

基于 FARIMA 过程的网络业务预报与应用¹

刘嘉焜 金志刚 薛飞 舒炎泰

(天津大学电子信息学院计算机系 天津 300072)

摘要 在高速网络控制与带宽分配研究中,网络业务量预报是一个重要的问题。本文首先介绍自回归分数整合滑动平均(FARIMA, Fractal Autoregressive Integrated Moving Average)过程的概念及其具体形式,并给出了 FARIMA 过程的预报方法,然后基于 FARIMA 过程的最小均方误差预报方法,提出了一个具有补偿功能的网络自相似业务的预报方法,最后给出这种预报方式在网络控制研究中的应用。

关键词 FARIMA 过程, 最优线性预报, 具有补偿功能的网络业务预报

中图分类号 TN913.2

1 前言

随着宽带网络服务需求的激增,高速网络传输技术成为目前网络研究的热点。90年代初期,由美国贝尔实验室和 Boston 大学研究人员所组成的小组采集了大量关于以太网 Ethernet^[1]和可变比特率 VBR(Variable Bit Rate) 视频服务^[2]的业务数据,对这些数据的分析证明:实际网络业务普遍存在统计上的自相似性及长相关特性。近年来人们发现^[1-3],业务的自相似性及长相关特性对队列系统和网络性能都产生了一些意想不到的影响,而基于传统模型下的网络性能结论和网络设计策略不能完全适用于具有自相似业务的高速网络中。最新研究表明利用 FARIMA(p, d, q)过程构造网络业务模型成为描述实际网络的有力工具^[4-6]。本文在介绍 FARIMA 过程的概念后,讨论了 FARIMA 过程的最优线性预报方法和预报效果。在此基础上,提出了一个具有补偿功能的网络业务预报方法,并讨论了该预报方法在动态带宽分配算法中的应用。由于 FARIMA 模型较准确地描述了业务的重要特性^[3],所以基于该模型的预报方法在网络控制、动态带宽分配等的研究中有较好的应用前景。

2 FARIMA 模型及其最优预报

随机过程 $\{W_t\}$ 称为服从 $d \in (-0.5, 0.5)$ 的 FARIMA(p, d, q) 模型,如果 W_t 是平稳(零均值)的,且满足差分方程

$$\Phi(B)\nabla^d W_t = \Theta(B)a_t \tag{1}$$

其中 $\{a_t\}$ 为白噪声序列, $\Phi(B)$ 和 $\Theta(B)$ 分别是 p 阶和 q 阶多项式, $\nabla = 1 - B$, 这里 B 是延迟算子,即 $BW_t = W_{t-1}$ 。显然, $\{W_t\}$ 是 $d \in (-0.5, 0.5)$ 的 FARIMA(p, d, q) 过程当且仅当 $\nabla^d W_t$ 是一个 ARMA(p, q) 过程。

设 $\{W_t\}$ 是 $d \in (-0.5, 0.5)$ 的 FARIMA(p, d, q) 过程,且(1)式中多项式 $\Theta(B)$ 与 $\Phi(B)$ 无公共的根,则

(1) 如果 $\Phi(B) \neq 0, |B| \leq 1$, 则(1)式有唯一的平稳解,形如

$$W_t = \sum_{j=-\infty}^{\infty} \psi_j \nabla^{-d} a_{t-j} \tag{2}$$

其中 $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j = \Theta(z)\Phi^{-1}(z)$ 。

(2) $\{W_t\}$ 具有物理可实现的单边滑动平均表示的充分必要条件是 $\Phi(B) \neq 0, |B| \leq 1$ 。

¹ 1998-09-14 收到, 1999-06-26 定稿

国家自然科学基金资助课题(项目号: 69672031)

(3) $\{W_t\}$ 可逆的充分必要条件是 $\Theta(B) \neq 0, |B| \leq 1$ 。

(4) 如果 $\{W_t\}$ 具有物理可实现的单边滑动平均表示且可逆, 则对 $d \neq 0$, 其自相关函数和谱密度有

$$r(h) \sim Ch^{2d-1}, \quad h \rightarrow \infty \quad (3)$$

其中 $C \neq 0$ 为常数。

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\Theta(e^{-i\lambda})}{\Phi(e^{-i\lambda})} \right|^2 |1 - e^{-i\lambda}|^{-2d} \sim \frac{\sigma^2}{2\pi} \left[\frac{\Theta(1)}{\Phi(1)} \right]^2 \lambda^{-2d}, \quad \lambda \rightarrow 0 \quad (4)$$

因此, 当 $0 < d < 0.5$ 时, FARIMA(p, d, q) 过程是长相关的、渐进二阶自相似过程, 且其 Hurst 参数 $H = d + 1/2$ 。

将 (1) 式表示为

$$W_t = \sum_{j=0}^{\infty} \pi_j a_{t-j}, \quad a_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}$$

其中 $\sum_{j=0}^{\infty} \pi_j B^j = \Theta(B)\Phi^{-1}(B)(1-B)^{-d}$, $\sum_{j=0}^{\infty} \psi_j B^j = \Phi(B)\Theta^{-1}(B)(1-B)^d$, $|B| < 1$ 。这样一来^[7], 我们已把 FARIMA(p, d, q) 过程表示成线性过程形式了。由此可得到向前 h 步预报值为 $\hat{W}_t(h) = \hat{E}(W_{t+h}|W_j, j \leq t) = -\sum_{j=1}^{\infty} \psi_j^{(h)} W_{t+1-j}$ 或 $\hat{W}_t(h) = \sum_{j=h}^{\infty} \pi_j a_{t+h-j}$, 其中 $\psi_j^{(1)} = \psi_j$, $\psi_j^{(h)} = \psi_{j+h-1} - \sum_{i=1}^{h-1} \psi_i \psi_j^{(h-i)}$, ($h > 1$)。

预报均方误差为

$$\hat{\sigma}_t^2(h) = E[W_{t+h} - \hat{W}_t(h)]^2 = \sigma^2 \sum_{j=0}^{h-1} \pi_j^2 \quad (5)$$

下面以贝尔实验室提供的 pAug.TL 数据为例介绍预报效果。我们首先根据 2000 个样本点建立了 FARIMA(1, d , 1) 模型与 AR(8) 模型^[6]。我们得到了 pAug.TL 的 FARIMA 模型为 $(1 + 0.1714523B)\nabla^{0.2949911}(Z_t - 3.8995) = (1 + 0.373878B)a_t$, $\sigma_a^2 : 5.091458$ 。对 AR(8) 模型, 估计所得的自回归参数 $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_8$ 依次为 0.48407406, -0.03356442, 0.06929456, 0.08137728, 0.06636146, -0.05099568, 0.04781558, 0.04714202, $\hat{\sigma}_a^2 = 5.027846$ 。

作为比较, 我们利用传统的 AR(8) 模型^[6]和上述 FARIMA 模型的预报方法来预报后续的 100 个值, 结果见图 1。从图 1 可以看出, 依 AR 模型作预报, 其预报值很快接近原过程的平均值, 即对后续的较远的值无法预报。这是由 AR 模型的短相关特性决定的。而用 FARIMA 模型预报时, 由于该模型能够同时刻画实际业务的长相关和短相关特性, 其预报值并未很快接近原过程的均值, 而是更接近网络业务的实际值。

3 具有补偿功能的网络业务预报与控制应用

在宽带网络控制中, 能有效地预报网络业务在下一时间周期上的行为是有意义的。这个预报值将为网络控制提供重要信息, 用于监控和分配网络业务所需的带宽或缓冲区。业务预报的基础是如何来描述网络业务。在传统的业务描述中, 由于忽略了业务中重要的长相关特性和突发特性, 导致不能准确反映业务特性, 从而使基于这些描述的预报值也欠准确, 严重影响了控制效果和网络性能。自相似业务的 FARIMA 模型较为全面地描述了业务间的相关特性, 所以基于 FARIMA 模型的业务预报技术可应用在网络控制的研究中。

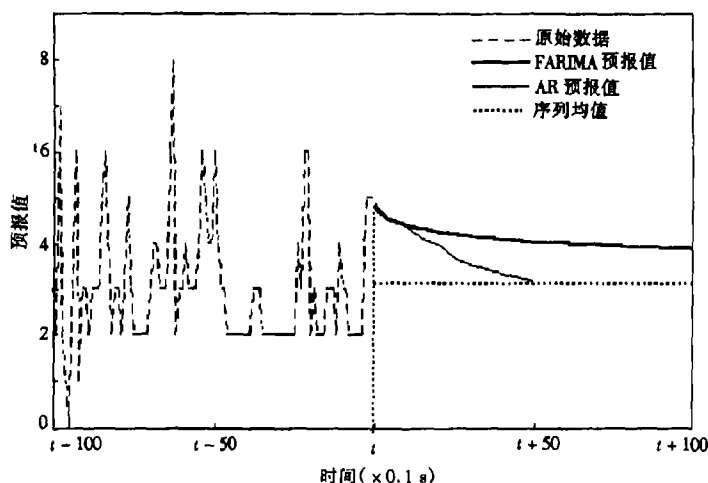


图1 不同模型的预报效果比较

3.1 具有补偿功能的网络业务预报 第2节中所介绍的最小均方误差预报技术是进行自相似业务预报的重要基础。但在实践中,对网络业务来说,如果预报值比真实业务值小,那么所分配的带宽或缓冲区就不能满足实际业务的需要,将导致数据包或信元的丢失。这将影响网络服务的质量,并降低网络性能。为了减少这种影响,我们在最小均方误差预报技术的基础上,提出一种具有补偿功能的预报方法。该方法的补偿作用由补偿系数 u 决定,其中 $0.5 \leq u < 1$ 。

在利用 FARIMA 模型对网络业务 $\{z_t\}$ 建立模型时,我们可以得到拟合的残差 $\{a_t\}$ 。记 $\hat{z}_t(r)$ 是基于无穷个过去数据 z_t, z_{t-1}, \dots 的 r 步最小均方误差预报值。

定义 z_{t+1} 补偿系数为 u 的补偿预报值 z_{t+1}^u 是一步最小均方误差预报值 $\hat{z}_t(1)$ 与 ξ_u 之和,即 $z_{t+1}^u = \hat{z}_t(1) + \xi_u$, 其中 $p\{a_t \leq \xi_u\} = u, 0.5 \leq u < 1$ 。

从定义可以看出, ξ_u 就是残差 $\{a_t\}$ 分布的单侧 u 分位点值。为了进行补偿系数为 u 的一步预报,我们有如下算法:(1)建立网络业务的 FARIMA 模型;(2)利用 FARIMA 模型拟合网络业务的残差 $\{a_t\}$,得到残差分布的单侧 u 分位点值,记为 ξ_u ;(3)根据 FARIMA 模型得到一步最小均方误差预报,记为 $\hat{z}_t(1)$;(4) z_{t+1} 补偿系数为 u 的补偿预报值为 $\hat{z}_t(1) + \xi_u$ 。

补偿预报值 z_{t+1}^u 有如下性质: $P\{z_{t+1} \leq z_{t+1}^u\} = u$ 。这是因为 $e = z_{t+1} - \hat{z}_t(1)$, e 的分布不依赖于 z 的历史值, $P\{z_{t+1} \leq z_{t+1}^u | z_t, z_{t-1}, \dots\} = P\{e \leq \xi_u | z_t, z_{t-1}, \dots\} = P\{e \leq \xi_u\} = u$ 。

由于残差 $\{a_t\}$ 服从正态分布,其分布的对称点在零值附近,零就是残差的单侧 0.5 分位点值,所以最小均方误差预报就是 $u = 0.5$ 的补偿预报。由补偿预报的性质可知,对于给定的 u 值,补偿预报值有 $100u\%$ 的可能高于或等于实际观测值,而有 $100(1-u)\%$ 的可能低于实际观测值。所以, u 的选取依赖于不同的网络服务类型。若在传输中数据包或信元的丢失对网络的影响较大(如数据传输等),那么 u 就应选取较大的值;反之,对那些在传输中数据包或信元的丢失不敏感的服务类型(如视频服务)来说, u 就没必要选取太大的值。 u 值越高,补偿预报值低于真实值的可能性就越小。但当 u 接近 1 时,存在着降低网络效率的问题。因此, u 的选取需要在网络效率和信元丢失率之间进行折衷。

为了简化 $\{a_t\}$ 的计算,我们可以直接应用模型 $AR(p)$ 拟合网络业务的残差来近似 $\{a_t\}$ 。另外,由于残差近似为正态分布 $N(0, \sigma^2)$,其中 σ^2 为残差的方差,所以 ξ_u 可以由标准正态分布 $N(0, 1)$ 的 $100u$ 百分位点值乘以 σ 得到。

3.2 基于具有补偿功能预报值的动态带宽分配算法 在宽带网络环境中, 如何进行业务控制和性能管理值得研究。由于网络业务相当复杂, 不同业务带宽需求的许多特性是不可预知的, 导致在业务传输之前就建立一个完善的控制机制是不可能的。另外由于 ATM 网络中信息传输的速度很高, 若采用传统的反应式拥塞控制方法, 当系统感知到拥塞并拟作出反应时, 已来不及进行有效的控制了。因此 ATM 网络主要采用预防式拥塞控制, 其基本思想是在拥塞发生前就对网络业务过程进行预报, 并加以控制, 从而避免网络进入拥塞状态。这就要求网络系统在运行时, 应具有动态识别业务特性, 自适应地进行决策的功能。业务预测被证明对预防式拥塞控制能起很大作用。利用业务的预报值进行带宽的动态分配就是基于上述观点的一种策略。

我们给出一个动态带宽分配算法, 该算法以具有补偿功能的业务预报值为基础进行带宽分配。

动态带宽分配算法

For cycle=1 to Total-cycle do

(1) 带宽分配阶段

For source=1 to N do

(a) 利用补偿系数为 $u[\text{source}]$ 的补偿预报器预报下一周期业务的带宽需求 $\text{predict}[\text{source}]$;

(b) 如果预报带宽的总和大于可分配带宽即 $\text{sum}(\text{predict}[\text{source}]) > \text{capacity}$, $\text{allocate}[\text{source}] = \text{predict}[\text{source}] - (\text{sum}(\text{predict}[\text{source}]) - \text{capacity}) / N$

(c) 如果预报带宽的总和并不大于可分配带宽即 $\text{sum}(\text{predict}[\text{source}]) \leq \text{capacity}$, $\text{allocate}[\text{source}] = \text{predict}[\text{source}] + (\text{capacity} - \text{sum}(\text{predict}[\text{source}])) / N$

(2) 传输阶段

For source=1 to N do

(a) 按阶段 1 中所分配的带宽传输到达业务;

(b) 数据包或信元在缓冲区中排队或离去, 若缓冲区已满, 则数据包或信元丢失。

动态带宽分配控制器负责周期性地为每个业务到达所在的虚信道 VC(Virtual Channel) 分配所需的带宽。利用补偿预报器所预报的业务值, 控制器按上述算法中的分配策略为每个 VC 分配带宽。每个业务到达按所分配的带宽进行传输。该分配算法的性能主要受周期长度(即更新带宽分配的时间间隔)、预报的运算时间等的影响, 有关分析需要深入进行。这里我们只讨论补偿系数 u 对算法的影响。

当补偿系数 u 取较大值时, 补偿预报值大于真实值的可能性较高, 这意味着其所要求的带宽偏高。但由于上述算法中带宽分配阶段的策略, 网络可以均衡地分摊不足带宽或追加剩余带宽, 所以根据预报值仍可以为业务分配一个合理的带宽。因此, 我们认为具有补偿功能的预报算法在网络控制上具有应用价值。

在实际网络应用中, 业务预报应是实时在线的。因此上述算法还需要结合业务的实时建模分析进行深入研究。

4 结论

本文由 FARIMA 过程出发, 给出了利用自回归分数整合滑动平均过程拟合实际网络业务的最优线性预报公式和方法。通过与传统的基于短相关的 AR 模型的预报效果的比较, 说明了 FARIMA 过程对长相关过程的预报能力优于传统模型。为了将基于 FARIMA 过程的预报方法应用于实际网络环境, 本文提出了具有补偿功能的网络业务预报。本文还提出了具体可行的基于业务预报的网络资源控制算法。

参 考 文 献

- [1] W. Leland, *et al.*, On the self-similar nature of Ethernet traffic (extended version), IEEE/ACM Trans. on Networking, 1994, 2(1), 1-15.
- [2] J. Beran, *et al.*, Long-range dependence in variable bit rate video traffic, IEEE Trans. on Commun., 1995, COM-43(2/3/4), 1566-1579.
- [3] W. Willinger, *et al.*, Self-similarity in high-speed packet traffic, Analysis and modeling of Ethernet traffic measurements, Statistical Science, 1995, 10(1), 67-85.
- [4] C. W. J. Granger, R. Joyeux, An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis, 1980, 1(2), 159-169.
- [5] J. R. M. Hosking, Fractional differencing, Biometrika, 1981, 68(1), 165-176.
- [6] 薛飞, 自相似网络业务的建模分析与性能评价, 博士论文, 天津, 天津大学, 1998年6月.
- [7] G. Box, G. Jenkins, Time Series, Forecasting and Control, revised edn., San Francisco, Holden-day, 1976, Ch.5.

TRAFFIC PREDICTION AND
ITS APPLICATION USING FARIMA MODELS

Liu Jiakun Jin Zhigang Xue Fei Shu Yantai

(Tianjin University, Tianjin 300072, China)

Abstract Network traffic prediction is important for network control and bandwidth allocation. This paper first introduces how to get optimal forecasting values, and provides prediction procedure for FARIMA models. And then provides a prediction method for self-similar traffic with compensation function and also gives out an example to demonstrate how to use this method in network control.

Key words FARIMA, Optimal linear prediction, Traffic prediction with compensation

刘嘉焜: 男, 1944年生, 副教授, 研究方向为概率论和数理统计.
金志刚: 男, 1972年生, 博士后, 研究方向为计算机网络及其性能评价.
薛飞: 男, 1970年生, 博士, 研究方向为计算机网络.
舒炎泰: 男, 1942年生, 教授, 研究方向为计算机网络、CIMS和实时系统.