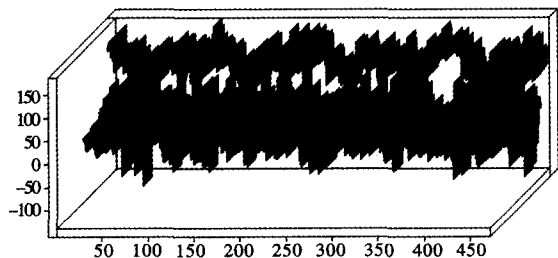


图 5 网络流量的小波分解结果

4.3 预测结果与分析

采用 SVM 分别对各分量的训练样本进行学习,并采用粒子群算法对 SVM 模型的参数进行优化,然后对测试样本集进行预测,得到各分量的预测结果,最后对各分量预测结果进行重构,得到的网络流量最终预测结果,其如图 6 所示。

从图 6 可知,模型的预测结果与实际值相当的接近,仿真结果表明,小波分解和 SVM 的结合,可以利用两者的优势,全面、准确地刻画网络流量的变量规律,从而有效地提高网络流量的预测精度。

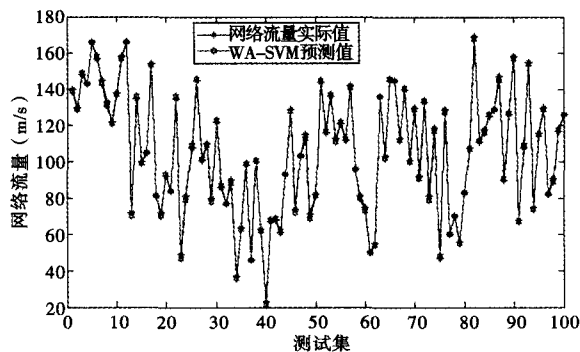


图 6 网络流量时间序列的拟合结果

为了对比本文模型(WA-SVM)与其它模型的优劣性,采用单一 SVM、小波分解和 BP 神经网络相结合(WA-BPNN)作为对比模型,采用预测结果的均方误差(MSE)作为模型的评价指标。具体结果如表 1 所示。

表 1 不同模型的预测结果

预测模型	MSE
SVM	9.765
WA-BPNN	8.755
WA-SVM	1.268

从表 1 的对比结果可知,WA-SVM 的预测精度最高,预测误差最小,WA-SVM 和 SVM 的预测结果对比可知,对网络流量进行小波分解可以很好对网络流量各分量进行分离,有利于提高网络流量的预测结果。同时将 WA-SVM 和 WA-BPNN 相比,WA-SVM 的预测相应较高,说明 SVM 的非线性预测能力要优于 BPNN,因此本文将小波分解和 SVM 相结合起来,能进一步提高了网络流量的预测准确率。

5 结束语

网络流量预测和分析是网络管理的基础,因其受到多种因素共同作用,具有非平稳性、非线性变化特点,传统预测方法预测准确率低。文中提出一种小波分解和 SVM 相结合的网络流量预测模型。仿真实验结果表明,本文算法能有效地提高网络流量预测准确率,是一种泛化能力强、鲁棒性好的网络流量预测模型。

(下转第 200 页)

表 1 两种检测方式准确率

算法	关键词抽取 样本数目	漏检率 (%)	检测准确率
传统 TFIDF 算法	200	23.1%	53%
加入贝叶斯统计 的 TFIDF 算法	200	16%	67%

4 结束语

本文提出一种基于贝叶斯推理和 TFIDF 算法的中文关键词智能抽取方法. 利用贝叶斯统计原理对文本信息进行概率化的统计, 运用贝叶斯决策理论对 TFIDF 算法进行优化, 克服传统的 TFIDF 算法存在着缺陷. 实验结果表明, 优化后的 TFIDF 算法在进行中文关键词智能抽取中, 抽取的准确性大幅提高. 实验证明, 本文提出的检测方法, 取得了很好的效果, 有较大的使用价值和应用前景.

参考文献:

- [1] Brendan Kitts, David Freed, Martin Vrieze. Cross-sell: A Fast Promotion Tunable Customer-Item Recommendation Method Based on Conditionally Independent Probabilities[C]// Proceedings of KDD, [S. L.]: Based on ACM Press, 2000: 437-446.

- [2] Mladenec D. Personal web watcher: design and implementation [M]. Department for Intelligent System, J. Stefan Institute, 2000.
- [3] Susan Gauch, Jason Chaffee. Ontology-based personalized search and browsing[J]. Web Intelligence and Agent Systems 2003, 1(3): 219-234.
- [4] Konstan JA, Miller B N Maltz, Herlocker D, et al. GroupLens: applying collaborative filtering to usenet news[J]. Communications of the ACM, 1997: 165-169.
- [5] YK Woon, WK Ng, EP Lim. Online and incremental mining of separately—grouped Web access Logs[C]// Proceedings of the 3rd International Conference on Web Information Systems Engineering . singapore: IEEE, 2002 :1-10.
- [6] 李艳美; 张卓奎. 基于贝叶斯网络的数据挖掘方法[J]. 计算机仿真, 2008 (02): 117-119
- [7] 陈作聪 孙树峰 基于混沌算法的无线安全认证技术研究[J]. 微电子学与计算机, 2007(12): 115-117.

作者简介:

胡局新 男, (1979—), 硕士, 讲师. 研究方向为数据库技术、智能计算.

鞠训光 男, (1966—), 博士, 副教授. 研究方向为智能计算、数据挖掘.

(上接第 196 页)

参考文献:

- [1] Juva I, Vaton S, Virtamo J. Quick traffic matrix estimation based on link count covariance[C]// proceedings of the IEEE Intl Conf on Icc, Istanbul: IEEE Communication Society, 2006: 603-608.
- [2] 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.
- [3] 洪飞, 吴志美. 基于小波的多尺度网络流量预测模型[J]. 计算机学报, 2006, 29(1): 166-170.
- [4] 曹建华, 刘渊, 戴悦. 一种基于灰色神经网络的网络流量预测模型[J]. 计算机工程与应用, 2008, 44(5): 155-157.
- [5] 冯海亮, 陈涤, 林青家, 等. 一种基于神经网络的网络流量组合预测模型[J]. 计算机应用, 2006, 26(9): 2206-2208.

- [6] 李海燕, 李咚. 改进的神经网络在网络流量预测中的应用研究[J]. 计算机仿真, 2011, 28(9): 132-135.
- [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] 田妮莉, 喻莉. 一种基于小波变换和 FIR 神经网络的广域网网络流量预测模型[J]. 电子与信息学报, 2008, 30(10): 2499-2502.
- [9] 蔡晓丽, 宁慧. 基于最大熵算法网络流量预测模型研究[J]. 计算机仿真, 2011, 28(10): 107-110.
- [10] 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.

作者简介:

段谟意 男, (1964—), 本科, 副教授. 研究方向为计算机网络.