软件可靠性预测的小波分析方法

韩峰岩， 覃 征， 王 昕
(西安交通大学 电子与信息工程学院，陕西 西安 710049)

摘 要:软件可靠性预测对于软件可靠性评估以及软件工程项目的过程控制具有重要作用。本文提出的软件可靠性预测模型,通过小波分解将非平稳的软件可靠性随机序列分解成趋势项、周期项和一系列时间调制平稳随机序列,对趋势项和周期项按常规的预测方法进行预测,而时间调制平稳随机序列再用小波方法预测,最后合成得到软件可靠性预测。实例分析表明,本方法是可行的。

关键词:软件;可靠性;预测;小波变换;时间序列
中图分类号:TP311 **文献标识码**:A **文章编号**:1009-3516(2004)06-0072-04

软件可靠性预测主要通过建立软件可靠性模型来实现,目前已经有很多软件可靠性模型[1-2]。这些模型大多建立在软件失效率服从某种分布的假设条件下,比如J-M模型假设失效率与软件中的剩余缺陷数成正比,G-O模型假设失效率呈指数衰减。由于软件逻辑结构的复杂性,测试行为的复杂性,以及失效模式的复杂性,软件可靠性模型的基本假设一直存在许多争论,模型在实际应用中也存在预测精度低、一致性差等问题。Soyer[3]将软件失效作为时间序列来处理,在对软件失效机理不作任何假设的条件下建立了可靠性模型,但由于软件失效随机序列数据离散性大,高度非平稳,分析预测也有一定困难。小波分析是时间序列数据分析的有力工具,已应用于时间序列的拟合预测、密度估计等方面[4]。对于某些非平稳随机序列,可以通过小波多尺度分解将随机序列分解为趋势项、周期项和平稳随机项,然后进行分析预测[5]。但对于像软件可靠性随机序列这样高度不平稳的随机序列,很难简单分解出平稳随机项。本文针对小波分析用于高度非平稳随机序列预测进行探讨,并将小波分析具体应用于软件可靠性预测。

1 软件可靠性数据与小波多分辨率分析(MRA)

记软件系统第 i 次失效时间为 t_i,定义 $t_0=0$,则失效间隔时间 $T(i)=t_i-t_{i-1}, i\leq I$。$T(i)$ 构成一个随机序列。$T(i)$序列的均值函数,就是我们最关心的系统失效间隔时间随测试时间的演化规律。对于测试过程中的一个软件系统,测试中发现错误即进行修正,系统处于不断变化之中,$T(i)$构成一个非平稳随机序列。

图1是Musa[2]描述的系统T1的测试数据,记为DT1。本文采用此数据集作为分析对象。

根据函数空间逐级剖分的概念,对于任意函数$f(t)\in V_{j-1}$,可在V_{j-1}空间展开为[6]$f(t)=\sum_k c_{j-1,k}2^{(-j+1)/2}\phi(2^{-j+1}t-k)$。也可在$V_j$,$W_j$空间展开,分解为$f(t)=A_j(t)+D_j(t)$。其中:$A_j(t)=\sum_k c_{j,k}2^{-j/2}\phi(2^{-j}t-k)$,$D_j(t)=\sum_k d_{j,k}2^{-j/2}\psi(2^{-j}t-k)$。分别称为$j$尺度空间的概要信号和细节信号。$2^{-j/2}\phi(2^{-j}t-k)$ 和 $2^{-j/2}\psi(2^{-j}t-k)$ 分别为尺度函数和小波函数,$c_{j,k}$ 和 $d_{j,k}$分别称为j尺度空间的尺度系数和小波系数。

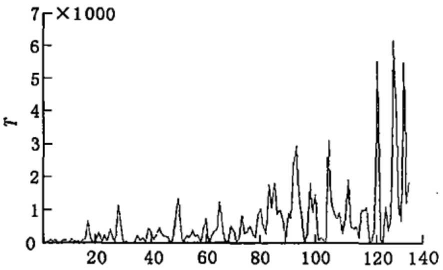

图1 软件失效数据 DT1

收稿日期:2004-07-15
基金项目:国防"十五"计划资助项目(20020416)
作者简介:韩峰岩(1963-),男,河北肃宁人,博士生,主要从事软件工程、可靠性工程研究.

尺度系数和小波系数可以通过逐级递推而求出。$c_{j,k} = \sum_n h_0(n-2k)c_{j-1,n}$;$d_{j,k} = \sum_n h_1(n-2k)c_{j-1,n}$。其中:$h_0(n) = \langle \phi(t), \phi_{-1,n}(t) \rangle$,$h_1(n) = \langle \psi(t), \phi_{-1,n}(t) \rangle$。分别称为低通和高通滤波器系数。滤波器系数及尺度函数、小波函数满足以下方程:$\phi(t/2) = \sqrt{2}\sum_n h_0(n)\phi t - n$;$\psi(t/2) = \sqrt{2}\sum_n h_1(n)\phi t - n$;$\sum_n h_0(n) = \sqrt{2}$;$\sum_n h_1(n) = 0$。

通过以上方法,可以将一个函数(序列)逐次分解,得到包含平稳、趋势信息的概要信号和一系列包含突变、随机信息的细节信号。

2　小波软件可靠性预测

对随机序列预测,可以通过小波多分辨率分析将非平稳随机序列分解成容易预测的简单结构,然后进行分析和预测。

根据小波多分辨率分析理论,对于随机序列 $T(i)$,取初始的尺度系数 $c_{0,k} = T(k)$。则通过 N 层多分辨率分解,可将原始随机序列 $T(i)$ 分解为 $A_N(i), D_1(i), D_2(i), \cdots D_N(i)$,且有 $T(i) = A_N(i) + D_1(i) + D_2(i) + \cdots\cdots + D_N(i)$。

在分解结构中,概要信号代表了原始序列的趋势信息;低层细节信号是序列的剧烈变化部分,$A_N(i)$ 代表了原始序列的随机信息;高层细节信号代表了原始序列的周期信息。可以对趋势部分、周期部分和随机部分分别预测,然后进行合成。

2.1　趋势项的预测

适当选取小波函数,随着分解层数的增加,概要信号中的随机信息将逐步被滤除,当分解到一定层次时,$A_N(i)$ 代表的趋势信息的变化规律比较明显,可以根据曲线的形状,按多项式或指数式进行最小二乘拟合,然后进行预测。

2.2　周期项的预测

高层次细节信号往往出现周期趋势。对周期项可以用周期图法进行拟合与预测。设周期序列$(x_1, x_2, \cdots, x_I)$可用一簇三角函数表示,$x_i = \sum_{k=1}^{k}(a_k\cos(2\pi f_k i) + b_k\sin(2\pi f_k i))$。用最小二乘法可求得系数:$\hat{a}_k = \frac{2}{I}\sum_{i=1}^{I} x_i\cos(2\pi f_k i)$;$\hat{b}_k = \frac{2}{I}\sum_{i=1}^{I} x_i\sin(2\pi f_k i)$。由此即可对周期项进行预测。

2.3　随机项的预测

细节信号中主要成分是随机项,尤其是低层次的细节信号,包含了复杂的随机成分。随机项的分析预测应根据具体情况进行处理。对于变化幅度很小的随机项可以忽略不计;对于平稳的随机项可用 AR、MA、ARMA 或 ARIMA 模型进行分析和预测。许多随机序列经小波分解后,随机项较小,或者变化平稳;而软件失效随机序列规律复杂,数据离散性较大,经小波分解,仍然不能得到平稳序列,也不能忽略。研究表明,多数软件可靠性随机序列经小波分解后的随机项满足时间调制平稳随机过程(time - modulated stationary processes)的条件,可以用时间调制平稳随机过程[7](随机序列)的理论进行分析和预测。

2.4　时间调制平稳随机序列的预测

设 Y_i 为零均值平稳随机序列,$\sigma(z)$ 是定义在$[0,1]$上的满足 Lipschitz 连续条件的函数,即 $|\sigma(z) - \sigma(z_0)| < k|z - z_0|$,$k$ 为常数。如果具有时变方差 $\sigma^2(i/I)$ 的随机序列 X_i 可以表示成 $X_i = \sigma(i/I)Y_i$。则称 X_i 为时间调制平稳随机序列。

对时间调制平稳随机序列 X_i 的分析和预测,可以通过对标准差 $\sigma(i/I)$ 的估计以及对平稳随机序列 Y_i 的分析和预测而解决。

2.4.1　时变方差的估计与预测

时变方差可以首先用最简单的无偏估计 $\sigma^2(i/I) = X_i^2$。然后用小波分析方法进行回归和平滑,便得到对时变方差的估计。也可以用小波分析直接对时变方差进行估计。用多分辨率分析技术,定义实验小波系数为[7] $d_{jt} = \sum_{s=0}^{I} X_S\psi_{jt}(S)$。则 $\sigma_{jt}^2 = \sum_{i=-j}^{-1} 2^j |d_{jt}|^2$。

时间调制的平稳随机序列,由于 $\sigma^2(i/I)$ 满足 Lipschitz 连续条件,属于局部平稳随机序列。对局部平稳

随机序列,其方差是渐变的,在对某点进行预测时,可将前一点的方差作为预测值,即$\sigma((i+1)/i)=\sigma(1)$。

2.4.2 平稳随机序列的预测

对平稳随机序列,通过观察样本自相关函数及样本偏相关函数的拖尾及截尾状况,判断适用模型的种类及阶数。比如用 p 阶自回归模型 AR(p) 可表示为 $Y_i=\sum_{j=1}^{p}\phi_j Y_{i-(p+1)+j}+a_i$。其中,参数 $\phi_1,\phi_2,\cdots\phi_p$的估计值,可根据样本数据,通过求解 Yule - Walker 方程或采用递推算法[8]得到。参数估计值确定后,就可以根据此式对后续点进行预测。

3 实例分析

用双正交 bior3.5 小波对数据 DT1 进行多分辨率小波分解,可得其 1 ~6 级概要信号 $A_1\sim A_6$,图 2 所示。

从图 2 容易看出,随着分解层数的增加,随机成分逐步被滤除,趋势越来越明显。当分解到第 6 层时,变化趋势信息的变化规律已经比较明了。因此总的分解层数可以定为 6。

图 3 所示为 DT1 的 1 ~6 级细节信号 $D_1\sim D_6$。

从图 3 可以看出,D_4 呈明显的周期变化趋势,D_5、D_6 也呈周期变化趋势;而 D_1、D_2、D_3 则包含了主要的随机成分。

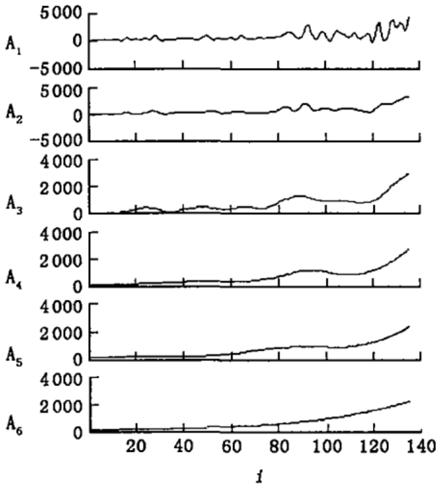

图 2 数据 DT1 的概要信号

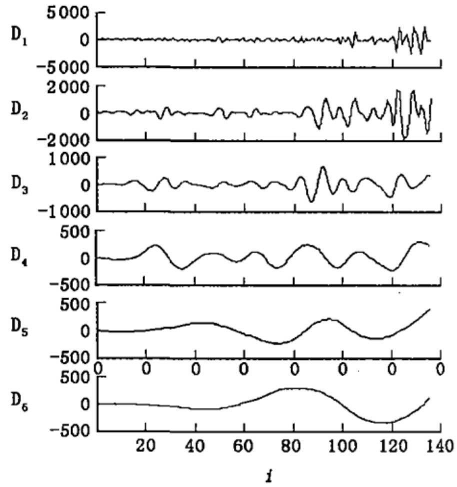

图 3 数据 DT1 的细节信号

趋势项 A_6 可以用多项式模型近似。用最小二乘法,通过试验可以确定满足精度要求的多项式的阶数和多项式系数。图 4 是用 10 阶多项式模拟预测的结果。其中实线代表预测曲线,虚线代表原始曲线。

图 5 为用中心频率为 1/21 的周期图法对 D_4 的预测结果。实线代表预测曲线,虚线代表原始曲线。

对随机项 D_3 可以近似按平稳随机序列进行分析和预测。对 D_1、D_2 按时间调制平稳随机序列处理。

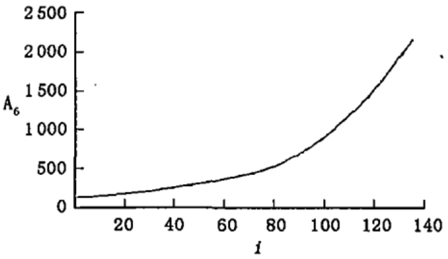

图 4 趋势项的预测结果

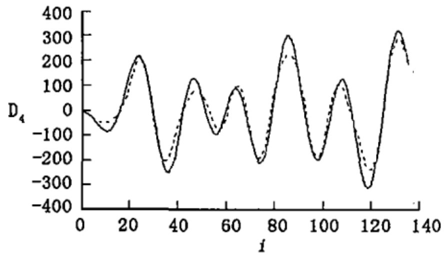

图 5 周期项的预测结果

图 6 为从 D_1 中分解出的时变标准差 σ 和平稳随机序列 Y_1。

经检验 Y_1 可以按平稳随机序列处理。通过自相关和偏相关函数分析,我们采用 AR(3) 自回归模型对 Y_1 进行预测。对 D_2 采用相同的步骤进行预测。

将分项预测结果进行合成,即得对系统 T_1 可靠性的预测,见图 7。实线为预测曲线,虚线为原始曲线。

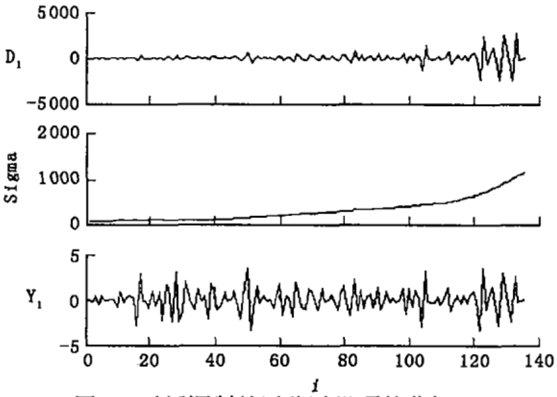

图6 时间调制的平稳随机项的分解

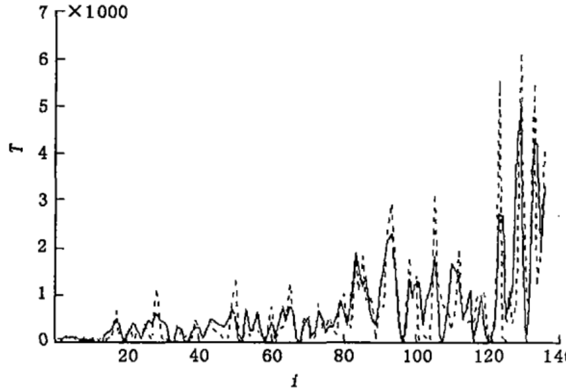

图7 DT1 随机序列的预测结果

4 结论和讨论

1)基于小波变换的软件可靠性预测方法,将高度非平稳的软件可靠性随机序列分解为易于分析和预测的简单结构,再应用时间调制平稳随机过程的理论,可以比较方便地进行预测,具有比较广泛的适用性。

2)预测精度与小波函数的选取及分解的层数有关。分解层数越多,趋势项的结构越简单,便于精确预测;但由于层数增多,累积误差也会加大,因此,分解层数应适当,一般以4~6层为宜。应适当选取小波函数以在尽量少的层次,将原始信号分解为简单结构。

3)本文提出的模型适用于失效数据较多的情况,失效数据太少时误差较大,甚至根本不能使用。

参考文献:

[1] Lyu M R. Handbook of Software Reliability Engineering[M]. New York:McGraw Hill,IEEE Computer Society Press, 1996.

[2] Musa J D,Iannino A,Okumoto K. Software Reliability: Measurement, Prediction, Application[M]. New York:McGraw - Hill, 1987.

[3] Soyer R. Applications of Time Series Models to Software Reliability Analysis, In State of the Art Report: Software Reliability[R]. Pengamon Infotech Ltd,1986.

[4] Herrick D R M,Nason G P,Silverman B W. Some New Methods for Wavelet Density Estimation[R]. Technic Report of University of Bristol, 2001.

[5] 马社祥,刘贵忠,曾召华. 基于小波分析的非平稳时间序列分析与预测[J]. 系统工程学报,2000,15(4):305 -311.

[6] 彭玉华,小波变换与工程应用[M]. 北京:科学出版社,2000.

[7] Fryzlewicz P,Van Bellegem S ,Von Sachs R. Forecasting Non - stationary Time Series by Wavelet Process Modelling, Discussion paper No 0208[R]. Universit'e catholique de Louvain: Institut de statistique,2002.

[8] Bowerman B L,Connell R T. Forecasting and Time Series: An Applied Approach[M]. Beijing:China Machine Press, 2003.

(编辑:姚树峰)

A Wavelets Analysis Approach to Software Reliability Forecasting

HAN Feng - yan, QIN Zheng, Wang Xin
(School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

Abstract:Software reliability forecasting is very important for improving and controlling a software engineering process. This paper presents a wavelet analysis based on forecasting model for software reliability. By the model, the time series of software faults can be decomposed into trend term, periodic term and time - modulated stationary processes, and then forecast them separately. The experience shows that the wavelet analysis approach to software reliability forecasting is feasible.

Key words:software; reliability; forecasting; wavelet transformation; time series