第 5卷第 4期　南京师范大学学报 (工程技术版)　Vol 5 No 4
2005年 12月　JOURNAL OF NANJING NORMAL UNIVERSITY (ENGINEERING AND TECHNOLOGY)　Dec, 2005

在确定软件最佳投放时间时的风险控制

周秀轻[1], 赵　进[2]

(1. 南京师范大学 数学与计算机科学学院, 江苏 南京 210097;
2. 南京大学 数学系, 江苏 南京 210093)

[摘要]　为了克服平均成本最小原则无法控制风险的缺点, 达到规避风险的目的, 针对软件的最佳投放时间的确定, 提出了“安全第一”的原则, 该原则将风险控制作为首要目标. 将安全第一准则与平均成本最小原则相结合, 提出了一种新的决定软件最佳投放时间的混合策略, 使用该混合策略, 既可以保证较低的平均成本, 又可以适当地控制风险。在这种新的策略下, 软件最佳投放时间的决定转化为一个带概率约束的最优化问题, 该问题是可解的. 并且在 J-M 模型下, 将该约束随机规划问题近似地转化成了常规的最优化问题. 最后, 用一个实例说明了这种新的策略的应用.

[关键词]　软件投放策略, 安全第一准则, J-M 模型, 约束随机规划

[中图分类号] O211. 3;　[文献标识码] A,　[文章编号] 1672-1292(2005) 04-0038-04

Optimal Release Policies for Software Systems under Principles of Incorporating Safety

ZHOU Xiuqing[1], ZHAO Jin[2]

(1. School of Mathematics and Computer Science, Nanjing Normal University, Jiangsu Nanjing 210097, China;
2. Department of Mathematics, Nanjing University, Jiangsu Nanjing 210093, China)

Abstract To overcome the shortcomings of the principle of minimizing the cost expectation, and to take the risk control into consideration in deciding the optimal release time of software, a principle of "safety-first" is proposed. A new hybrid principle is proposed by incorporating the cost expectation principle with the safety-first principle. By the new hybrid principle, the cost expectation and the risk can be controlled at the same time. Under the new principle, the optimal release time problems turn out to be probabilistic constrained programs in stochastic programming and the solution procedures are available. The stochastic programming problem is reduced to a normal programming problem when the hybrid principle is applied to J-M model. Finally, a numerical example is presented.

Key words software release policy, safety-first principle, J-M model, probabilistic constrained program

0　引言

对软件开发者来说, 决定软件的最佳投放时间是非常重要的. 一方面, 软件的质量很大程度上依赖于测试时间的长短, 过短的测试时间会使软件的质量受到影响, 导致软件的可靠性过低, 从而增加后期修复错误的费用, 甚至失去客户的信任; 另一方面, 过长的测试时间会增加开发成本, 而且延迟交货时间会引起用户的不满, 同时也会降低市场竞争力. 因此, 在这两方面之间寻找一种平衡是必要的, 这就是软件的最佳投放时间的决定. Okumoto 和 Goelcite[7] 首先考虑了该问题, 并提出了基于成本指标的最佳投放策略. Dalal和 Mallows[2], May 等[5], Morali等[6], Singpurwalla[10], Yamada和 Osaki[12]也对软件最佳投放时间作了研究. 这些文献在考虑成本时, 针对不同的可靠性模型, 采用的往往都只是使平均成本达到最小的原则.

最小平均成本原则在很多决策问题中都是常用的, 也是有一定道理的. 但是, 该原则没有把风险考虑在内, 而且平均成本最小原则的使用要求测试是有重复的, 而事实上, 软件投放只能进行一次, 没有重复性, 这就增加了风险. 如下节所述, 在某些情况下, 平均成本的最小化就意味着成本方差的最大化, 而方差在一定程度上就表现了风险的大小. 因

收稿日期: 2005-06-28

作者简介: 周秀轻(1975-), 女, 讲师, 博士, 主要从事数理统计方面的教学与研究. E-mail: zhouxiuqing@nju.edu.cn

此, 单独使用平均成本最小原则是不够的, 必须同时适当地控制风险.

在金融领域, 避免风险的一个常用方法就是"安全第一"原则, 即要求成本超出某个范围的概率不大于某个给定的值. "安全第一"原则在经济领域有着广泛的应用, 可以起到很好的规避风险的作用, 见 Buhlman[1], Pyle和 Turnovsky[8]以及 Sengupta[9]等. 软件投放的时间问题与软件成本密切相关, 可以看作是一个经济问题, 为了规避风险, 基于上述原则, 本文提出了决定软件最佳投放时间的"安全第一"的原则, 并将该原则与最小平均成本原则相结合, 提出了一种新的决定软件最佳投放时间的混合策略, 使用该策略确定的投放时间, 既能够保证较低的平均成本, 又能够在一定程度上控制风险.

1 风险控制

记 t 为软件测试时间, $L(t)$ 表示测试时间为 t 时软件的成本, 决定软件最佳投放时间的方法通常是: 找到一个合适的测试时间 t, 使软件的成本达到最小, 则该 t 即作为软件最佳投放时间. 由于对任意给定的 t, $L(t)$ 是一个随机变量, 一个自然的方法就是使 $EL(t)$ (其形式根据软件可靠性的不同而不同) 达到最小, 从而求得一个最佳的投放时间 T, 这种方法称为平均成本最小原则.

平均成本最小原则是决定软件最佳投放时间的一个基本的, 也是最常用的原则. 但是, 该原则没有考虑风险, 而且, 因为软件的投放是不可重复的, 故单独考虑平均成本就有很大的风险. 考虑 Okumoto和 Goel[7] 中的软件可靠性模型, 记:

$$L(t) = c_1 t + c_2 N(t) + c_3 (N - N(t)),$$

式中, c_1 为每增加一个单位的测试时间所增加的成本, 包括由于延迟软件投放时间所造成的损失等; c_2 为在测试阶段每去除一个错误所需的成本; c_3 为在软件投放之后每修复一个错误所需的成本, 且 $c_3 > c_2$; $N(t)$ 为测试到 t 时刻为止累计被查出的错误数; N 为软件中总的错误数.

通常假定 $N(t)$ 为非齐次 Poisson过程, 其均值函数记为 $\Lambda(t)$. 此时成本的均值和方差分别为:

$$EL(t) = c_1 t + c_3 N - (c_3 - c_2)\Lambda(t),$$

$$\mathrm{Var}L(t) = (c_3 - c_2)^2 \Lambda(t).$$

当 $c_1 = 0$ 时, 成本均值的极小化就意味着成本方差的极大化, 而较大的方差就意味着较大的风险. 当然, $c_1 = 0$ 只是一个极端情况, 但由此可以看出只考虑平均成本将会带来一定风险, 尤其是在 c_1 相对较小的情况下.

规避风险的一个方法就是使用"安全第一"原则, 即: 对给定的水平 $0 < \alpha < 1$, 找到一个合适的 t 值, 使得对于该 t 值, $u(t)$ 的值达到最小, 其中 $u(t)$ 为成本函数 $L(t)$ 的上侧 α 分位数. 该问题可以表示为如下的约束随机规划问题:

$$\begin{aligned} &\min_t \quad u(t) \\ &\mathrm{s.t.} \quad P(L(t) \geqslant u(t)) \leqslant \alpha, \end{aligned} \tag{1}$$

其中, $0 < \alpha < 1$ 为一个给定的概率值. 使用"安全第一"原则, 可以找到一个最小的 $u(t)$, 使得成本 $L(t)$ 不超过 $u(t)$ 的概率大于或等于 $1-\alpha$, 这实际上就在一定程度上控制了风险.

将"安全第一"原则与最小平均成本原则综合考虑, 此时, 软件最佳投放时间的决定可以表示成如下的约束随机规划问题:

$$\begin{aligned} &\min_t \quad a u(t) + (1-a) EL(t) \\ &\mathrm{s.t.} \quad P(L(t) \geqslant u(t)) = \alpha, \end{aligned} \tag{2}$$

其中, $0 \leqslant a \leqslant 1$ 是一个给定的权值. 显然 (2) 是问题 (1) 当 $a = 1$ 时的特殊情况. 在实际应用中, 可以根据问题的具体要求确定合适的权重 a, 使得在决定软件投放时间时, 既能够保证较低的成本, 又能够适当的规避风险.

问题 (1) 和 (2) 的解法可以参考随机规划方面的文献, 如 Wets[11], Kall和 Wallace[4]等. 解问题 (1) 和 (2) 的主要困难在于对随机约束的处理. 对于很多常用的软件可靠性模型, $P(L(t) \geqslant u(t))$ 可以直接计算, 从而问题 (1) 和 (2) 中的概率约束可以用一个包含 $u(t)$ 的明确的不等式表示. 这样, 随机规划问题 (1) 和 (2) 就转化成了一个不含任何随机性的非线性规划问题, 其解可由常规的数值方法得到.

2 在 J-M 模型中的应用

J-M 模型[3] 是软件可靠性中的一个重要模型, 对以后的许多模型都有重要的影响. 本节将针对 J-M 模型给出问题 (1) 和 (2) 的近似解.

设软件中总的错误数为 N. 记 $0 = t_0 < t_1 < t_2 < \cdots < t_i < \cdots$ 为软件测试过程中故障发生时间, $x_i = t_i - t_{i-1}$ $(i = 1, 2, \cdots)$ 表示相邻两次故障的间隔时间, 在 J-M 模型下, $\{x_i\}$ 为独立的随机变量序列, 且 x_i 服从参数为 $\lambda(N - i + 1)$ 的指数分布, 即: $P(x_i < t) = 1 - e^{-\lambda(N-i+1)t}$, 其中 λ 表示单个错误的故障率.

将 $L(t) = c_1 t + c_2 N(t) + c_3 (N - N(t))$ 带入

随机规划问题(2)的约束条件,得:

$$\alpha = P(L(t) \geqslant u(t)) = P\left(N(t) \leqslant \frac{c_1 t + c_3 N - u(t)}{c_3 - c_2}\right).$$

为了进一步计算上述概率,需要下面性质 1.

性质 1　在 J-M模型下,$N(t)$服从参数为N和$1-e^{-\lambda t}$的二项分布.即$N(t) \sim B(N, 1-e^{-\lambda t})$.

证明　首先容易看出

$$P(N(t)=0) = e^{-N\lambda t}.$$

于是

$$\begin{aligned} P(N(t)=i) &= \int_0^t P(N(t-t_1) = i-1 \mid t_1) N\lambda e^{-N\lambda t_1} \mathrm{d}t_1 = \\ &\int_0^t \binom{N-1}{i-1}(1-e^{-\lambda(t-t_1)})^{i-1} e^{-(N-i)\lambda(t-t_1)} N\lambda e^{-N\lambda t_1} \mathrm{d}t_1 = \\ &e^{-N\lambda t}\binom{N}{i}\int_0^t \lambda i(1-e^{-\lambda x})^{i-1} e^{i\lambda x} \mathrm{d}x = \\ &\binom{N}{i} e^{-N\lambda t}(e^{\lambda t}-1)^i = \binom{N}{i}(1-e^{-\lambda t})^i e^{-(N-i)\lambda t}, \end{aligned}$$

其中第一个等号成立是因为 t_1服从参数为$N\lambda$的指数分布,第二个等号是运用数学归纳法得到的.

由上述性质及中心极限定理可知,当N充分大时,$\dfrac{N(t) - N(1-e^{-\lambda t})}{\sqrt{N(1-e^{-\lambda t})e^{-\lambda t}}}$近似服从标准正态分布$N(0,1)$.于是有:

$$P(L(t) \geqslant u(t)) = P\left(\frac{N(t) - N(1-e^{-\lambda t})}{\sqrt{N(1-e^{-\lambda t})e^{-\lambda t}}} \leqslant \frac{c_1 t + c_3 N - u(t) - (c_3 - c_2)N(1-e^{-\lambda t})}{(c_3 - c_2)\sqrt{N(1-e^{-\lambda t})e^{-\lambda t}}}\right) = \alpha$$

由此可得$u(t)$的近似表达式:

$$u(t) \approx c_1 t + c_3 N - (c_3 - c_2)[N(1-e^{-\lambda t}) - \mu_\alpha \sqrt{N(1-e^{-\lambda t})e^{-\lambda t}}], \tag{3}$$

其中,μ_α是标准正态分布的上侧 α分位数.而

$$EL(t) = c_1 t + c_2 N + (c_3 - c_2)N e^{-\lambda t}. \tag{4}$$

将(3)和(4)带入(2)式,可将约束随机规划问题(2)近似地转化成为常规的无约束最优化问题

$$\min_t c_1 t + c_2 N + (c_3 - c_2)[N e^{-\lambda t} + a\mu_\alpha \sqrt{N(1-e^{-\lambda t})e^{-\lambda t}}]. \tag{5}$$

当 $a = 1$时,(5)即为(1)的近似问题.

3　实例

考虑 J-M 模型的一个具体例子.数据来自 Jelinski和 Moranda[3],是 NTDS(Naval Tactical Data System,一个大型的军事软件系统)的测试数据.其中共有 34个出错时间,分别属于 4个阶段:生成阶段、测试阶段、使用阶段及再测试阶段,数据详见表 1.

表 1　NTDS测试数据

出错序号(i)	x_i	t_i	出错序号(i)	x_i	t_i
生成阶段			生成阶段		
1	9	9	20	1	105
2	12	21	21	11	116
3	11	32	22	33	149
4	4	36	23	7	156
5	7	43	24	91	247
6	2	45	25	2	249
7	5	50	26	1	250
8	8	58	测试阶段 1		
9	5	63	27	87	337
10	7	70	28	47	384
11	1	71	29	12	396
12	6	77	30	9	405
13	1	78	31	135	540
14	9	87	使用阶段		
15	4	91	32	258	798
16	1	92	测试阶段 2		
17	3	95	33	16	814
18	3	98	34	35	849
19	6	104			

令成本函数 $L(t)$中的系数 $c_1 = 0.01$,$c_2 = 1.0$,$c_3 = 5.0$,概率水平 $\alpha = 0.05$.使用第一阶段的数据,可得参数N和 λ的极大似然估计值分别为$\hat{N} = 31.2$和 $\hat{\lambda} = 0.00685$,将其带入(5)可得最优化问题

$$\min_t 0.01t + 31.2 + 4[31.2e^{-0.00685t} + 1.65a\sqrt{31.2e^{-0.00685t}(1-e^{-0.00685t})}],$$
$$\text{s.t.}\quad t \geqslant 0.$$

其中,a为权重系数.当 a变化时,最佳投放时间 T也随之变化,具体如图 1所示.当 $a = 0$时,表示只考虑最小平均成本原则,此时 $T = 649.4$;当$a = 1$时,表示只考虑“安全第一”原则,此时 $T = 835$.当 a由 0增大到 1时,最佳投放时间 T从 649.4增加至 835.由图 1可以看出,在 $a \in [0,1]$上 T的变化趋势近似于直线,这是由于单个错误的故障率 λ过小

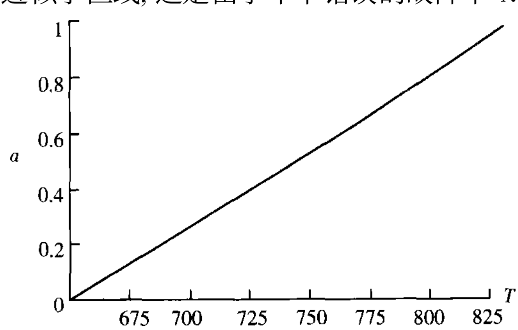

图 1　NTDS的最佳投放时间 T与权重系数 a的变化关系

造成的, 因为在本例中, λ的估计值 λ = 0 00 685 事实上, 假设 λ的估计 λ = 0 2, 而保持其它参数值不变, 则在 $a \in [0, 1]$ 时, 最佳投放时间 T 的变化趋势如图 2所示. 而在实际确定最佳投放时间时, 可以根据问题的具体情况, 选择适当的权重系数 a, 使得既能够保证较低的软件成本, 又能够在一定程度上控制风险.

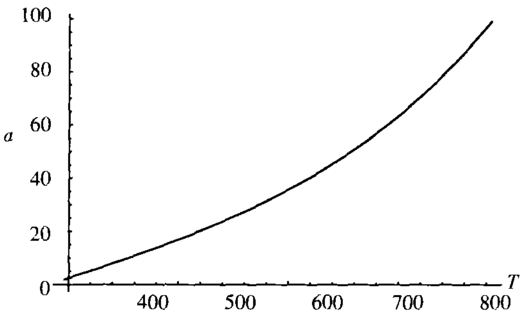

图 2 假设 $\hat{\lambda}$=0.2 时,最佳投放时间 T 与权重系数 a 的变化关系

[参考文献]

[1] Buh lman H. M athematicalM ethods in R isk Theory [M]. Berlin: Springer-V erlag 1970.

[2] DalalS R, M allows C L. When should one stop testing softw are [J]. J Am Statist Assoc, 1988, 83(403): 872 - 879.

[3] Jelinski Z, M oranda P B. Softw are reliability research [C] / / Freiberger Statistical Computer Performance E-valuation. N ew York A cademic Press, 1972: 465 - 484.

[4] KallP, W allace S. Stochastic Programming [M]. Chichester W iley, 1994

[5] M ay J, H ughes G, Lunn A D. Reliability estimation from appropriate testing of plant protection softw are [J]. Software Eng J, 1995, 10(6): 206 - 218.

[6] M oraliN, Soyer R. Optimal stopping in software testing [J]. N av Res Logi, 2003, 50(1): 88 - 104.

[7] Okumoto K, GoelA L. Opitmal release time for software systems based on reliability and cost criteria [J]. J Systems Software, 1980, 1(4): 315 - 318

[8] Pyle D, Turnovsky S. Safety-first and expected utility maximization in mean standard deviation portfolio analysis [J]. Rev Econ and Stat, 1970, 52(1): 75 - 81.

[9] Sengupta J. Safety-first rules under chance-constrained programming [J]. Operations Res, 1969, 17(1): 112 - 132.

[10] Singpurwalla N D. Determining an optimal time interval for testing and debugging software [J]. IEEE Trans Software Eng, 1991, 17(4): 313 - 319.

[11] W ets R. Stochastic programming solution techniques and approximation schemes [C] / / Bachem A, GrotschelM, Korte B. M athematical Programming Berlin: Springer-V erlag 1983: 566 - 603

[12] Yamada S, Osaki S. Optimal software release policies with simultaneous cost and reliability requirement [J]. Eur J Opre Res, 1987, 31(1): 46 - 51.

[责任编辑: 刘 健]

(上接第 3页)

3 结束语

由于目前在理论研究方面还很不够, 特别是重要的粒界效应研究方面更是如此. 虽然本文中介绍了粒界的导电特性, 但它忽略了诸如偏析等许多重要因素, 粒界的结构复杂, 还有待于人们继续探究.

总之, 迄今人们对 CTR 陶瓷的认识仍只是初步的. 今后应从缺陷化学、熔体化学及粒界形成的热力学方面作进一步的深入研究. 在这方面理论和应用研究的进展, 将会促进我国电子陶瓷的生产和电子科学技术的进步.

[参考文献]

[1] 孙健, 甘朝钦. V-P-Fe系 CTR的研制 [J]. 应用科学学报, 2000, 18(2): 186 - 188

[2] 莫以象, 李标荣, 周国良. 半导体陶瓷及其敏感元件 [M]. 上海: 上海科学技术出版社, 1983.

[3] 杨文, 杨邦朝. 阳离子分布对尖晶石型热敏陶瓷电性能的影响 [J]. 功能材料, 2000, 31(5): 513 - 515.

[4] 刘思科. 半导体物理学 [M]. 北京: 国防工业出版社, 1979.

[5] 孙生才. 常温 NTC 热敏电阻的几种材料 [J]. 传感器技术, 1991(5): 9 - 13.

[6] 小西良弘, 迁改郎. 电子陶瓷基础和应用 [M]. 王兴斌, 译. 北京: 机械工业出版社, 1983

[责任编辑: 严海琳]