

基于贝叶斯的软件可靠性评估研究

柴哲丽, 林佳齐, 朱金平, 唐 平

(广东工业大学自动化学院, 广州 510006)

摘 要: 以贝叶斯公式为基础, 根据软件正确性的先验概率密度函数以及正确通过软件测试案例的数量分布服从二项分布的数学特征, 求得软件正确性概率的后验分布, 并提出一种改进的软件可靠性评估方法, 从而解决了软件测试可靠性评估过程复杂且计算量较大的问题。在 Matlab 平台上对软件系统(中文学习平台)的测试可靠性进行评估, 实验结果表明, 该方法具有较高的实用性。

关键词: 软件测试; 贝叶斯公式; 软件可靠性; 二项分布

Research on Software Reliability Evaluation Based on Bayes

CHAI Zhe-li, LIN Jia-qi, ZHU Jin-ping, TANG Ping,

(Faculty of Automation, Guangdong University of Technology, Guangzhou 510006)

【Abstract】 On basis of Bayes formula, a modified method about evaluation of software reliability is presented according to the probability-density function of the correctness of the software and mathematical characteristics of the number of test cases passing software obeying binomial distribution. The probability distribution function of the correctness of the software is deduced. Compared with the process evaluating software reliability by defining parameter, the modified method is more simply and has less calculated amount. The reliability of the platform of learning for Chinese study is evaluated with Matlab platform. Experimental results show this method is more practical.

【Key words】 software test; Bayes formula; software reliability; binomial distribution

1 概述

测试在软件开发过程中备受关注, 在软件工程中有个明确独立的测试阶段。由于软件危机的频频出现以及人们对软件本质的进一步认识, 测试的地位得到很大提高^[1]。随着软件测试理论不断完善, 软件测试的评价和软件测试质量的度量获得很大进步。选择科学合理的评价方法成为软件测试质量评价的重要问题, 也是当前软件测试质量评价研究中的难点。

目前, 评估软件可靠性一般以贝叶斯理论为基础, 文献[2]提出基于输入域的软件可靠性评估的贝叶斯方法, 上海计算机软件技术开发中心研究了软件测试的不确定性以及解决方法, 中国科学技术研究院提出基于贝叶斯理论评估软件测试可靠性过程中参数选取的依据^[3]。这些方法以软件的测试次数作为总体信息, 为基于输入域的软件测试可靠性评估提供了理论的依据, 但是这些软件测试可靠性的评估方法都比较复杂、计算过程比较繁琐、误差较大。

2 贝叶斯测试模型

贝叶斯理论可以粗略地描述为: 根据结果, 预测原因。贝叶斯理论的陈述如下^[4]: 样本空间为 S , A 为 E 的事件, B_1, B_2, \dots, B_n 为 S 的一个划分, 且 $P(B_i) > 0 (i=1, 2, \dots, n)$, 则贝叶斯公式为

$$P(B_i | A) = \frac{p(B_i)P(A/B_i)}{\sum_{j=1}^n p(B_j)P(A/B_j)} \quad i=1, 2, \dots, n \quad (1)$$

贝叶斯公式的意义在于: 设事件 A 已发生, 需要判断引起 A 发生的“原因”。如果已知 A 发生的可能“原因”共有 n 个: B_1, B_2, \dots, B_n , 且两两互不相容, 那么希望知道其中某个“ B_i ”的概率, 即条件概率 $p(B_i/A)$ 。在实际应用中, 往

往要求出每个 $p(B_i/A) (i=1, 2, \dots, n)$, 然后找出其中最大的一个 $p(B_i/A)$, 则 B_i 就是引起事件 A 发生的最可能的“原因”。

利用贝叶斯公式, 建立以下软件测试有效性的评估数学模型:

$$p(\bar{T} | S) = \frac{P(S | \bar{T}) \times P(\bar{T})}{P(S | T) \times P(T) + P(S | \bar{T}) \times P(\bar{T})} \quad (2)$$

其中, \bar{T} 表示软件存在缺陷; T 表示软件正确; \bar{S} 表示缺陷未被侦测出; S 表示软件缺陷被侦测出。

根据贝叶斯公式, 可以得到软件在测试环境中缺陷被侦测出的条件下, 软件存在缺陷的概率, 即 $p(\bar{T} | S)$, 也就是说, 软件在测试过程中根据经验知道式(2)右边的各个事件的概率值, 就可以求得 $p(\bar{T} | S)$, 而这些经验值在测试过程中存在随机性, 可以通过随机测试过程的数学特征来改善软件测试可靠性的评估。

3 服从二项分布的软件可靠性评估

软件测试是保证和提高软件质量的重要方法, 在采用一定的软件测试方法对软件测试之后, 可以利用软件的测试结果对软件的可靠性进行评估。软件的可靠性评估可以建立在概率统计以及数理推论基础上, 即利用软件测试的结果数据以及测试过程的信息对相应软件进行可靠性评估。

假设软件测试的结果为: 软件正确的概率 $P(T)=r$, 即 $r=t/n$, t 表示程序的输入域中所有使得程序正确运行的输入域的总和, 在随机测试过程中 t 代表测试用例成功通过测

基金项目: 广州市越秀区科技计划基金资助项目(2006-GX-028)

作者简介: 柴哲丽(1984-), 女, 硕士研究生, 主研方向: 软件测试, 软件工程; 林佳齐、朱金平, 硕士研究生; 唐 平, 教授

收稿日期: 2009-10-25 **E-mail:** zheli1201@sina.com

试的数量； n 表示程序总的输入域，随机测试中 n 表示测试用例总数。现在需要考虑的就是 r 的可靠性。按照贝叶斯统计推断原理，对 r 的可靠性的评估需要先确定 r 的先验分布。

软件正确的概率用 r 表示，在软件的正确性没有评估之前，软件的正确性在(0,1)之间有可能取任意值，那么 r 在(0,1)服从均匀分布(uniformly distribution)，则可以假设 r 的先验分布的概率密度函数为^[4]

$$f(r) = \begin{cases} 1 & 0 < r < 1 \\ 0 & \text{其他} \end{cases} \quad (3)$$

r 的后验分布指：假设总体样本有 n 条测试用例，其中有 t 条测试样本通过测试条件下的概率密度函数 $f(r|t)$ ，根据贝叶斯公式 $f(r|t)$ 可以表示为

$$f(r|t) = \frac{p(t|r)f(r)}{\int_0^1 p(t|r)f(r)dr} \quad (4)$$

式(4)中的 $p(t|r)$ 表示软件正确的概率为 r 的条件下， n 条测试用例中有 t 条用例通过测试的概率。随机测试过程是在输入域中随机选择输入，通过程序的运行检查程序中是否存在错误的过程，这一过程可以用随机试验表示^[4]。在每次测试中，对于选取的某一测试用例，程序在该输入下只有2种状态：即正确和不正确。

软件的正确性可以通过测试案例通过被测试软件输出的结果与软件需求分析的要求比较来判定：输入某一测试案例，输出的结果与该需求分析要求一致，则定义软件正确，否则软件错误。如果每次都是从所有的输入中选取一个测试用例，程序正确的概率为 r ，用 n 表示的实验输入的总次数，用随机变量 t 表示 n 次试验中程序正确的次数，则随机变量 t 服从二项分布，即在 n 次测试中，每次测试正确性概率为 r 时，程序正确运行次数为 t 的概率为

$$p(t|r) = \binom{n}{t} r^t (1-r)^{n-t} \quad (5)$$

将式(5)代入式(4)得：

$$f(r|t) = \frac{C_n^t r^t (1-r)^{n-t}}{\int_0^1 C_n^t r^t (1-r)^{n-t} dr} = \frac{\tau(n+2)}{\tau(t+1)\tau(n-t+1)} r^t (1-r)^{n-t} \quad (6)$$

式(6)表明软件正确性的后验概率服从参数为 $t+1$ 和 $n-t+1$ 的 β 分布，对于 β 分布参数的确定中国科学技术研究院做过详细的推导，过程比较繁琐，本文不考虑该处参数的确定，仅利用描绘函数图像和求函数值的软件(本文使用Matlab)直接获得 $f(r|t)$ 的函数图像和相应的函数值，得到函数的集中分布的区间。

软件正确性的后验概率分布函数 $F(r|t)$ 可以通过对式(6)积分获得：

$$F(r|t) = \int_0^r \frac{\tau(n+2)}{\tau(t+1)\tau(n-t+1)} x^t (1-x)^{n-t} dx \quad (7)$$

同样，可以利用数学软件获得 $F(r|t)$ 函数图像和函数值，软件正确性的概率可以直接由函数值获得。

软件的正确性 r 在(0,1)区间上的取值是连续的，对于连续型(continuous)的随机变量，它的概率分布函数和概率值有如下关系： $F(x) = p(x \leq x)$ ，则

$$p(x \leq x) = 1 - p(x > x) \quad (8)$$

通过式(8)，可以获得软件正确性 r 大于任何一个估计点的概率，也就可以得到评估软件可靠性的量化结果。

通常，在评估软件可靠性时，视软件的用途等条件会对该软件的可靠性有一个特定的要求，若软件可靠性评估的结

果优于这个特定的要求，就可以定义软件是可靠的，也就是说测试是有效的，否则，软件是不可靠的。

4 测试可靠性评估实例

近年来，汉语正越来越受到各国的青睐，在国外引起一股学习汉语的热潮，也出现了各样的中文学习平台，中文学习平台使用人数多，且使用者的文化层次多样，这就要求开发中文学习平台的同时，对其做大量且可靠的测试。本文以一个中文学习平台为例说明软件测试的可靠性评估过程。

4.1 依据经验计算测试可靠性

依据中文学习平台在测试过程中的经验可以大致估计： $P(T) = 0.90$ ， $P(\bar{T}) = 0.081$ ， $P(S|T) = 0.0207$ ， $P(S|\bar{T}) = 0.990$ 。根据贝叶斯公式，可得：

$$P(\bar{T}|S) = \frac{P(S|\bar{T}) \times P(\bar{T})}{P(S|T) \times P(T) + P(S|\bar{T}) \times P(\bar{T})} = 0.81 \quad (9)$$

即中文学习平台的测试过程能正确地测试缺陷(或可靠测试)的概率为0.81，该中文学习平台测试软件正确性为0.90的可靠性为0.81。

4.2 Matlab 计算软件正确性的概率分布密度

随机使用110个测试用例测试该中文学习平台，结果有100例的结果与中文学习平台的需求分析一致，即100条测试用例通过了测试。那么在110条测试用例中有100例成功通过测试的条件下，软件正确性的概率密度函数图像以及函数值可以通过Matlab工具获得。

根据测试的结果，把 $n=110$ ， $t=100$ 代入式(6)，利用Matlab的仿真结果 $f(r|t)$ 的图像如图1所示，其中，横轴表示 r 的取值，纵轴表示 $f(r|t)$ 的函数值，同时在Matlab界面可以得到在不同 r 的取值条件下 $f(r|t)$ 与自变量一一对应的关系。

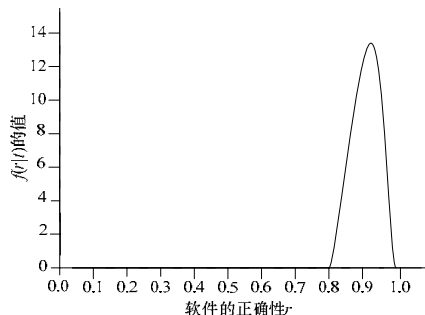


图1 当 $n=110, t=100$ 时 $f(r|t)$ 的取值

Matlab 求得的 $n=110$ ， $t=100$ 时， $f(r|t)$ 的部分取值如下(0.905 r 0.920)：

```
Columns 905 through 912
12.817 9    12.893 1    12.952 7    12.996 3    13.023 4
13.033 6    13.026 5    13.001 7
Columns 913 through 920
12.959 0    12.898 3    12.819 4    12.722 2    12.606 7
12.473 2    12.321 7    12.1524
```

通过Matlab画 $f(r|t)$ 函数和分析Matlab求得的 $f(r|t)$ 函数值的结果可知：当 $r=0.910$ 时， $f(r|t)$ 值最大，0.881 r 0.930 概率密度最集中(函数值大于10)。

4.3 Matlab 计算软件正确性的概率

在Matlab平台上画出函数的图像，并记录函数的自变量与函数值的对应的结果。软件正确性的后验概率分布的图像如图2所示。(下转第77页)