

消失矩对小波分析求解自相似参数 Hurst 的影响研究

任勋益^① 王汝传^{①②} 祁正华^①

^①(南京邮电大学计算机学院 南京 210003)

^②(南京大学计算机软件新技术国家重点实验室 南京 210093)

摘要: 研究了消失矩对小波求解自相似参数 Hurst 值的影响,给出了理论推导和分析,间接论证了消失矩对 Hurst 参数求解的影响趋势,通过大量的实验结果给出了详细情况。理论和实验结果一致表明,在局部上, Hurst 参数随着消失矩的增大呈现周期性变化,但总的趋势是随着消失矩的增大 Hurst 值减小,误差在增大,满足条件的最小消失矩是最佳选择。该结论能为基于自相似和小波分析的大量应用提供理论依据。

关键词: 消失矩; 小波; 自相似; Hurst

中图分类号: TP393.06

文献标识码: A

文章编号: 1009-5896(2007)09-2257-05

Study on Vanishing Moments Influence on Estimation Hurst Parameter of Self-similar Traffic

Ren Xun-yi^① Wang Ru-chuan^{①②} Qi Zheng-hua^①

^①(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

^②(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

Abstract: A theoretic analysis is proposed which indirectly proves how vanishing moments influence on estimation Hurst parameter of self-similar traffic, and a lots of experimental results show the Hurst's value changes periodicity with the vanishing moments, but total way is the Hurst's value decrease and its variance raises with the increase of the vanishing moments. This conclusion is different from common view, and could provide reason for choice vanishing moments on wavelet analysis method for Hurst parameter.

Key words: Vanishing moments; Wavelet; Self-similarity; Hurst

1 引言

1994 年, Leland 等人^[1]对 Bellcore 的局域网测试与分析的结果显示,实际网络流量模型具有统计自相似性。1995 年, Beran 等^[2]通过对大量的不同类别的可变比特率视频流数据的统计发现,它们也同样表现出一种长相关特性。另外,对 WAN 等网络的测量,同样发现这些网络业务量表现出长相关的特性^[3,4]。网络流量的这种特性给许多应用带来了新的方法,如在异常流量检测方面,文献[1,5,6]根据 DDoS (Distributed Denial of Service)攻击会造成网络的自相似性能降低,从 Hurst 值的变化,检测是否发生了 DDoS 攻击,提出了基于自相似性检测 DDoS 攻击;在网络性能优化方面,中国科学院软件研究所将广域网流量的长相关特性应用于千兆交换路由器的缓存控制。这些基于网络流量自相似性的

应用,关键问题是对采集来的数据进行自相似性分析。因此关于自相似参数 Hurst 的准确求解变成了一个重要的研究课题。

大量的文献在研究 Hurst 参数的求解问题^[7-9]。常用的求解 Hurst 参数的方法有聚集方差法、周期图法、R/S 分析法(rescaled range analysis)、Whittle 法等。如 Leland 等人在文献[1]中采用 R/S 方法和聚集方差法分别对自相似参数求解,文献[2]利用聚集方差法求解 DDoS 攻击对自相似性的影响。这些方法存在需要样本容量较大,计算过程较缓慢,准确性不高等缺点。目前最具有代表性的是利用小波分析法求解 Hurst 参数。文献[10,11]探讨了将小波分析法用于求解 Hurst 的情况。在中国科学院软件研究所的“信息网前沿技术研究:通信网络动态仿真”的项目中对科技网流量进行分析,也采用了小波分析的方法分析整理采集的数据。然而,小波分析的精度往往受到一些数学特征因素的影响,如消失矩、对称性、紧支性等,其中,消失矩是选择小波分析的最重要的因素^[11-13]。比如我们要研究检测 DDoS 攻击,不同的小波消失矩对 DDoS 攻击造成的 Hurst 值变化的敏感程度不一样,选择不合适的消失矩会造成 DDoS 攻击检测不及时的问题^[11]。然而,上面这些文献在应用小波分析自相似性时并未给出选用小波种类和消失矩的深入研究与分析。在文献[8,11]中给出

2006-02-26 收到, 2006-08-29 改回

国家自然科学基金(6057314, 70271050), 江苏省自然科学基金(BK2005146), 江苏省高技术研究计划(BG2005037, BG200600 1), 国家高科技 863 项目(2005AA775050), 国家通信技术重点实验室基金(9140C1101010603), 江苏省计算机信息处理技术重点实验室基金(kjs050001, kjs06)和江苏省高校自然科学研究计划(05KJB520092)资助课题

了不同小波和消失矩求解Hurst的情况，但是它们没有给出理论推导，而且文献[8]的结果分析建立在拟合误差之上。

本文研究了消失矩对小波求解自相似参数 Hurst(H)值的影响，给出了理论推导分析和大量的实验结果。其余部分组织如下：第2节介绍了消失矩。第3节首先给出了自相似过程和小波系求解 Hurst 方差法，在此基础上理论推导和分析了小波正则度对 H 值影响，间接证明了消失矩对 H 值影响。第4节是实验。最后是结束语。

2 消失矩

定义1 小波函数 $\psi(t)$ 如果满足 $\int_{\mathbb{R}} t^k \psi(t) dt = 0, 0 \leq k$

$< m$ ，则称 $\psi(t)$ 具有 m 阶消失矩。从消失矩的定义中可以推知， m 阶消失矩意味着小于 m 次的多项式与小波 $\psi(t)$ 内积作用的结果都是零。当 $m = 1$ 时，小波函数具有一阶消失矩，这是小波允许条件的基本要求。当 $m > 1$ 时，称小波函数具有高阶消失矩。消失矩越高，小波逼近光滑函数的能力越强，因此在一般的应用中总是选择消失矩比较高的小波基。

由于正则度刻画了函数的光滑程度，正则度和消失矩紧密相关，它也是自相似模型的一个参数，影响着消失矩的选择。

3 基于自相似和小波理论分析消失矩对 H 值影响

3.1 自相似过程

$X = (X_i, i = 0, 1, 2, 3, \dots)$ 为一个广义平稳过程， X_i 表示第 i 个单位时间内到达的网络流量单元的数目(如到达的数据包的个数，或到达的字节数)，记 $\mu = E[X_i]$ ， $\sigma^2 = E[(X_i - \mu)^2]$ ，自相关函数 $r(k) = E[(X_i - \mu)(X_{i+k} - \mu)] / \sigma^2$ 。令 $X_k^{(m)} = (X_{km-m+1} + \dots + X_{km}) / m, k = 1, 2, 3, \dots$ ，称为 $\{X_i\}$ 的 m 阶的聚合过程，对每个 m ， $X^{(m)}$ 都表示一个广义平稳随机过程， $r^{(m)}$ 为其对应的自相关函数。如果对所有的 m ， m 阶聚合过程 $X^{(m)}$ 都具有与原过程同样的相关函数结构，即 $r^{(m)}(k) = r(k), 0 < \beta < 1, k \rightarrow \infty$ ，则称为自相似参数为 $H = 1 - \beta/2$ 的严格自相似过程；如果 $r^{(m)}(k) \rightarrow r(k) \sim k^{-\beta}, m \rightarrow \infty, k = 0, 1, 2, \dots$ ，则称为自相似参数为 H 的渐进二价自相似过程。其中 H 是表述自相似特性的唯一参数， H 值越大，自相似程度越高，其取值范围是 $(1/2, 1)$ 。

3.2 小波系数方差法求解 Hurst

小波分析法求解 Hurst 系数方法主要有小波系数方差法、谱估计法、能量法。正如文献[8]所述，这些方法殊路同归，其原理都是基于能量的分析方法。

若 $\{x(t)\}$ 为一统计自相似过程，其功率谱为 $P_x(\omega)$ 。对 $x(t)$ 作小波变换，得到小波系数

$$d_l^{(j)} = \langle x(t), \psi_{j,l}(t) \rangle = 2^{j/2} \int_{\mathbb{R}} x(t) \psi(2^j t - l) dt \quad (1)$$

其中 $\psi_{j,l}(t) = 2^{j/2} \int \psi(2^j t - l) dt$ 为二进正交小波，其正则度为 R 。

$$E[d_l^{(j)}] = E[x(t)] 2^{j/2} \int \psi(2^j t - l) dt = 0 \quad (2)$$

根据相关系数的定义，任意两个小波变换系数 $d_l^{(j)}, d_{l'}^{(j')}$ 间的相关值为

$$E[d_l^{(j)}, d_{l'}^{(j')}] = \iint E[x(t)\psi_{j,l}(t)x(t')\psi_{j',l'}(t')] dt dt' \quad (3)$$

作傅里叶变换并用 Parseval 公式得到

$$E[d_l^{(j)}, d_{l'}^{(j')}] = \frac{2^{-(j+j')/2}}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma_x^2}{|\omega|^\gamma} \widehat{\psi}(2^{-j}\omega) \overline{\widehat{\psi}(2^{-j'}\omega)} e^{-i[l2^{-j} - l'2^{-j'}]\omega} d\omega \quad (4)$$

由式(2)，式(3)得到 $d_l^{(j)}$ 的方差为 $\text{Var}[d_l^{(j)}] = E[d_l^{(j)}]^2 = \frac{2^{-j\gamma}}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma_x^2}{|\omega|^\gamma} \left| \widehat{\psi}(\omega) \right|^2 d\omega$ 。记 $\sigma^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma_x^2}{|\omega|^\gamma} \left| \widehat{\psi}(\omega) \right|^2 d\omega$ ，则

$$\text{Var}[d_l^{(j)}] = \sigma^2 2^{-j\gamma} \quad (5)$$

通过对式(5)两边的数据取对数后，在均方误差最小的意义下进行线性拟合得到以 j 为自变量，以 $\log_2 \text{Var}[d_l^{(j)}]$ 为函数的直线，斜率即为 γ 。利用 $\gamma = 2H + 1$ 便可求得 H 值。

3.3 小波正则度对 H 值影响的理论分析

在小波系数方差法中要求基本小波正则度 $R \geq \gamma/2$ 。其原因是 σ^2 须取得有限值，即

$$\begin{aligned} \sigma^2 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma_x^2}{|\omega|^\gamma} \left| \widehat{\psi}(\omega) \right|^2 d\omega = \frac{\sigma_x^2}{2\pi} \int_{-\infty}^{+\infty} \frac{|\omega|^{2R}}{|\omega|^\gamma} \left| \widehat{\psi}_0(\omega) \right|^2 d\omega \\ &= \frac{\sigma_x^2}{2\pi} \int_{-\infty}^{+\infty} |\omega|^{2R-\gamma} \left| \widehat{\psi}_0(\omega) \right|^2 d\omega \end{aligned} \quad (6)$$

由于 $\gamma = 2H + 1$ ，所以 R 的取值对于 H 值有着直接的影响。下面我们通过归一化相关系数考察正则度对 H 值影响的总趋势。归一化相关系数定义为^[12]：

$$\rho(l, l', j, j') = \frac{E[d_l^{(j)} d_{l'}^{(j')}]}{\sqrt{\text{Var}[d_l^{(j)}] \text{Var}[d_{l'}^{(j')}]}} \quad (7)$$

把式(4)，式(5)代入式(7)式得到

$$\begin{aligned} \rho(l, l', j, j') &= \frac{\sigma_x^2 2^{-(j+j')/2}}{\sigma^2 2\pi 2^{-(j+j')/2}} \int_{-\infty}^{+\infty} \frac{1}{|\omega|^\gamma} \widehat{\psi}(2^{-j}\omega) \\ &\quad \cdot \overline{\widehat{\psi}(2^{-j'}\omega)} e^{-i[l2^{-j} - l'2^{-j'}]\omega} d\omega \\ &= \frac{\sigma_x^2 2^{-(j+j')(1-\gamma)/2}}{\sigma^2 \pi} \int_0^{+\infty} \frac{1}{|\omega|^\gamma} \widehat{\psi}(2^{-j}\omega) \\ &\quad \cdot \overline{\widehat{\psi}(2^{-j'}\omega)} \cos[(l2^{-j} - l'2^{-j'})\omega] d\omega \end{aligned}$$

令 $l = l'$ ，可以得到

$$\rho(l; j, j') = \frac{\sigma_x^2 2^{-(j+j')(1-\gamma)/2}}{\sigma^2 \pi} \int_0^{+\infty} \frac{1}{|\omega|^\gamma} \widehat{\psi}(2^{-j}\omega) \cdot \overline{\widehat{\psi}(2^{-j'}\omega)} \cos[l(2^{-j} - 2^{-j'})\omega] d\omega$$

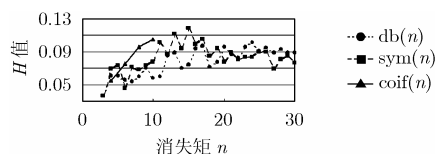
令 $j = j' + 1$ ，可以得到

$$\rho(l; j, j-1) \approx \frac{\sigma_x^2}{\sigma^2 \pi} \int_0^{+\infty} \frac{1}{|\omega|^\gamma} \left| \widehat{\psi}(\omega) \right| \cos[l(2^{-j-1})\omega] d\omega \quad (8)$$

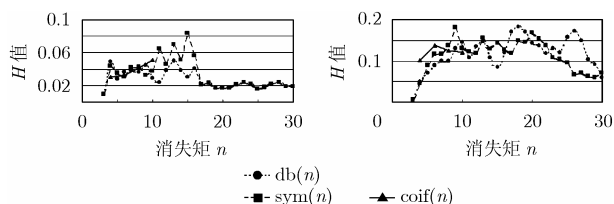
由式(8)可知随着 j 的增大，同一分解级数 ($l = l'$) 对应

表 4 不同小波消失矩求解的 H 值(信号长度为 1024, 信号 H 为 0.62)

消失矩	3	4	5	6	7	8	9	10	11	12	13	14	15	16
db(n)	0.6133	0.5694	0.5496	0.5308	0.5210	0.5201	0.4889	0.4740	0.5111	0.4838	0.4726	0.5107	0.5340	0.4997
sym(n)	0.6133	0.5751	0.5298	0.5041	0.5029	0.4828	0.4380	0.4895	0.4973	0.5016	0.4631	0.4907	0.4757	0.4929
coif(n)		0.5175	-	0.4827	-	0.4949	-	0.4982	-	-	-	-	-	-
消失矩	17	18	19	20	21	22	23	24	25	26	27	28	29	30
db(n)	0.4482	0.4375	0.4483	0.4765	0.4829	0.5003	0.4882	0.4984	0.4618	0.4478	0.4693	0.5157	0.5287	0.5595
sym(n)	0.5013	0.4693	0.4744	0.4521	0.4667	0.4827	0.5126	0.5168	0.5246	0.5533	0.5493	0.5581	0.5597	0.5502
coif(n)	-	-	-	-	-	-	-	-	-	-	-	-	-	-

图 3 不同消失矩求解的 H 值误差

为 0.56, 长度为 8192 的信号实验结果, 图 4 是表 3 对应的不同消失矩求解的 H 值误差, 表 4 是 H 为 0.62, 长度为 1024 的信号实验结果, 图 5 是表 4 对应的不同消失矩求解的 H 值误差。图 4、图 5 与图 3 实验结果一致。

图 4 不同消失矩求解的 H 值误差(长度为 8192, 信号 H 为 0.56)图 5 不同消失矩求解的 H 值误差(长度为 1024, 信号 H 为 0.62)

由于随着消失矩的增大, 分解级数越来越小, 实验拟合的点数越来越少, 此时求出的值不能代表正常情况, 如在表 4 中, 消失矩大于 27 时, 分解级数等于 4, db(n)与 sym(n)所求的 H 不减少反而一直增大。可见分解级数对于 H 值的影响也很大。消失矩越小, 分解级数越大, 所求的 H 值越准确。本文选用最大消失矩为 30, 当分解级数较大时(大于 7), 反映了消失矩周期性的减少这一规律。不难发现, 同一分解级数对应的消失矩所求得的 H 值呈现周期性变化, 如在表 2 中, db(n)在分解级数[5,8]范围内, 消失矩为 7 时, H 值最大, 两边的则变小, 在[9,16]内, 消失矩为 10 时, H 值最大。可见消失矩在一定的分解级数内所求得的 H 值呈现周期性变化, 而这些最高点连接起来又使得整体上呈现减少的变化。

5 结束语

小波分析方法求解 H 值比一般的方法(如 R/S 法)具有更高的精度, 但是选择不同的小波和消失矩往往会得到不同的结果, 这会对基于自相似和小波分析的应用必然带来影响, 如在基于小波检测 DDoS 攻击应用中, 如果选择最适合

于发现 H 值变化的消失矩, 这对于及时和准确地发现入侵异常会产生很好的效果。本文正是为这些应用提供小波选择的依据, 我们从理论和实验方面对小波消失矩选择进行了研究, 并做了分析, 可以看出, 实验结果和理论分析基本一致。总的规律是: 随着消失矩增大, 局部上其 H 值时大时小, 总体上 H 值在减小, 消失矩越小, 分解级数越大, 求解的 H 值越准确, 具有最小消失矩的小波误差最小, 是小波分析法求解 H 值的最佳选择。在实践中, 我们还发现, 随着消失矩的增大, 求解 H 值的所花费的时间越来越大, 这主要是因为计算量增大导致的。相信这些研究结果对基于自相似和小波分析的应用能提供有价值的参考, 未来的工作是研究如何使用小波系数方差法在线快速求解 H 值。

参考文献

- [1] Leland W, Taqqu M, and Willinger W, *et al.* On the self-similar nature of Ethernet traffic (Extended Version)[J]. *IEEE/ACM Trans on Networking*, 1994, 2(1): 1-15.
- [2] Beran J, Sherman R, and Traqu M S, *et al.* Long range dependence in variable bit rate video traffic [J]. *IEEE Trans. on Communication*, 1995, 43(2/3/4), 1566-1579.
- [3] Paxson V and Floyd S. Wide area traffic: The failure of Poisson modeling[J]. *IEEE/ACM Trans on Networking*, 1995, 3(3): 226-244.
- [4] Feldmann A, Gilbert A C, and Willinger W, *et al.* The changing nature of network traffic: Scaling phenomenon[J]. *Comp. Commun. Rev.*, 1998, 28(2): 5-29.
- [5] Xiang Y, Lin Y, Lei W L, *et al.* Detecting DDoS attack based on network self-similarity[J]. *IEE Proc Communications*, 2004, 151(3): 292-295.
- [6] Lan Li and Gyungho Lee. DDoS attack detection and wavelets[C]. Proc IEEE ICCCN (International Conference on Computer Communications and Networks) 2003, Oct. 2003: 421-427.
- [7] Ritke R, Hong X, and Gerla M. Contradictory relationship between Hurst parameter and queuing performance (extended version) [J]. *Telecommunication Systems*. 2001, 16(1-2): 159-175.

- [8] 李永利, 刘贵忠, 王海军, 等. 自相似数据流的 Hurst 指数小波求解分析法[J]. 电子与信息学报, 2003, 25(1): 100-105.
Li Yong-li, Liu Gui-zhong, and Wang Hai-jun, et al. Wavelet-based analysis of hurst parameter estimation for self-similar traffic[J]. *Journal of Electronics & Information Technology*, 2003, 25(1): 100-105.
- [9] Roghan M, Veitch D, and Abry P. Real-time estimation of the parameters of long-range dependence[J]. *IEEE/ACM Trans on Networking*, 2000, 8(4): 467-478.
- [10] Ma S and Ji C. Modeling heterogeneous network traffic in wavelet domain[J]. *IEEE/ACM Trans on Networking*, 2001, 9(5): 634-649.
- [11] 任勋益, 王汝传, 王海艳. 基于自相似检测 DDoS 攻击的小波分析方法[J]. 通信学报, 2006, 27(5): 6-11.
Ren Xun-yi, Wang Ru-chuan, and Wang Hai-yan. Wavelet analysis method for detection of DDoS attack based on self-similar[J]. *Journal on Communications*, 2006, 27(5): 6-11.
- [12] Daubechies I. Ten Lectures on Wavelets[M]. Philadelphia, PA: SIAM, 1992.
- [13] 李弼程, 罗建书. 小波分析及其应用[M]. 北京: 电子工业出版社, 2003.
Li Bi-cheng and Luo Jian-shu. Ananlysis and Application of Wavelet[M]. Beijing: Publishing House of Electronics Industry, 2003.
- 任勋益: 男, 1973 年生, 博士, 研究方向为计算机网络和网络计算、移动代理和信息安全技术等.
- 王汝传: 男, 1943 年生, 教授, 博士生导师, 主要研究方向为计算机软件、计算机网络和网络、信息安全、移动代理和虚拟现实技术等.
- 祁正华: 女, 1975 年生, 硕士, 助教, 研究方向为计算机软件、计算机网络和网络、信息安全技术等.