软件可靠性评价的 Hazard Rate 模型☆

郭建英，丁喜波，王天荣

（哈尔滨理工大学， 黑龙江 哈尔滨 150080）

摘 要： 研究软件的可靠性问题，首先要考虑如何测试与评价，作为基础的信息和数据，主要是在各开发阶段通过调试发现。介绍了一类比较简单实用的软件可靠性评价模型，并给出了解析方法。

关键词： 软件可靠性；故障率模型；可靠性评价

中图分类号： O 213.2 **文献标识码：** A

Hazard Rate Model for Software Reliability Evaluation

GUO Jian－ying, DING Xi－bo, WANG Tian－rong

(Harbin University of Science and technology, Harbin 150080, China)

Abstract: To research software reliability, it should be considered as the most important thing that how to measure and appraise. As a basis, information and data are derived mainly through debugging in each development stage. A simple and practical model of software reliability is introduced, and analysis method is given.

Keywords: software reliability; Hazard Rate Model; appraisal of reliability

1 软件可靠性的内涵

软件是一种智力创造，随着计算机科学的发展，已成为系统中独立主宰着系统功能的关键部分。它所提供的形式是程序及相应文档，如何保证在开发阶段实现软件的预期质量与可靠性已是倍受关注的问题。为此，软件可靠性研究应贯穿于从拟定计划任务书开始直到最终调试及交货前的试运行的全过程。核心的问题是防止、发现并纠正串人软件内的各种故障因素、缺陷及隐患。依据全过程的信息和数据记录，最终评价和确认软件的可靠性。

软件可靠性定义为“在规定的环境下，规定的运行期间内，软件正确实现规定功能的能力”。如果用概率描述其能力，就是软件可靠度。为了充分满足用户对软件可靠性的要求，软件开发商往往投入大量的人力资源，在全过程实施对一切故障隐患及缺陷的对抗。查找缺陷、排除隐患、避免故障已是软件开发调试过程的永恒主题。

2 工程上常用的 Hazard Rate 模型

Hazard Rate 模型是基于软件故障发生时间的软件可靠性评价模型。

设随机变量 $T_k(k=1,2,\cdots n)$为第 $k-1$ 和第 k 个软件故障发生的时间间隔，在(0, t]区间内，软件第 k 个故障发生的概率为

$$F_k(t)=P[T_k\leq t] \tag{1}$$

☆国家自然科学基金资助项目(79970045)

收稿日期：2001－08－09

作者简介：郭建英(1943－)，男，河北唐山人，哈尔滨理工大学测试计量技术与仪器专业教授，博士生导师，曾两次赴日本研修；从“六五”～“九五”期间，承担国家科技攻关项目等课题十五项，一项获机械部科技进步二等奖，四项获三等奖，主要从事可靠性工程理论及应用研究。

软件第 k 个故障不发生的概率为

$$R_k(t)=1-F_k(t)=P[T_k>t] \tag{2}$$

在软件第 k 个故障尚未发生的条件下，即刻将会发生故障的概率为

$$Z_k(t)=\frac{f_k(t)}{1-F_k(t)} \tag{3}$$

式(1)中的 $f_k(t)$ 是概率分布函数 $F_k(t)$ 的导函数 $f_k(t)=dF_k(t)/dt$，称作概率密度函数。$Z_k(t)$ 反映了软件第 k 个故障发生在 t 时刻的瞬间速率。式(3)便是基本的故障率(Harzard Rate)模型。

工程上常用的有代表性的 Harzard Rate 模型有以下 3 种：

(1)Jelinski－Moranda 模型

$$Z_k(t)=H[N-(k-1)];\ (N>0;\ H>0;\ k=1,2,\cdots,N) \tag{4}$$

式中：N：调试前软件内潜在的故障隐患总数

H：残存隐患的故障率。

(2)Moranda 模型

$$Z_k(t)=DC^{k-1};\ (D>0;\ C>0;\ k=1,2\cdots N) \tag{5}$$

式中：D：初期故障；

C：故障率减少系数。

(3)Xie 模型

$$Z_k(t)=\lambda_0[N-(k-1)]^a;\ (N>0;\ \lambda_0>0;\ a>0;\ k=1,2,\cdots N) \tag{6}$$

式中：N:调试前软件内潜在的故障隐患点数；

λ_0:故障率系数；

a:常数。

上述 3 个模型，若设 $Z_k(t)=\lambda_k$ 为常数，则

$$f_k(t)=\lambda_k\exp[-\lambda_k\cdot t] \tag{7}$$

$$f_k(t)=1-\exp[-\lambda_k\cdot t] \tag{8}$$

$$R_k(t)=\exp[-\lambda_k\cdot t] \tag{9}$$

这时，随机变量 T_k 服从指数分布。

3 模型参数的极大似然估计

设 t_k 为第 $k-1$ 和第 k 个故障的时间间隔的观测值($k=1,2,\cdots n$)，获取几个观测值数据的概率的似然函数可表示成

$$L=\prod_{k=1}^{n}f_k(t_k) \tag{10}$$

对式(10)两边取对数

$$\ln L=\sum_{k=1}^{n}\ln f_k(t_k) \tag{11}$$

当 $Z_k(t)=\lambda_k$ 为常数时，可将式(7)代入式(11)，得到

$$\ln L\sum_{k=1}^{n}\ln\lambda_k-\sum_{k=1}^{n}\lambda_k\cdot t_k \tag{12}$$

针对式(4)的 Jelinski－Moranda 模型，可表示成

$$\ln L=\sum\ln\{H[N-(k-1)]\}-\sum_{k=1}^{n}H[N-(k-1)]t_k \tag{13}$$

令

$$\begin{cases}\dfrac{\partial\ln L}{\partial N}=0\\[2mm]\dfrac{\partial\ln L}{\partial H}=0\end{cases} \tag{14}$$

称其为似然方程组，经整理可得

$$\begin{cases}H=\dfrac{n}{\sum\limits_{k=1}^{n}(N-k+1)\cdot t_k}\\[4mm]\sum\limits_{k=1}^{n}\left(\dfrac{1}{N-k+1}\right)=\dfrac{n\cdot\sum\limits_{k=1}^{n}t_k}{\sum\limits_{k=1}^{n}(N-k+1)\cdot t_k}\end{cases} \tag{15}$$

将故障记录数据代入式(15)，可求解出 N、H 的极大似然估计 $\hat{N}$ 和 $\hat{H}$。

工程上还希望对下一个软件故障的到来时间进行推测。第 n 和第 $n+1$ 个软件故障的时间间隔 T_{n+1} 仍然是随机变量，其期望值即 $(\mathrm{MTBF})_{n+1}$，以及其可靠度函数可用下面的公式求解。

$$E[T_{n+1}]=(\mathrm{MTBE})_{n+1}=\frac{1}{\hat{H}(\hat{N}-n)} \tag{16}$$

$$R_{n+1}(t)=\exp[-\hat{H}(\hat{N}-n)t] \tag{17}$$

针对式(5)的 Moranda 模型，其对数似然函数为

$$\ln L=\sum_{k=1}^{n}\ln(D\cdot C^{k-1})-\sum_{k=1}^{n}D\cdot C^{k-1}\cdot t_k \tag{18}$$

分别对其待估参数 D・C 求导，令其为 0，可得似然方程组

$$\begin{cases}D=\dfrac{n}{\sum\limits_{k=1}^{n}C^{k-1}\cdot t_k}\\[4mm]\sum\limits_{k=1}^{n}\dfrac{k-1}{C}=\sum\limits_{k=1}^{n}D(k-1)C^{k-2}\cdot t_k\end{cases} \tag{19}$$

将 $t_k(k=1,2,\cdots,n)$观测值代入，便可求解出参数 D 和 C 极大似然估计$\hat{D},\hat{C}$。同理，对于下一个软件故障的时间间隔，可用下面公式推测，即

$$E(T_{n+1})=(\mathrm{MTBF})_{n+1}=\frac{1}{DC^n} \tag{20}$$

$$R_{n+1}(t)=\exp[-D\cdot C^n\cdot t] \tag{21}$$

针对式(6)的 Xie 模型，其对数似然函数为

$$\begin{aligned}\mathrm{In}L=&\sum_{k=1}^{n}\{ln\lambda_0+\alpha\ln[N-(k-1)]\}\\&-\sum_{k=1}^{n}\lambda_0[N-(k-1)]^\alpha\cdot t_k\end{aligned} \tag{22}$$

分别对其待估参数 λ_0, N, α 求导，令其为 0，整理后的方程组如下：

$$\begin{cases}\lambda_0=\dfrac{n}{\sum\limits_{k=1}^{n}(N-k+1)^\alpha\cdot t_k}\\[2ex]\dfrac{a}{\sum\limits_{k=1}^{n}(N-k+1)}=\dfrac{n\cdot\alpha\sum\limits_{k=1}^{n}(N-k+1)^{\alpha-1}t_k}{\sum\limits_{k=1}^{n}(N-k+1)^\alpha\cdot t_k}\\[2ex]\sum\limits_{k=1}^{n}In(N-k+1)=\dfrac{\sum\limits_{k=1}^{n}(N-k+1)\cdot\mathrm{In}(N-k+1)\cdot t_k}{\sum\limits_{k=1}^{n}(N-k+1)^\alpha\cdot t_k}\end{cases} \tag{23}$$

将 $t_k(k=1,2\cdots n)$观测值代入式(23)，可求解出参数 λ_0, α, N 的极大似然估计$\hat{\lambda},\hat{a},\hat{N}$。同理，还可以求出

$$E(T_{n+1})=(\mathrm{MTBF})^{n+1}=\frac{1}{\lambda_0(\mathrm{N}-\mathrm{n})^\alpha} \tag{24}$$

$$R_{n+1}(t)=\exp[-\lambda_0(N-n)^\alpha\cdot\mathrm{t}] \tag{25}$$

式(15)，(19)，(23) 3 个方程组存在收敛速度慢的问题时，应先给定误差值，再用数值方法迭代求解。

4 数值列

某软件在调试中，记录了 20 次故障，其故障间隔时间依次为 0.5, 1, 1, 1.5, 2, 2.5, 3, 3, 3.5, 4, 4, 3, 3, 5, 4, 4, 5, 6, 6, 8(小时)，试用 Jelinski－Moranda 模型解析。

将数据代入式(15)的方程组中可得：

$$H=\frac{20}{70\cdot n-863}$$

$$\sum_{k=1}^{20}\left(\frac{1}{N-K+1}\right)=\frac{1\ 400}{70\ N-863}$$

因该式收敛缓慢，给定一个小数 $\varepsilon=0.01$，令

$$\sum_{k=1}^{20}\left(\frac{1}{N-k+1}\right)-\frac{1\ 400}{70\ N-863}\leq\varepsilon$$

采用迭代法求解，得

$$\hat{N}=80,$$

将 N 代入上式

$$\hat{H}=4.22\times10^{-3}(1/\text{小时})$$

再将$\hat{N}$和$\hat{H}$代入式(16)，(17)，则

$$E(T_{n+1})=(\mathrm{MTBF})_{n+1}=3.95(\text{小时})$$

$$R_{n+1}(t)=\exp[-0.25/t]$$

$E(t_{n+1})$给出了第 $n+1$ 个故障到来的平均时间，当调试时间 $t=3.95$ 时，$R_{n+1}=0.37$。

参考文献：

[1]山田茂．ソフトウェア信赖性モデル——基础と应用[M]．东京：日本科技连盟，1994.

[2]山田茂．ソフトウェア信赖性评价技术[M]．东京：HBJ 出版局，1989.

[3] Musa, J, Iannino, A, Odumoto, K. Software Reliability; Measurement, Prediction, Application[J]. MCGraw－Hill, 1987.