

基于混沌特征的网络延迟预测模型

李 超,赵 海,葛 新,张 君

(东北大学信息科学与工程学院,辽宁沈阳 110004)

摘 要: 结合非线性时间序列分析方法与 CAIDA 组织授权的真实 Internet 海量样本数据,计算了网络延迟演化序列的混沌特征量,结果表明演化序列具有混沌特征.在此基础上,对混沌系统中典型的 Logistic 模型加以改进,提出了一种基于 Logistic 模型的以正弦函数作为指数衰减因子的模型,以描述网络延迟的演化态势.使用微粒群算法根据实际数据,分别从算法收敛性、模型的拟合准确度及预测准确度等方面对备选模型参数选优.实验结果表明最终优选模型在结构选择上比较合理,能够准确反映网络延迟的变化情况.

关键词: Internet 测量;网络延迟;Logistic 模型;微粒群算法

中图分类号: TP393.6 **文献标识码:** A **文章编号:** 0372-2112 (2009) 12-2657-05

Network Delay Prediction Model Based on Chaotic Characteristic

LI Chao, ZHAO Hai, GE Xin, ZHANG Jun

(School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China)

Abstract: Together with the nonlinear time series analysis and the giant data samples authorized by CAIDA, the chaotic character of network delay time series indicates that its evolvement process existed chaotic identity. Then, we proposed a revised Logistic model with sine and cosine functions to describe the evolvement state of network delay. Moreover, particle swarm optimization (PSO) algorithm is adopted for the parameters estimation of the revised model, which is evaluated from the perspective of convergence, fitting accuracy and forecast accuracy. The result reviews that the structure of the optimized model is reasonable, and reflects the movement of network delay accurately.

Key words: Internet measurement; network delay; Logistic model; PSO

1 引言

继对 Internet 的性能指标,如流量、延迟、跳数等进行测量并对其做分析之后^[1],目前已有不少相关工作围绕网络特征量的行为预测进行展开.文献[2]根据实际采集的非单播包数的观测值序列,建立了网络流量 AR-MA 模型.文献[3]在小范围内测量到往返延迟,将 AR 模型与自适应线性神经网络相结合,通过建立模型对网络延迟进行预测估计,以保证网络的正常服务.然而,Internet 作为一个开放的复杂巨系统^[4],其子系统数量庞大,网络中各种局部交互作用错综复杂,如果仅从局部范围对网络性能进行测量,往往难以衡量网络整体的发展趋势.因此,从宏观角度对 Internet 表现出的整体行为进行预测分析就显得十分必要.

本文以 CAIDA^[5]授权的海量数据为分析样本,针对网络延迟在演化涨落过程中具有的混沌特征,引入了混沌系统中典型的 Logistic 方程加以改进,使其符合网络延

迟的生长态势.在此基础上,使用微粒群算法对模型的参数取优,实验结果表明选优模型可以准确描述网络延迟演化序列的生长态势,并在短期内对其做出准确预测.

2 网络演化样本获取

2.1 基于演化的网络延迟定义

为提取网络延迟的演化规律,需要以长时间跨度的网络样本数据作为演化分析基础.由于网络规模过于庞大,考虑到每天探测所得的网络样本数据集大小接近 1G,如果以网络探测周期为单位分析网络延迟的演化行为,其计算负担将是不可想象的^[6].相对的,以月为单位对网络延迟演化行为进行考察能够从宏观角度满足网络整体的演变速度的分析需求,故此本文以月为单位统计网络样本数据.为了表征网络整体传播行为,真实地反映网络延迟演化规律,本文从统计角度出发,对基于表现网络整体客观属性的延迟行为特征,给出如下定义.

在 Internet 中,针对某条路径,如果将发送探测数据包的时间与收到目的端返回响应的时间之差定义为该条路径的延迟,那么,大量数据包在 Internet 中从任一源 IP 地址到任一目的 IP 地址所经过延迟的统计均值,称为网络延迟. 设一条路径的延迟为 $T_d(k)$, 样本总量为 n , 则网络延迟为

$$T = \frac{1}{n} \sum_{d=1}^n T_d(k) \quad (1)$$

尽管本文以统计均值的概念来衡量网络的整体有效性可能会带来误差,但是从分析的角度以及大规模采样获取的样本数据规模量等多方面综合考虑,基于演化的网络延迟不失为衡量网络整体发展趋势的有效手段之一. 对于刻画网络整体的发展趋势,并对未来网络性能趋势做出整体的预测,网络延迟是有效的特征参量. 因此对网络延迟的演化行为进行研究是具有实际意义的.

2.2 网络访问时间演化序列

本文从时间维度上选取 riesling 监测节点于 2003 年 4 月到 2007 年 11 月共计 56 个月对 Internet 进行探测的有效样本,其样本总数共计近 1980 万条. 根据网络延迟的定义,得到网络延迟的演化特征如图 1 所示.

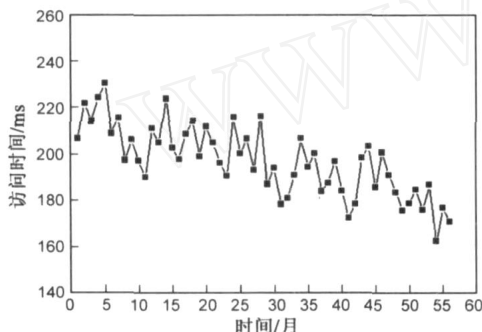


图1 网络延迟随时间(2003年4月到2007年11月共56个月)变化的情况

针对海量样本数据表现的网络延迟发展趋势,图 1 在宏观上(全球范围内)描述了网络延迟在时间维度下的变化情况. 从图中可以看到,网络延迟在整体上表现出缓慢下降的趋势,并且在演化过程中带有准周期性震荡的衰减. 尽管无法准确确定其衰减周期,但是为了描述震荡变化规律,定义网络延迟的变化率如下:

$$r(i+1) = \frac{T(t_{i+1}) - T(t_i)}{T(t_i)} \quad (2)$$

其中 $T(t_i)$ 为第 i 个月($1 \leq i \leq 55$)的网络延迟,对 r 的变化情况进行计算,其变化率如图 2 所示.

图 2 呈现的网络延迟变化率的震荡特性,说明其演化过程具有较强的随机性. 由于路径上的延迟在测量过程中不可避免地受到其他随机噪声的干扰,因此在其影响下,网络延迟表现出局部不稳定的随机特性,使

得对其较难预测. 根据非线性动力学,一些貌似随机的过程是非线性确定系统内在随机性的表现,系统复杂的外在行为可由内在简单的混沌规律加以解释^[7]. 由于 Internet 作为一个开放的复杂巨系统,在各种错综复杂的局部交互作用,某一特征量的演化行为包含了系统其他参变量的演化信息,利用有限数据重构动力系统将是解决问题的关键. 因此,从混沌动力学的角度研究网络延迟的演化行为不失为一种可行的方法.

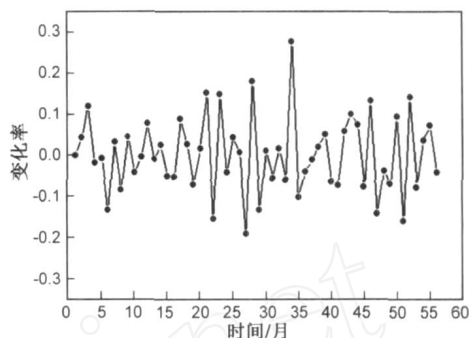


图2 网络延迟的变化率情况

文献[8]利用相空间重构技术及 G-P 算法,对时间维度的网络延迟演化序列进行相空间重构,得到其混沌吸引子饱和和关联维数为 2.8308,证实了网络延迟在演化过程中具有混沌特征. 针对网络延迟的混沌特性,考虑到混沌系统耗散性和初值敏感的特点,常用的全局建模方法如神经网络、多项式等无法给出简明直观的模型表达形式^[9],因此可以尝试采用混沌系统中的模型对网络延迟进行描述,试图得到网络延迟关于时间 t 的演化模型.

3 网络延迟预测模型

3.1 模型建立

系统模型建立的两个基本问题是模型结构的选择以及模型参数的估计. 目前,已经发现网络延迟在演化过程中具有混沌特征,而网络延迟的演化行为与混沌系统中典型的 Logistic 方程描述的虫口增长方式有着相似之处. 从图 2 看到,网络延迟演化序列总体上呈现出带有周期性的衰减振荡. 由于网络基础设施建设不断得到完善,网络延迟从总体演化趋势上必然呈现衰减趋势,但是随着网络业务的不断扩展,网络延迟在演化过程中同时存在着准周期性的振荡衰减特性. Logistic 生长曲线预测模型则用于模拟生物的成长过程,一般经历发生、发展、成熟等几个阶段. 发生期成长速度较慢,并逐渐由慢到快;发展期成长速度较快,此种快速增长将维持相当长的一段时间,随后进入成熟期;在成熟期,其增长速度将逐渐变慢,最后增长速度趋于零. 由于数据包在网络传播过程中受到各种软硬件条件的制约,网络延迟存在下限,不可能无限制衰减. 因此有

必要对原始 Logistic 生长曲线做适当变形,以适合其描述网络延迟的演化态势^[10]. Logistic 人口发展模型的直接形式是非线性微分方程

$$\frac{dT}{dt} = rT \left[1 - \frac{T}{K} \right] \quad (3)$$

其中, T 为 t 时刻的网络延迟, t 是以月为周期单位的时间, r 为网络延迟的变化率. 对上式作积分, 得到

$$T = \frac{k}{1 + me^{-rt}} \quad (4)$$

观察该式可以看到, 在 $r > 0$ 的条件下, T 随 t 的增长呈单调递增变化. 图 1 表明网络延迟随着时间的推移呈缓慢衰减, 因此对 Logistic 方式作如下变换

$$T = d - \frac{k}{1 + me^{-rt}} \quad (5)$$

其次, 考虑式中 r 为常数, 无法刻画演化过程中的震荡行为, 因此有必要加入震荡衰减因子, 使模型更符合网络延迟的变化趋势. 本文选取带有正余弦的指数线性组合来反映演化过程中的震荡特征, 得到 Logistic 方程的改进模型如下:

$$T = d - \frac{k}{1 + me^{-(p+r_1 e^{\sin(\frac{u_1}{m_1} + v_1)t} + r_2 e^{\cos(\frac{u_2}{m_2} + v_2)t})t}} \quad (6)$$

式中 T 为模型输出值, r_1, r_2 为震荡幅值, u_1, u_2 为震荡半周期, v_1, v_2 为初始幅角. 考虑到指数函数在不能反映实际震荡趋势的情况下, 为保证系统模型按照演化趋势进行变化, 有必要引入校正系数 p .

为描述方便, 将方程表示为 $T = f(d, k, m, p, r_1, u_1, v_1, r_2, u_2, v_2, t)$. 选择该模型的意义在于保证模型简单性原则的基础上尽可能地反映出网络延迟演化曲线中的震荡衰减特性. 模型中的参数由实际数据确定, 并需要通过实验结果检验模型的有效性.

3.2 参数确定算法

在模型结构初步建立之后, 进一步的问题就是模型参数的选择. 一般来说, 参数的确定对预测结果往往有较大的影响. 本文采用微粒群算法 (PSO) 对改进的 Logistic 模型参数进行优化选择. 相比其他的优化算法如遗传算法 (GA), 微粒群算法具有运算量小, 可调参数较少的优点, 且更适合对实数进行操作.

本文首先选取前 48 个月的网络延迟样本进行学习以得到模型参数, 并通过对模型输出的预测值与后 8 个月的数据进行比较, 来评价模型预测准确度. 基于 PSO 的改进 Logistic 模型参数优化算法描述如下:

输入: 网络延迟演化序列 (从 2003 年 4 月到 2007 年 3 月共 48 个月)

输出: 最优适应度个体 $T = f(d, k, m, p, r_1, u_1, v_1, r_2, u_2, v_2, t)$

(1) 设置微粒群算法的初始参数, 如: 群体规模、加

速度、最大迭代次数等.

(2) 确定模型中参数的取值范围.

(3) 定义适应度函数, 在文本中, 为保证预测模型的网络延迟 $T^*(i)$ 与实际值 $T(i)$ 拟合程度最好, 定义适应度函数:

$$S(x) = \frac{1}{n} \sum_{i=1}^n |T(i) - T^*(i)| \quad (7)$$

n 为按月统计的时间跨度, 适应度函数越小, 模型输出值越接近真实数据.

(4) 在限定的参数范围内随机生成初始群体, 根据适应度函数计算群体中每个微粒的适应值, 用粒子的适应值衡量粒子的优劣, 并与其之前所经历过最好位置的适应值进行比较, 若小于所经历过的最好位置的适应值, 则将其位置作为局部最好位置. 对每个微粒将其适应值与所有群体所经历过的最好位置的适应值进行比较, 若小于所有群体经历过的最好位置的适应值, 则将其位置作为全局最好位置.

(5) 判断粒子适应值是否超过预设值或是否到达最大进化代数, 若不满足则转到步骤 4, 否则结束输出结果.

4 实验结果及讨论

4.1 选择模型

本文按照所描述的算法根据实际数据对改进的 Logistic 模型参数作优化. 针对网络延迟演化序列的数据规模, 在微粒群初始参数设置中, 设群体规模 $S = 50$, 加速度 $c_1 = c_2 = 2$, 最大迭代次数 1000 次. 对于模型参数取值范围, 从图 1、图 2 可以看到, 网络延迟分布主要集中在 $[50\text{ms}, 250\text{ms}]$ 范围内, 网络延迟主要集中在 $[140\text{ms}, 260\text{ms}]$ 区间. 另外, 网络延迟的变化率幅值在 $[0, 0.35]$ 之间, 说明其变动范围不大. 考虑到误差因素的影响, 不妨将参数选择的范围适当扩大. 综合以上分析因素, 各参数的取值范围如表 1. 为使结果分析不失一般性, 本文选取三组 PSO 输出的优化值如表 1, 收敛过程见图 3.

表 1 参数取值范围和 PSO 优化结果

待优化参数	参数取值范围	优化值		
		组 1	组 2	组 3
k	$[160, 300]$	235.2113	228.0559	250.2074
m	$[0, 120]$	61.4113	52.4800	76.4446
p	$[0, 100]$	65.3886	84.6781	91.4229
r_1	$[0, 5]$	3.3778	2.9810	3.9865
u_1	$[0, 1]$	0.2494	0.4967	1.0000
v_1	$[0, 2]$	1.2265	1.0720	1.0630
r_2	$[0, 0.4]$	0.2642	0.2860	0.3519
u_2	$[0, 1]$	0.5649	0.1280	0.5557
v_2	$[0, 2]$	1.2476	1.8811	0.2293
分数		7.2340	7.1436	7.0102

算法运行初期, PSO 可能会产生大量无意义的解.

但随着种群的演化, 它们的比例会降低. 从图 3 可以看到, 三组实验的进化代数超过 700 之后, 群体将较难产生更优的个体, 说明算法收敛性好, 参数范围选择较合理.

由于系统本身的复杂性、数据采样过程中存在的随机噪声以及算法本身的有偏性等影响, PSO 输出的解

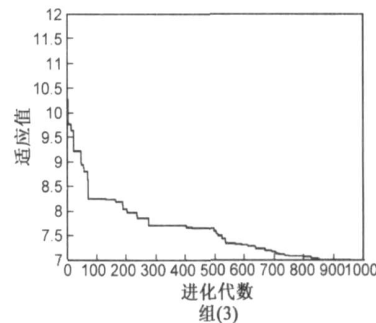
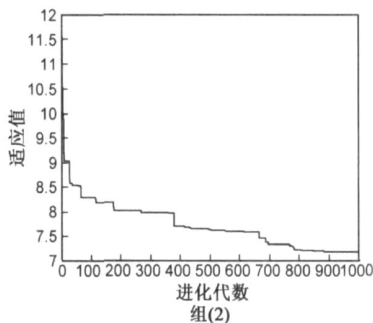
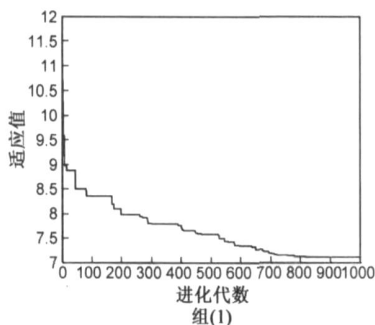


图3 PSO进化收敛过程

$T = 250.2074 -$

$$\frac{76.4446}{1 + 91.4229e^{-\left(\frac{1}{3.9865 + 0.0343} \times e^{\sin\left(\frac{t+1.0630}{t}\right)} + 0.3519 \times e^{\cos\left(\frac{0.5557 + 0.2295}{t}\right)}\right) // t} \quad (8)$$

模型是否能够准确刻画网络延迟的演化行为, 需要通过实验数据做进一步验证.

4.2 结果分析

根据客观实践经验, 如果一个模型对数据的拟合程度在 95% 以上, 那么该模型能较真实的反映数据的变化情况. 以模型计算得到的拟合值及预测值与真实值进行对比如图 4 所示:

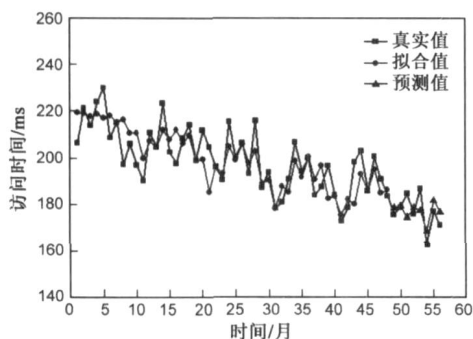


图4 网络延迟真实值和模型预测值比较

从图 4 可以看到, 原始数据所表现的变化态势以及模型表达式得到的结果均表明网络延迟演化序列在总体上呈现带有局部震荡的缓慢衰减. 由于微粒群体在初始阶段需要学习过程, 因此直观上看 20 个月之前的预测值在时间维度上与真实值有较大差距, 但在第 28 个月之后模型的拟合值基本上与真实值重合. 根据定义的适应值函数, 模型拟合值与真实值(从 2003 年 4 月到 2007 年 3 月共 48 个月)的绝对平均误差 7.0102. 为更完备的评价模型, 引入相对平均误差作为模型优劣

未必是最优解. 事实上, 从完备解空间中找出最优解是不可能的, 也是不需要的. 如果从非完备空间中找到的一个次优解能够满足某种客观判定准则, 那么这个解就是可接受的. 根据计算得到的以上三组实验的适应值分数及进化收敛过程, 选取第 3 组优化值作为网络延迟演化模型的参数:

的评判标准.

$$= \frac{\frac{1}{n} \sum_{i=1}^n |T(i) - T^*(i)|}{\frac{1}{n} \sum_{i=1}^n T^*(i)} = \frac{7.0102}{199.3262} = 0.0352 \quad (9)$$

因此模型的拟合准确度为 $1 - 0.0352 = 96.48\%$

更进一步, 计算模型的预测值与真实值(从 2007 年 4 月到 2007 年 11 月)的相对平均误差

$$= \frac{\frac{1}{n} \sum_{i=1}^n |T(i) - T^*(i)|}{\frac{1}{n} \sum_{i=1}^n T^*(i)} = \frac{5.3073}{175.1584} = 0.0303 \quad (10)$$

模型的预测准确度为 $1 - 0.0303 = 96.97\%$. 模型的拟合准确度和预测准确度均能达到评价准则要求, 故模型可以接受.

以上分析可以看到, 模型预测结果比较接近于真实值, 说明模型在短期内可以对网络延迟进行较好的预测. 但是随着时间的增长, 网络延迟的混沌特性使其对初始值比较敏感, 伴随预测过程中产生累计误差的增大, 势必对长期预测结果有影响, 预测准确度将迅速降低. 由于采样的样本数据量较小, 因此增加网络样本量, 采用非线性混沌预测方法对网络延迟的演化进行分析是本文未来进一步的工作.

5 结论

本文基于 CAIDA 授权的真实 Internet 海量样本数据, 从宏观角度考察了网络延迟的整体发展趋势, 在对网络延迟的演化行为进行分析的过程中, 得到以下结论:

在各种错综复杂局部交互作用的影响下,网络延迟的演化行为表现出带有随机干扰的震荡特征.在采用非线性时间序列分析方法对网络延迟演化序列进行分析之后,证实网络延迟演化序列具有混沌特性.在此基础上,引入了混沌动力学中经典的 Logistic 方程,并对其加以改进,使其符网络延迟的涨落演化态势.通过 PSO 算法对模型参数进行选择之后,得到的优化模型能够较好的反映网络延迟的变化情况,并能够对其演化行为做出准确预测.

参考文献:

- [1] 苏威积,赵海.基于 hops 的 Internet 复杂网络分割度分析[J].通信学报,2005,26(9):1-8
SU Wei Ji,ZHAO Hai,XU Ye. Internet complex network separation degree analysis based on hops[J].Journal on Communications,2005,26(9):1-8. (in Chinese)
- [2] 邹柏贤,刘强.基于 ARMA 模型的网络流量预测[J].计算机研究与发展,2002,39(12):1645-1652
Zou Bo Xian,Liu Qiang. ARMA-based traffic prediction and overload detection of network[J].Journal of Computer Research and Development,2002,39(12):1645-1652. (in Chinese)
- [3] LI Hong Yan,WANG Hong,GUI Chao. Internet time - delay prediction based on autoregressive and neural network model [A]. Proceedings of International Conference on Communications,Circuits and Systems [C]. Guilin. Guangxi:IEEE,2006. 1758-1761.
- [4] 戴汝为,操龙兵. Internet 一个开放的复杂巨系统[J].中国科学,2003,33(4):289-296.
Dai Ru Wei,Cao Long Bin. Internet-an open complex giant systems[J]. Science in China,2003,33(4):289-296. (in Chinese)
- [5] CAIDA skitter Project[EB/OL]. http://www.caida.org.
- [6] 李超,赵海,张昕,袁绍谦. Internet 的访问时间分析[J].通信学报,2007,28(7):54-60.
Li Chao,Zhao Hai,Zhang Xin,Yuan Shao Qian. Analysis on traveling time of Internet [J]. Journal on Communications,2007,28(7):54-60. (in Chinese)
- [7] 吕金虎,陆君安.混沌时间序列分析及其应用[M].武汉:武汉大学出版社,2002.
- [8] 李超,赵海,张昕,袁绍谦. Internet 访问时间的混沌性解析[J].计算机科学,2008,35(5):85-86.
Li Chao,Zhao Hai,Zhang Xin,Yuan Shao Qian. Chaotic analysis of Internet traveling time [J]. Computer Science,2008,35(5):85-86. (in Chinese)
- [9] 张伟,吴智铭,杨根科.混沌时间序列的遗传演化建模[J].电子学报,2005,33(4):748-751.
Zhang Wei,Wu Zhi Ming,Yang Gen Ke. Genetic programming modeling on chaotic time series [J]. Acta Electronica Sinica,2005,33(4):748-751. (in Chinese)
- [10] 吴淑玲.利用 Logistic 模型预测我国数字图书馆的发展趋势[J].情报方法,2004,23(4):56-57.
WU Su Ling. Forecast of development of China Numerical Library by Logistic model [J]. Journal of Information,2004,23(4):56-57. (in Chinese)

作者简介:



李超男,湖北洪湖人,东北大学信息科学与工程学院博士,主要研究方向为计算机网络、复杂网络。

E-mail:lich.mail@gmail.com

赵海男,辽宁沈阳人,东北大学信息科学与工程学院教授、博士生导师,主要研究方向为计算机网络、复杂网络、数据融合。