文章编号:1006-2467(2006)11-1951-04

基于强局部加权回归算法的软件老化趋势提取

蒋乐天, 徐国治, 周玲玲 (上海交通大学电子工程系,上海 200240)

摘 要:针对软件老化引起系统资源损耗、系统性能恶化以及软件失效的现象,提出利用强局部加权回归算法进行软件老化趋势的提取方法.该方法利用强局部加权回归算法对系统资源的损耗过程进行提取,得到软件老化的趋势估计,建立软件老化模型.从而可为由于软件老化所引起的软件失效提供前期预测,为系统软件的及时维护提供科学依据.通过对一实际软件系统的实验分析,验证了该方法的可行性和有效性.

关键词: 软件老化; 强局部加权回归; 趋势提取

中图分类号: TP 302.8

文献标识码: A

Abstraction of Software Aging Trend Based on Robust Locally Weighted Regression Algorithm

JIANG Le-tian, XU Guo-zhi, ZHOU Ling-ling
(Dept. of Electronic Eng., Shanghai Jiaotong Univ., Shanghai 200240, China)

Abstract: Software in operational phase often experiences an aging phenomenon, characterized by either progressive performance degradation or transient failures or both, which is caused by aging-related software faults. In order to deal with software aging, a method based on robust locally weighted regression algorithm was presented to abstract aging trend from system resource usage. With the trend, software failure can be predicted and avoided. An experiment for a practical system verifies that the presented method is feasible and effective.

Key words: software aging; robust locally weighted regression; trend abstraction

随着软件技术的发展,软件在一个系统中的作用越来越重要.但近年来的研究和大量的数据表明,一个系统发生失效更多的是由于软件缺陷造成的^[1-3].因此,对软件质量问题的研究已成为一个重要的、前沿的,但同时又是一个非常具有挑战性的研究领域,而软件缺陷的研究更是其中的一个重点.

近年来,人们观察到了一类特殊的软件现象——软件老化.一个软件一旦开始运行,潜在的缺陷条件就会随着时间不断积累,并引起软件系统性

能的恶化. 当经过长时间的连续运行,积累的错误条件达到一定的程度就可能使系统发生失效. 文献[2]中把引起老化的缺陷称为与老化有关的缺陷. 软件老化现象不仅存在于广泛使用的软件中,如Netscape、Windows、Unix等,而且也存在于一些高可用性和安全性系统的应用软件中,如通信系统软件、军事武器系统软件等. 因此老化现象一提出就引起了广泛的关注. 目前,国外很多科研机构和学校,如 IBM、贝尔实验室、Motorola 全球软件中心、Duke

大学、马里兰大学等,都有研究人员在从事研究,并取得了一些进展. 他们的研究成果已开始在实际系统中进行应用,如 IBM 的 XSeries 服务器、Motorola公司的一些产品,并取得了很好的效果^[2,4]. 而国内尚未见相关的文献报道.

与老化有关的软件缺陷引起系统资源的损耗,如内存泄漏、打开文件未被释放、数据破坏等.但与老化有关的缺陷即使被触发也并不会使系统立刻发生失效,它总是表现在系统性能的逐步恶化,直至失效.软件老化现象可以通过系统资源的使用来进行研究,从资源使用中提取出损耗的趋势. Garg 等[5]用 Sen 提出的斜率估计方法对软件的老化趋势建立模型,Castelli等[6]提出了线性回归的方法.但这些方法失真较大,只能在一定程度上判断软件是否存在老化现象,并不能对损耗趋势进行很好的估计.本文利用强局部加权回归算法(Robust Loess)对系统资源使用情况进行分离,将系统资源的损耗过程同正常使用过程进行分离,从而得到软件老化趋势的估计,建立软件老化模型.

1 强局部加权回归算法

强局部加权回归算法由 Cleveland^[7]提出,主要利用局部观测数据对欲拟合点进行多项式加权拟合,并用最小二乘法进行估计.它综合了传统的局部多项式拟合,局部加权回归以及具有强鲁棒性的拟合过程.在每一个拟合点处的拟合数据宽度称为窗

口宽度,该窗口宽度对结果影响很大. 当宽度较大时可以获得大范围的趋势;当宽度较小时,拟合结果比较接近原始数据.

对于一系列观测点 $\{x_i, y_i\}$, $i=1,2,\dots,n$,给出一个如图 1(a)所示的权值函数 W(x),满足:

- (1) 当|x|<1 时 W(x)>0;
- (2) W(-x) = W(x);
- (3) 对于 $x \ge 0$, W(x) 是一个非增函数;

定义 x_i 的 r 次最近点. 令 $0 < f \le 1$, r 为离 fn 最近的整数值. x_i 的 r 次最近点定义为将权值函数 W(x) 以 x_i 为中心放置, 然后对 W(x) 进行尺度变换(见图 1(b)), 使得在离 x_i 点第 r 个近的点处 W(x) 首次变为 0, 就称该点为 x_i 的 r 次最近点. r 相当于对数据进行局部回归时所选取的数据的宽度 (窗口宽度), f 称为平滑系数.

如图 1(c) 所示,对每一个点 x_i ,得到权值 $w_k(x_i)$, $k=1,2,\cdots,n$,使用加权最小二乘估计对 x_i 点进行 d 阶多项式拟合,得到拟合值 \hat{y}_i .然后根据残差 $y_i-\hat{y}_i$ 确定 δ_i 值,当残差较大时取较小的 δ_i ,当残差较小时取较大的 δ_i 值. 利用 $\delta_i w_k(x_i)$ 代替原来的权值 $w_k(x_i)$ 再次计算得到新的拟合值. 重复进行多次得到最后的拟合值 \hat{y}_i . 利用 $w_k(x_i)$ 得到 \hat{y}_i 称为局部加权回归,利用 $w_k(x_i)$ 和 $\delta_i w_k(x_i)$ 进行多次迭代得到 \hat{y}_i 称为强局部加权回归.

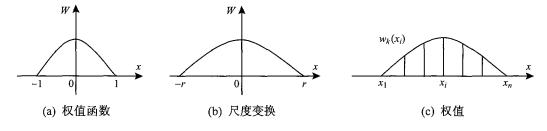


图 1 权值函数 Fig. 1 Weight function

2 软件老化趋势的提取

软件老化过程是一个渐近的过程,是错误条件不断积累的过程,它总是通过系统资源的不断损耗表现出来.而一个软件系统的老化缺陷引起的系统资源的损耗并不会使系统立刻发生失效,它总是表现在系统性能的逐步恶化,直至失效.因此可以通过系统资源数据的损耗来确定软件是否存在老化,以及来确定软件老化的趋势.由于软件存在老化过程,而系统资源的使用主要包括两部分:系统资源的损

耗过程;系统正常的使用部分.系统正常的使用部分随系统负载的变化而变化,其损耗过程具有明显的上升或下降的趋势.对软件老化过程进行分析,就是要从系统资源的使用中将损耗过程提取出来.本文利用强局部加强回归算法实现对系统资源损耗的分离提取.

对某个系统资源参数进行定期监测得到该参数的时间序列 (t_i, P_i) , t_i 为监测时间点, P_i 为 t_i 时刻的观测值.对 P_i 进行拟合,令 h_i 为从 t_i 到 t_j 的 r 次最近点的距离,即 h_i 为 $|t_i-t_i|$ 中的第 r 个最小值,j

 $=1,2,\dots,n$,其中 r=fn. 对于 $k=1,2,\dots,n$,令权值:

$$w_k(t_i) = W(h_i^{-1}(t_k - t_i))$$
 (1)

(1) 对每个 i 计算 P_k 对 t_k 的 d 阶多项式回归 参数的估计值

 $\hat{\beta}_l(t_i) =$

$$\min\left(\sum_{k=1}^{n} w_{k}(t_{i}) (P_{k} - \beta_{0} - \beta_{1} t_{k} - \cdots - \beta_{d} t_{k}^{d})^{2}\right) (2)$$

$$l = 0, 1, \cdots, d$$

此时得到利用 d 阶局部加权回归算法平滑的点(t_i , \hat{P}_i), \hat{P}_i 为在 t_i 点的拟合值,

$$\hat{P}_{i} = \sum_{l=0}^{d} \hat{\beta}_{l}(t_{i}) t_{i}^{l} = \sum_{k=1}^{n} r_{k}(t_{i}) P_{k}$$
 (3)

式中, $r_k(t_i)$ 为由回归得到的 P_k 的系数,不取决于 P_j , $j=1,2,\cdots,n$.

(2) 令 B 为由下式定义的 4 次方权值函数

$$B(z) = \begin{cases} (1-z^2)^2 & |z| < 1\\ 0 & |z| \ge 1 \end{cases}$$
 (4)

并令 $e_i = P_i - \hat{P}_i$ 为拟合值的残差,s 为 $|e_i|$ 的中位值. 定义

$$\delta_b = B(e_b/(6s)) \tag{5}$$

- (3) 对每一个 i,在(t_k , P_k)点用 $\delta_k w_k(t_i)$ 代替原来的权值,利用加权最小二乘法进行 d 阶多项式拟合,计算新的 \hat{P}_i 值.
- (4) 重复执行(2)、(3) T 次,最终得到的 \hat{P}_i 就是强局部加权回归拟合值.

利用强局部加权回归算法对观测序列进行拟合时,多项式的阶数 d、权值函数 W、迭代次数 T 以及 f 是很重要的参数,对结果的影响较大. 前 3 个参数 都可以预先确定,一般来说 d=1, T=2 已经足够,而 f 必须根据数据的需要来确定,因为 f 决定了回归曲线在 t_i 点处的平滑程度. 对所有的观测值进行拟合后,最终得到一个平滑的回归曲线,即

$$P_i = \hat{P}_i + \varepsilon_i \tag{6}$$

式中, ϵ 。为具有零均值和固定方差的的随机变量. 具体到软件的资源使用, \hat{P}_i 代表系统资源的损耗过程,而 ϵ 。为系统正常的使用部分. f 决定了用来进行回归运算的数据宽度的大小. 增大 f 就增大了邻近观测值对 \hat{P}_i 值的影响,从而增大了平滑度. 关于该4 个参数的选择可以参考文献[7].

3 实验

实验条件:实验室内的局域网;服务器配置:奔腾 133 MHz,32 MB 内存,安装 Winroute,该服务器 只充当网关作用,不作其他任何工作.现象:每次刚

启动时,系统运行正常,当运行一段时间后(几个小时),系统运行速度及局域网内的计算机通过该服务器上网速度变得很慢.对该服务器进行重新启动后,又正常.基本上一天之内要重启 2、3 次.对该服务器的运行情况进行了监测.图 2 所示为所采集到的数据,是系统的可用内存.由图 2 可见,虽然系统可用内存是一个快速变化的随机量,但在整个运行阶段,它有一个明显的下降趋势.

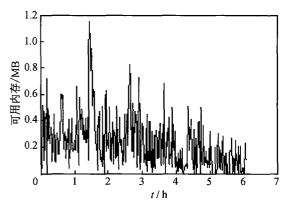
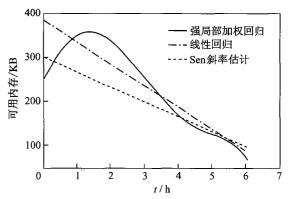
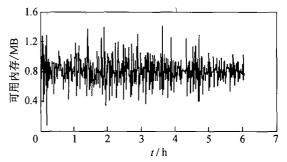


图 2 可用内存 Fig. 2 Available memory

对该可用内存参数分别利用强局部加权回归、 线性回归和 Sen 斜率估计法进行趋势提取得到图 3(a)所示的结果. 由图可见,线性回归和Sen斜率估



(a) 系统资源的损耗过程



(b) 负载使用的系统资源部分 图 3 系统资源使用情况的分离

Fig. 3 Separation of the system resource usage

计法可以对资源的损耗进行趋势估计,但不能较好地反映趋势的变化. 而强局部加权回归方法却得到了较好的结果. 因此可以利用线性回归和 Sen 斜率估计法对系统的某一个资源参数进行是否存在老化现象的初步判断,然后用强局部加权回归算法对它的趋势进行估计,从而得到老化过程. 图 3(b)所示为可用内存去除损耗趋势后的结果,反映了负载所使用的系统资源.

4 结 语

与老化有关的软件缺陷是近年来被认识到的一种软件缺陷,它引起系统资源的损耗而造成软件系统性能的不断恶化,直至发生失效.为了处理与老化有关的缺陷引起的软件老化现象,本文提出了基于强局部加权回归算法的系统资源损耗提取方法.通过从系统资源的使用中提取分离出资源的损耗过程,从而对软件老化过程进行研究.通过实验分析可以看出,该方法确实可以分离资源的损耗过程.对于一个存在老化的软件来说,资源损耗过程的提取可以为由软件老化引起的软件失效提供前期预测.

参考文献:

[1] Huang Y, Kintala C, Funton N D. Software rejuvenation: Analysis, module and applications[C]// Proc

- 25th IEEE Int'l Symp On Fault Tolerant Computing. Los Alamitos, CA: IEEE Computer Society Press, 1995;381—390.
- [2] Vaidyanathan K, Trivedi K S. Classification and treatment of software faults[EB/OL]. (2002-10-06)[2004-05-03]. http://www.software-rejuvenation.com.
- [3] Gray J. Why do computers stop and what can be done about it? [C]// Proc of 5th Symposium on Reliability in Distributed Software and Database Systems. Los Angeles, CA: IEEE Computer Society Press, 1986:3—12
- [4] Network World. IBM's software rejuvenation [EB/OL]. (2001-05-02) [2003-02-27]. http://www.itworld.com/appdev/904/Nww00380627.
- [5] Garg S, Moorsel A V, Vaidyanathan K, et al. A methodology for detection and estimation of software aging[C]// Proc 9th Int'l Symp on Software Reliability Eng. Los Alamitos, CA: IEEE Computer Society Press, 1998;282-292.
- [6] Castelli V, Harper R E, Heidelberger P, et al. Proactive management of software aging[J]. IBM Journal of Research & Development, 2001, 45(2):311-332.
- [7] Cleveland W S. Robust locally weighted regression and smoothing scatterplots [J]. Journal of the American Statistical Association, 1979, 74(368):829-836.

(上接第 1950 页)

- [4] Ruan M, Shen J, Wheeler CB. Latching micro electromagnetic relays [J]. Sensors and Actuators J, 2001, 91: 346-350.
- [5] 张永华,丁桂甫,赵小林,等. 电磁型双稳态射频开关 的微机械结构设计[J]. 上海交通大学学报,2004,38 (5):725-728.
 - ZHANG Yong-hua, DING Gui-fu, ZHAO Xiao-lin, et al. Micromechanical structural design of electromagnetically actuated bistable RF micro-electromechanical systems switches [J]. Journal of Shanghai Jiaotong University, 2004, 38(5): 725-728.
- [6] Gretillat M A, Thiebaud P, Linder C, et al. Integrated-circuit compatible electrostatic polysilicon microrelays [J]. J Micromech and Microeng, 1995, 5 (2): 156-160.
- [7] Fullin E, Gobet J, Tilmans H A C, et al. A new basic technology for magnetic micro-actuators [C]//
 Int Workshop on MEMS, Heidelberg: IEEE, 1998:
 143-147.

[8] ZHANG Yu-feng, LI De-shen. Fabrication and simulation of an electromagnetic microrelay [J]. Chinese Journal of Semiconductors, 2002, 23 (12): 1298—1302.

- [9] 李团结,曹 炎,李世俊,等. 平面双稳态柔性微机构 的优化设计[J]. 机械科学与技术,2004,23(6):709
 - LI Tuan-jie, CAO Yan, LI Shi-jun, *et al*. Optimal design of in-plane bistable compliant micromechanisms [J]. **Mechanical Science and Technology**, 2004, 23(6):709-711.
- [10] 蒋振新,丁桂甫,杨春生,等. 微弯曲面成型技术一多 线成面技术 [J]. **电子工艺技术**,2003,24(5):210-213.
 - JIANG Zhen-xin, DING Gui-fu, YANG Chun-sheng, et al. A novel microfabrication technique of microbuckling surface-forming surfaces with a series of lines [J]. Electronics Process Technology, 2003, 24 (5): 210-213.