第 46 卷 第 2 期
2007 年 3 月
厦门大学学报(自然科学版)
Journal of Xiamen University (Natural Science)
Vol.46 No.2
Mar. 2007

信息流量预测模型及其研究

计国君
(厦门大学管理学院,福建 厦门 361005)

摘要:信息流量的实时监测对信息源的统计、信息系统设计等有重要价值.因此预测信息流量过载的发生及其发生的时间,从而提高系统及设备的规划、优化和控制的安全性和可靠性.本文利用已经得到的信息流量预处理方法,建立了信息流量预测模型,根据 ARMA 模型的最小线性均方误差预测方法,预测信息流量,得到了在未来超过阈值的概率,所应用方法改善了预测效果,同时根据对预测结果进行分析和评价,结论表明总体趋势特性良好.

关键词:信息流量;ARMA;预处理
中图分类号:O 236　　文献标识码:A　　文章编号:0438-0479(2007)02-0164-06

信息流量的实时监测是信息管理的一个重要方面,对信息源的统计、信息系统设计等有重要意义.从理论上而言,一般是通过设定流量的阈值对流量进行监测.当网络系统出现异常情况时,通常根据管理系统得悉发出告警(诸如电话、邮件通知等方式)通知之后,由系统管理人员着手解决出现的问题,这是一种滞后响应式的行为,即先问后答的方式.该方式因适时性难以保证常常会使服务受到影响,从而造成诸多用户的不满意.因为当系统管理人员发现这样的警报时,往往没有足够的时间来分析并采取措施,很可能会影响整个系统的正常运行.但如果能够预测流量,则在流量过载发生之前,有充分的时间与准备去分析问题,并考虑应该采取的防范应对措施,避免严重问题的出现,或者减轻对网络及系统的影响,从而提高网络及系统的安全性和可靠性.这样的流量预测的方式可从根本上改变以往管理中被动响应的方式,即实现信息流量过载问题的预警功能,或者称为预先的信息管理方式(Proactive Information Management, PIM)[1].

所谓预测,是以历史的已知状况作为输入,在可能的预测变换(或者算子、方法)的作用下,得到未来输出结果的过程,如图 1 所示.一般而言,预测不存在惟一及确定的方法.

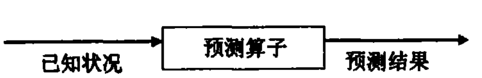

图 1　预测问题示意图

Fig.1　Forecasting problems sketch

从数学上说,预测就是从一个时间序列的历史的数据估算整个系统的统计参数,确定预测算子,应用统计的方法进行预测.统计预测方法在众多学科领域都有广泛的应用,它是以统计学理论为基础、以统计量(平均值、方差等)为对象进行的预测,在一定程度上可以反映客观现象的规律性.应用统计预测的方法预测网络流量,是一种定量的、定时的预测,即对某个流量在未来某个时刻的量做出预测,是一种短期预测的行为,它直接影响到当前的系统管理者对网络运行情况的判断,需要较高的综合评价的精确度[2].应用到信息流量预测的方法也很多,主要的预测技术有最小均方误差预测、神经网络、自回归滑动平均模型预测、FARIMA 模型预测等,每种预测方法都有各自的优点和缺陷.本文不讨论有关具体系统设计及方法,而是以理论方法来描述该问题,即利用已经得到的信息流量预处理方法[3],建立信息流量预测模型,根据 ARMA 模型的最小线性均方误差预测方法,来预测信息流量,并对流量的过载进行预测,得到在未来超过阈值的概率,并对预测结果进行分析和评价,结论表明总体趋势特性良好.

1　建立信息流量模型

为进行信息流量的预测,我们将建立流量预测模型.预测理论表明,任何预测都是不准确的,存在预测误差和预测误差方差,在基于 ARMA 模型的最小均方误差预测中,可以得到预测误差方差.由于实际测量参数的方差会随均值的增加而增大,这样的测量值不能简单地刻画为标准正态随机变量.为减小预测误差,必须减小测量参数的均值,实践与理论表明取对数是减少标准差的一种最有效方法.因此,首先对流量的观

收稿日期:2006-04-11
Email:jiking@xmu.edu.cn

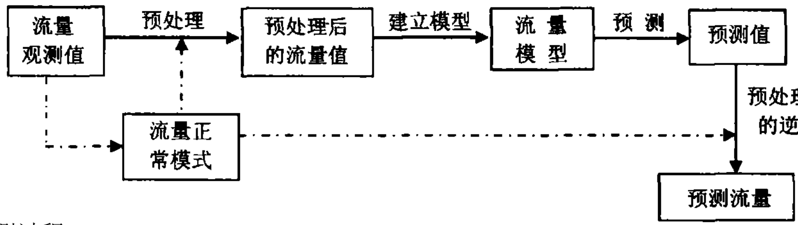

图 2　流量预测过程

Fig. 2　Forecasting process of information flux

测值取对数, 然后在对取对数的模型利用预处理办法. 通过取对数的办法, 对流量的观测值进行转换, 这样可以减小预测误差的标准差, 使得流量的预测更准确.

如图 2 是流量预测的过程. 首先是对采集的实际观测值进行预处理, 然后建立模型 ARMA(2, 1), 再进行真实流量的预测, 最后进行预处理的逆过程, 得到实际预测流量.

对真实信息的流量取自然数为底的对数, 利用预处理方法及拟合自回归滑动平均模型, 从而得到网络流量的 ARMA(2, 1)模型. 假定 S_{ij} 是流量正常行为模式序列, 用 μ 表示 $\ln S_{ij}$ 的总平均, a_i 表示在第 i 个时刻 $\ln S_{ij}$ 的平均值与总平均值 μ 的偏差, β_j 为第 j 天 $\ln S_{ij}$ 的平均值与总平均值 μ 的偏差, 这样有:

$$\mu = \frac{\sum_{j=1}^{5}\sum_{i=1}^{288}\ln S_{ij}}{5\times 288},$$

$$\alpha_i = \frac{\sum_{j=1}^{5}\ln S_{ij}}{5} - \mu,$$

$$\beta_j = \frac{\sum_{i=1}^{5}\ln S_{ij}}{288} - \mu.$$

假设流量的正常行为模式基于前 8 周的流量观测值而得, 现在对下一周的流量观测值 X_{ij} 建立 ARMA(2, 1) 模型. 依据实测数据分析, 建立关于 X_{ij} 的如下对数模型:

$$\ln X_{ij} = \mu + \alpha_i + \beta_j + y_{ij} \tag{1}$$

即:

$$y_{ij} = \ln X_{ij} - \mu - \alpha_i - \beta_j \tag{2}$$

对观测值序列 X_{ij} 经过预处理后的结果 y_{ij} 如图3 与图 4 所示. 图 5 是预处理前后的流量观测值的自相关函数, 从图中可见, 对整个预测处理过程前后, 自相关函数发生了很大的变化, 处理后的序列已接近于平稳序列, 可以用 ARMA 模型拟合.

对处理后的时间序列 y_{ij}, 建立自回归滑动平均混合模型 ARMA(2, 1), 即:

$\varphi(B)y_{ij} = \theta(B)a_{ij}$,

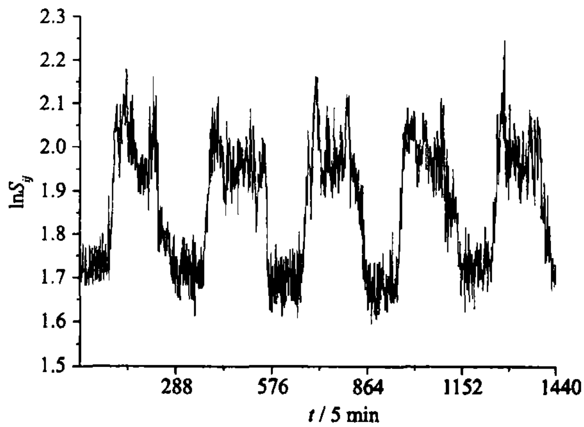

图 3　采样数据取对数后结果图

Fig. 3　Logarithmic sampling data sketch

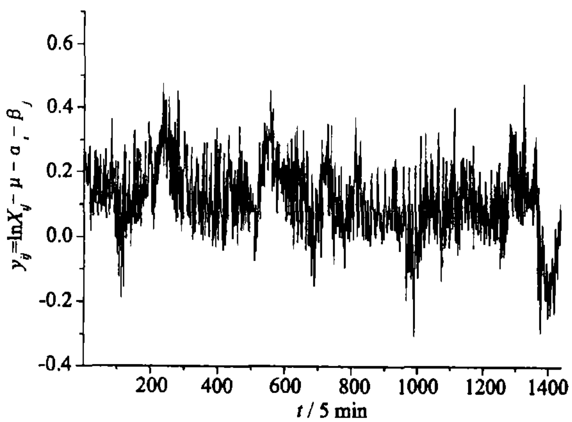

图 4　X_{ij} 预处理后得到的序列

Fig. 4　Sequence based on X_{ij} pretreated

其中, a_{ij} 是高斯白噪声, B 是后向算子满足:

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2$;

$\theta(B) = 1 - \theta_1 B$.

依据样本数据, 利用矩估计方法, 可得到其中的参数如下:

$\varphi_1 = 0.98861, \varphi_2 = -0.08235,$

$\theta_1 = 0.67572, \sigma_a^2 = 0.00161,$

其中 φ_1 和 φ_2 同时满足以下条件:

$\varphi_1 + \varphi_2 = 0.90626 < 1$;

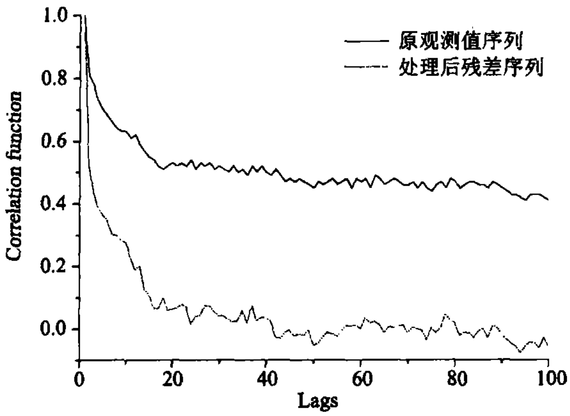

图 5　X_{ij} 预处理前后的自相关函数

Fig. 5　Auto relative function based on fore and aft pretreatment

$\varphi_2 - \varphi_1 = -1.070\ 96 < 1,\ |\varphi_2| < 1.$

根据平稳性判定准则, 得到其为一个平稳的时间序列. 于是拟合的 ARMA(2, 1) 模型为:

$$y_t = 0.988\ 61 y_{t-1} - 0.082\ 35 y_{t-2} + a_t - 0.675\ 72 a_{t-1},$$

其中 t 与式(1)、(2) 中的 i, j 相关, 即 $i = t$ 被 288 除后所得余数, 且当 t 是 288 的整数倍时, $i = t/288$; $j = t/288 + 1$, 这里$[x]$ 表示小于 x 的最大整数.

2　预　测

时间序列的预测是利用一个时间序列在 t 时刻的有效观测值来预测在某个未来时刻 $t + l$ 该序列的值, 假定观测值是在离散的、等间隔区间上得到的. 如在一个销售预测问题中, 当上月份 t 的销售额 x_t 和历史月份的销售额 $X_{t-1}, X_{t-2}, \cdots$ 已知时, 可用来预测未来 $L = 1, 2, \cdots, 12$ 个月的销售额. $X_t(l)$ 记作在原点 t 对未来某个时刻 $t + l$ 的销售额 X_{t+l} 所做的预测值, 也就是提前期为 l 的预测. 函数 $X_t(l), l = 1, 2, \cdots$, 给出了在原点 t 对所有未来提前期的预测值, 也称为在原点 t 的预测函数. 目的是构建这样的一个预测函数, 并使得对于每一提前期 l, 实际值与预测值之间偏差 $X_{t+l} - X_t(l)$ 的均方值尽可能小.

对应于流量模型, 用线性最小方差预测方法预测流量, 即用 X_{t+l} 的条件期望值 $X_t(l)$ 作为预测值, 此时预测均方误差最小. 由于对流量观测值序列 X_{ij} 进行式(2) 的预处理后得 y_t, 然后根据 y_t 的 ARMA 模型, 可以得出预测值 $\hat{y_t}$(或 $\hat{y_{ij}}$), 因此需要进行预处理的逆过程才能得出流量的预测值, 即:

$$\ln X_{ij} = \mu + \alpha_i + \beta_j + \hat{y_{ij}} \tag{3}$$

从而:

$$X_{ij} = \exp\left\{\mu + \alpha_i + \beta_j + \hat{y_{ij}}\right\} \tag{4}$$

下面将考虑一步和多步流量预测. 用 ARMA 模型预测时, 有两种方法进行一步预测, 这两种方法分别是模型递推法和逆函数法, 它们的预测均方误差是一致的.

若线性最小均方误差为 η^2, l 为向后预测步数, 则 l 步预测最小均方误差满足:

$$\eta^2(l) = \sigma_a^2\left(G_0^2 + G_1^2 + \cdots + G_{l-1}^2\right),$$

其中 σ_a 是随机干扰 a_t 的标准差, $G_0, G_1, G_2, \cdots, G_j, \cdots$ 是 ARMA 模型的格林函数, 满足关系:

$$G_0 = 1, G_1 = \varphi_1 - \theta_1,$$

$$G_j = \varphi_1 G_{j-1} + \varphi_2 G_{j-2} (j \geqslant 2),$$

其中 φ_1、φ_2 和 θ_1 是 ARMA 模型的参数. 从 η 的表达式中看出, η 与预测时间 t 无关, 只与预测的步长 l 有关, 预测的时间越长(即 l 越大), 预测的最小均方误差 η 就越大, 表明对长期越不可靠或不可预测, 这与实际情况相符. 下面先计算 ARMA 模型序列一步预测, 分别采用如下两种方法计算一步预测值.

(i) 模型递推法

模型递推法预测是根据模型自身的关系式, 一步一步逆向递推得出随机干扰项 a_t, 来进行预测值的估计. 根据 ARMA(2, 1) 模型有:

$$y_t = \varphi_1 y_{t-1} - \varphi_2 y_{t-2} + a_t - \theta_1 a_{t-1} \tag{5}$$

即:

$$a_t = y_{t-1} - \varphi_1 y_{t-2} + \varphi_2 y_{t-3} + \theta_1 a_{t-2} \tag{6}$$

也即:

$$a_{t-1} = y_{t-1} - \varphi_1 y_{t-2} + \varphi_2 y_{t-3} + \theta_1 a_{t-2} \tag{7}$$

$$a_{t-2} = y_{t-2} - \varphi_1 y_{t-3} + \varphi_2 y_{t-4} + \theta_1 a_{t-3} \tag{8}$$

令

$$a_{t-3} = 0 \tag{9}$$

其中 ARMA 模型参数是已知的. 当然上述的递推过程还可以继续下去, 考虑到实际问题前后的相关性, 这里只递推到得出 a_{t-2}. 在预测时, 可按照上述转换过程相反的顺序, 往逆方向递推, 如将式(9) 代入式(8) 得:

$$a_{t-2} = y_{t-2} - \varphi_1 y_{t-3} + \varphi_2 y_{t-4} \tag{10}$$

利用式(10), 即可求出 a_{t-2}, 将 a_{t-2} 代入(7) 得 a_{t-1}, 将 a_{t-1} 代入式(6) 中得 a_t, 最后将 a_t、a_{t-1}、y_t 和 y_{t-1} 代入式(5), 即得到一步预测值 $\hat{y_t}(1)$. 易见模型递推法预测只用到过去的三个观测值 $y_{t-1}, y_{t-2}, y_{t-3}$.

(ii) 逆函数法

利用逆函数法得一步预测值关系如下:

$$\hat{y_t}(1) = \sum_{j=1}^{m} I_j y_{t+1-j} \tag{11}$$

其中 m 是自然数, $I_1, I_2, \cdots, I_j, \cdots$ 是 ARMA 模型的逆函数, 满足:

$$I_1 = \varphi_1 - \theta_1, I_2 = \varphi_2 + I_1\theta_1, I_j = I_{j-1}\theta_1 (j \geqslant 3),$$

当 ARMA 模型是平稳时, 逆函数 I_j 将趋于零(当 $j\to\infty$ 时), 因此可根据精度要求来确定 m 的值, 这一点在模型递推预测方法中是难以做到的, 该预测方法需要 m 个实际观测值才能计算. 对于一个确定的 ARMA 模型, 它的逆函数 $I_1, I_2, \cdots, I_j, \cdots$ 是确定的, 只要把预测时的观测值 $y_t, y_{t-1}, y_{t-2}, \ldots, y_{t+1-m}$ 直接代入式(11) 中, 便可得出一步预测值. 当预测步数大于或等于 2 时, 就要考虑多步预测, 即将 y_{t+l} 的条件期望值 $\hat{y_t}(l)$ 作为预测值. 由于 $y_{t+l}=\varphi_1 y_{t+l-1}-\varphi_2 y_{t+l-2}+a_{t+l}-\theta_1 a_{t+l-1}$, $E(y_{t+l})=\varphi_1 E(y_{t+l-1})-\varphi_2 E(y_{t+l-2})+E(a_{t+l})-\theta_1 E(a_{t+l-1})$. 而在 t 时刻, 当 $l>1$ 时, 正态分布的 a_t 的期望值是零, 即: $E(a_{t+l})=0$, $E(a_{t+l-1})=0$. 这样:

$$\hat{y_t}(l)=E(y_{t+l})=\varphi_1 E(y_{t+l-1})-\varphi_2 E(y_{t+l-2})=\varphi_1\hat{y_t}(l-1)-\varphi_2\hat{y_t}(l-2),$$

所以, 当 $l=2$ 时有:

$$\hat{y_t}(2)=\varphi_1\hat{y_t}(1)+\varphi_2 y_t \tag{12}$$

当 $l>2$ 时有:

$$\hat{y_t}(l)=\varphi_1\hat{y_t}(l-1)+\varphi_2\hat{y_t}(l-2) \tag{13}$$

因为在 t 时刻, y_t 是已知的, 一步预测中已经得到 $\hat{y_t}(1)$, 所以代入式(12) 可以得到 $\hat{y_t}(2)$, 再代入式(13), 从而可逐步计算出 $\hat{y_t}(3)$, $\hat{y_t}(4)$, …, 也即可得到多步预测值. 考虑实际信息流量的一个例子, 分别用两种预测方法预测 2004 年 7 月 23 日(星期一) 7:00 到 9:20 的流量, 共 24 个采样值(采样时间间隔仍然是 5 min), 这期间是上班时间, 所以流量呈现逐渐上升趋势. 把同一时刻的预测值和实际流量观测值一起进行比较, 如图 6 所示. 其中 L 表示预测步数, “R” 表示实际观测值, “M” 表示模型递推预测, “N” 表示逆函数预测.

从图 6 可见, 当预测步数较小时, 两种预测方法预测的结果略有差别, 但是随着预测步数的增大, 两种方法计算得到的预测值越来越接近, 它们的预测均方误差是一样的.

3　信息流量过载预测

现在讨论信息流量过载的预测问题. 即判断流量在未来某一时刻是否超越某一给定的阈值, 我们提出一种概率预测方法. 对预先给定一个阈值 U, 计算在未来流量超过 U 的概率. 假设预先给定上限阈值为 U, 它经过与类似的流量预处理转换后为 $T_t(l)$(与时间有关, 对于同一个阈值 U, 经过转换后, 不同 t 的阈值 $T_t(l)$ 也不同), 即:

$$T_t(l)=\ln U-\mu-\alpha_i-\beta_j \tag{14}$$

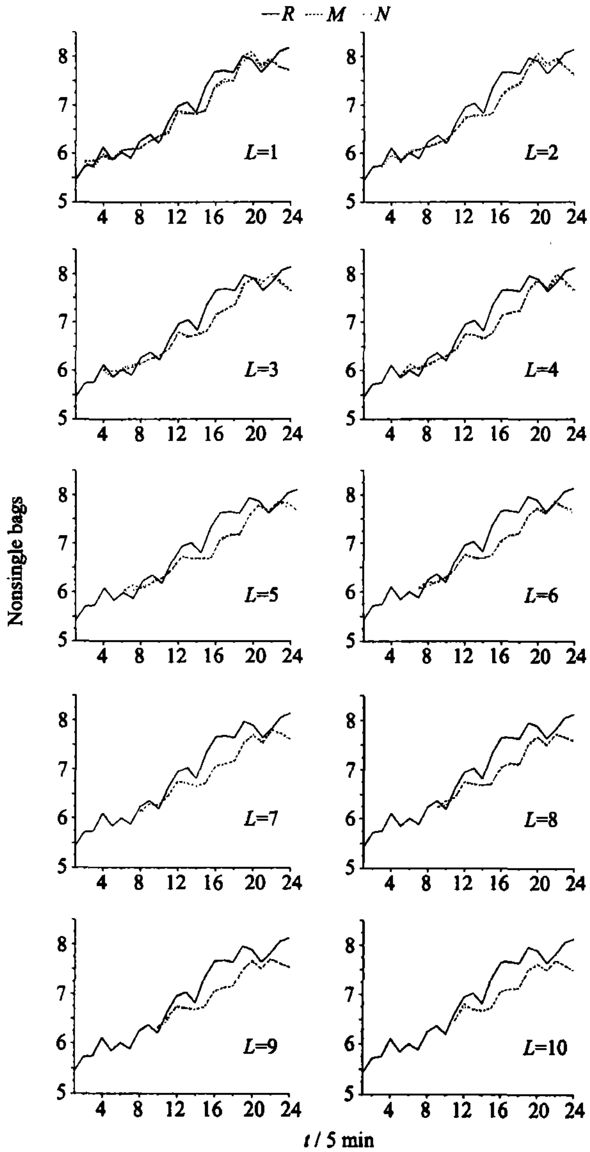

图 6　两种方法预测结果的比较
$L=i$, i 步预测, $i=1, 2, \cdots, 10$

Fig. 6　Comparison between two methods

其中 i, j 与 $t+l$ 有关, 例如, 当 $t+l=289$ 时, 它表示周二的第一个值, 则 $i=1$, $j=2$, 它们的关系可以由下式确定: (1) $i=t+l$ 被 288 除后所得余数, 且当 $t+l$ 是 288 的整数倍时, $i=\frac{t+l}{288}$; (2) $j=\left[\frac{t+l}{288}\right]+1$, 　这里 $[x]$ 表示小于 x 的最大整数.

如果在 t 时刻预测 $t+l$ 时的观测值超越阈值 $T_t(l)$ 的概率 $P_t(l)$, 则 $P_t(l)$ 可由下式表示:

$$P_t(l)=P\Big(y_{t+l}>T_t(l)\,|\,y_t, y_{t-1}, \cdots, y_{t-m}\Big),$$

其中 m 的取值因一步预测方法的不同而不同, 它表示在计算预测值时所依赖的历史观测值的个数, 与逆函数预测方法中的 m 大小相同, 而在模型递推法预测中 $m=3$. 分析 ARMA(2,1) 模型:

$$y_t = 0.98861y_{t-1} - 0.08235y_{t-2} + a_t - 0.67572a_{t-1}$$

从模型本身可以看出，在 $t-1$ 时刻只有 a_t 是随机变量，其余 3 个量 y_{t-1}、y_{t-2} 和 a_{t-1} 是已知的，在假定 a_t 服从正态分布情况下，所以 y_t 也服从正态分布，这样 $y_{t+1}, y_{t+2}, \cdots, y_{t+l}$ 也服从正态分布. 因为上述预测是观测值的条件期望预测，而且有线性最小均方误差预测的方差 $\eta^2(l)$，即 $\eta^2(l) = [y_{t+l} - \hat{y}_t(l)]^2$，因此为了计算 $P_t(l)$，可以对 $y_t(l)$ 进行正态标准化处理，记 $x_t(l)$ 为标准化后的随机变量，则 $x_t(l) = \dfrac{y_{t+l} - \hat{y}_t(l)}{\eta(l)}$，这样 $x_t(l)$ 的期望值满足：

$$E[x_t(l)] = E\left[\frac{y_{t+l} - \hat{y}_t(l)}{\eta(l)}\right] = \frac{E[y_{t+l} - \hat{y}_t(l)]}{\eta(l)} = \frac{E\left(y_{t+l}\right) - \hat{y}_t(l)}{\eta(l)} = 0,$$

$x_t(l)$ 的方差是：

$$Var[x_t(l)] = E\left[\left(\frac{y_{t+l} - \hat{y}_t(l)}{\eta(l)} - E[x_t(l)]\right)^2\right] = E\left[\left(\frac{y_{t+l} - \hat{y}_t(l)}{\eta(l)}\right)^2\right] = \frac{E\left[\left(y_{t+l} - \hat{y}_t(l)\right)^2\right]}{\eta^2(l)} = \frac{E[\eta^2(l)]}{\eta^2(l)} = 1.$$

所以 $x_t(l)$ 是服从均值为零、方差为 1 的标准正态分布，且：

$$P_t(l) = P\left(y_{t+l} > T_t(l)\right) = 1 - P\left(y_{t+l} \leqslant T_t(l)\right) = 1 - P\left(\frac{y_{t+l} - \hat{y}_t(l)}{\sigma(l)} \leqslant \frac{T_t(l) - \hat{y}_t(l)}{\sigma(l)}\right) = 1 - P\left(x_t(l) \leqslant \frac{T_t(l) - \hat{y}_t(l)}{\sigma(l)}\right) = 1 - \Phi\left(\frac{T_t(l) - \hat{y}_t(l)}{\sigma(l)}\right).$$

此处的 $\Phi(x)$ 表示标准正态分布的累积分布函数. 上式表明，当给定流量阈值时，就可以计算出流量在未来某个时刻过载的概率.

从式(14) 可以看出，当阈值 U 越大，转换后的阈值 $T_t(l)$ 也越大，在其它参数保持不变时，$\Phi(x)$ 就越大，$P_t(l)$ 就越小，也就是说预测值超越上限(阈值) 的可能性越小. 当其它参数不变，预测值越大，那么 $\Phi(x)$ 越小，相应地，$P_t(l)$ 就越大，或者说预测值超越阈值的可能性就越大，另外，$P_t(l)$ 随 $\sigma(l)$ 的增大而增大，这些与实际情况是相符合的. 考虑一种极端情况：当阈值 U 趋于正无穷大时，$T_t(l)$ 同样也趋于无穷大，$\hat{y}_t(l)$ 是预测值，显然它是有限的，那么 $\Phi(x)$ 的值趋近于 1，于是 $P_t(l) = 0$，即任何时候的预测值都不可能超越无穷大. 上面只讨论了预测值超越上限的问题，对下限超越阈值问题也可类似处理. 比如下限阈值设为 V，当然这里的网络流量不可能为负数，所以只需考虑 $V > 0$ 的情形，进行阈值转换：

$$L_t(l) = \ln V - \mu - \alpha - \beta_j \tag{15}$$

在 t 时刻预测 $t+l$ 时的流量超越阈值 $T_t(l)$ 的概率满足：

$$P_t(l) = P\left(y_{t+l} < L_t(l) \mid y_t, y_{t-1}, \cdots, y_{t-m}\right),$$

所以：

$$P_t(l) = P\left(y_{t+l} < T_t(l)\right) = P\left(y_{t+l} \leqslant T_t(l)\right) = P\left(\frac{y_{t+l} - \hat{y}_t(l)}{\sigma(l)} \leqslant \frac{L_t(l) - \hat{y}_t(l)}{\sigma(l)}\right) P\left(x_t(l) \leqslant \frac{L_t(l) - \hat{y}_t(l)}{\sigma(l)}\right) = \Phi\left(\frac{L_t(l) - \hat{y}_t(l)}{\sigma(l)}\right).$$

从上式中可以看出，当阈值 $V(V > 0)$ 减小时，经过式(15) 转换后的阈值 $L_t(l)$ 也减小，若其它参数保持不变，$\Phi(x)$ 也将会减小，$P_t(l)$ 也将减小，也就是说预测值超越阈值的概率就相应地减小；同样，当其它参数不变，预测值越大，$\Phi(x)$ 就越小，即相应的 $P_t(l)$ 就越小，或者说预测值超越阈值的可能性就越小，同时 $P_t(l)$ 随 $\sigma(l)$ 的增大而减小.

这种情况下的极端情形是，当阈值 $V(V > 0)$ 趋于零时，根据式(15)，$L_t(l)$ 同样也趋于负无穷，$\Phi\left(x\right) \to 0$，那么 $P_t\left(l\right) \to 0$.

4 结 论

在已有的对信息流量的预测中，通常是把得到的预测结果与实际流量进行对比，或者计算概率限，即预测信息流量将以一给定概率 P 在什么样的范围[a, b]，即是对给定的概率预测流量的范围. 而信息流量的过载预测问题正好与之相反，可以归结为对给定的一个流量范围[a, b]，预测信息流量在未来某一时刻超出此范围的概率 P，即对给定流量的范围[a, b]，求概率 P，这正是本文所提出的信息流量过载的预测方法. 即通过建立流量的 ARMA 模型，预测网络信息流量及其超过阈值的概率，对流量进行预处理，改善了预测效果. 当然，为了适应信息资源不断增加的特点，可以定期(每周、1 个月或 2 个月)刷新信息流量的正常行为模式及信息流量模型. 从对实际预测效果来看，自回归滑动平均模型在短期内的流量预测取得较好的效果，但是长期预测效果不很理想，这有待于将来做更深入的研究.

利用建立的信息流量模型，根据预测超过概率的大小，可以预测信息流量过载的发生及其发生的时间，对网络、系统及设备的规划、优化和控制等都有重要的

意义. 在系统维护过程中, 如果提供了这样的功能, 系统管理员可以根据实时预测的概率, 预测信息流量过载的可能性, 必要时采取预防措施或及时修复, 尽可能地在信息服务受到严重影响之前就得以恢复, 还可以确定超过阈值的时间, 这就为信息的实时监测提供了预警的功能.

参考文献:

[1] 计国君. 全面提升高校图书馆信息服务水平的创新体系[J]. 大学图书馆学报, 2003, 3: 1－7.

[2] 蔡美德. 预测与决策[M]. 北京: 科学技术文献出版社, 1992.

[3] 计国君. 信息流量预处理方法的对策研究[C]//知识化信息服务: 第 17 届全国计算机信息管理学术研讨会论文集. 北京: 科技文献出版社, 2003: 55－65.

Information Flux Forecasting Model and Its Discussion

JI Guo-jun

(School of Management, Xiamen University, Xiamen 361005, China)

Abstract: In real time measure of information flux has an important value for information statistics, information system design etc. Therefore, forecasting information flux over loading occurred and occurred time, are important meanings system and programming, optimization and control and so on. In this paper, by the known pretreatment method, the information flux forecasting model is constructed. Based on the least linear error technique of the ARMA model, information flux is forecasted and the exceeded threshold value in future is presented. The approaches improve the forecasting effect, and the conclusions show the better characteristic in overall tendency based on analysis and evaluation.

Key words: information flux; ARMA; pretreatment