

# 基于小波的 Hurst 指数自适应估计方法\*

洪飞<sup>1+</sup>, 吴志美<sup>2</sup>

<sup>1</sup>(北京航空航天大学 计算机学院,北京 100083)

<sup>2</sup>(中国科学院 软件研究所,北京 100080)

## Adaptive Hurst Index Estimator Based on Wavelet

HONG Fei<sup>1+</sup>, WU Zhi-Mei<sup>2</sup>

<sup>1</sup>(School of Computer Science, BeiHang University, Beijing 100083, China)

<sup>2</sup>(Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-10-82317612, Fax: +86-10-82311939, E-mail: hongfei@buaa.edu.cn, http://www.buaa.edu.cn

Received 2003-07-28; Accepted 2004-11-22

**Hong F, Wu ZM. Adaptive Hurst index estimator based on wavelet. *Journal of Software*, 2005,16(9): 1685-1689. DOI: 10.1360/jos161685**

**Abstract:** The measurement studies show that the burstiness of packet traffic in LAN as well as WAN is associated with self-similar and long-range dependency, and Hurst index is the key value of this model representing the burstiness of traffic. With the analysis in discrete wavelet domain, the nature of the wavelet coefficients and their statistical properties are proposed. Then an adaptive, efficient unbiased estimator of Hurst index based on multiresolution wavelet analysis and weighted regression is presented. Simulation results based on fractal Gaussian noise and real traffic data reveal the proposed approach shows more adaptiveness, accuracy and robustness than traditional estimators which has only  $O(N)$  computation. Thus this estimator can be applied to the application of traffic management and real-time control in high-speed networks.

**Key words:** self-similar; LRD (long-range dependent); wavelet; Hurst index; adaptive

**摘要:** 对局域网和广域网上大量突发网络流量的分析结果表明,网络流量普遍存在着自相似性和长相关性,其中 Hurst 指数是表征网络流量突发特性的重要参数.通过在小波域内对网络流量这种特性的分析,给出了其小波系数的本质和统计特性.针对基于小波的 Hurst 指数估计方法的自适应问题,结合方差分析给出了一种有效的解决方法,从而提出了自适应的参数估计方法,并且该方法在一般意义上是无偏的.分形高斯噪声和真实突发网络数据的仿真结果均表明,自适应方法比传统估计方法具有更高的估计精度,能够自适应地选择最优尺度区间,而且仅具有  $O(N)$  的计算复杂度,因此可望应用于高速网络的网络管理和实时控制.

**关键词:** 自相似性;长相关性;小波;Hurst 指数;自适应

中图法分类号: TP393 文献标识码: A

\* Supported by the National Grand Fundamental Research 973 Program of China under Grant No.G1998030407 (国家重点基础研究发展规划(973)); the Beijing Science and Technology Committee Program under Grant No.H011710010123 (北京市科委项目)

作者简介: 洪飞(1974 - ),男,山东临沂人,博士,讲师,主要研究领域为网络行为学,网络模型,网络测量;吴志美(1942 - ),男,研究员,博士生导师,CCF 高级会员,主要研究领域为多媒体通信,网络模型.

通信领域中对高速网络突发网络流量建模时,通常假定网络到达过程为泊松流.然而 Bellcore 对于 LAN 上数据包网络流量的监测<sup>[1,2]</sup>以及许多研究机构对 WAN 上 Internet 网络数据和 VBR 视频流的分析<sup>[3-5]</sup>表明,突发网络到达过程更符合渐近或严格的自相似模型,Hurst 指数是表征自相似网络流量突发特性的重要参数.通常,突发网络流量的 Hurst 指数在 0.5~1 之间,表示网络具有正的相关结构,Hurst 指数越大说明网络的突发程度越剧烈.因此,对 Hurst 指数的估计直接影响到网络流量建模的准确程度以及高速网络传输中流量控制和统计复用的性能.现有文献中给出的估计方法主要分为时域法和频域法,这些方法在分析较长时间范围的网络流量数据时,需要大量数据样本,计算复杂度较高,对 Hurst 指数的时变特性不能进行精确描述,因而文献[6-8]将小波变换引入到网络流量自相似性和长相关性的研究以及 Hurst 指数的估计中,研究表明,在小波域内能够快速准确地估计 Hurst 指数.但是,小波域内的估计结果依赖于小波尺度区间的选择,文献[9]给出了该方法的性能评价以及尺度区间的选择对 Hurst 指数估计值的影响,并仅在 Hurst 指数较小时给出了一个选择方法,因此对尺度区间的自适应选择是一个需要解决的问题.

本文简要叙述了自相似模型的数学描述,主要分析了自相似过程小波变换系数的本质和相关的统计特性.在此基础上针对基于小波的 Hurst 指数估计方法,提出了一种有效的方法,较好地解决了小波尺度区间的自适应选择.对分形高斯噪声和真实网络流量数据的仿真结果说明了自适应方法简单、快速、有效,能够选择最优小波尺度区间的特点,适用于高速网络突发网络流量的实时监测.

## 1 自相似网络模型

自相似过程是在统计意义上具有尺度不变性的一种随机过程,从这一点上来说,自相似过程实际上是在随机过程中引入了分形的概念.

定义 1. 一个随机过程  $\{X(t), t \in \mathbb{R}\}$ , 如果  $X(t)$  在时间上进行压缩或扩展时,统计特性不变,即满足  $X(t) \stackrel{D}{=} a^{-H} X(at)$ , 其中  $H > 0$  是 Hurst 指数,等号表示统计意义上的相等,那么称这个过程是统计自相似的.如果  $X$  的各阶矩存在,则符合指数规律,即  $E|X(t)|^q = E|x(1)|^q |t|^{qH}$ , 进一步地,如果对于任意有限维的联合概率分布,等号成立,那么称这个过程是严格自相似的;如果只是均值和方差相等,也就是说,当

$$M_X(t) = E[X(t)] = a^{-H} M_X(at),$$

$$R_X(t, s) = E[X(t)X(s)] = a^{-2H} R_X(at, as)$$

时,称这个随机过程是宽自相似的.

定义 2. 一个随机过程  $\{X(t), t \in \mathbb{R}\}$ , 如果其增量函数  $Y(\delta, t)$  存在并满足:

$$\{Y(\delta, t) := Y_\delta(t) = X(t+\delta) - X(t), t \in \mathbb{R}\} \stackrel{D}{=} \{X(\delta) - X(0)\}, \forall \delta,$$

称这个随机过程为具有平稳增量  $Y(\delta, t)$  的随机过程.

虽然具有自相似特性的随机模型很多,但目前通过与真实网络流量数据相比较,认为适于作为突发网络建模的随机模型只有分形布朗运动(fractional Brownian motion, 简称 FBM)和分形 ARIMA 过程,而且它们都基于分形高斯噪声(fractional Gaussian noise, 简称 FGN).其中分形布朗运动的独立增量过程即为分形高斯噪声,而分形 ARIMA 过程可视为分形高斯噪声经过一个以 ARMA 参数为滤波器系数进行滤波的结果.当  $1/2 < H < 1$  时,分形高斯噪声是严格二阶自相似过程,因其参数简单而成为自相似网络流量建模的主要工具.

## 2 自相似过程的多分辨率小波分析

本节主要描述自相似过程小波系数的统计特性,以及小波域中的尺度不变性,假设读者熟悉离散小波变换和多分辨率分析.如果  $X(t), t \in \mathbb{R}$  代表一个随机过程,  $d_{j,k}$  为其小波系数.如果小波  $\psi$  在时间上消逝很快并且满足以上的条件,则小波系数  $d_{j,k}$  也是一个随机变量.结合文献[6,7,10],得到以下结论:

对于具有平稳增量的随机过程  $\{X(t), t \in \mathbb{R}\}$ , 也就是  $\{X(t+h) - X(t), t \in \mathbb{R}\}$  的有限维分布概率不依赖于  $t$ . 因此,对于固定的  $j \in \mathbb{Z}, \{d_{j,k}, k \in \mathbb{Z}\}$  是一个平稳序列.很容易得出其一维边缘分布不依赖于  $k$ . 即

$j=0, d_{0,k+k_0} \stackrel{D}{=} d_{0,k}, \forall k_0$ , 因此对于任意的整数  $n \geq 1$  和实常数  $\theta_i$  可得:

$$\sum_{i=1}^n \theta_i d_{0,k_i+k_0} = \sum_{i=1}^n \theta_i d_{0,k_i} \tag{1}$$

也就是说,两者的特征函数是相等的,因此  $n$  维分布不依赖于  $k_0$ , 即  $d_{j,k}, k \in \mathbb{Z}$  关于  $k$  是平稳的.

如果  $\{X(t), t \in \mathbb{R}\}$  是一个 Hurst 指数为  $H$  的自相似过程, 则对于固定的  $j \in \mathbb{Z}$ , 由于  $X(2^j u) \stackrel{d}{=} 2^{jH} X(u)$ , 并进行变量代换可得:  $d_{j,k} = \int X(2^j u) 2^{-j/2} \psi(u-k) 2^j du \stackrel{d}{=} 2^{j(H+1/2)} d_{0,k}$ , 也就是  $d_{j,k} \stackrel{D}{=} 2^{j(H+1/2)} d_{0,k}$  表示为  $k$  的一个随机过程. 如果  $\{X(t), t \in \mathbb{R}\}$  是一个平稳增量的随机过程,  $0 < H < 1$ , 具有零均值和有限方差,  $Ed_{j,k} = 0$ , 由上式容易得到:  $Ed_{j,k}^2 = C 2^{j(2H+1)}, C = Ed_{0,0}^2$ . 对上式两边取对数可以得到  $j$  的斜率为  $2H+1$  的线性方程, 这是在小波域内对 Hurst 指数进行估计的基础.

### 3 自适应参数估计方法

前已指出, Hurst 指数是表征具有长相关特征网络突发程度的重要参数, 根据一定时间范围内网络流量的监测数据估计, Hurst 指数对自相似网络的建模和流量控制都是十分重要的. 目前估计 Hurst 指数的已有方法可分为时域和频域两类, 其中时域法包括 R/S 分析、方差-时间分析、IDC 分析, 频域法包括 Whittle 的最大似然估计、基于周期图的半参数估计和小波域内的估计等. 时域法直接对原始网络数据序列进行处理, 作出分析曲线, 用最小二乘拟合估计 Hurst 指数; 频域法则利用 FFT 对原始网络数据序列的谱密度在频域进行估计以及文献[9]中在小波域内基于 EM 估计的迭代方法.

首先介绍估计中采用的一维加权线性回归方法. 对于随机变量  $y_k$  和  $x_k, k = j_1, \dots, j_2$ , 基本的假设回归为  $Ey_k = bx_k + a$ . 定义以下变量:  $S = \sum 1/\sigma_k^2, S_x = \sum x_k/\sigma_k^2, S_{xx} = \sum x_k^2/\sigma_k^2$ , 这里  $\sigma_k^2$  是与  $y_k$  相关的任意权值. 通常的  $(b, a)$  无偏估计  $(\hat{b}, \hat{a})$  为

$$\hat{b} = \frac{\sum y_k (S_{x_k} - S_x) / \sigma_k^2}{SS_{xx} - S_x^2} \equiv \sum w_k y_k \tag{2}$$

$$\hat{a} = \frac{\sum y_k (S_{xx} - S_x^2) / \sigma_k^2}{SS_{xx} - S_x^2} \equiv \sum v_k y_k \tag{3}$$

这里, 权值  $w_k$  和  $v_k$  满足  $\sum w_k = \sum jv_k = 0, \sum jw_k = \sum v_k = 1$ .

由上述对小波系数的讨论可知:

$$\log Ed_{j,k}^2 = (2H+1)j + \log C \tag{4}$$

其中, 尺度区间  $[j_1, j_2]$  的选择对拟合结果有很大影响, 这里采用方差分析进行拟合优度分析, 分析不同尺度区间对 Hurst 指数估计的拟合程度, 从而自适应地选择最优的尺度区间. 在保证  $j_2 - j_1 \geq 3$  的前提下: 参差平方和  $Q = \sum (y_i - \hat{y}_i)^2$ , 回归平方和  $U = \sum (\hat{y}_i - \bar{y})^2$ , 其中  $\bar{y}$  和  $\hat{y}_i$  为  $y_i$  的均值和回归值; 利用统计量  $F = \frac{U/1}{Q/(n-2)} \sim F(1, n-2)$  来进行最优度检验, 得到最优的尺度区间  $[j_1, j_2]$ .

下面给出基于小波变换的 Hurst 指数估计方法. 假定网络到达过程的自相似模型不包含噪声, 具体估计方法如下:

输入参数: 原始数据  $X[n]$ , Daubechies 小波消逝矩大小.

第 1 步: 根据消逝矩的大小初始化多分辨率小波变换.

第 2 步: 进行离散小波分解, 通过 Daubechies 带通小波得到相应各尺度的小波系数  $d_{j,k}$ .

第 3 步: 对不同的尺度区间进行方差拟合度检验, 得到最优的尺度区间  $[j_1, j_2]$ .

第 4 步: 根据最优尺度区间进行参数估计, 应用公式(2)和公式(3).

第 5 步: 计算 Hurst 指数估计值.

在上述方法中, 对 Hurst 指数的估计是无偏的. 通过下面的仿真结果可以看出, 尺度区间的选择对方法的精

确性影响较大,自适应方法的尺度区间选择具有最优性.同时,对 Hurst 指数一致性较好的网络源, $N$  可取小一些,在减少方法复杂度的同时,保持估计的准确性.

#### 4 仿真结果

在仿真实验中,分别采用分形高斯噪声和 Bellcore 于 1989 年在 LAN 上监测得到的网络流量数据作为原始数据,估计它们的 Hurst 指数.前者是严格的二阶自相似过程,后者是渐近自相似过程.本文采用基于快速 Fourier 变换的快速生成方法生成具有不同 Hurst 指数期望值的分形高斯噪声.仿真中选择 R/S 分析、方差-时间分析、基于周期图法的半参数估计以及文献[9]中的方法与自适应方法进行比较.以上方法基本包括了目前时域和频域内 Hurst 指数的有效估计方法.有关它们各自的原理和方法实现参见文献[1,2,9].

针对 Hurst 指数期望值不同的分形高斯噪声分别生成 4 096 点数据序列,具体实验数据不再列出.首先对自适应方法中的尺度区间选择进行实验分析,根据文献[8]中的研究结果,主要考虑 Hurst 指数较小的网络流量,也就是短相关性较强的网络流量.在表 1 和表 2 中,对相同 Hurst 指数值(0.55,0.60)的分形高斯噪声,给出不同的尺度区间下的 Hurst 指数估计值及相应的拟合优度,由于篇幅所限,只列出较为典型的尺度区间.从表中可以很直观地说明不同的尺度区间对 Hurst 指数的估计结果有很大影响.表中具有\*标志的区间和估计值是本文的尺度区间自适应选择的结果,从中可以说明本文的尺度区间选择的最优性.而文献[8]表明,在 Hurst 指数较小时,需要选择较小的尺度区间才能得到较好的估计结果,然而自适应方法在 Hurst 指数较小时所选的尺度区间基本是所有估计区间的最小区间,这与文献[8]的结论是一致的.从表 3 中可以看出,无论 Hurst 指数大小如何,自适应方法都可以给出一个最优的尺度选择,因而在区间的自适应性选择上要优于文献[8]的选择.

**Table 1** The goodness of fit and estimation result to different scale interval ( $H=0.55$ )

表 1 不同尺度区间下的拟合优度和估计值( $H=0.55$ )

$[j_1, j_2]$	[5,8]	[2,9]	[6,9]	[3,9]	[6,8]*
$H$	0.517	0.537	0.602	0.534	0.553
Goodness	0.449	0.358	0.343	0.252	0.216

**Table 2** The goodness of fit and estimation result to different scale interval ( $H=0.60$ )

表 2 不同尺度区间下的拟合优度和估计值( $H=0.60$ )

$[j_1, j_2]$	[2,9]	[2,8]	[2,7]	[3,6]	[2,6]*
$H$	0.590	0.588	0.585	0.596	0.597
Goodness	0.351	0.301	0.291	0.283	0.272

表 3 给出了分别用 R/S 分析、方差-时间分析、基于周期图法的半参数估计、文献[9]中的方法和自适应方法对分形高斯噪声进行 Hurst 指数估计的结果,同时还给出了自适应方法的最优尺度区间 $[j_1, j_2]$ .在自适应方法中,选择 Daubechies 的消逝矩是 3,样本点数决定了最大尺度阶数为 12,回归分析尺度区间列在表中.分析实验数据结果可得,当 Hurst 指数较大时,也就是突发性较高的网络流量,自适应方法选择较宽的尺度区间;反之,对选择较窄的尺度区间,显示了拟合度检验的优越性.时域法虽然直观,但估计的精度不如频域和小波域方法,这是因为当数据样本有限时,“块”大小的选取对估计结果影响很大.在表 3 中还给出了自适应方法和文献[9]方法估计结果的误差.通过比较可以看出,在本文最优尺度选择下,Hurst 指数估计结果的精度要好于文献[9]的方法.这里需指出一点,多分辨率小波分解方法的复杂性接近  $O(N)$ ,因此,自适应方法无论在估计精度还是在计算效率上都优于传统的频域方法.

**Table 3** Hurst estimation results with various algorithm  
表 3 不同方法对分形高斯噪声序列 Hurst 指数估计结果

$H$	Conf. interval	R/S	V-T	Period plot	Method in Ref.[9] and error	Adaptive $[j_1, j_2]$ and error
0.55	[.537,.559]	0.550	0.540	0.552	0.548 (0.002)	0.553[6,8] (0.003)
0.60	[.589,.607]	0.607	0.606	0.596	0.597 (0.003)	0.597[2,6] (0.003)
0.65	[.641,.660]	0.652	0.655	0.648	0.650 (0.000)	0.649[2,9] (0.001)
0.70	[.693,.705]	0.685	0.689	0.693	0.697 (0.003)	0.698[2,9] (0.002)
0.75	[.738,.764]	0.739	0.749	0.759	0.750 (0.000)	0.750[2,9] (0.000)
0.80	[.789,.814]	0.817	0.826	0.806	0.805 (0.005)	0.796[3,9] (0.004)
0.85	[.836,.863]	0.867	0.843	0.858	0.852 (0.002)	0.850[2,9] (0.000)
0.90	[.879,.916]	0.899	0.886	0.892	0.898 (0.002)	0.901[2,9] (0.001)

为了说明自适应方法对真实网络数据在一定观察时段内 Hurst 指数的估计也有效,选取 Bellcore 的 3 段数据(pAug.TL,pOct.TL 和 OctExt.TL),同样提取 4 096 个点作为样本点.与方差-时间分析相同,Whittle 估计方法与自适应方法得到的结果分别列于表 4 中.自适应方法的结果与传统方法较为接近,计算复杂度较低,在  $N$  较小时可以得到较为准确的估计结果,符合实时性要求.因此,自适应方法可望用于对真实网络数据的实时监测分析.

**Table 4** Hurst index estimation results of real traffic

**表 4** 对真实网络数据 Hurst 指数估计结果

Real traffic	R/S	V-T	Period plot	Adaptive $[j_1, j_2]$
pAug.TL	0.842	0.821	0.834	0.856[2,9]
pOct.TL	0.776	0.762	0.784	0.781[3,9]
OctExt.TL	0.756	0.801	0.763	0.762[3,9]

## 5 结 论

针对基于小波的 Hurst 指数估计方法的自适应问题,本文给出了一个有效的解决方法,提出一种基于多分辨率小波分析的 Hurst 指数自适应快速估计方法.采用分形高斯噪声和真实突发网络数据的仿真结果均表明,在 Hurst 指数一致性较好时,自适应方法比传统时域估计方法具有更高的估计精度,回归尺度区间的自适应和仅具有  $O(N)$  的计算复杂度满足了实时性的要求.因此,自适应方法可望用于高速网络的网络管理和拥塞控制.进一步的研究工作将是如何将自适应方法应用到接入控制和拥塞控制以及网络建模中.

致谢 感谢 Melbourne 大学 Darryl Veitch 博士提供了基本的 MATLAB 算法.

## References:

- [1] Leland WE, Taqu MS, Willinger W, Wilson DV. On the self-similar nature of ethernet Traffic (extended version). IEEE/ACM Trans. on Networking, 1994,2(1):1-15.
- [2] Paxson V, Floyd S. Wide area traffic: the failure of Poisson modeling. IEEE/ACM Trans. on Networking, 1995,3(3):226-244.
- [3] Garrett MW, Willinger W. Analysis, modeling and generation of self-similar VBR video traffic. In: Proc. of the ACM SIGCOMM'94. London: ACM Press, 1994. 269-280.
- [4] Abry P, Veitch D. Wavelet analysis of long range dependent traffic. IEEE Trans. on Information Theory, 1998,44(1):2-15.
- [5] Abry P, Flandrin P, Taqu MS, Veitch D. Wavelets for the analysis, estimation and synthesis of scaling data. In: Park K, Willinger W, eds. Self Similar Network Traffic Analysis and Performance Evaluation. New York: John Wiley & Sons Press, 2000. 134-156.
- [6] Abry P, Flandrin P, Taqu MS, Veitch D. Self-Similarity and long-range dependence through the wavelet lens. In: Taqu MS, ed. Theory and Applications of Long Range Dependence. Boston: Birkhauser Press, 2002. 345-379.
- [7] Veitch D, Abry P. A wavelet based joint estimator for the parameters of LRD. IEEE Trans. on Information Theory, 1999,45(3): 878-897.
- [8] Xue F, Ljiljana T. Performance analysis of a wavelet-based hurst parameter estimator for self-similar traffic. In: Proc. of the SPECTS 2000. Vancouver: Canada: SCS Press, 2000. 294-298.
- [9] Chen HM, Cai H, Li YD. Self-Similar traffic: Hurst parameter estimation based on multiresolution sampling and wavelet analysis. Acta Electronica Sinica, 1998,26(7):88-93 (in Chinese with English abstract).
- [10] Daubechies I. Ten Lectures on Wavelets. New York: SIAM, 1992.

## 附中文参考文献:

- [9] 陈惠民,蔡弘,李衍达.自相似网络:基于多分辨率采样和小波分析的 Hurst 系数估计方法.电子学报,1998,26(7):88-93.