Markov模型的软件可靠性测试充分性问题的研究

雷 航[1]，马成功[2]

(1. 电子科技大学示范性软件学院 成都 610054; 2. 电子科技大学计算机科学与工程学院 成都 610054)

【摘要】在分析现有Markov模型的软件可靠性的充分性判别的基础上，定义了严格转移概率的概念，提出了基于严格转移概率的测试充分性判别方法。将Markov模型转化为严格Markov模型，在每个边或者状态的差异计算时引入严格转移概率，对可达相异转移集中所有可达相异转移k的转移概率进行原转移概率对应计算后进行求和再平均运算，得到的结果作为每个边或者状态的差异值。实验表明严格测试充分性判别方法比非严格Markov模型方法更稳定、有效。

关 键 词 Markov模型; 可靠性测试; 严格转移概率; 测试充分性

中图分类号 TP311.5 文献标识码 A doi:10.3969/j.issn.1001-0548.2010.01.023

Testing Adequacy of Software Reliability in Markov Model

LEI Hang[1] and MA Cheng-gong[2]

(1. School of Software, University of Electronic Science and Technology of China Chengdu 610054;
2. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract By analyzing the original testing adequacy determination in Markov model, the strict transfer probability is defined and then testing adequacy based on strict transfer probability is advanced. Markov model is transformed to strict Markov model, and the strict transfer probability is used to calculate the differentiation of every edge. The primal transfer probability of all of accessibility discriminative transfer probability is computed for accessibility discriminative transfer set, and then the probability is obtained. Experiments declares that new method is more stabile and efficient.

Key words Markov model; reliability test; strict transfer probability; testing adequacy

软件测试的技术多种多样，但都要解决同一个问题，即测试进行到什么程度停止。

作为理想的情况，测试应该进行到找出并排出程序中的所有错误为止，但这是不现实的。实际的情况是抽取输入域中的有限子集作为测试的输入集，并根据测试结果推断程序的正确性或可靠性。这是研究测试充分性的根本意义。

软件测试理论研究的中心问题是软件测试充分性问题，即如何得到一个测试充分性准则，使得在此准则基础上得到的测试数据能对软件进行充分的测试。软件测试的充分性是指软件在有限测试数据时的表现能够代表软件在所有输入数据上的表现的性质[1]，其度量是充分程度的定量表示，也是测试执行程度的定量描述。

软件可靠性测试是提高软件可靠性、定量评定可靠性水平的关键技术，也面临测试充分性问题。该问题是软件可靠性的关键问题和难点问题之一。

1 软件可靠性测试的测试充分性

1.1 软件可靠性测试的特点

为了满足用户对软件的可靠性要求、评价软件可靠性水平及验证软件产品是否达到可靠性要求，软件可靠性测试是一个最有效的途径[2]。

1983年，IEEE计算机学会软件工程技术委员会对软件可靠性的定义如下：

(1) 在规定的条件下，在规定的时间内，软件不引起系统失效的概率，该概率是系统输入和系统使用的函数，也是软件中存在的错误的函数；系统输入将确定是否会遇到已存在的错误(如果错误存在的话)。

(2) 在规定的实践周期内，在所述条件下程序执行所要求的功能的能力。

定义中提到的“规定的条件”和“规定的时间”就是软件使用环境的表述。因此，可靠性测试的主要

收稿日期：2008 – 07 – 15; 修回日期：2008 – 10 – 20
作者简介：雷 航(1960 –)，男，教授，博士生导师，主要从事实时软件工程和软件可靠性工程等方面的研究.

特点就是按照实际的软件使用环境测试软件[3]。

1.2 软件可靠性测试的测试充分性

软件可靠性测试是一种面向使用的测试。程序存在错误是必然的，并不需要将所有的错误都找出来。制约测试的根本因素是能花费多少人力、物力和时间资源。资源太少而测试不充分，也可能资源充足使得测试过头，导致资源浪费。不充分的测试是不负责任的；过分的测试浪费资源，也是不负责任的。针对用户使用得多的、易出错的功能就多测，集中力量先发现对可靠性影响大的错误。这样就意味着软件可靠性测试的目的和出发点不是多发现错误，而是多发现对可靠性影响大的错误。

可靠性测试的测试充分性描述如下[4]：

OK(d)定义为一个谓词，表示对于输入域D中的点d，当在该点程序的输出值$F(d)$等于功能规范S定义的预期值$S(d)$时，有OK(d)成立，即$F(d)=S(d)\Rightarrow$ OK(d)。

Reliable(T)也同样定义为一个谓词，表示测试数据集合T是输入域D的一个子集，并且通过T的测试结果得到的软件可靠度的评估值Reval与软件真实的可靠度Rreal，满足 $\lim_{\|T\|\to\infty} |\text{Reval}-\text{Rreal}|\to 0$，并且Reval ≧ Rreq，其中Rreq为要求软件达到的可靠性的最小值。

理想的可靠性测试的充分测试集合可以描述为具有如下性质的测试集合$T\subseteq D$，T的统计特征与D的统计特征一致，Reliable(T)$\Rightarrow$ Reliable(D)成立。即通过程序在该有限测试集合上的可靠能够代表其在整个输入域上的可靠。

1.3 研究现状

目前，可靠性测试方法主要有基于使用模型的统计测试(Markov模型是最主要的使用模型表达方式)和基于操作剖面的可靠性测试[5]。可靠性测试的测试充分性也主要是在这两种方法中进行研究。文献[4]借鉴已有的软件测试充分性方面的理论，对软件可靠性测试的充分性问题进行了研究尝试。文献[6-7]则分别提出了不同的基于Markov模型的测试充分性表述。在已有的研究中，基于Markov模型的可靠性测试充分性研究相对多一些[8]。本文研究Markov模型的可靠性测试充分性问题。

2 Markov模型的可靠性测试充分性

2.1 Markov模型

使用模型是软件使用过程中软件形态的精确刻划，它把软件的使用方式以模型的方法表示出来，描述软件的使用特性。

Markov模型用Markov过程来描述软件的使用模型。在Markov模型中，使用模型由状态和边组成。状态表示软件使用过程中的内部环境，边表示状态间的转移关系。每条边都有一个激励输入与之对应，表明在当前状态下输入这种激励使软件转移到下一个状态。每条边都有一个转移概率，转移概率标志状态转移发生的可能性。特定状态的所有退出边的转移概率之和应该为1。每一个Markov模型都有唯一的初态和终态。初态是Markov模型的初始状态，它是每一次软件使用的开始；终态是Markov模型的终止状态，它是软件每一次使用的终结。软件的每一次使用或者说每一次操作都从初态开始，经过若干个中间状态，最后到达终态。测试用例就是从初态到终态的一系列状态和边的序列[9]。

图 1 是一个简单的 Markov 模型。

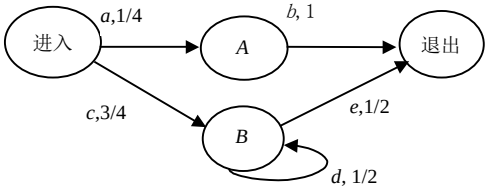

图 1 一个 Markov 模型

2.2 测试充分性的量化

在Markov模型中，测试的充分性是通过测试过程中的使用链和测试链的比较来衡量的。测试过程中，会产生使用链和测试链。测试之初，Markov模型被称为使用链，这是相对测试链来说的。测试链是从使用链产生的，它把使用链各边所对应的转移概率替换为一个处置为0的计数器。随着测试的进行，每当一个测试用例经过该边时计数器就加1，根据每一条边计数器的值计算出该边的相对转移概率就形成了测试链。使用链代表使用环境，测试链代表测试环境。当使用环境和测试环境的差异足够小时，可靠性测试就充分了，此时从测试环境中计算出的软件可靠性可以代表实际使用时的软件可靠性。

现有的基于Markov模型的可靠性测试充分性理论，通过计算欧氏距离(Euclidean Distance[6]) 和Discriminant差值[7]对使用链和测试链的差异进行定量分析。

2.2.1 欧氏距离

使用链和测试链归根结底是两个具有不同转移概率的有向图，因此可以用两个有向图的欧氏距离来比较使用链和测试链的差异程度，其计算公式为：

$$\text{Euclidean Distance} = \sqrt{\sum_{i,j}(u_{i,j} - t_{i,j})^2} \quad (1)$$

式中　$u_{i,j}$和$t_{i,j}$分别为使用链和测试链中从状态i到状态j的转移概率(下同)。Euclidean Distance不是一种精确可靠的计算方法[5]。

2.2.2　Discriminant差值分析

使用链和测试链的比较还可以通过Discriminant差值来进行，它是两个随机过程似然度的期望值，其计算公式为：

$$D(U,T) = \sum_{i=1}^{u}\pi_i \sum_{j=1}^{u} u_{i,j} \lg(u_{i,j} / t_{i,j}) \quad (2)$$

式中　U为使用链；T为测试链；π_i为状态i长时间运行中的占有率，即长时间运行中各状态所占有的比例。Discriminant Value可以提供比较精确的结果，但它并不总是可计算的。在测试中，只有在使用模型的所有边都覆盖以后，它才有意义。因此就产生了以下的变体计算方法，使得在任何情况下D(U, T)都是可计算的：

$$\hat{\mathrm{D}}(U,T) = \sum_{i=1}^{u}\pi_i \sum_{j=1}^{u} u_{i,j} \lg\left[\frac{u_{i,j}}{\varepsilon - \varepsilon(\mathrm{sgn}(t_{i,j})) + t_{i,j}}\right] \quad (3)$$

式中　ε为一个很小的正数；sgn(x)为符号函数。当x=0时，sgn(x)=0；当x<0时，sgn(x)=−1；当x>0时，sgn(x)=1。显然，当所有边都覆盖以后，$D(U, T)$=$\hat{D}(U, T)$。

2.3　测试充分性度量方法的分析

考虑如图2所示的Markov模型片断。

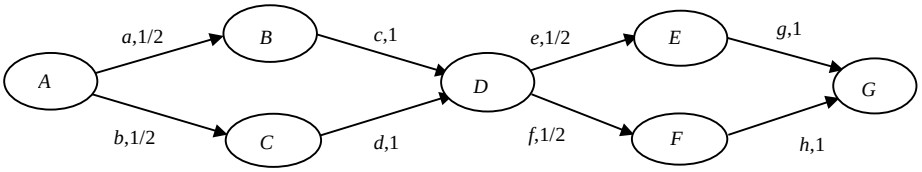

图2　Markov模型片段

假设根据图2的使用链生成了相同数目的测试用例*ABDEG*和*ACDFG*片段。此时，计算测试充分性，有Euclidean Distance和Discriminant差值均为0，即测试已充分。但事实上测试用例*ABCFG*和*ACDEG*片段遗漏，测试是不充分的。本文称这种现象为“早熟”。

进一步分析图2、Euclidean Distance和Discriminant差值的计算公式，可以得到Markov模型中转移概率约束不严格导致“早熟”的结论。在Markov模型中，转移概率仅仅约束了此转移的头状态的概率分布，却没有约束此转移的所有可达相异转移的概率分布。以图2中的转移*DE*为例，其转移概率为1/2，表示由状态*D*到达*E*的转移*DE*概率为1/2，但没有对其可达相异转移*ABD*和*ACD*到达状态*E*的概率均为1/2做出约束。正是由于这个原因，导致了上述测试用例遗漏却得到测试充分结论的悖论。为了解决这个问题，本文给出了严格转移概率的定义，且基于这一定义，给出了Strict Euclidean Distance和Strict Discriminant差值的计算方法。

3　严格转移概率和软件可靠性测试充分性

3.1　严格转移概率

严格转移概率不仅表示了状态转移的概率分布，也表示了所有可达相异转移的概率分布。

以图2为例，如果$n(X)$表示状态X或边X的访问次数，$p(X)$表示边的转移概率，则有$n(D)$=$n(ABD)$+$n(ACD)$、$n(DE)$=$n(ABDE)$+$n(ACDE)$，这是访问次数的转移和可达相异转移的关系式。在Markov模型中$p(DE)$=1/2，在严格转移概率定义中则有$p(DE|ABD)$=$p(DE|ACD)$=1/2。由于$p(DE|ABD)$=$p(DE|ACD)$=1/2是$p(DE)$=1/2的充分非必要条件，即$p(DE)$=1/2成立不能保证$p(DE|ABD)$=$p(DE|ACD)$=1/2成立，但$p(DE|ABD)$=$p(DE|ACD)$=1/2成立就有$p(DE)$=1/2成立，所以，这样的转移概率被称为严格转移概率。支持严格转移概率的Markov模型称为严格Markov模型，由严格转移概率表述的测试充分性称为严格测试充分性。

3.2　严格测试充分性

严格转移概率的定义是为了解决现有的测试充分性判别方法的不足。

分析充分性测试计算公式，可以看出公式的计算是对每个边或者状态的差异进行运算后求和而得到整个测试链的测试充分性特征。由于$n(i)$=$\sum_{k=1}^{U} n(k)$，式中$n(k)$表示状态k的访问次数，U为能到达状态i的可达相异转移集，$k \in U$。因此，在每个边或者状态的差异计算时引入严格转移概率，通过将集合U中所有可达相异转移k的转移概率$u_{k,i,j}$进行原转移概率$u_{i,j}$对应计算后求和再平均运算，得到的结果作为每个边或者状态的差异值，最后进行求和得

到整个测试链的测试充分性。

严格测试充分性计算公式为：

$$\text{Euclidean Distance}=\sqrt{\sum_{i,j}\frac{1}{n}\sum_{k=1}^{n}(u_{k,i,j}-t_{k,i,j})^2} \quad (4)$$

$$D(U,T)=\sum_{i=1}^{u}\pi_i\sum_{j=1}^{u}\frac{1}{n}\sum_{k=1}^{n}u_{k,i,j}\lg(\frac{u_{k,i,j}}{t_{k,i,j}}) \quad (5)$$

$$\hat{D}(U,T)=\sum_{i=1}^{u}\pi_i\sum_{j=1}^{u}\frac{1}{n}\sum_{k=1}^{n}u_{k,i,j}\lg(\frac{u_{k,i,j}}{\varepsilon-\varepsilon(\text{sgn}(t_{k,i,j}))+t_{k,i,j}}) \quad (6)$$

式中 n表示集合U的势；$u_{k,i,j}$表示使用链中经由可达相异转移k到达状态i且转移至j的严格转移概率；$t_{k,i,j}$表示测试链中经由转移k到达状态i处转移j的严格转移概率。

4 实验分析

4.1 实现过程

在 Markov 模型的可靠性测试中实现严格测试充分性判别有以下 3 个过程：

(1) 生成严格 Markov 模型。

严格 Markov 模型由 Markov 模型生成，其全部的工作就是将 Markov 模型中所有的转移概率转换成严格转移概率。这是实现严格测试充分性判别的基础，也是难点。

完全的严格 Markov 模型根据所有状态的完整可达相异转移集进行严格转移概率的生成。所谓完整可达相异转移集就是所有的可达相异转移的集合。随着模型复杂度的增加，完全严格 Markov 模型生成的复杂度将呈超线性的增长。

非完全的严格 Markov 模型则根据所有状态的可达相异转移集进行严格转移概率的生成。相比于完全的严格 Markov 模型，非完全的严格 Markov 模型不要求状态可达相异转移集的完整性，降低了模型生成的复杂度，同时在一定程度上降低了测试充分性判别的精确度。另一方面，它也要求所有 Markov 模型中可达相异转移集的势大于 1 的状态在非完全的严格 Markov 模型中其可达相异转移集必须大于 1。这是保证生成的模型为严格 Markov 模型的最基本的要求。

(2) 生成测试链。

根据可靠性测试生成算法生成测试链。现有的生成算法有随机生成、模拟退火算法[10]、遗传算法[11]等。本文主要研究测试充分性的判别，生成算法就不多述。

(3) 计算严格测试充分性。

根据严格测试充分性的计算结果判别测试链的充分性。

4.2 结果分析

通过实现非完全的严格Markov模型，采用式(1)、式(2)、式(4)和式(6)对严格测试充分性判别进行实现。实现时要考虑式(2)和式(6)中ε的取值。ε是一个很小的正数，其意义在于使式(2)和式(6)在任何情况下可用。当$\varepsilon \ll \min(t_{i,j})$或$\varepsilon \ll \min(t_{k,i,j})$时，$\varepsilon$对计算结果产生的影响可以忽略。一方面，当$\varepsilon \ll \min(t_{i,j})$或$\varepsilon \ll \min(t_{k,i,j})$时，$\varepsilon$对计算结果精确度的影响可以忽略。当然，$\varepsilon$越小，计算结果越精确。另一方面，当选定一个很小的正数ε后，计算结果仍然会如实地反映出随测试链变化的测试充分性的变化方向。

将图 2 中的状态 *A* 和 *G* 分别作为状态进入和退出，作为实验 Markov 模型，且ε取值 0.000 000 1 (0.000 000 1 ≪ 0.5 成立)。其实验结果如表 1 所示。

表1 图2的测试充分性判别

增加的测试用例	欧氏距离	Discriminant 差值	严格欧氏距离	严格 Discriminant 差值
空	2.236 067 977 499 79	0	2.236 067 977 499 79	0
ABDEG	1.732 050 807 568 88	1.599 485 002 168 01	1.732 050 807 568 88	2.036 985 002 168 01
ACDFG	0	0	0.707 106 781 186 55	0.799 742 501 084 01
ABDFG	0.333 333 333 333 33	0.012 788 130 611 85	0.552 770 798 392 57	0.406 265 315 847 93
ACDEG	0	0	0	0

初始时，测试链为空。对所有边有 $t_{i,j}=0$ 和 $t_{k,i,j}=0$ 成立。此时，Euclidean Distance 和严格 Euclidean Distance 相等。对所有状态，有 π_i=0 成立。此时 Discriminant Value=0，严格 Discriminant Value=0。当测试链为(*ABDEG*, *ACDFG*)时，Euclidean Distance 和严格 Euclidean Distance 出现了“早熟”现象，而严格 Euclidean Distance 和严格 Discriminant Value 则很好地反映了此时的测试充分性水平。当测试链为(*ABDEG*、*ACEFG*、*ABDFG*、*ACDEG*)时，测试链和使用链完全吻合，测试充分。

纵观整个测试过程，Euclidean Distance 和严格 Euclidean Distance 出现了“早熟”现象，而严格 Euclidean Distance 和严格 Discriminant Value 则很好地度量了测试充分性。可见，严格的测试充分性判别更加稳定和有效。

5 结 束 语

Markov 模型是进行可靠性测试最主要的测试模型，对其进行可靠性测试充分性的研究具有重大的意义。本文通过对原模型中测试充分性的分析，定义了严格转移概率，将 Markov 模型转化为严格 Markov 模型，进行测试充分性的优化，最后进行了验证。

参 考 文 献

[1] GOODENOUGH J B, GERHART S L. Toward a theory of test data selection[C]//Proceedings of the International Conference on Reliable Software. Los Angeles, California: ACM, 1975.

[2] LYU M R. 软件可靠性工程手册[M]. 刘喜成, 钟婉懿, 译. 北京: 电子工业出版社, 1996.
LYU M R. Handbook of software reliability engineering[M]. Translated by LIU Xi-cheng, ZHONG Wan-yi. Beijing: Publishing House of Electronics Industry, 1996.

[3] 周新蕾, 缭峥红. 安全性关键软件的可靠性分析[J]. 载人航天, 2005, (6): 54-57.
ZHOU Xin-le, LIAO Zheng-hong. Reliability analyais for saftey-critical software[J]. Manned Space Flight, 2005, (6): 54-57.

[4]李秋英, 阮　镰, 刘　斌. 软件可靠性测试充分性研究[J]. 测控技术, 2003, 22(11): 49-52.
LI Qiu-ying, RUAN Lian, LIU Bin. Research on software reliability testing adequacy[J]. Measurement & Control Technology, 2003, 22(11): 49-52.

[5] JOHN D M. Software reliability engineering[M]. New York: The McGraw-Hill Companies, Inc, 1999.

[6] KIRK S. Improved techniques for software testing based on Markov chain usage models[D]. Knoxiville: University of Tennessee, 1999.

[7] JAMES A W, MICHAEL G T. A Markov chain model for statistical software testing[J]. IEEE Trans on software Engineering, 1994, 20(10): 812-824.

[8] 颜　炯, 王　戟, 陈火旺. 基于模型的软件测试综述[J]. 计算机科学, 2004, 31(2): 184-187.
YAN Jiong, WANG Ji, CHEN Huo-wang. Survey of model-based software testing[J]. Computer Science, 2004, 31(2), 184-187.

[9] STACY J P, CARMEN J T, RICHARD C L, 等. 净室软件工程: 技术与过程[M]. 贲可荣, 张志祥, 张秀山, 等译. 北京: 电子工业出版社, 2001.
STACY J P, CARMEN J T, RICHARD C L, et al. Clean room software engineering: Technology and process[M]. Translated by BEN Ke-rong, ZHANG Zhi-xiang, ZHANG Xiu-shan, et al. Beijing: Publishing House of Electronics Industry, 2001.

[10] 江加和, 宋子善, 沈为群, 等. 模拟退火算法在连续变量全局优化问题中应有[J]. 北京航空航天大学学报, 2001, 27(5): 556-559.
JIANG Jia-he, SONG Zi-shan, SHEN Wei-qun. Application of simulated annealing to global optimization problems with continuous variables[J]. Journal of Beijing University of Aeronautics and Astronautics, 2001, 27(5): 556-559.

[11] 周献中, 孙勇成, 江金龙. 基于使用模型和遗传算法的测试数据自动产生技术[J]. 兵工学报, 2006, 27(6): 1051-1055.
ZHOU Xian-zhong , SUN Yong-cheng , JIANG Jin-long. Automatic test data generation based on usage model and genetic algorithm[J]. Acta ARMAMENTAR, 2006, 27(6): 1051-1055.

编　辑　税　红