

一种基于粒子群的实际业务流预测算法

李文泽¹, 盛光磊²

(1 河南广播电视大学, 河南 郑州 450008; 2 郑州升达经贸管理学院, 河南 郑州 451191)

摘 要: 为了解决因业务流迅速增长而造成的网络拥塞现象, 提出了一种新的业务流预测算法 PSATP (Particle Swarm-ARIMA model based Traffic Prediction). 该算法首先利用 ARIMA 模型和 Kalman 滤波建立预测方法, 并且结合粒子群进行优化, 使预测精度得以提高. 同时, 以实际数据进行仿真实验, 深入研究了时延、丢包率与缓冲区、利用率之间的关系. 相比于其他算法, 仿真结果表明 PSATP 算法具有较好性能.

关键词: 拥塞; 预测; 精度; 粒子群; ARIMA

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2014)01-0056-04

An Actual Traffic Prediction Algorithm Based on Particle Swarm

LI Wen-ze¹, SHENG Guang-lei²

(1 Henan Radio & Television University, Zhengzhou 450008, China;

2 Shengda Trade Economics & Management College of Zhengzhou, Zhengzhou 451191, China)

Abstract: In order to mitigate network congestion caused by the rapid growth of traffic, a novel traffic prediction algorithm PSATP (Particle Swarm-ARIMA model based Traffic Prediction) is proposed. In this algorithm, the prediction method is presented by ARIMA model and Kalman filter at first, and the prediction accuracy is improved with particle swarm optimization. Then, a simulation with actual data was conducted to study the relationship between delay, dropping rate buffer and utilization rate. The simulation results show that, compared to other algorithm, PSATP has better performance.

Key words: congestion; prediction; accuracy; particle swarm; ARIMA

1 引言

随着通信网络的飞速发展, 业务流拥塞现象^[1]日趋严重. 在有限的网络资源下准确地预测下一时刻业务流状态, 成为目前通信网络研究的热点. 国内外学者对此进行大量研究工作. 温祥西等^[2]针对业务流的时变性和长周期性, 基于最小二乘支持向量机提出了一种最优样本子集在线模糊预测算法. 胡治国等^[3]为了解决传统预测方法受干扰大、辨识过程繁杂等问题, 基于随机网络模型提出了多步网络时延预测方法, 使其能够反映网络的动态负载状态. 文献[4]中基于神经网络模型提出了软件可靠性的预测方法. Farsi H 等^[5]采用神经网络方法研究了非参数、非线性网络的预测问题. 单伟等^[6]基于自相

关系数和偏自相关系数的拖尾方法, 利用具有超线性收敛性的变尺度法对 ARMA 模型进行改进, 以此克服了样本数据的长度有限时的估计误差, 并进行业务流预测和异常检测.

在上述工作基础上, 本文结合 ARIMA(p, d, q)模型和粒子群(Particle Swarm, PS)算法提出了一种新的业务流预测算法. 该算法利用 Kalman 滤波和粒子群技术对下一时刻的业务流进行估计, 并且通过仿真实验对比了与其它算法之间的性能.

2 业务流预测方法

2.1 预测算法介绍

在如图 1 所示的网络 $G(V, E)$ 中, V 代表节点集合, E 代表链路集合. 令某节点到达业务流服从

收稿日期: 2013-03-27; 修回日期: 2013-05-12

基金项目: 国家自然科学基金项目(10901144)

ARIMA(p, d, q)模型,并在时间间隔 $t_1 < t_2 < \dots < t_n$ 上存在 n 个观测值.结合 ARIMA(p, d, q)模型,本文首先提出如下预测算法(ARIMA model based Traffic Prediction, ATP):

Step 1 在开始时刻 $t = 0$, 初始化网络和 ARIMA(p, d, q)模型参数,确定当前业务量 X_t :

Step 2 由 Kalman 滤波计算 $X(t)$ 的估计值 $\hat{X}_{t|t}$:

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + \beta_t(z_t - \eta_t \hat{X}_{t|t-1}) \quad (1)$$

$$\beta_t = \Gamma_{t|t-1} \eta_t' (\eta_t \Gamma_{t|t-1} \eta_t' + \sigma_N^2)^{-1} \quad (2)$$

式中, $\Gamma_{t|t-1}$ 为误差协方差矩阵;

Step 3 根据式(3)获得真实值和估计值之间的误差:

$$\Gamma_{t|t} = (1 - \beta_t \eta_t) \Gamma_{t|t-1} = \Gamma_{t|t-1} - \Gamma_{t|t-1} \eta_t (\eta_t \Gamma_{t|t-1} \eta_t' + \sigma_N^2)^{-1} \eta_t \Gamma_{t|t-1} \quad (3)$$

Step 4 由 $\Gamma_{t|t-1}$ 和 $\hat{X}_{t|t}$, 通过式(4)和式(5)对下一时刻业务流 X_{t+1} 进行预测:

$$X_{t+1} = \gamma^A \hat{X}_{t|t} \quad (4)$$

$$\Gamma_{t+1|t} = \gamma^A \Gamma_{t|t} (\gamma^A)' \quad (5)$$

Step 5 令 $t = t + 1$, 跳转到步骤(1), 重新进行预测下一时刻业务流, 直至运行完成;

Step 6 算法结束.

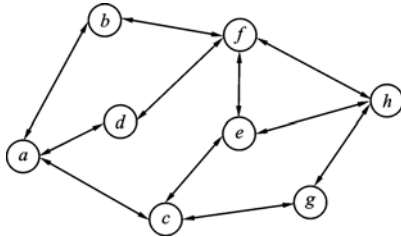


图1 网络拓扑结构图

2.2 算法改进

实际业务流具有各种复杂特性, 仅仅采用 ARIMA(p, d, q)模型进行预测可能导致精度不高, 并且 ARIMA(p, d, q)模型的计算复杂度太大, 所以考虑结合粒子群(Particle Swarm, PS)算法^[7]进行联合预测, 以提高预测精度. 其改进算法(Particle Swarm and ARIMA model-based Traffic Prediction, PSATP)描述如下:

Step 1 将某时刻 t 将到达业务流视为粒子, 并确定粒子群规模 n 、搜索空间维数 d 、初始位置 $x_i(t)$ 和粒子到达速度 $v_i(t)$;

Step 2 如果粒子 i 不满足约束条件, 那么根据式(6)获得更新值, 以此替换 $x_i(t)$;

$$x_i(t+1) = x_i(t) + v_i(t) \times (\tau \times \text{randn}() + 1) \quad (6)$$

式中, τ 为启发式因子, $\text{randn}()$ 为 0 到 1 之间的正态分布数;

Step 3 按照式(7)所示的适应度函数计算粒子 i 的适应值, 并将粒子 i 的最佳位置 z_i 确定为当前位置, 同时令 z 为种群的最佳位置:

$$f = (n \log(\frac{1}{n} \sum_{i=1}^n (z_i(t) - x_i(t)))) + \epsilon \quad (7)$$

式中, ϵ 为一正常数;

Step 4 根据式(8)和式(9)更新粒子的到达速度和位置:

$$v_i = v_{i-1} + \alpha(z_i - x_i) + \beta(z - x_i) \quad (8)$$

$$x_{i+1} = x_i + v_i \Delta t \quad (9)$$

其中, $\alpha + \beta = 1$;

Step 5 如果粒子 i 不满足条件则以 $x_i(t)$ 作为初始点, 并根据式(7)得到的结果替换 $x_i(t)$, 同时计算其适应值; 如果其适应值优于 z_i 的适应值, 则令 z_i 为粒子 i 的适应值, 否则令 z 为粒子 i 的适应值;

Step 6 输出当前最佳位置 z_i 对应的 $x_i(t)$, 即为粒子群预测结果 y_1 ;

Step 7 同时根据 ATP 算法利用 ARIMA(p, d, q)模型进行预测, 获得其结果 y_2 ;

Step 8 实际业务流的预测结果 y 可表示为

$$y = \xi y_1 + \chi y_2 \quad (10)$$

其中, ξ 和 χ 为权重系数, 并且 $\xi + \chi = 1$;

Step 9 令 $t = t + 1$, 并跳转到 Step 3 重复执行, 直至运行结束;

Step 10 算法结束.

3 数学仿真

本文首先在 OPENT 中建立如图 1 所示的网络拓扑结构图. 其参数设为: 各链路带宽为 10 Mb/s, 延时 15 ms, 各节点缓存区大小为 500 kb, 数据包大小为 500 B. 为了体现 PSATP 算法的优势, 将 PSATP 算法与文献[8]中提出的 FARIMA 算法, 以及文献[9]中提出的 PSR-LSSVM 算法作对比分析. 这里将 a 作为源节点, h 作为目的节点, 进行了 2 000 s 的仿真实验. 取节点 h 前 1 000 s 到达的业务流作为先验信息, 分别利用三种算法来预测前 1 000 s 业务流, 并将预测结果和后 1 000 s 的实际业务流进行比较, 如图 2 所示. 从图 2 来看, 本文提出的 PSATP 算法与实际业务流较为接近, 其次是 PSR-LSSVM 算法, 性能最差的是 FARIMA 算法. 通过数据分析, PSATP 算法、SR-LSSVM 算法和 FARIMA 算法的残差分别是 18.36%、27.73% 和

29.29%.

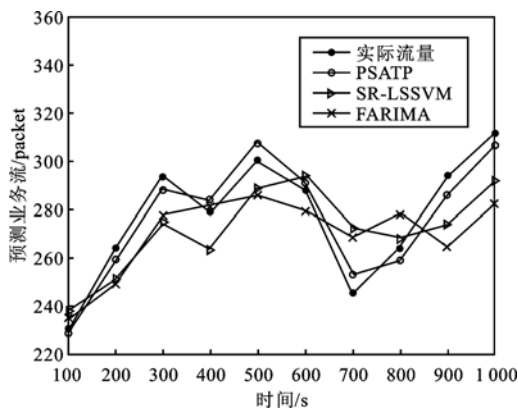


图2 预测业务流比较

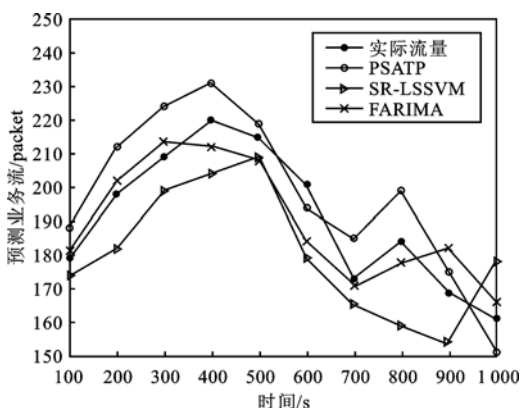


图3 分形业务流预测结果比较

由于实际业务流具有复杂特性,其中分形特性是比较有代表性的,所以这里利用 FGN 模型在节点 a 产生分形业务流,然后分别利用三种算法预测节点 h 到达的业务流,结果如图 3 所示.从图 3 可以看出,本文提出的 PSATP 算法和 FARIMA 算法对于分形业务流具有较好的预测能力,而 SR-LSSVM 算法的预测精度则较低. FARIMA 算法自身对于刻画长、短相关特性有着明显优势,所以在预测分形业务流方面有着较高精度.

同时,这里将三种算法预测结果的时延与缓冲区、利用率(利用率定义为业务流达到率与链路带宽之比)进行比较.随着缓冲区的增加时延是随之增加的,但是 SR-LSSVM 算法增加的幅度较大,而 PSATP 算法和 FARIMA 算法对于时延的控制较好.实验结果表明了时延随着利用率变化的情况.随着利用率的增加,时延的整体趋势是下降的,并且 PSATP 算法和 FARIMA 算法的时延低于 SR-LSSVM 算法.

将预测结果的丢包率与缓冲区、利用率进行比较,如图 4 和图 5 所示.从图 4 可以看出,丢包率随着缓冲区的增加开始呈现出减少趋势,随后呈现出增加

状态,PSATP 算法性能较 FARIMA 算法和 SR-LSSVM 算法更优.由于缓冲区增加在起始阶段可以缓解拥塞的发生,从而降低了丢包数量,但是一旦到达的业务流增多,缓冲区增大并不能缓解拥塞状况,反而会使被堵塞的业务流增加,从而造成丢包率增加.而从图 5 来看,丢包率随着利用率的增加而减少.

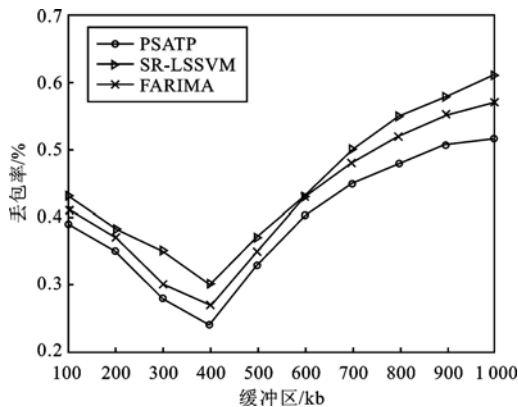


图4 丢包率与缓冲区之间的关系

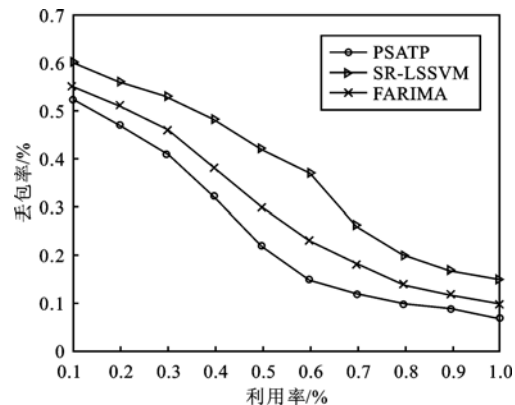


图5 丢包率与利用率之间的关系

4 结束语

本文结合 ARIMA(p, d, q)模型和粒子群算法提出了一种新的预测算法 PSATP.该算法首先基于 ARIMA(p, d, q)模型建立了初步的预测方法 ATP,根据 ATP 方法的预测结果,利用粒子群算法对其进行优化,使其预测精度得到进一步提高.同时,通过仿真实验对比分析了与 SR-LSSVM 算法、FARIMA 算法之间的性能差异,并且验证了影响 PSATP 算法的关键因素(如丢包率、利用率、缓冲区大小和时延等).在今后研究中,可以考虑结合小波变换、神经网络和灰色模型等建立一套完整的网络业务流预测模型.

参考文献:

- [1] Zheng G, Choo L, Wong K. Optimal cooperative jam-

- ming to enhance physical layer security using relays [J]. IEEE Transactions on Signal Processing, 2011, 59(3):1317-1322.
- [2] 温祥西, 孟相如, 马志强, 等. 小时间尺度网络流量混沌性分析及趋势预测[J]. 电子学报, 2012, 40(8): 1609-1616.
- [3] 胡治国, 张大陆, 侯翠平, 等. 基于随机网络的多步网络时延预测模型[J]. 计算机科学, 2009, 36(7): 85-87.
- [4] Li S, Yin Q, Guo P. A hierarchical mixture of software reliability model for prediction[J]. Applied Mathematics and Computation, 2007, 185(2):1120-1130.
- [5] Farsi H, Gopal F. Artificial neural network simulator for supercapacitor performance prediction[J]. Computational Materials Science, 2007, 39(3): 678-683.
- [6] 单伟, 何群. 基于非线性时间序列的预测模型检验与优化的研究[J]. 电子学报, 2008, 36(12):2485-2489.
- [7] 程祥, 张忠宝, 苏森, 等. 基于粒子群优化的虚拟网络映射算法[J]. 电子学报, 2011, 39(10): 2240-2244.
- [8] 潘乔, 罗辛, 王高丽, 等. 基于 FARIMA 模型的流量抽样测量方法[J]. 计算机工程, 2010, 36(15): 7-8.
- [9] 陈卫民, 陈志刚. 基于 PSR-LSSVM 的网络流量预测[J]. 计算机科学, 2012, 39(7): 92-94.

作者简介:

李文泽 男, (1979—), 硕士, 讲师. 研究方向为网络优化.

盛光磊 男, (1981—), 硕士, 讲师. 研究方向为网络安全.

(上接第 55 页)

- [5] 朱浩. 云环境下 MapReduce 容错技术的研究[D]. 上海: 上海交通大学, 2012.
- [6] 王晋川, 何宏, 张福临. MapReduce 框架与调度容错机制研究[J]. 中国储运, 2010(12): 90-91.
- [7] Cheng D. GridMix introduction [EB/OL]. <http://dongxicheng.org/mapreduce/hadoop-gridmix-benchmark/>.

作者简介:

张治斌 男, (1953—), 教授. 研究方向为云计算、分布式计算等.

李燕歌 女, (1987—), 硕士研究生. 研究方向为云计算、MapReduce 并行编程模型.