基于可用性的一致性测试套选择

邢 熠[1,2,3] 叶新铭[2] 谢高岗[1]
(中国科学院计算技术研究所 北京 100190)[1] (内蒙古大学计算机学院 呼和浩特 010021)[2]
(中国科学院研究生院 北京 100190)[3]

摘 要 一致性测试套选择是一致性测试中非常重要的阶段。测试生成的测试套集受到测试成本、覆盖度、可用性等因素的影响,只有经过测试选择后得到的测试套子集才可以用于实际的一致性测试。在测试选择研究的基础上,根据可用性标准,提出了一种基于统计检验的测试选择方法。该方法把可用性标准转化为假设检验模型,使用形式化方法构造了测试例执行的接受域;然后对测试例执行进行采样并且统计成功的次数;如果统计结果落入接受域,就认为该测试例是可用的。该方法取得了非常好的应用效果。

关键词 一致性测试,测试选择,接受域,采样,可用性
中图法分类号 TP306.2 **文献标识码** A

Conformance Test Selection Based on Availability

XING Yi[1,2,3] YE Xin-ming[2] XIE Gao-gang[1]
(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)[1]
(Inner Mongolia University, Hohhot 010021, China)[2] (Graduate University of Chinese Academy of Sciences, Beijing 100190, China)[3]

Abstract Conformance test selection is a very important stage in conformance test. Test suite from test generation can't be suitable for practical test because of test cost, coverage, availability etc. Test selection is necessary to generate subset of test suite. We gave a method to carry out test selection based on statistical test. It included three stages. The first was to construct acceptance region using formal method. The second step was to sample test execution and count the number of success. The last was to judge the availability of test case based on whether the number of success lie in acceptance region. The method has been used in our test selection application.

Keywords Conformance test, Test selection, Acceptance region, Sample, Availability

1 背景

一致性测试生成算法所产生的抽象测试套,经过数据选择后就转化为可执行测试套。它的数量是非常大的,甚至是无限的,这意味着执行所有测试例是不可能与不可行的。需要在测试套中选择合适的测试例子集,从而形成测试套的最小子集,以满足错误覆盖要求,便于实际执行。

现有的测试选择方法可以从语义模型进行分类:一种基于状态机,一种以标记变迁系统为模型。在标记变迁系统模型基础上,文献[1]中给出了一种测试选择框架。该框架对测试套构造两个参数:价值和成本。价值反映测试套的测试能力,成本反映选择的代价。根据这两个参数,测试选择的标准就转化为具有最低成本同时具有最高价值的测试套子集。但是该框架并没有给出如何获得价值和成本的方法,只是一个指导性框架。文献[2]提出了一种基于PICS信息进行TTCN[3-9]描述的测试套选择算法。但是该算法忽略了协议规范的模型要求,而只是从实现规范说明进行选择,缺乏准确性。而在状态机模型上,文献[10]使用部分w方法产生了完全错误检测和长度更短的测试套;而文献[11]使用通用wp方法在非确定性状态机上生成完全错误覆盖和长度更小的测试套。

经过上述测试选择方法所生成的测试套在实际执行时仍然有可能失败和有可能成功。这种现象可以归结为测试执行的随机性。但是已有的测试选择都忽略了测试套随机执行时所造成的不可用性。如何评价这些不可用的测试例集,就是本研究的重点。

一致性测试是通过在被测实现上执行测试例,同时进行观察,最后根据观察的结果来判断是否满足一致性关系。在上述过程中,一般假定测试的执行过程中每一步都可以正确地执行。但是实际进行测试时,测试器与被测实现通过一定的通讯介质进行通讯,通讯的质量可能影响测试的执行。测试器若发送了正确编码的消息,但是由于通讯介质的问题,可能会成功发送到对端,也可能会失败,不能发送到对端。这说明测试的每一步执行都是一个随机行为,以一定的概率成功

到稿日期:2009-01-15 返修日期:2009-03-25 本文受国家自然科学基金项目(90604015)(60863015),内蒙古自然基金重点项目(200711020803),教育部春晖项目(Z200-1-01032)资助。

邢 熠(1971-),博士研究生,主要研究方向为网络协议的测试,E-mail:csxingyi@imu.edu.cn;叶新铭(1943-),教授,博士生导师,主要研究方向为网络协议形式验证与测试;谢高岗(1974-),研究员,CCF高级会员,主要研究方向为对等计算、网络测量与服务质量。

或以一定的概率失败。整个测试执行就是一个随机过程。

文献[12,13]讨论了 ISO96 中测试执行的概率框架，主要包括 3 个方面：一个是假定测试执行的结果满足一定的概率分布，从而建立测试套接受或拒绝实现的概率；再一个方面是对被测实现赋予权值，以表现被测实现错误的严重程度；最后是被测实现的概率分布。上述 3 个方面的结合最终形成了归一化的测度指标，它可以量化测试套检测错误实现的程度和测试套接受正确实现的程度，同时可以进行测试选择的指导。只选择那些可以检测大部分具有严重错误实现的测试套，这些测试套同时也能接受大多数正确实现。

ISO96 中对一致性测试进行了很多假设。第一个是测试实现假设，测试实现可以建模为 $i_{IUT} \in MODS$。第二个假设是关于测试执行函数 exec。测试执行通常由很多测试运行组成。由于被测实现的不确定性，测试执行就会形成不同的观察结果。测试执行具有有限的特点，因此不可能获得所有的观察结果。第三，假设测试例都是被正确实现的，这样测试执行就可以获得正确的观察结果。

文献[12,13]中并没有改变上述假设，而是把测试运行看作一个随机过程，运行结果是个随机变量，被测实现仍然是确定的 LTS 模型。这种随机特点实际是由被测实现引起的，可能是被测实现进行了随机选择或非确定性选择。这就要求测试器具有更强的观察能力，不仅可以观察确定的行为，而且可以观察行为的概率分布。

测试器需要具有一个 reset 按钮，可以从初始状态重新运行测试器。这样，观察者就可以记录测试器多次运行的执行迹。这些执行可以假定是独立进行的，即一次执行的随机选择与下一次执行选择是不相关的。因此，每次运行都取决于一个可能不同的随机分布。

我们的工作是在概率自动机 PA(probabilistic automata)的框架下进行的。PA[14]扩展了 LTS 模型，使之具有离散概率分布特性。该模型也应用在分析分布式算法中[15-18]。在 PA 基础上，定义了执行路径的概率分布(即概率执行)和执行迹上的概率分布(即迹分布)。使用统计的方法来验证测试例是否满足迹分布所确定的接受域。

文献[19-23]提出了几种基于随机代数的测试前序和测试等价关系，所有这些都是在文献[24]中提出的框架下进行的。即测试被定义为与规范或实现进行交互的随机过程，并且返回成功或失败，通过比较成功的概率就可以得到各种不同的测试关系。但是这些研究并没有描述如何从外部观察到成功的概率。我们的方法更接近于文献[25,26]，其中随机互模拟关系通过基于假设检验的测试场景来刻画。相近的方法也出现在统计模型验证中[27-30]。通常情况下，概率模型验证器需要遍历状态空间，同时计算所有相关的概率；而统计模型检验的思想是收集模型的运行样本。感兴趣的性质被表述为测试假设，通过增加样本运行的数量就可以产生对所求问题产生错误应答的概率。目前统计模型检验主要用于离散和连续时间马尔可夫链[2,29]、半马尔可夫过程[29]、随机离散事件系统[27,30]。相关的研究也出现在随机系统研究中[31-33]。应更多地关注系统的行为方面和状态空间的概率分布，这些分布以事件的出现为条件。

本文讨论了基于可用性的一致性测试选择问题。首先介绍了相关的研究背景。然后给出了概率自动机模型，在此模型基础上推出了行为迹空间的概率分布；通过假设检验的方法给出了测试例的接受域，从而测试选择问题转化为接受域的判定问题。最后通过实例说明了测试选择的方法和今后的研究方向。

2 基本概念

2.1 测度空间与概率空间

一个由研究对象全体构成的集合称为样本空间(sample space)，记为 Ω。Ω 中的元素称为样本点，记为 ω。

定义 1 空间 Ω 上的一个非空子集类 $\mathcal{F}$ 若满足条件：

1) $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$，

2) $A_1 \in \mathcal{F}, A_2 \in \mathcal{F} \Rightarrow A_2 \cup A_2 \in \mathcal{F}$，

则称 $\mathcal{F}$ 为一个域(field)，又称为一个代数(algebra)。

定义 2 设 $\mathcal{F}$ 为 Ω 上的一个域，且满足 $A_j \in \mathcal{F}, j=1,2,\cdots \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$，则称 $\mathcal{F}$ 为一个 σ 域或 σ 代数。σ 域中的元素又称为事件(event)。

定理 1 设 S 是 Ω 上的任一子集簇，则存在 Ω 上的最小 σ 域 $\mathcal{F}$，使得 $S \subseteq \mathcal{F}$。我们称 $\mathcal{F}$ 由 S 生成。

定义 3 空间 Ω 及其上的一个 σ 域 $\mathcal{F}$ 构成一个可测空间(measurable space)，记作 $\langle \Omega, \mathcal{F} \rangle$。

定义 4 设 $\langle \Omega, \mathcal{F} \rangle$ 为可测空间，μ 为定义在 $\mathcal{F}$ 上的非负函数，满足

1) $\mu(\emptyset)=0$，

2) 若 $A_n \in \mathcal{F}, n=1,2,\cdots$，且 $A_n A_m = \emptyset, n \neq m$，则 $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} (A_n)$，则称 μ 为 $\mathcal{F}$ 上的测度，称 $\langle \Omega, \mathcal{F}, \mu \rangle$ 为测度空间(measure space)。

定义 5 若测度空间 $\langle \Omega, \mathcal{F}, \mu \rangle$ 满足 $P(\Omega)=1$，则称该测度空间为概率空间(probability space)，称 P 为 $\langle \Omega, \mathcal{F}, \rangle$ 上的概率测度，简称概率。

2.2 随机变量

随机变量是用来将随机现象进行量化描述的一个数学概念，它实质上是从一个概率空间到另一个可测空间的映射，即可测函数。

定义 6 设 Ω 是一个样本空间，X 是定义在 Ω 上的实函数，即对任一样本点 $\omega \in \Omega$，$X(\omega)$ 为一实数，则称 X 为一个随机变量(random variable)。

随机变量一般用大写字母来表示，如 X,Y,Z 等。随机变量的取值一般用小写字母表示，如 x,y,z 等。根据随机变量取值的可能性，可以将随机变量分为 3 种类型：离散型随机变量、连续型随机变量和混合型随机变量。离散型随机变量只能在有限或可数无穷多个实数点上取值；连续型随机变量只能在一个或多个非退化的实数区间连续取值，在单个离散点上的取值为零；混合型随机变量则在某些离散点上取值大于零，而在其他地方是连续取值。本讨论仅涉及离散型随机变量。

离散型随机变量通常用它的分布律来表示。设离散型随机变量 X 所有可能的取值为 $\{x_k, k=1,2,\cdots\}$，X 取各个可能值的概率为 $P\{X=x_k\}=p_k, k=1,2,\cdots$。$\Omega$ 上所有离散随机分布的集合记为 DIstr(Ω)。若 $\sum_k p_k \leqslant 1$，则称该概率分布为子概率分布，所有子概率分布记为 SubDistr(Ω)。

定义 7 随机变量 X 的数学期望 $E[X]$：

$E[X]=\sum_k x_k p_k$

定义 8 随机变量 X 的方差 $\mathrm{Var}[X]$：

$\mathrm{Var}[X]=E[(X-E[X])^2]=E[X^2]-E[X]^2$

2.3 统计与假设检验

研究对象的全体所构成的集合称为总体，总体中某些元素的集合称为样本。根据不同的获取方法，样本分为方便样本和概率样本。本文只考虑随机概率样本。

定义 9 容量为 n 的随机样本是指一组 n 个独立同分布的随机变量序列。

定义 10 一个统计量是将样本空间中的样本点映射到实数上的函数，其中样本空间中的样本点是一些多元随机变量的所有可能值。换句话说，统计量就是几个随机变量(如样本均值、样本方差)的函数。

统计量的一个基本用处是估计总体的性质。用来做估计的统计量自然叫做估计量。样本均值、方差可以作为总体均值、方差的估计量。一个好估计量的标准是无偏性。

假设检验是根据样本来推断总体的一些给定陈述是否成立的过程，这些陈述称为假设。可以分为统计假设检验和非统计假设检验。本文主要使用统计假设检验。

定义 11 临界域(critical region)是样本空间中使得拒绝零假设全体样本点的集合，有时临界域亦称为拒绝域(rejection region)。所以很明显，样本空间中不在临界域的全体样本点的集合称为接受域(acceptance region)。

在假设检验中有可能做出两种类型的错误判决。如果零假设为真，而我们错误地拒绝了它，那么所犯的错误是第一种错误，亦称第一类错误(type Ⅰ error)。也就是说，当 H_0 为真，而我们试验的结果却落在临界域内时，即发生了第一类错误。

定义 12 第一类错误是拒绝了正确零假设的错误。

假设检验中的另外一类错误是指当零假设为假时，却接受了零假设，这类错误是第二种错误，亦称第二类错误(type Ⅱ error)。

定义 13 第二类错误是接受了不正确零假设的错误。

这两类错误可以和一定的犯错误概率联系在一起。首先考虑犯第一类错误的概率。

定义 14 显著性水平(level of significance)α 是拒绝正确零假设的最大概率。

在统计假设检验中，了解在零假设成立时检验统计量的概率分布是非常必要的，这称为检验统计量的零分布(null distribution)。

定义 15 检验统计量的零分布是当零假设成立时检验统计量的概率分布。每个统计假设检验的显著型水平 α 都可以由检验统计量的零分布得到。显著性水平有时亦称为临界域的大小。若 H_0 成立，拒绝 H_0 的最大概率是 α，则接受 H_0 的最小概率是 $1-\alpha$。

犯第二类错的概率用 β 表示。显然，在假设检验中我们希望 α 和 β 都接近于零。在实际应用中，样本容量可以帮助我们决定 α 和 β 会有多小。只有当样本包含了总体所有的信息时，犯错误的可能性才可能被完全消除。

3 概率自动机

3.1 概率自动机

由于测试执行具有随机性的特点，我们引入概率自动机对该随机过程进行建模。

定义 16 概率自动机 PA 是四元组 $\mathscr{A}=(S,\Sigma,T,s^0)$ 其中，S：非空状态集；Σ：可观察动作集；$T\subseteq S\times\Sigma\times\mathrm{Distr}(S)$：概率变迁关系集；$s^0\in S$：初始状态。

为了讨论方便，动作集 L 只包含外部可观察动作，没有内部动作。而且 L 是可数的集合，$\Sigma=\{a_i\mid i\in N\}$。$\Sigma_l$ 代表动作序列 $a_0,\cdots,a_{l-1}$。有限迹的集合记为 $\Sigma^{<\omega}$，无限迹的集合记为 Σ^{ω}，所有迹的集合记为 $\Sigma^{\leqslant\omega}$，ε 表示空迹。

记 $S\xrightarrow{\alpha}\mu$ 表示 $(s,a,\mu)\in T$ 从状态 s 经过动作变迁 a 到达状态分布 μ。$S\xrightarrow{a,\mu}t$ 表示 $S\xrightarrow{\alpha}\mu$ 并且 $\mu(t)>0$。有时使用 $S_{\mathscr{A}},\Sigma_{\mathscr{A}},T_{\mathscr{A}},S^0_{\mathscr{A}}$ 来表示概率机 $\mathscr{A}$ 的组成部分。

一般来说，概率变迁关系 T 中的目标状态分布就是一种概率分支，即 $S\xrightarrow{a,\mu}t$ 表示一种非确定的变迁 $s\xrightarrow{\alpha}u$ 与概率变迁 $\mu\xrightarrow{\alpha}t$ 的综合，但我们经常会忽略这种概率变迁，而只表示非确定变迁。这样概率机就具有非确定性的性质。

定义 17 概率机 $\mathscr{A}$ 的路径是满足下列条件的有限或无限序列 $\pi=s_0a_1\mu_1s_1a_2\mu_2s_2\cdots$。其中，

- s_i 代表状态，a_i 代表动作，μ_i 代表状态上的概率分布；
- s_0 表示初始状态；
- 若路径 π 有限，那么 π 结束于状态；
- $S_i\xrightarrow{a_{i+1},\mu_{i+1}}S_{i+1}$，对于每个非终止的 i。

有限路径 π 的长度是路径上发生变迁的数量。$\mathscr{A}$ 上所有路径的集合记为 $\mathrm{Path}(\mathscr{A})$，而 $\mathrm{Path}^{<\omega}(\mathscr{A})$ 代表有限路径的集合，$\mathrm{Path}^{\leqslant k}(\mathscr{A})$ 表示路径长至多为 k 的路径集。路径的最终状态记为 $\mathrm{last}(\pi)$。路径 π 的迹 $\mathrm{Tr}(\pi)$ 定义为路径 π 上的动作序列 $\mathrm{Tr}(\pi)=a_1a_2a_3\cdots$。若 $F\subseteq\mathrm{Path}^{<}0,a\in\Sigma$，那么 $\mathrm{Succ}(F,a)$ 表示所有路径 π' 的集合，满足条件 $\pi'=\pi\alpha\mu s$；同样可以定义 $\mathrm{Succ}(F,\beta),\beta\in\Sigma^{<\omega}$。

定义 18 称概率机 $\mathscr{A}$ 是有限的(可数的)，如果对每个状态 s，集合$\{\langle s,\mu\rangle\mid s\xrightarrow{\alpha}\mu\}$是有限的(可数的)。称 $\mathscr{A}$ 是象可数的，如果对每个状态 s 和动作 a，集合$\{\mu\mid s\xrightarrow{\alpha}\mu\}$是有限的。

本文将假设概率自动机是象有限的。由于动作集 Σ 是有限的，因此每个象有限的概率自动机也是可数分支的，进而 $\mathrm{Path}^{<\omega}(\mathscr{A})$ 是可数的，这样就可以进行计数操作。

3.2 概率执行

下面要定义概率自动机的行为。非概率情形下，路径的执行可以通过确定的方式解决非确定性问题。而在概率自动机中，引入调度器来解决非确定性问题，即它是随机的、历史依赖的、部分的。

定义 19 $\mathscr{A}$ 的调度器 E 是一个函数 $E:\mathrm{Path}^{<\omega}(\mathscr{A})\to\mathrm{SubDistr}(\Sigma\times\mathrm{Distr}(S_{\mathscr{A}}))$。满足对每条有限路径 π，$E(\pi)(a,\mu)>0\Rightarrow\mathrm{last}(\pi)\xrightarrow{a}\mu$。

$\mathrm{Adv}(\mathscr{A})$ 表示 $\mathscr{A}$ 上所有调度器函数的集合。直观上来讲，一个调度器 E 在每一步 $\mathscr{A}$ 的计算中，通过投硬币来选择下一条变迁，这样 E 就生成了一棵概率执行树。

定义 20 设 E 是 $\mathscr{A}$ 的调度器，由 E 产生的概率执行函

数 Q_E:$\mathrm{Path}^{<\omega}(\mathscr{A})\to[0,1]$递归定义如下:

$$\begin{cases}Q_E(s_0)=1\\ Q_E(\pi a\mu s)=Q_E(\pi)\cdot E(\pi)(a,\mu)\cdot\mu(s)\end{cases}$$

$\mathscr{A}$上所有概率执行的集合记为 $\mathrm{ProbExec}(\mathscr{A})$。函数 Q_E 根据调度器 E 的决定给路径赋予概率。$Q_E(\pi)=p$ 表示在调度器 E 的控制下,$\mathscr{A}$以概率 p 执行路径 π。记 $\pi\sqsubseteq\pi'$,如果事件 $\mathscr{A}$沿路径 π'蕴含事件 $\mathscr{A}$沿路径 π。

定义 21 设 $\pi\in\mathrm{Path}^{<\omega}(\mathscr{A})$,由路径 π 产生的锥 $C_\pi=\{\pi'\in\mathrm{Path}(\mathscr{A})\mid\pi\sqsubseteq\pi'\}$。

设 $\Omega_{\mathscr{A}}=\mathrm{Path}(\mathscr{A})$代表样本空间,$\mathscr{F}_{\mathscr{A}}$表示由集合$\{C_\pi\mid\pi\in\mathrm{Path}^{<\omega}(\mathscr{A})\}$所产生的最小 σ 域。下面定理说明了函数 Q_E 在 $\mathscr{F}_{\mathscr{A}}$ 产生了唯一的概率测度。

定理 2 设 E 是 $\mathscr{A}$ 的调度器,则存在 $\mathscr{F}_{\mathscr{A}}$ 上的唯一测度 μ_E,满足对于所有路径 $\pi\in\mathrm{Path}^{<\omega}(\mathscr{A})$,$\mu_E[C_\pi]=Q_E[\pi]$。

证明:见文献[26]。

由该定理,可以导出概率空间$(\Omega_{\mathscr{A}},\mathscr{F}_{\mathscr{A}},\mu_E)$。

3.3 迹分布

概率自动机 $\mathscr{A}$的外部行为可以通过删除概率执行中的不可见元素获得,这就产生了 $\mathscr{A}$ 上的迹分布。为了定义迹分布,我们采用迹算子 $\mathrm{Tr}:\mathrm{Path}^{<\omega}(\mathscr{A})\to\Sigma^{<\omega}$。

定义 22 设 E 是 $\mathscr{A}$的调度器,而且存在函数 Q_E,$\mathrm{Path}^{<\omega}(\mathscr{A})\to[0,1]$,那么由 E 生成的迹分布是函数 $\mathrm{Tr}(Q_E):\Sigma^{<\omega}\to[0,1]$,定义如下:$\mathrm{Tr}(Q_E)(\beta)=\Sigma_{\pi\in Tr^{-1}(\beta)}Q_E(\pi)$。

通常记 $D_E=\mathrm{Tr}(Q_E)$,使用 D,K 等表示迹分布变量。$\mathscr{A}$上所有迹分布的集合记为 $\mathrm{TrDist}(\mathscr{A})$。同样,分布 D_E 也可以产生概率测度。样本空间 $\Omega=\Sigma^{<\omega}$;σ 域 $\mathscr{F}=\{C_\beta\mid\beta\in\Sigma^{<\omega}\}$,其中 $C_\beta=\{\beta'\in\Omega\mid\beta\sqsubseteq\beta'\}$。

定理 3 设 E 是 $\mathscr{A}$的调度器,在 $\mathscr{F}$ 上存在唯一的测度 μ^E,使得 $\mu^E[C_\beta]=D_E(\beta)$。

这样,测度 μ^E 又产生了一个概率空间$(\Omega,\mathscr{F},\mu^E)$。

定义 23 迹分布包含关系 $\mathscr{A}\sqsubseteq_{\mathrm{TD}}\mathscr{B}$ 当且仅当 $\mathrm{TrDist}(\mathscr{A})\subseteq\mathrm{TrDist}(\mathscr{B})$。

本文将集中研究有限调度器。$\mathrm{Adv}(\mathscr{A},k,l),k,l\in N$ 表示深度 k、长度 l 的所有调度器的集合。对于给定的路径 π,我们说 π 是调度器 E 可达的,如果 $Q_E(\pi)\neq0$。若调度器 E 在每条长度为 k 路径后都终止,就说 E 深度为 k。如果对所有 E 可达路径 π,$\mathrm{Tr}(\pi)\in(\Sigma_l)^{<\omega}$,我们说 E 宽度为 l,即 E 分支的程度为 l。如果 E 有有限的深度和宽度,E 是有限的。$\mathrm{ProbExec}(\mathscr{A},k,l),k,l\in N$ 表示由 $E\in\mathrm{Adv}(\mathscr{A},k,l)$所产生的深度 k、长度 l 概率执行的集合。$\mathrm{TrDist}(\mathscr{A},k,l),k,l\in N$ 代表由 $E\in\mathrm{Adv}(\mathscr{A},k,l)$所产生的深度 k、长度 l 迹分布的集合。有限迹包含关系 $\mathscr{A}\sqsubseteq^{k,l}_{\mathrm{TD}}\mathscr{B}$ 当且仅当 $\mathrm{TrDist}(\mathscr{A},k,l)\subseteq\mathrm{TrDist}(\mathscr{B},k,l)$。

4 统计观察

由于确定性一致性测试选择方法在随机测试方面存在不足,因此采取了基于统计的方法来研究随机测试选择。下面首先从测试实验来收集测试执行的样本,然后根据概率机迹包含关系判断样本是否可接受。

4.1 采样

我们通过实验来收集样本,即采样。使用$\langle k,l,m\rangle$说明采样的性质,即在概率机上进行深度为 k、宽度为 l、广度为 m 的实验:

1)测试器按 reset 键启动测试;

2)每次执行时,只可以执行宽度 l 的动作集 Σ_l 中的动作;

3)记录执行中出现的动作;

4)当执行结束或测试器已经记录到长度为 k 的序列时,重置测试器,开始新一次的测试;

5)当测试器已经记录了 m 次测试,实验结束。

在实验过程中,测试器记录了执行迹序列 $\beta_0,\beta_1,\cdots,\beta_{m-1},\beta_i$ 代表宽度为动作集 Σ_l 并且深度不超过 k 的动作序列。使用 $O=\{\beta_0,\beta_1,\cdots,\beta_{m-1}\}$代表参数$\langle k,l,m\rangle$的一个样本。$O=((\Sigma_l)^k)^m$ 是参数$\langle k,l,m\rangle$的所有可能样本集,同时也是概率机 $\mathscr{A}$的观察结果。由参数$\langle k,l,m\rangle$确定的样本 O 对应着一个可能的迹分布序列 $D_0,D_1,\cdots,D_{m-1}$,同时也是该分布的一个可接受的结果。使用 D 来代表迹分布序列 $D_0,D_1,\cdots,D_{m-1}$。

在每次测试执行中,迹 β 的选择由概率机 $\mathscr{A}$迹分布 D 控制。当按下重置键后,测试器又开始了新的测试,并且选择了另外不同的迹分布。由于参数$\langle k,l,m\rangle$规定的有限性,迹分布 D 只能从 $\mathrm{TrDist}(\mathscr{A},k,l),k,l\in N$ 中选择。

在单独的一次测试执行中,迹 β 的长度可能严格小于 k。如果 $\beta_0\neq\beta_1$,事件观察到 β_0 与事件观察到 β_1 是相互独立的。

定义 24 一次测试执行中,恰好观察到 β 的概率为

$$P_{D,k}[\beta]=\begin{cases}D(\beta), & \text{如果 }\beta\text{ 的长度恰是 }k\\ D(\beta)-\Sigma_{E\in\Sigma_1}D(\beta\alpha), & \text{否则}\end{cases}$$

M 次测试运行中,产生深度 k 的一个样本 $O=\{\beta_0,\beta_1,\cdots,\beta_{m-1}\}$的概率可以表示为 $P_{D,k}[O]=\sum_{i=0}^{m-1}P_{D_i,k}[\beta_i]$。对于样本集 $\mathcal{O}$,$P_{D,k}[\mathcal{O}]=\Sigma_{O\in\mathcal{O}}P_{D,k}[O]$。

定义 25 每条迹 β 出现在样本 O 中的频率 $\mathrm{freq}(O)(\beta)=\frac{m_i(\beta)}{m}$,表示样本 O 中与迹 β_i 相等的迹 β 的数量占样本集的百分数。

固定参数 $k,l,m,D,\beta\in(\Sigma_l)^{\leqslant k}$,对于 $0\leqslant i\leqslant m-1$,如果在第 i 次运行时观察到迹 β,则第 i 次运行的结果是成功的。所以 $P_{D_i,k}[\beta_i]$表示第 i 次运行成功的概率,这可以看作具有参数 $P_{D_i,k}[\beta_i]$的贝努利分布。

设 X_i 代表成功次数的随机变量,随机变量 $Z=\frac{1}{m}\sum_{i=1}^{m}X_i$ 表示由分布 D 控制的 m 次运行中成功的频率,所以频数的期望值为

$$E^{D,k}_\beta=E[Z]=E(\frac{1}{m}\sum_{i=0}^{m-1}X_i)=\frac{1}{m}\sum_{i=0}^{m-1}E[X_i]=\frac{1}{m}\sum_{i=0}^{m-1}P_{D_i,k}[\beta]$$

4.2 接受域

根据假设检验的思想,首先定义零假设:样本由分布列 D 产生。假设显著性水平 $\alpha\in(0,1)$,实验参数$\langle k,l,m\rangle$,集合 $\mathrm{Obs}(D,k,l,m,\alpha)$表示接受域,如果满足下列条件:

1)$P_{D,k}[\mathrm{Obs}(D,k,l,m,\alpha)]\geqslant1-\alpha$;

2)$P_{D',k}[\mathrm{Obs}(D,k,l,m,\alpha)]$在不同的分布 D' 中是最小的。

条件 1)说明拒绝正确零假设(第一类错误)的最大概率是 α。条件 2)指出错误接受(第二类错误)的概率应尽可能

小。

根据区间估计的思想，从采样获得的样本来猜测迹分布的区间。假设 m 次运行是等分布的，即满足同样迹分布 D，频数 $\mathrm{freq}(O)(\beta)$ 可以作为观察到迹 β 的概率 $P_{D,k}[\beta]$ 的估计值。由于进行完全正确的估计的概率是非常小的，因此包含频数 $\mathrm{freq}(O)(\beta)$ 的区间可以保证猜测正确的概率是 $1-\alpha$。也就是，如果观察到频数 $\mathrm{freq}(O)(\beta)$，那么猜测以概率 $P_{D,k}[\beta]$ 落入区间 $[\mathrm{freq}(O)(\beta)-r,\mathrm{freq}(O)(\beta)+r]$，$r$ 取决于参数 α。

如果固定概率 $P_{D,k}[\beta]$，我们可以得到频数 $\mathrm{freq}(O)(\beta)$ 的区间估计 $[P_{D,k}[\beta]-r,P_{D,k}[\beta]+r]$。如果这个区间中的一个频数被实际观察到，那么对于概率 $P_{D,k}[\beta]$ 的猜测将是正确的。频数向量 $\mathrm{freq}(O)$ 是可以接受的，如果任取一条迹 β，频数 $\mathrm{freq}(O)(\beta)$ 都在适当的区间 $[P_{D,k}[\beta]-r,P_{D,k}[\beta]+r]$。由于实际中无法获得单独一条迹 β 的概率值 $P_{D,k}[\beta]$ 的估计，因此使用频数 $\mathrm{freq}(O)(\beta)$ 作为 m 次运行迹分布平均值 $E_\beta^{D,k}=\frac{1}{m}\sum_{i=0}^{m-1}P_{D_i,k}[\beta]$ 的估计值。

根据上述思想，我们接受一个样本 O，如果它的频数落入区间 $[E_\beta^{D,k}-r,E_\beta^{D,k}+r]$。下面的任务就是找到合适的参数 $r\in[0,1]$，使得上述条件 1)和条件 2)，即拒绝正确假设和接受错误得到满足。

定义 26

$$B_r(E^{D,k})=\{\nu\in[0,1]^{(\Sigma_1)^{\leqslant k}}\mid\forall\beta\in(\Sigma_1)^{\leqslant k},|\nu(\beta)-E_\beta^{D,k}|\leqslant r\}$$

表示以 $E_\beta^{D,k}$ 为中心、r 为半径的邻域。$[0,1]^{(\Sigma_1)^{\leqslant k}}$ 表示长度为 $(\Sigma_1)^{\leqslant k}$ 的 0,1 序列。由于 $\nu,E_\beta^{D,k}$ 都可以看作度量空间中的点，因此 $\nu(\beta)-E_\beta^{D,k}$ 表示两点间的距离。$\mathrm{freq}^{-1}(B_r(E^{D,k}))$ 表示频数偏离平均值 r 的样本集。

定义 27 设 $k,l,m\in N$，$D=\{D_i\in\mathrm{TrDist}(\mathscr{A},k,l)\mid k,l\in N,i\in[0,m-1]\}$

$$\bar{r}=\inf\{r\mid P_{D,k}[\mathrm{freq}^{-1}(B_r(E^{D,k}))]>1-\alpha\}$$

表示距离 r 的下确界。

定义 28 具有参数 $\langle k,l,m\rangle$ 和显著性水平 α 的迹分布 D 的接受域为

$$\mathrm{Obs}(D,k,l,m,\alpha)=\mathrm{freq}^{-1}(B_{\bar{r}}(E^{D,k}))=\{O\mid\mathrm{dist}(\mathrm{freq}(O),E^{D,k})<\bar{r}\}$$

定义 29 具有参数 $\langle k,l,m\rangle$ 和显著性水平 α 的概率机 $\mathscr{A}$ 的接受域为

$$\mathrm{Obs}(\mathscr{A},k,l,m,\alpha)=\bigcup_{D\in(\mathrm{TrDist}(\mathscr{A},k,l))^m}\mathrm{Obs}(D,k,l,m,\alpha)$$

5 概率测试选择示例

考虑下面协议规范概率自动机 $\mathscr{A}_{s1}=(S_S,\Sigma_S,T_S,S^0)$，如图 1(a)所示。其中，

$S_S=\{S^0,S_1,S_2,S_3,S_4\}$

$\Sigma_S=\{L?\ x,L!\ y,L!\ z\}$

$$T_S=\{s^0\overset{L?x,\mu_1}{\rightsquigarrow}s_1,s^0\overset{L?x,\mu_2}{\rightsquigarrow}s_2,s_1\overset{L!y}{\rightarrow}s_3,s_2\overset{L!z}{\rightarrow}s_4\}$$

s^0 是初始状态。

概率机 $\mathscr{A}_s$ 规范要求在收到消息 x 后，以概率 $\mu_1=1/2$ 执行动作输出消息 y，以概率 $\mu_2=1/2$ 执行动作输出消息 z。

图 1(b)是又一个协议规范概率机 $\mathscr{A}_{s2}$。

为了进行测试选择，首先需要进行采样。步骤如下：假定实验参数 $\langle k,l,m\rangle=\langle 2,2,100\rangle$，

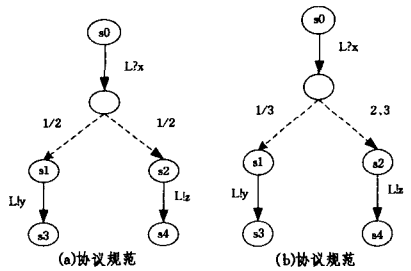

图 1 协议规范概率自动机

1)测试器按 reset 键启动测试；

2)每次执行时，只可以执行宽度 l 的动作集 Σ_l 中的动作；

3)记录执行中出现的动作；

4)当执行结束或测试器已经记录到长度为 k 的序列时，重置测试器，开始新一次的测试；

5)当测试器已经记录了 m 次测试，实验结束。

假设显著性水平 $\alpha=0.05$，在实验 $\langle 2,100\rangle$ 和协议规范 $\mathscr{A}_{s1}$ 的迹 $L?xL!y$ 分布为 1/2 的条件下，迹 $L?xL!y$ 的接受域为 $[41,59]$，即观察到成功的次数应该在接受域范围内；而在在实验 $\langle 2,100\rangle$ 和协议规范 $\mathscr{A}_{s2}$ 的迹 $L?xL!y$ 分布为 1/3 的条件下，迹 $L?xL!y$ 的接受域为 $[24,42]$。如果观察到迹 $L?xL!y$ 成功的次数是 50，那么协议规范 $\mathscr{A}_{s1}$ 接受它，而协议规范 $\mathscr{A}_{s2}$ 拒绝它。

结束语 随机性存在于各种系统行为中，如何基于随机性进行测试套选择对传统选择方法提出了新的挑战。本文受到随机系统统计验证的启发，提出了一种基于可用性的测试选择方法，即把可用性判别标准转化为统计检验问题。首先使用概率自动机模型建立了测试例的接受域；然后对测试例执行进行采样；如果采样结果符合接受域，就认为该测试例是可用的，即可以被选择；否则就是不可用的，即不能被选择。

当然现有的研究只是集中于通用的随机过程，希望在接下来的工作中，把该方法应用到具体的随机过程，如马尔可夫过程；还有就是如何计算迹分布等。

参 考 文 献

[1] Tretmans J. A Formal Approach to Conformance Testing[D]. Enschede, The Netherlands: University of Twente, 1992

[2] Ostapenko V V. Test selection based on implementation specification

[3] Information technology—Open Systems Interconnection—Conformance testing methodology and framework — Part 1: General concepts[S]. ISO/IEC 9646-1. 1994

[4] Information technology — Open Systems Interconnection -- Conformance testing methodology and framework -- Part 2: Abstract Test Suite specification[S]. ISO/IEC 9646-2. 1994

[5] Information technology -- Open Systems Interconnection — Conformance testing methodology and framework — Part 3: The Tree and Tabular Combined Notation (TTCN)[S]. ISO/IEC 9646-3. 1998

[6] Information technology -- Open Systems Interconnection — Conformance testing methodology and framework — Part 4: Test re-

alization[S]. ISO/IEC 9646-4. 1994

[7] Information technology -- Open Systems Interconnection -- Conformance testing methodology and framework -- Part 5: Requirements on test laboratories and clients for the conformance assessment process[S]. ISO/IEC 9646-5. 1994

[8] Information technology --Open Systems Interconnection--Conformance testing methodology and framework -- Part 6: Protocol profile test specification[S]. ISO/IEC 9646-6. 1994

[9] Information technology -- Open Systems Interconnection -- Conformance testing methodology and framework -- Part 7: Implementation Conformance Statements[S]. ISO/IEC 9646-7. 1995

[10] Fujiwara S, Bochman G. Test selection based on finite state machine

[11] Luo G, Bochman G. Test selection based on Communicating nondeterministic finite state machine using a generalized wp-methods

[12] Heerink L, Tretmans J. Formal methods in Conformance testing: a probabilistic refinement[C]// Baumgarten B, Burkhardt H, Giessler A, eds. Ninth International Workshop in Testing and Communication System. Volume IX, 1996: 261-276

[13] Goga N. A probabilistic coverage for on-the-fly test generation algorirhms[C]// Automated Verification of Critical Systems (AVoCS '03). Southampton, UK, 2003

[14] Segala R. Modeling and Verification of Randomized Distributed Real-time Systems[D]. Dept. of Electrical Engineering and Computer Science, MIT, 1995

[15] Aggarwal S. Time optimal self - stabilizing spanning tree algorithms[D]. Department of Electrical Engineering and Computer Science, Massschusetts Institute of Technology, May 1994

[16] Lynch N A, Saias I, Segala R. Proving time bounds for randomized distributed algorithms[C]// Proceedings of the 13th Annual ACM Symposium on the Principles of Distributed Computing. 1994: 314-323

[17] Pogosyants A, Segala R, Lynch N A. Verification of the randomized consensus algorithms of Aspnes and Herlihy: a case study [J]. Distributed Computing, 2000, 13(3): 155-186

[18] Stoelinga M I A, Vaandrager F W. Root contention in ieee 1394 [C]// Proceedings 5th International AMAST Workshop on Formal methods for Real-time and Probabilistic Systems. volume 1601 of Lecture Notes in Computer Science. Springer-Verlag, 1999: 53-74

[19] Segala R. Testing Probabilistic Automata[C]// Proceedings of the 7th International Conference on Concurrency Theory. LNCS 1119. 1996: 299-314

[20] Christoff I. Testing .equivalence and fully abstract models of probabilistic processes[C]// Proceedings CONCUR 90. volume 458 of Lecture Notes in Computer Science. Springer-Verlag, 1999

[21] Gregorio - Rodrigez C, Nunez M. Denotational semantics for probabilistic refusal testing[C]// Proceedings of ProbMIV 98. volume 22 of Electronic Notes in Theoretical Computer Science. 1998

[22] Cleaveland R, Dayar Z, Smolka S A, et al. Testing preorders for probabilistic processes[J]. Information and Computation, 1999, 154(2): 93-148

[23] Josson B, Yi W. Compositional testing preorders for probabilistic processes[J]. Theoretical Computer Science, 2001

[24] De Nicola R, Hennessy M. Testing equivalence for processes [J]. Theoretical Computer Science, 1984, 34: 83-133

[25] Larsen K G, Skou A. Bisimulation through probabilistic testing [J]. Information and Computation, 1991, 91: 1-28

[26] Cheung L, Stoelinga M, Vaandrager F. A Testing Scenario for Probabilistic Processes[J]. Journal of the ACM, 2007, 54(6)

[27] Younes H L S, Simmons R G. Probabilistic verification of discrete event systems using acceptance sampling[C]// Computer-Aided Verification. volume 2404 of Lecture Notes in Computer Science. 2002: 223-235

[28] Younes H L S, Kwiatkowska M Z, Norman G, et al. Numerical vs. statistical probabilistic model checking: an empirical study [C]// Tools and Algorithms for the Constuction and Analysis of Systems. volume 2988 of Lecture Notes in Computer Science. 2005: 46-60

[29] Sen K, Viswanathan M, Agha G. Statistical model checking of black-box probabilistic systems[C]// Computer-Aided Verification. volume 3114 of LNCS. 2004: 202-215

[30] Younes H L S. Probabilistic verification for black-box systems [C]// Proceedings CAV 2005. 2005: 253-265

[31] Blute R, Desharnais J, Edalat A, et al. Bisimulation for labeled markov processes[J]. Information and Computation, 2002, 179 (2): 163-193

[32] Baier C, Kwiakowska M. Model checking for a probabilistic branching time logic with fairness[J]. Distributed Computing, 1998, 11(3): 125-155

[33] Edalat A. Domain theory in stochastic processes[C]// Proceedings LICS 95. 1995: 244-254

(上接第 96 页)

[2] Waharte S, Boutaba R, Iraqi Y, et al. Routing protocols in wireless mesh networks: challenges and design considerations[J]. Multimed Tools Appl, 2006(29): 285-303

[3] 姜红旗，康凯，林孝康. 拓展宽带接入的无线 Mesh 网技术[J]. 电信科学，2005(1): 24-30

[4] 沈强，方旭明. 无线 Mesh 网中一种基于综合准则的 DSR 扩展路由方法[J]. 电子学报，2007, 35(4): 614-620

[5] 沈强，方旭明，宋文. 无线 Mesh 网络路由协议研究[J]. 数据通信，2005(4): 30-33

[6] Jacquet P, Muhlethaler P, Clausen T, et al. Optimized Link State Routing Protocol for Ad Hoc Networks[C]// Proceedings of IEEE international Multi Topic Conference. Pakistan: IEEE, 2001: 62-68

[7] Qayyum A, Viennot L, Laouiti A. Multipoint relaying: An efficient technique for flooding in mobile wireless networks[C]// Proceedings of the 35th Annual Hawaii International Conference. Hawaii: IEEE, 2001

[8] http://www.isi.edu/nsnam/ns