

# 基于奇异谱方法的网络行为分析

吴桦 丁伟

(东南大学计算机科学与工程系, 南京 210096)

**摘要:** 运用数字信号处理中的奇异谱分析方法, 讨论了如何提取网络流量的周期和趋势特征, 并结合最大熵谱方法对吞吐量趋势做出预测. 将该方法运用于中国教育和科研网 CERNET 华东(北)地区网的主干网, 分析了若干天的网络流量行为特征, 分析结果与使用情况相吻合. 对主干的流量进行了预测, 并将预测结果与实际情况做了比较. 定义了评价预测准确性的公式, 将预测的准确性与国际上具有代表性的成果进行多方面的比较, 结果说明该方法准确性较好.

**关键词:** 网络测量; 网络行为; 性能分析; 奇异谱分析

中图分类号: TP393.07 文献标识码: A 文章编号: 1001-0505(2002)06-0889-06

## Study of network behavior based on singular-spectrum analysis

Wu Hua Ding Wei

(Department of Computer Science and Engineering, Southeast University, Nanjing 210096, China)

**Abstract:** A method borrowed from digital signal processing is applied to identify and retrieve low-frequency variability and trend components from history metrics. Combined with maximum entropy method, predictions of network throughput can be made. This arithmetic was applied to the backbone throughput of China Education and Research Network Eastern China(North) Regional network, and the results were consistent with the practice. Predictions of the backbone throughput were made and the results were compared with the practice. A function which can evaluate the veracity of prediction results is defined, the veracity of this arithmetic is compared with the results of other international research communities. It is shown that this arithmetic has a better veracity.

**Key words:** network measurement; network behavior; performance analysis; singular-spectrum analysis

因特网已成为社会最重要的基础设施之一, 它的规模越来越大, 结构也越来越复杂. 但目前对网络结构和行为的研究却远远落后于它的技术发展水平, 致使很多新应用的服务质量不能得到保证, 因此, 对网络行为的研究一直是人们关注的问题. 网络行为主要表现为它的可用性的变化规律, 因此传统的网络行为研究主要是针对性能而言, 在这方面比较有代表性的研究成果是 IETF 的 IPPM (IP performance metrics) 工作组在测度定义方面的一些工作和 CAIDA (cooperative association for internet data

analysis) 所做的一些针对网络行为的测量研究.

目前在数据测量方面已经有很多方法和研究成果, 但对数据的分析却远远落后于数据采集技术. 最普遍的分析方法是根据特定的要求进行统计和摘要, 为管理员和用户实时或历史的数据汇总. 但是从当前网络运行管理的要求看, 管理员需从已有的数据中提取更多的信息. 为了规划网络的发展, 网络管理员要求得到网络的运行规律报告以及对网络行为的预测报告. 有了这些信息, 管理员才能保证网络的服务质量, 特别是在服务质量下降前通过政策调整或资源升级避免其发生. 于是一些时间序列模型被用于分析网络行为, 由于传统的宏观流量时序模型只能处理平稳过程和特殊的非

收稿日期: 2002-07-24.

基金项目: 国家 863 资助项目(2001AA112060)、国家自然科学基金重点课题资助项目(90104031).

作者简介: 吴桦(1973—), 女, 硕士, 助教; 丁伟(联系人), 女, 博士, 教授.

平稳过程,所以在描述流量行为时误差较大.如:AR<sup>[1]</sup>(auto regressive)模型、MA(moving average)模型和ARMA<sup>[2]</sup>(auto regressive moving average)模型用于解决平稳过程,ARIMA<sup>[2]</sup>(auto regressive integrated moving average)模型和ARIMA<sup>[3]</sup>季节模型、小波分解法<sup>[4]</sup>用于处理齐次的非平稳性过程等.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,数据中既含有多种周期类波动,又呈现非线性升、降趋势,还受到未知随机因素的干扰,而这些特点难以用传统模型来描述.

在本文中,将对网络主干网信道行为进行分析和预测.

对网络行为的预测主要是基于对历史数据的分析.由于天文学、海洋学、气象学等自然科学在时间序列的分析方面已经积累了很多模型,对网络行为的研究可以根据它的特点,从其他领域中挑选一些合适的模型进行拟合.网络行为有以下 2 个特点:

- 1) 有比较明显的周期性规律;
- 2) 经常发生各种意外的事件干扰网络的正常运行.

这 2 个特点是互相对立的,也就是说,一方面网络的运行是有规律可循的,另一方面各种突发事件会对这些规律造成干扰.考虑到这 2 个特征,一种首先运用在数字信号分析中的奇异谱(singular spectrum analysis, SSA)方法在本文中被用来分析网络变化规律.SSA 的功能是对于事先未知物理本质的系统,可以从它的包含噪声的有限长观测序列中提取尽可能多的可靠的信息,并依据这些信息建立预报模型.这一方法早已在数字信号处理、海洋学、气象学中得到有效的应用,它的好处是能够提炼出主要成分,滤去非周期性的异常现象,结合最大熵谱方法(maximum entropy spectral estimates, MEE),可以对网络行为的参数进行分析、预测.下面首先介绍 SSA 和 MEE,然后就如何利用 SSA 提取周期、趋势并结合 MEE 对未来进行预测,最后对预测结果的准确性进行了评价.

## 1 奇异谱方法和最大熵谱方法

### 1.1 奇异谱方法

有关 SSA 的较完整的介绍可见文献[5],现简单介绍.

SSA 分析的对象是一维时间序列,记为  $x_i, i = 1, 2, \dots, N$ ,要求是中心化的,它的  $M$  阶迟后协方差

矩阵记为  $T_x$ .  $T_x$  的第  $k$  个特征向量  $E^k$  称为第  $k$  个时间经验正交函数(T-EOF),分量记为  $E_j^k, j = 1, 2, \dots, M$ .第  $k$  个时间主成分(T-PC)为

$$a_i^k = \sum_{j=1}^M x_{i+j} E_j^k \quad 0 \leq i \leq N - M; 1 \leq k \leq M \tag{1}$$

任意 T-EOF 的  $M$  个分量构成一个时间序列,反映  $x$  序列中的时间演变型,时间主成分  $a_i^k$  是  $E^k$  表示的时间型在原序列的  $x_{i+1}, x_{i+2}, \dots, x_{i+m}$  时段的权重.

SSA 的重要功能由重建成分(reconstruction components, RC)实现,RC 用于在分析和预报中提取感兴趣的信息,过滤掉噪声.所谓重建是由 T-EOF 和 T-PC 重建一个长度为  $N$  的序列.由第  $k$  个 T-EOF 和 T-PC 重建  $x_i$  的成分记为  $x_i^k$ ,即

$$x_i^k = \begin{cases} \frac{1}{M} \sum_{j=1}^M a_{i-j}^k E_j^k & M \leq i \leq N - M + 1 \\ \frac{1}{i} \sum_{j=1}^i a_{i-j}^k E_j^k & 1 \leq i \leq M - 1 \\ \frac{1}{N - i + 1} \sum_{j=i-N+M}^M a_{i-j}^k E_j^k & N - M + 2 \leq i \leq N \end{cases} \tag{2}$$

RC 具有叠加性,所有 RC 之和等于原序列

$$x_i = \sum_{k=1}^M x_i^k \tag{3}$$

在求 T-EOF 的过程中,先将求得的迟后协方差矩阵  $T_x$  的特征值按绝对值大小进行排列,再依次求出相应的特征向量,即 T-EOF.经过这样的处理,建立重建成分的时候,反映原序列主要特征的重建成分排在前面.利用部分主要 RC 进行预报,再将预报结果进行叠加产生预报序列.通过降低噪声的干扰,可达到提炼主要趋势的目的.

对主要的 RC 进行趋势成分和周期成分的识别,还可以提炼出原序列的趋势成分和振荡周期成分.

检验某个 RC 是否属倾向成分可采用 Kendall 非参数检验.考查第  $k$  个 RC 时,计数满足  $i < j, x_i^k < x_j^k$  的指标  $(i, j)$  数  $k_y$ , 统计量

$$\tau = \frac{4k_y}{N(N-1)} - 1 \tag{4}$$

在  $x^k$  不是倾向成分的假设成立时服从均值为零,均方差为

$$S = \sqrt{\frac{2(2N+5)}{9N(N-1)}} \tag{5}$$

的正态分布,若取信度  $\alpha = 0.05$ ,当样本  $\tau$  值落在

$(-1.96S, 1.96S)$  以外时, 拒绝原假设, 认为第  $k$  个 RC 是倾向成分,  $\tau > 1.96S$  和  $\tau < -1.96S$  分别对应上升和下降倾向。

按 SSA 原理, 若样本中存在一个周期振荡成分, SSA 将得到一对 RC, 它们对应的特征值相等, T-EOF, T-PC 正交. 实际应用时, 这些条件不可能严格满足. 因此利用 Vautard 和 Ghil 提出的 3 个判据:

1) 对应的特征值接近相等 ( $\lambda_k \approx \lambda_{k+1}$ ).

2) T-EOF $_k$ , T-EOF $_{k-1}$  频率相近. 确定的方法是在  $f = 0.0 \sim 0.5$  之间 500 个等间距  $f$  上计算出  $|E^k(f)|^2$  和  $|E^{k+1}(f)|^2$ ,  $|E^k(f)|^2$  和  $|E^{k+1}(f)|^2$  达最大值的频率分别记为  $f_k$  和  $f_{k+1}$ , 要求差值  $\mathcal{G}_k = |f_k - f_{k+1}|$  最小. 由于纯红噪声有关系  $\mathcal{G}_k \approx 1/(2M)$ , 因此至少要求  $\mathcal{G}_k < 1/(2M)$ , 又由于要识别的是周期成分, 故此加强标准, 要求  $2M\mathcal{G}_k < 0.75$ .

3)  $|\tilde{E}^k(f)|^2$ ,  $|\tilde{E}^{k+1}(f)|^2$  足够大. 具体的判据是, 如果原序列中频率介于  $f_k$  和  $f_{k+1}$  之间的一个频率  $f^*$  的成分完全由这一对 RC 表示出来, 则

$$\frac{1}{M} [ |E^k(f^*)|^2 + |E^{k+1}(f^*)|^2 ] = 1$$

由于实际上不可能严格对应, 特别是对于一个孤立的频率, 因此取

$$\frac{1}{M} [ |E^k(f^*)|^2 + |E^{k+1}(f^*)|^2 ] > \frac{2}{3}$$

这意味着原序列中频率为  $f^*$  的振荡的方差至少有  $2/3$  被这一对 RC 表示出来.

满足以上 3 个条件的一对 RC 之和是原序列的一个周期成分.

在 SSA 中一个关键的问题对  $M$  的选择. 一般来说,  $M$  不要超过  $N/3$ . 对于如何准确地选择  $M$  尚无定论, Vautard 在 1992 年提出 SSA 分析介于  $(M/5, M)$  区间内的周期成分比较成功.

## 1.2 最大熵谱方法

熵是信息论中表示信息量大小的一个度量, 是信息量的加权平均. J. P. Burg<sup>[6]</sup> 提出了一种量测数据从一阶模型开始逐步增加阶数的方法, 通过使自回归模型序列  $x_i$  的前向预测误差和后向预测误差的平均为最小的方法求得 AR 模型的各参数值. 在规定最高阶数后, 程序从 1 阶开始按伯格递推算算法求出各种阶数下的自回归系数和残差方差, 并根据指定的 FPE (final prediction error 最小最终预报误差准则, 由模型的预报误差来判明自回归模型的结束是否恰当) 算式指出在上述阶数范围内的最佳阶数. 得到最佳阶数时的预报函数, 就此可向前预报.

## 1.3 建立预报模型

将 SSA 方法结合到 MEE 中, 通过 SSA 的过滤作用, 可降低噪声影响, 消除虚假谱峰. 具体做法是, 先对时间序列  $x$  做 SSA, 选出若干个有意义的时间主成分 (T-PC), 用最大熵谱的伯格算法估计这些 T-PC 的功率谱  $p_k(f)$ , 由此得到原序列的功率谱  $p_x(f)$  的近似估计, 同时得到每个 T-PC 序列的最优阶数的自回归模型的回归系数, 因此可对这些 T-PC 作出延伸预报, 然后用预报出的 T-PC 和原来计算出的 T-EOF 计算出原变量的预报. 这样得到的预报值利用主成分对信息进行过滤, 保留下来的是主要趋势, 因此结果比较有意义.

## 2 对网络流量进行分析

### 2.1 数据来源

在各种网络行为测度中, 流量的定义比较明确, 测量技术也比较成熟, 流量能够反映用户对网络的使用情况和需求, 因此, 可以用流量来验证上文所述的 SSA 分析预测方法. 现以自 2002-01-01—2002-01-18 中国教育和科研计算机网 CERNET 华东(北)地区网 (NJNET) 与 CERNET 主干之间的流量做为试验数据. NJNET 是 CERNET 中最大的地区网, 包含江苏、山东、安徽三省, 现有接入学校 169 所, 每天与 CERNET 间的流量交互达到 2 000 Gbit, 对这样大的数据进行分析, 可以得到 NJNET 中用户网络行为的统计特性. 按小时排列流量即可得到时间序列, 图 1 所示为流量曲线,  $T$  表示吞吐量.

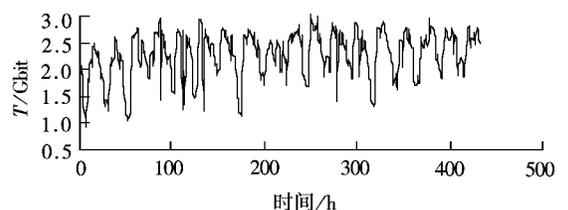


图 1 2002-01-01—2002-01-18 流量曲线

### 2.2 对趋向成分的提炼

下面用 SSA 对上述数据进行分析. 根据 1.1 中所叙的 Kendall 方法, RC1 的  $\tau$  值为 0.586,  $1.96S$  为 0.063, 满足  $\tau > 1.96S$ , 有上升的趋势. 图 2 为 RC1 的曲线图,  $T_{RC1}$  为 RC1 的值.

由于第一特征值的贡献率为 97%, RC1 反映了原序列的主要信息. 在滤去了周期成分后, 原序列的趋势变化主要被 RC1 反映出来了. 从曲线图上看, 在 2002-01-01—2002-01-04 间流量的增幅较大, 考虑到元旦的庆祝活动, 在元旦期间流量较少,

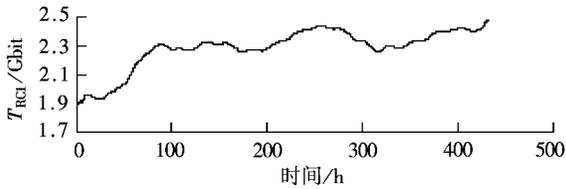


图 2 2002-01-01—2002-01-18 主要趋势图

元旦以后以较快速度回升。2 次下降都出现在周末, 几次峰值都出现在周三左右。虽然有升有降, 从峰值的大小比较可以直观地看出整个趋势还是上升的。这与 Kendall 检测得出的结论是一致的。

### 2.3 对网络流量振荡周期的分析

网络资源的使用有很明显的周期性, 特别是峰

表 1 主要成对 RC 的周期检验及参数

$k$	$k+1$	$\lambda_k$	$\lambda_{k+1}$	$2M\hat{\sigma}_k$	$\frac{1}{M} [  E^k(f^*) ^2 +  E^{k+1}(f^*) ^2 ]$	周期/h
2	3	$1.84 \times 10^{12}$	$1.53 \times 10^{12}$	0.000	0.968	23.91
4	5	$5.13 \times 10^{11}$	$4.99 \times 10^{11}$	0.160	0.952	12.74
6	7	$1.50 \times 10^{11}$	$1.44 \times 10^{11}$	0.240	0.690	7.95
14	15	$5.08 \times 10^{10}$	$4.95 \times 10^{10}$	0.240	0.824	2.50
16	17	$4.89 \times 10^{10}$	$4.75 \times 10^{10}$	0.240	0.692	3.00

一定的误差, 可以认为主要周期是 24 h。

虽然主要周期接近 24 h 是预料之中的, 但它体现了具体时段(线路忙或线路闲)的划分, 这对管理员制定管理策略有很大的帮助。图 3 为 RC2+RC3 所构成的周期曲线图, 图中  $T_{RC2+RC3}$  表示 RC2+RC3 的值。

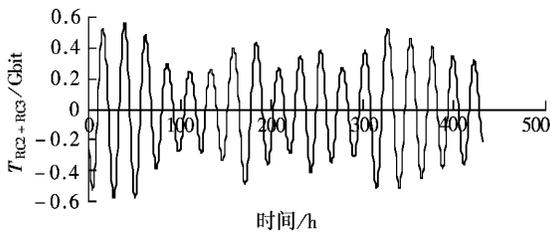


图 3 2002-01-01—2002-01-18 主要周期图

从图 3 可以看出休息日的振幅较大, 工作日振幅较小。此外, 通过对 RC2+RC3 数据的分析可以发现相对于日常生活中的白昼之分, NJNET 上的用户要迟后 3~4 h。最低值一般在每天凌晨 3:00~5:00, 然后持续上升, 峰值在下午 16:00~17:00 出现。这个周期特征的出现与 NJNET 的特殊用户组成有很大关系。由于高校学生普遍有晚睡晚起的生活习惯, 比之正常的作息, 上网时间也迟后了。因此, 如果用户有较大的下载任务, 应该避开下午的高峰期。

### 2.4 预报分析

结合最大熵谱方法, 对得出的前几个有意义的成分用最大熵谱的伯格算法同时递推得到每个 T-

值和低谷时的流量可能相差很大, 为了能充分利用网络资源, 合理安排网络负载, 必须了解网络资源的周期特性。应用奇异谱分析可以得到网络性能参数的周期特性。

根据上文介绍的 Vautard 和 Ghil 提出的 3 个周期判别方法对上文中的数据用 SSA 分析, 得到表 1, 发现满足上文所述 3 个条件而构成主要周期成分的成对 RC 分别为 RC2 与 RC3(周期为 23.91 h), RC4 与 RC5(周期为 12.74 h), RC6 与 RC7(周期为 7.95 h) 等。从特征值的大小来看, 最重要的周期是 23.91 h, 很接近 24 h, 由于存在着测量及辨认上的限制, 允许

PC 序列的最优阶数的自回归模型的回归系数, 对前几个 T-PC 做出延伸预报, 然后根据重建公式, 用预报出的 T-PC 和原来计算出的 T-EOF 重建出原变量的预报值。采集 2001-12-01—2001-12-27 NJNET 至 CERNET 主干的每小时流量数据, 以 2001-12-01 00:00—2001-12-26 16:00:00 的数据作为原始序列, 向后预报 24 h 的流量值, 并将结果与测量值做比较, 可以看出预报曲线是比较接近测量曲线的(见图 4)。

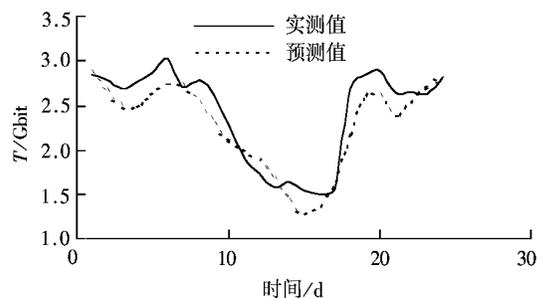


图 4 测量值与预值的比较

### 2.5 预报结果评价

图 4 得到的只是直观的印象, 不能作为严格的评价标准, 必须将预报的准确性进行量化处理, 并与国际上该领域内的成果进行比较。先简要介绍 CAIDA 的预报方法及其评价方法。

CAIDA 的网络天气服务(network weather service, NWS)<sup>[7]</sup>系统在分布式系统中定时测量资源的性能, 并用数学模型动态预测将来的性能水平。这个系统比较有开拓性的是它的预测模块, 这个模块

用了几种不同的预测算法对性能进行预测. 将监测到的数据按时标进行排列, 就可将数据按时间序列进行处理. 每一种特定的算法是该模块的一个独立组成部分. 需要进行预测时, NMS 评价这些不同的预测算法并根据最近预测的准确性选择一个最准确的方法.

目前该系统使用的预测算法主要有以下几种:

1) 基于平均值的预测(mean-based methods)

用自测试以来的所有数据的平均值做为预测值;

2) 基于中间值的预测(median-based methods)

用最近测试的  $k$  个数据的中间值作为预测值;

3) 自回归模型(autoregressive models)

利用最近测试的  $k$  个数据建立自回归模型.

CAIDA 使用了多种评价方法, 其中的一种易于进行比较的评价方法为 MIN\_MPE(minimum mean percentage prediction error). 首先定义:  $v(t)$  为  $t$  时刻的实际测量值,  $p_f(t)$  为对  $t$  时刻的预测值,  $e_f(t) = v(t) - p_f(t)$  为用  $f$  方法在  $t-1$  时刻所做的  $t$  时刻的预测值与实际测量值之间的绝对差. 然后定义  $M_f(t)$  为

$$M_f(t) = \frac{1}{t+1} \sum_{i=0}^t \frac{|e_f(i)|}{v(i)} \quad (6)$$

从 CAIDA 提供的报告来看, 以上 3 种方法向前预报一步的  $M_f(t)$  值基本上保持在 0.089 至 0.107 之间.

类似于  $M_f(t)$  值, 作者定义了以下公式评价预测准确性:

$$n_f(\tau) = \frac{1}{N} \sum_{i=1}^N \frac{|x(k+t+\tau) - p(k+t+\tau)|}{x(k+t+\tau)} \quad (7)$$

式中,  $x(k+t)$  为预报的起始点;  $x(k+t+\tau)$  和  $p(k+t+\tau)$  分别为从起始点开始向后  $\tau$  步的实际测量值和预报值;  $N$  为预报的次数. 与 CAIDA 的 MPE 评价方法不同的是多了参数  $\tau$  表示向前预报的步数, 这样该公式可以表示预报多步时的平均预报差错, 而 CAIDA 只向前预报一步, 因此没有这个参数. 对 2001-12-01—2001-12-27 NJNET 至 CERNET 主干的流量数据进行向前预报, 从不同的起始点开始进行 10 次 ( $N=10$ ) 预报, 每次预报其后 24 步的值, 由式 (7) 所得的值见表 2.

从式 (7) 可以发现, 相对于 CAIDA 所做的预测工作, 这种方法一步预测平均预报相对误差为 0.063. 前文提过 CAIDA 使用的各种预报方法的平

均相对预报误差为 0.08~0.10 之间, 这种方法的准确性要高出 2%~3%. 说明在滤去噪声的影响后, 对于近期的预报准确性比较高.

表 2 各步预报平均误差

$\tau$	$n_f$	$\tau$	$n_f$
1	0.063	6	0.204
2	0.075	7	0.206
3	0.118	⋮	⋮
4	0.152	23	0.214
5	0.183	24	0.227

同时可以看出远期预报的准确性比较差. CAIDA 没有进行多步预报的试验, 无法与之比较. 在本试验用的数据是每隔半小时采集的数据. 如果采集了长期的数据, 例如几年中每个月的吞吐量, 从表 2 看, 向后预报 3 个月的近期吞吐量的准确性还是可以接受的. 这就为管理员规划网络升级提供了充裕的时间.

以上比较是在对不同的实验环境中不同的算法进行比较. 为了使比较的结果更有说服力, 将 CAIDA 的算法在本地的实验环境中应用, 对同样的数据段进行预报实验(预报次数皆为 40 次). 将这 3 种方法和本文中方法的  $M_f(t)$  值进行比较, 得到的  $M_f(t)$  值见表 3. 对该表的结果进行观察, 可以得出 2 个结论:

1) 平均值预测法和中间值预测法的  $M_f(t)$  值偏大, 平均值达到 0.262 和 0.324, 自回归模型预测法的  $M_f(t)$  值为 0.078.

CAIDA 的预测结果显示上述 3 种方法的  $M_f(t)$  值基本上为 0.090 左右, 这说明环境不同, 预测方法的适用性完全不同. 对于 NJNET 来说, 比较适用的是自回归模型, 此外从对 SSA 与 MEE 的介绍可以看出, 这种方法其实是优化的自回归模型, 它的预测准确性接近但优于自回归模型的结果.

2) 由于网络行为变化除了固有的规律外, 还受到很多突发事件的影响, 这些事件相互之间没有前后相关性, 会干扰对网络的研究. SSA 方法对异常数据的过滤作用使得预测的准确性提高了 1.6% 左右.

表 3 各种方法的  $M_f(t)$  值

算法	最大	最小	平均
平均值法	0.863	0.006	0.262
中间值法	1.645	0.006	0.324
自回归模型法	0.287	0.001	0.078
SSA 与最大熵谱法	0.158	0.002	0.061

通过以上两个方面的比较可见, 本文介绍的方法优于国际上同类研究. 这种方法在网络行为研究

方面的作用还可以引申到其他网络行为参数.

### 3 结 论

本文介绍了一种对网络行为进行周期和趋势分析的方法. 该方法得出的周期成分和趋势成分(包括线性的和非线性的)是依据数据变化的自身特征为依据(自适应), 而不是以任何假设为依据. 结果显示用 SSA 与 MEE 方法进行预测的结果与实际测量值之间还有一定的差距, 不准确性的来源有以下 3 个主要原因:

1) 网络行为是各种因素综合作用的结果, 同时受到各种突变因素如自然灾害, 重大政治事件的影响, 由于其中涉及的因素是如此之多, 它们之间的耦合又是那么紧密, 只能依据统计力学来研究, 不可能得到确定的解. 显然, 对网络行为的预测只能是不确定预测.

2) 对网络行为进行分析和预测是建立在对现有状态的测量上的, 无论用哪种测量方法, 都会产生误差. 这些误差必然会影响分析结果的准确性.

3) 用来预测的最大熵谱方法是线性自回归模型, 但网络的变化并不是线性的, 这样会造成对误差的放大作用.

针对以上分析, 可以采用下列方法进行改进:

1) 对最大熵谱方法进行调整. 用最大熵谱求自回归系数时, 用最终预报误差作为定阶标准, 而最终预报误差的计算有多种方法, 可以通过试验选取最佳方法.

2) 可用一些非线性模型进行预测, 如自适应模型, 门限自回归模型代替最大熵谱方法进行前向预报.

### 参考文献 (References)

- [1] Wolski Rich. Forecasting network performance to support dynamic scheduling using the network weather service [DB/OL]. <http://citeseer.nj.nec.com/wolski98dynamically.html>. 1998-04-07/2002-10-10.
- [2] Basu S, Mukherjee A. Time series models for internet traffic [A]. In: *Proc IEEE INFOCOM'96* [C]. San Francisco, CA, 1996, 2: 611-620.
- [3] Groschwitz N, Polyzos G. A time series model of long-term traffic on the NSFnet backbone [A]. In: *Proceedings of the IEEE International Conference on Communications (ICC'94)* [C]. New Orleans, LA, 1994. 1400-1404.
- [4] 徐科, 徐金梧, 班晓娟. 基于小波分解的某些非平稳时间序列预测方法 [J]. *电子学报*, 2001, 29(4): 566-568.  
Xu Ke, Xu Jinwu, Ban Xiaojuan. Forecasting of some non-stationary time series based on wavelet decomposition [J]. *Acta Electronica Sinica*, 2001, 29(4): 566-568. (in Chinese)
- [5] Vautard. SSA: a toolkit for noisy chaotic signals [J]. *Physica D*, 1992, 58: 95-126.
- [6] Burg J P. Maximum entropy spectral analysis [A]. In: Childers D G, ed. *Modern spectral Analysis* [C]. New York: IEEE Press, 1978. 34-41.
- [7] Wolski Rich. Dynamically forecasting network performance using the network weather service [J]. *Journal of Cluster Computing*, 1998, 1: 119-132.