

基于 AR 模型的主机负载预测

黄常青¹, 徐海水², 钟晖云¹, 廖志坚¹, 张红¹

(1. 广东工业大学 计算机学院, 广东 广州 510075;

2. 广东工业大学 网络信息与现代教育技术中心, 广东 广州 510075)

摘要:线性时间序列作为一种概率统计方法, 已经被运用到各个领域。AR 模型是最常见的一种时间序列模型, 是根据时间序列的自相似性质, 利用时间序列在过去时刻的观察值推算时间序列的出现规律, 预测今后可能出现的观察值。文中利用 AR 模型预测在分布式实时环境上主机的负载情况。对于有多台主机的分布式实时环境, 当有一个新的任务到达时, 如果能够较准确地预测出各台主机在今后的一段时间内的负载情况, 调度器可以有选择地将任务分配给适当主机, 不仅可以满足尽可能多数量的实时任务的最后时限, 并且可以提高系统的性能。

关键词:AR 模型; 主机负载; 预测

中图分类号:TP368

文献标识码:A

文章编号:1673-629X(2007)09-0038-03

Host Load Prediction Based on AR Module

HUANG Chang-qing¹, XU Hai-shui², ZHONG Hui-yun¹, LIAO Zhi-jian¹, ZHANG Hong¹

(1. Faculty of Computer, Guangdong Univ. of Techn., Guangzhou 510075, China;

2. Center of Network Info. & Modern Edu. Techn., Guangdong Univ. of Techn., Guangzhou 510075, China)

Abstract: As a probability and statistics methodology, linear time series has been applied to many fields. AR module is one of the most familiar linear time series, which based on self-similarity characteristic of time series, predict host load using past observed host load. This paper will use this statistical character and predictable characteristic of AR module to predict host load in distributed real-time system. In a system with several hosts involved, if the application can predict the exact load of each host, it will be helpful to not only the schedule but also the performance, for the scheduler can choose a host to perform.

Key words: AR module; host load; prediction

1 时间序列模型

时间序列是指按时间先后顺序排列的随机序列。

对时间序列参数化后得到的模型称为时间序列模型^[1]。时间序列模型能够反映时间序列向量观测值的具体物理过程或者动态系统的规律性, 又能够将相互关联的时间序列向量转化为相互独立的白噪声序列, 以便进行分析和处理, 为预测或者过程控制提供依据^[2]。

1.1 AR 模型(自回归模型)

对于一种实际的问题, 如果它所表现的随机过程或时序的观测值 $\{X_t, t = 0, \pm 1, \dots\}$ 与其自身的前一个或者前几个时刻的观测值 X_{t-1}, X_{t-2} 有关或者有依

赖性, 就可以使用 AR 模型进行建模。AR 模型的线性方程^[3]用

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + a_t \quad (1)$$

来描述, 等式(1)称为 p 阶自回归模型, 记做 $AR(p)$ 。其中实数 $\varphi_i (1 \leq i \leq p)$ 称为自回归参数, a_t 表示残差, a_t 的期望为 0, 方差为 σ_a^2 , 且

$$\sigma_a^2 = \frac{1}{n} \sum_{i=0}^{n-1} (a_{t+i}^2) \quad (2)$$

1.2 预测原理

在这个应用中使用 AR 模型的主要目的, 是为了将对主机负载的周期性采样随机过程序列 $\{X_t, t = 0, \pm 1, \dots\}$ 作为白噪声源, 使用线性过滤器进行过滤^[4,5]。过滤器的系数可以从过去的序列中得到, 得到的结果即是预测值。预测过程如图 1 所示。

从图 1 可以看到, 模型的输入是白噪声序列 a_t , 白噪声序列的期望为 0, 方差为 σ_a^2 。输出的负载序列 Z_t , 由一些期望值 μ 和一些方差 σ_z^2 表示。中间是系数为 ψ_i 的线性过滤器。每个输出的 z_t 值都是目前白噪声的输

收稿日期: 2006-11-08

基金项目: 广东省自然科学基金资助项目(032497)

作者简介: 黄常青(1982-), 女, 湖北武汉人, 硕士研究生, 研究方向为分布式实时系统; 徐海水, 副教授, 硕士生导师, 研究方向为分布式实时系统。

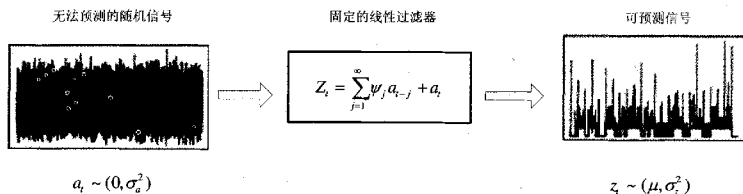


图1 线性时间序列模型

人和所有以前的噪声输入值以 ψ_j 为权重的和。据此可以得出, ψ_j 的最优系数是使白噪声的方差 σ_a^2 最小的数。

引入延迟算子, 令 $Bz_t = z_{t-1}$, 一般的 $B^k z_t = z_{t-k}$ ($k = 1, 2, 3, \dots$), B^k 为步延迟算子。

$$z_t = \psi(B)a_t \quad (3)$$

其中, $\psi(B) = 1 - \psi_1 B - \dots - \psi_p B^p$

对AR模型, 预测值

$$z_t = \frac{a_t}{\psi(B)} + \mu \quad (4)$$

2 实现

由于在分布式实时环境中, 各个主机相互关联, 一台主机的变化可能会引起其他主机的连锁变化。尽管分布式实时系统本身具有动态性和复杂性, 但是简单的线性模型相对于其他的模型在对主机负载的预测上却能够更好地工作^[6]。因此文中使用的AR模型具有计算简单、快速的特点, 模型开销小, 能够达到较好的预测效果^[7]。硬实时系统对满足任务时限的要求十分严格, 简单的预测无法满足硬实时系统的要求, 因此本应用是在软实时系统基础上实现的^[2]。

为了能够评价预测结果, 还使用了另外一种模型——MEAN模型——同AR模型同时进行预测。MEAN模型的预测结果等于过去的实际测量结果的期望值。如果使用AR模型预测结果的方差小于MEAN模型的方差, 则意味着AR模型能够有效地预测主机负载^[5,8]。

预测器包括计算模型、预测到的下一个序列值、一个有最后 p 个实际值的序列和一个包括最后 q 个预测误差值的序列。当模型得到下个序列值后, 它的实际值将被放入实际值序列的队列中, 它的预测误差也将被放入预测误差的序列队列中。

具体实现步骤如下:

1) 对某台处于分布式实时系统的主机负载进行固定间隔时间量化采样, 时间间隔从5秒到3分钟不等, 每次采样100000到300000次。

2) 获取了样本集后, 选取一个样本, 在样本中随机选择一个时间点 $t_{\text{Crossover}}$ 作为初始化预测器和预测样本的分界点。在 $t_{\text{Crossover}}$ 前的样本, 假设数量为 m , 这 m 个样本 $z_{t_{\text{Crossover}}-m}, z_{t_{\text{Crossover}}-m+1}, \dots, z_{t_{\text{Crossover}}-1}$ 用于初始化预测器; 包含了 $t_{\text{Crossover}}$ 并且在 $t_{\text{Crossover}}$ 后的样本, 假设数量为 n , 这 n 个样本 $z_{t_{\text{Crossover}}}, z_{t_{\text{Crossover}}+1}, \dots, z_{t_{\text{Crossover}}+n-1}$ 用于检测预测器是否能够准确地预测主机负载。

3) 根据AR模型编写预测器的算法。预测器的工作就是反复执行以下过程:

```
do
|
task newTask, newTaskPred;
newTask = getNewTask(); // 获得新的任务
newTaskPred = doPrediction(newTask); // 预测主机负载
calErrorSquare(newTask, newTaskPred); // 计算方差
sleepForSamplInterval();
| while(1=1)
```

4) 初始化预测器。从 $i = 0$ 到 m , 使用 $z_{t_{\text{Crossover}}-i}$ 初始化预测器。初始化工作的主要目的是利用等式(1)获得模型的 p 个 ψ_j 值。初始化完成后, 预测器就可以开始预测工作了。

5) 从 $i = 0$ 到 $n-1$, 将 $z_{t_{\text{Crossover}}+i}$ 带入预测器, 对 $k = 1, 2, \dots, 30$ 秒, 计算 $\hat{z}_{t_{\text{Crossover}}+i}^k$ 是利用已知的样本 $z_{t_{\text{Crossover}}-m}, z_{t_{\text{Crossover}}-m+1}, \dots, z_{t_{\text{Crossover}}+i}$ 计算出来的 $z_{t_{\text{Crossover}}+i+k}$ 的预测值。计算预测误差 $a_{t_{\text{Crossover}}+i}^k = \hat{z}_{t_{\text{Crossover}}+i}^k - z_{t_{\text{Crossover}}+i+k}$ 。

6) 对持续时间 $k = 1, 2, \dots, 30$ 秒, 计算 k 步以后的预测误差 $a_{t_{\text{Crossover}}+i}^k, i = 0, 1, \dots, n-1$ 。

7) 输出测试样本的参数和预测误差。

3 评价

这个应用里, 判断AR模型能否有效地预测主机负载的标准是将AR模型预测结果的方差和MEAN模型预测结果的方差进行比较, MEAN模型的方差实际上就是样本的方差, 因此, 如果AR模型的方差能够持续比MEAN模型的方差小, 则证明AR模型能够有效地预测主机负载。如图2所示, 使用AR模型预测30步以后的主机负载的结果的方差都比MEAN模型

的方差小,因此,AR 模型确实能够有效地预测主机负载。

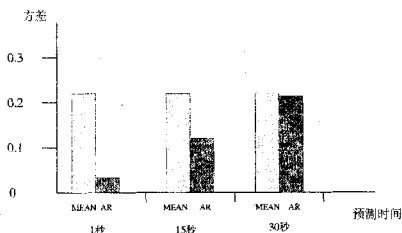


图 2 AR 模型和 MEAN 模型

此外,预测器的初始化时间和预测时间也是评价预测模型的一个重要因素^[9,10]。对在所用的几种模型的初始化时间和预测时间进行比较后,结果如图 3 所示。AR 模型能够在 0.1 秒的时间内初始化完毕,并且在能够接受的足够短的时间内完成预测任务。

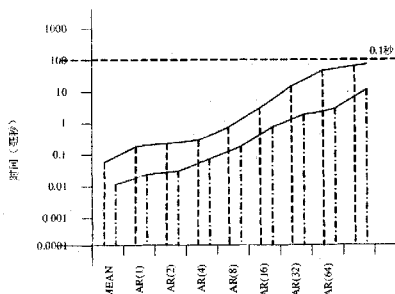


图 3 预测器初始化时间和预测时间比较图

4 结 语

在多种行业中得到广泛应用的线性时间序列^[11]同样也可以被运用到分布式实时系统中,根据线性时间序列的统计功能,利用过去时刻的负载在较准确的范围内计算将来可能发生的负载,使预测分布式实时系统的在未来一段时间内的状况成为可能^[12]。文中利用一种简单的线性时间模型——AR 模型在一个较小误差的范围内完成预测任务,还证实了 AR 模型能够在较短的时间内完成模型的初始化和预测任务,具

有高效性和稳定性的特点。

如果能够较准确地预测主机负载,分布式实时系统的调度器就能够为任务选择主机,为预测任务在某台主机上的运行时间提供了前提。利用主机负载预测任务在某台固定主机上的运行时间,根据预测结果调度任务,最终不仅能够满足尽可能多的任务的最终时限,并且能够很大程度上提高系统的性能。

参考文献:

- [1] 吴怀宇. 时间序列分析与综合[M]. 武汉: 武汉大学出版社, 2004.
- [2] Liu J W S. 实时系统[M]. 姬孟洛, 李 军, 王 馨, 译. 北京: 高等教育出版社, 2003.
- [3] 姚天任, 孙 洪. 现代数字信号处理[M]. 武汉: 华中理工大学出版社, 1999.
- [4] Dinda P A. A Prediction - based Real - time Scheduling Advisor[C]// Proceedings of the 16th International Parallel and Distributed Processing Symposium (IPDPS 2002). Washington D. C. USA : IEEE Computer Society, 2002.
- [5] Dinda P A, O'Hallaron D R. An Evaluation of Linear Models for Host Load Prediction[C]// Proceedings of the 8th IEEE Symposium on High - Performance Distributed Computing (HPDC - 8). Redondo Beach, CA: [s. n.], 1999.
- [6] 屈志坚, 陈剑云. 分布式运动调度系统中实时数据库的研究与实现[J]. 微机计算机信息, 2004(8): 115 - 116.
- [7] 徐浩峰, 应宏伟, 朱向容. 时间序列分析方法预报基坑支撑轴力[J]. 水利学报, 2004(1): 105 - 109.
- [8] Dinda P A. Online Prediction of the Running Time of Tasks [J]. Cluster Computing, 2002, 5(3): 225 - 236.
- [9] Ghysels Y, Valkanov R. Linear Time Series Processes with Mixed Data Sampling and MIDAS Regression Models [EB/OL]. 2006 - 07. <http://ssrn.com/abstract=920610>.
- [10] Dinda P A, Lowekamp B, Kallivokas L F, et al. The Case For Prediction - based Best - effort Real - time Systems[C]// Proceedings of the 7th International Workshop on Parallel and Distributed Real - Time Systems (WPDRTS 1999). San Juan: [s. n.], 1999: 309 - 318.
- [11] 王双强, 陈 强, 李 江. 基于 AR 模型的车辆车型自动分类技术 [EB/OL]. 2005 - 10. 中国科技论文在线, http://www.paper.edu.cn/paper.php? serial_number=200511-308.
- [12] 刘江群, 徐海水. 基于 RT - CORBA 的任务运行时间预测研究[D]. 广州: 广东工业大学, 2005.

(上接第 37 页)

- [6] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. Review of Modern Physics, 2002, 74: 47 - 97.
- [7] Strogatz S H. Exploring complex networks[J]. Nature, 2001, 410: 268 - 276.

- [8] Watts D J, Strogatz S H. Collective dynamics of 'small - world' networks[J]. Nature, 1998, 393: 440 - 442.
- [9] Lee D L, Chung H, Seamons K. Document ranking and the vector - space model[J]. IEEE Software, 1997, 14(2): 67 - 75.