

基于卡尔曼滤波和小波的网络流量预测算法研究

李捷 侯秀红 韩志杰

(河南大学计算机与信息工程学院 开封 475001)

摘要: 流量预测是流量工程, 拥塞控制和网络管理的核心问题。该文针对网络流量的特点, 将卡尔曼滤波和小波分析混合的预测算法引入到网络流量预测领域中, 对其进行了理论证明。仿真结果表明, 该算法与传统的算法相比, 具有较高的预测精度和较好的实时性与广谱性。

关键词: 流量预测; 小波; 卡尔曼滤波;

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2007)03-0725-04

Application of Kalman Filter and Wavelet in Traffic Prediction

Li Jie Hou Xiu-hong Han Zhi-jie

(College of Computer & Information Engineering, Henan University, Kaifeng 475001, China)

Abstract: Traffic prediction is the core of network quality of service problems, such as traffic engineering and congestion control, etc. According to the characters of traffic, a novel network traffic prediction algorithm in which Kalman filter and wavelet are mixed is presented and proved abstractly. The simulation results show that the proposed algorithm can guarantee higher precision and better real-time processing compared with traditional algorithm.

Key words: Traffic prediction; Wavelet; Kalman filter

1 引言

传统的网络流量管理已经不能适应网络规模的飞速发展, 业务管理(service management)以其支持 QoS, 优化网络配置和流量工程, 有效地提高了网络运行速度和利用率, 逐渐成为新一代网络管理的核心。网络流量预测是业务管理的关键问题, 网络流量预测的精确性, 实时性, 广谱性直接关系到业务管理的效率和性能。

现有的网络流量预测分为线性预测和非线性预测。其中 ARIMA^[1-3]作为线性预测方法的代表得到了广泛应用。ARIMA的理论前提是网络流量具有线性宽平稳过程特征。文献[4,5]提出并验证了网络流量具有多构性, 自相似性, 突然继发性。文献[6]提出并论证了网络流量在不同的时间频率尺度上具有自相似性和多尺度特征。因此ARIMA模型预测的精度低, 无法准确地描述出网络的全部特征。文献[7]提出了基于卡尔曼滤波的流量预测, 它引入了状态方程和测量方程, 有效地处理了系统噪声和测量噪声, 从而在一定程度上提高了预测精度, 但其模型的限制无法准确地描述出真实网络的自相似性, 多构性以及多尺度性特征。针对上述不足, 又出现了以小波分析^[8]和神经网络^[9]为代表的非线性预测研究。但是基于小波模型的预测实时性较差^[10]。基于神经网络的预测方法, 其缺点是收敛速度太慢, 并且容易陷入局部次

优, 并且时间空间复杂度太高。

文献[11]将卡尔曼滤波(KF)和小波分析相结合以充分地发挥各自的优秀特性, 成功地应用在电力预测中以及天气预报等动态随机预测中, 有效地提高了预测精度。在对上述文献研究的基础上, 本文基于文献[11]所提出的预测方法, 结合网络流量自身的特殊性, 提出了针对网络流量的预测模型, 很好地适应了网络流量的不确定性和突发性。

网络流量同时包含有线性变化部分, 非线性变化部分^[8,12]。对线性变化部分, 可利用卡尔曼滤波进行处理; 对非线性部分则采用小波分析进行处理。通过将目标状态的小波变换系数描述为KF方法的状态变量, 进而将KF和多尺度分析方法相结合, 提出了既具有实时性和递归性又具有多尺度分析能力的小波-卡尔曼滤波混合估计与预报方法 WKHEFA(Wavelet-Kalman filtering Hybrid Estimating and Forecasting Algorithm)。

下文组织如下: 第2节建立了 WKHEFA 的数学模型; 第3节对算法进行了描述和数学证明; 第4节对仿真结果进行比较分析; 第5节是结束语。

2 WKHEFA 的预测模型

初始网络流量预测模型^[7]如式(1)所示, 下面利用卡尔曼滤波和小波分析对其进行展开, 获取本算法的预测模型:

$$\frac{d}{dt} P_n(t) = (N+1-n)\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t) - [(N-n)\lambda + n\mu]P_n(t) \quad (1)$$

2005-8-16 收到, 2006-06-26 改回

国家自然科学基金重点项目(No.60434020)、河南省自然科学基金(0411014100)和河南大学校内重点基金(XK03YBSW0138)资助课题

其中 $P_n(t)$ 定义为报文到达的数目, λ 为 ON 时间均方指数, μ 为 OFF 时间指数。

根据式(1)得到 Kalman 预测模型:

$$\tilde{x}(k) = \hat{x}(k-1) + B\bar{u}(k-1) \quad (2)$$

$$\tilde{P}(k) = \tilde{P}(k-1) + B^2Q(n-1) \quad (3)$$

和 Kalman 估计模型:

$$\hat{x}(k) = \tilde{x}(k) + K(k)[y(k) - C(\tilde{x}(k))] \quad (4)$$

$$K(k) = \tilde{P}(k)C / [\tilde{P}(k)C^2 + R(k)] \quad (5)$$

$$\tilde{P}(k) = [1 - K(k)C]\tilde{P}(k) \quad (6)$$

其中 $B=D/C$

$$\left. \begin{aligned} C &= \frac{\gamma}{\lambda + \mu} [1 - e^{-(\lambda + \mu)T}] \\ D &= \frac{\gamma}{\lambda + \mu} [(\lambda + \mu)T - (1 - e^{-(\lambda + \mu)T})] \\ \bar{u} &= \frac{N\lambda}{\lambda + \mu} \end{aligned} \right\} \quad (7)$$

$Q(k)$ 和 $R(k)$ 分别是系统噪声和测量噪声, $\tilde{P}(k)$, $\hat{P}(k-1)$ 分别是预测误差和估计误差, $\bar{u}(k)$ 是系统噪声均方差, $K(k)$ 是滤波增益

根据式(1)–式(7), 建立网流量预测模型为

$$P_L(t) = B(t) + W(t) + v(t) \quad (8)$$

其中 $P_L(t)$ 表示网络流量总量, $B(t)$ 表示网络流量平稳部分, $W(t)$ 表示周期性变化部分, $v(t)$ 表示随机变化部分, 并且

$$B(t) = Cx(t), \quad W(t) = Du(t) \quad (9)$$

2.1 平稳变化模型

对 $B(t) = Cx(t)$ 进行离散小波分解, 即

$$B(t) = \sum_{m=1}^{2^L} bc_{s,m}^L \phi_{L,m}(t) + \sum_{j=L}^{M-1} \sum_{m=1}^{2^j} bd_{s,m}^j \psi_{j,m}(t) \quad (10)$$

其中 $\phi_{L,m}(t)$ 是尺度函数, $\psi_{j,m}(t)$ 是小波函数, 且 2^L 为最细尺度, 且

$$bc_{s,m}^L = \langle B(t), \phi_{L,m}(t) \rangle, \quad bd_{s,m}^j = \langle B(t), \psi_{j,m}(t) \rangle \quad (11)$$

根据网络流量的自相似特征, 对小波分析系数进行处理^[6], 这里定义处理过程为BWT(Base Wavelet Transform), 定义尺度因子 $u(k) = bc_{s,m}^L / (bd_{s,m}^j)$, 且 $u(k) \geq 1$ 。其中 $u(k)$ 的分布约为正态分布且均方为 0, 其分布概率函数定义为 $N(j)$, 则BWT过程为: 在最细尺度 j 上, 根据 $N(j)$ 生成随机变量 $u(j)$, bc_j , 得到

$$bd_j = bc_j / u(j) \quad (12)$$

重复以上步骤直到 $j = 0$, 最终得到

$$\mathbf{WB} = [bc_0^0 \quad bd_0^0 \quad bd_0^1 \quad bd_1^1 \quad bd_1^2 \quad bd_1^3 \quad bc_2^2 \cdots bc_{2^{M-1}-1}^{M-1}]^T \quad (13)$$

2.2 周期性变化模型

对 $W(t) = Du(t)$ 进行离散小波变换, 即

$$W(t) = \sum_{m=1}^{2^L} ac_{s,m}^L \phi_{L,m}(t) + \sum_{j=L}^{M-1} \sum_{m=1}^{2^j} ad_{s,m}^j \psi_{j,m}(t) \quad (14)$$

其中 $\phi_{L,m}(t)$ 是尺度函数, $\psi_{j,m}(t)$ 是小波函数, 且 2^L 为最细尺度, 且

$$bc_{s,m}^L = \langle W(t), \phi_{L,m}(t) \rangle, \quad bd_{s,m}^j = \langle W(t), \psi_{j,m}(t) \rangle \quad (15)$$

得到

$$\mathbf{WA} = [ac_0^0 \quad ad_0^0 \quad ad_0^1 \quad ad_1^1 \quad ad_1^2 \quad ad_1^3 \quad ad_2^2 \cdots ad_{2^{M-1}-1}^{M-1}]^T \quad (16)$$

2.3 KF 方程

根据上述模型可得

$$Z(k) = W^T \cdot \mathbf{WB} + W^T \cdot \mathbf{WA} + V(k) \quad (17)$$

其中 W_k^T 为 $\phi_{M,r}(t)$ 尺度函数的系数矩阵的逆阵, 据此求得 KF 方程为

$$\left. \begin{aligned} X(k) &= \Phi(k, k-1)X(k-1) + J(k) \\ Z(k) &= C(k)X(k) + V(k) \end{aligned} \right\} \quad (18)$$

其中转移阵 $\Phi(k, k-1)$ 为 $\mathbf{I}_{N \times N}$, \mathbf{I} 表示单位阵, $J(k)$, $V(k)$

分别为白噪声模型, $X(k) = \begin{bmatrix} \mathbf{WB} \\ \mathbf{WA} \end{bmatrix}$ 。

3 WKHEFA 算法

3.1 算法描述

记

$$Z(k) = [z(k_1), z(k_2), \dots, z(k_N)]^T \quad (19)$$

$$Z_1^k = [Z^T(1), Z^T(2), \dots, Z^T(k)] \quad (20)$$

$Z(k)$ 表示 k 周期内顺序获得的测量值序列 $z(k_1), z(k_2), \dots, z(k_N)$ 集合, Z_1^k 表示 $1, 2, \dots, k$ 周期所有时刻的测量值序列集合。

步骤 1 建立测量方程和状态方程如式(18);

步骤 2 获得初值 $= \hat{X}, P_0, K_0$;

步骤 3 在第 $(k-1)$ 周期

(1) 获得 $\hat{X}(k-1 | k-1)$, $P(k-1 | k-1)$

(2) 计算 $\hat{X}(k | k-1)$, $P(k-1 | k-1)$, 其中

$$\hat{X}(k | k-1) = E\{X(k) | Z_1^{k-1}\} \quad (21)$$

$$P(k | k-1) = E\{[X(k) - \hat{X}(k | k-1)] \cdot [X(k) - \hat{X}(k | k-1)]^T\} \quad (22)$$

(3) 对 $\hat{X}(k | k-1)$ 小波重构, 获得 k 周期 $S(k)$ 的预测值:

$$\hat{S}(k | k-1) = E\{S(k) | Z_1^{k-1}\} \quad (23)$$

步骤 4 在第 (k) 周期内, 依次取得 $Z(k)$ 的值

For $k = 1$ to N

Begin

(1) 用 $z(k_i)$ 对 $\hat{X}_e(k | k_1, k_2, \dots, k_{i-1})$ 进行更新, 得到状态 $X(k)$ 基于 Z_1^{k-1} 和观测信息 $z(k_1), z(k_2), \dots, z(k_i)$ 的估计值和相应的估计误差协方差阵:

$$\hat{X}_e(k | k_1, k_2, \dots, k_i) = E\{X(k) | Z_1^{k-1}, z(k_1), z(k_2), \dots, z(k_i)\} \quad (24)$$

$$P_e(k | k_1, k_2, \dots, k_i) = E\{\tilde{X}(k | k_1, k_2, \dots, k_i)\tilde{X}^T(k | k_1, k_2, \dots, k_i)\} \quad (25)$$

(2)对 $\hat{X}_e(k | k_i)$ 进行小波重构可得 $S(k)$ 基于 Z_1^{k-1} 和观测信息的 $z(k_1), z(k_2), \dots, z(k_i)$ 估计值

$$\hat{S}(k | k_1, k_2, \dots, k_i) = E\{S(k) | Z_1^{k-1}, z(k_1), z(k_2), \dots, z(k_i)\} \quad (26)$$

End

从而实现基于 Z_1^{k-1} 和观测信息 $z(k_1), z(k_2), \dots, z(k_i)$ 对 $s(k_i)$ ($l = i + 1, \dots, N$) 进行预报, 计算公式如式(27)和式(28)所示:

$$\hat{s}(k_l | k_1, k_2, \dots, k_i) = E\{s(k_l) | Z_1^{k-1}, z(k_1), z(k_2), \dots, z(k_i)\}, \quad l = i + 1, \dots, N \quad (27)$$

$$\tilde{X}(k | k_1, k_2, \dots, k_i) = X(k) - \hat{X}_e(k | k_1, k_2, \dots, k_i), \quad 1 \leq i \leq N \quad (28)$$

步骤5 根据以上步骤3和步骤4, 最终获得状态 $X(k)$ 基于 Z_1^k 的估计值和相应的估计误差协方差阵。即

$$\hat{X}(k | k) = \hat{X}_e(k | k_N) = E\{X(k) | Z_1^k\} \quad (29)$$

$P(k | k) = P_e(k | k_N)$ 及 $S(k)$ 基于 Z_1^k 的估计值:

$$\hat{S}(k | k) = \hat{S}(k | k_N) \quad (30)$$

步骤6 重复步骤3, 步骤4, 步骤5, 直到处理完所有周期。

算法流程如图1所示

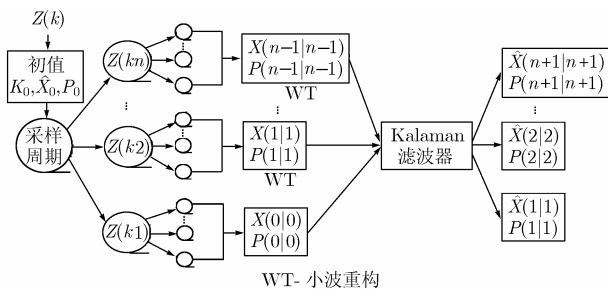


图1 WKHEFA 算法示意图

4 仿真实验

本次仿真采用NS2 和Matlab, 具体步骤为: (1)在NS中产生随机拓扑图, 然后随机选择一个节点。(2)在Matlab中, 实现ARIMA算法和WKHEFA算法¹⁾。(3)把上步骤中所得到的随机序列化为矩阵分别代入 ARIMA 预测算法和 WKHEFA预测算法。(4)利用Matlab绘图工具绘制成图。

4.1 算法预测网络流量尺度性仿真实验

在图2 中本文用 WKHEFA 对网络流量进行了4层的尺度分析, 其中 A1 表示为第1层的尺度系数, A2 为第2层尺度系数, A3 为第3层尺度系数, D4 表示为第4层小波系数, X 表示原始网络流量, 其中 $X = A1 \otimes A2 \otimes A3 \otimes A4 \otimes D4$ 。

¹⁾仿真中实现了经典ARIMA算法文献[1], WKHEFA采用了Haar小波分析

从图2 中, 可以看出 WKHEFA 把网络流量按照其尺度特征, 分别在4层尺度下予以解析, 其尺度特征如 A1, A2, A3, A4 所示, 同时 WKHEFA 对网络流量每一层的细节部分进行分析, 得到其小波系数如 D4 所示, WKHEFA 通过多尺度的分析和重构, 在不同的精度要求提取出了网络流量的平稳变化和周期性变化特征, 有效地提高了网络流量预测精度。

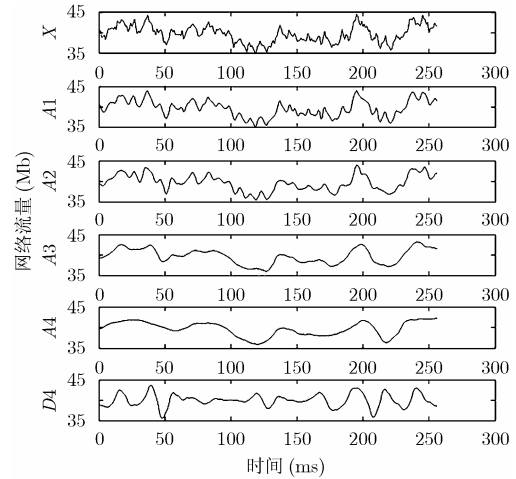


图2 网络流量的尺度分析图

4.2 WKHEFA 与 ARIMA 仿真结果的比较

图3 为 WKHEFA 与 ARIMA 的网络流量预测值对比。横坐标为采样时刻, 纵坐标为网络流量。将实际网络流量值与两种方法预测的值进行对比。从中可以看出 WKHEFA 真实网络流量的重合度远远优于 ARIMA 预测模型, 这是因为 ARIMA 预测模型的前提条件要求时间序列是宽平稳过程, 但是真实的网络流量往往无法符合这样的条件。WKHEFA 能够实时地跟踪网络流量的细节变化, 从图3 中可以看到对于网络流量的急剧变化的部分 WKHEFA 也能够以比较低的预测误差反映出来。

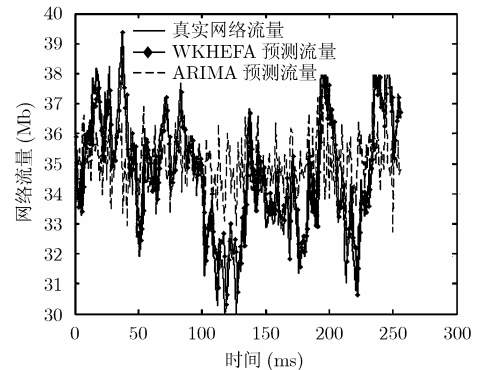


图3 WKHEFA 与 ARIMA 仿真结果的比较

图4 是 ARIMA 和 WKHEFA 的误差统计图。横坐标为误差值, 纵坐标为时刻。从中可以发现 WKHEFA 误差区间为 $(-2.6, +2.6)$, ARIMA 的误差区间是 $(-3.2, +5.4)$, 并且对于误差在 $(-1, +1)$ 之间的统计量, WKHEFA 明显高于 ARIMA, 这与图2 的结果相互对应, 因此 WKHEFA 网络流量模型的预测精度远远高于 ARIMA。

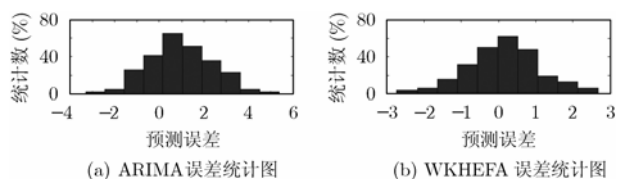


图4 ARIMA 和 WKHEFA 的误差统计图

5 结束语

对于传统的 Kalman 滤波预测模型, 测量值的每个分量只能对状态变量相对应的分量进行更新估计。但在 WKHEFA 中, 通过将目标状态的小波变换系数向量描述为 KF 的状态变量, 使测量值的每个分量能够对状态变量的所有分量进行更新估计, 进而利用小波重构就能够对一类非平稳随机过程的平稳部分的各个分量进行更新估计, 这样就实现了在一个周期内利用某个时刻及其以前的信息对其后所有时刻进行预报, 即实现了周期内实时的动态多步预报。同时, 与小波分析方法相比, WKHEFA 具有实时性和递归性, 可在时域中对网络流量进行实时的动态估计和分析。下一步的工作将对 WKHEFA 的模型进行改进, 以实现网络流量的长期预测。

参考文献

- [1] 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(自然科学版), 2003, 6(3): 6-10.
Jin qi, Pei Chang-xing, and Zhu Chang-hua. ARIMA analysis method in network traffic. *Journal of Xidian University(Natural Science)*, 2003, 6(3): 6-10.
- [2] Yu Guo-qiang and Zhang Chang-shui. Switching ARIMA model based forecasting for traffic flow[C]. ICASSP '04. Canana, 2004, Vol.(2): 429-432.
- [3] Sang Ai-min and Li San-qi. A predictability analysis of network traffic[C]. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Tel Aviv, Israel, 2000, Vol.(1): 342-351.
- [4] Leland W E, Taqqu M S, Willinger W, and Wilson D V. On the self-similar nature of Ethernet traffic [J]. *IEEE/ACM Transactions on Networking*, 1994, 2(1): 1-15.
- [5] Park Kihong and Walter Willinger. Self-Similar Network Traffic and Performance Evaluation[M]. Wiley, U.S.A, 2000.
- [6] Gao Chuan-shan, Han Liangxiu, Cen Zhiwei, and Chu Chun-bo. A new multi fractal traffic model based on the wavelet transform[C]. Proceedings of the ISCA 14th International Conference: Parallel and Distributed Computing Systems, Richardson, Texas USA, Aug. 8-10, 2001: 157-162.
- [7] Kolarov A, Atai A, and Hui, J. Application of Kalman filter in high-speed networks[C]. GLOBECOM '94. San Francisco, U.S.A, 28 Nov.-2 Dec. 1994 vol.2: 624-628.
- [8] 韩良秀, 丛锁, 阎华, 高传善. 基于小波技术的网络流量特性刻画[J]. 小型微型计算机系统, 2001, 22(9): 1110-1113.
Han Liang-xiu, Cong Suo, and Yan Hua, *et al.*. Characterizing network traffic based on the wavelet technique. *Mini-micro Systems*, 2001, 22(9): 1110-1113.
- [9] Yu E S, Chen C Y R. Traffic prediction using neural networks[C]. Global Telecommunications Conference, 1993, including a Communications Theory Mini-Conference. Technical Program Conference Record, IEEE in Houston. GLOBECOM '93., IEEE29 Nov.-2 Dec. 1993 vol.2: 991-995.
- [10] 文成林, 周东华. 多尺度估计理论及其应用[M]. 北京: 清华大学出版社, 2002: 98-112.
- [11] Zheng Tong-xin, Girgis A, Makram B. A hybrid wavelet-Kalman filter method for load forecasting[J]. *Electric Power Systems Research*, 2000, 54 (1): 11-17.
- [12] Kawahara R and Komatsu N. A scalable IP traffic control method for weighted bandwidth allocation per-flow. *IEICE Trans. Commun.*, 2001, E84-B(10): 2815-2829.

李捷: 男, 1975年生, 博士, 讲师, 研究方向为IP网络QoS、无线传感器网络等。

候秀红: 女, 1957年生, 副教授, 研究方向为人工智能、电子商务等。

韩志杰: 男, 1979年生, 硕士生, 研究方向为IP网络路由算法、网络建模等。