

对逻辑覆盖软件测试准则的公理化评估*

刘玲⁺, 缪淮扣

(上海大学 计算机工程与科学学院, 上海 200072)

Axiomatic Assessment of Logic Coverage Software Testing Criteria

LIU Ling⁺, MIAO Huai-Kou

(School of Computer Engineering and Science, Shanghai University, Shanghai 200072, China)

+ Corresponding author: Phn: +86-21-56337684, Fax: +86-21-56333061, E-mail: liuling@mail.shu.edu.cn, <http://www.shu.edu.cn>

Received 2003-10-22; Accepted 2004-02-02

Liu L, Miao HK. Axiomatic assessment of logic coverage software testing criteria. *Journal of Software*, 2004,15(9):1301~1310.

<http://www.jos.org.cn/1000-9825/15/1301.htm>

Abstract: Since formal specifications precisely describe the software requirements in a form that can be automatically manipulated, they can be used as the base for automatic test generation and software verification. Logic coverage criteria are the common criteria used in specification-based testing. The main problem of applying these criteria that test engineers face is how to appropriately select each criterion. Thus, the comparison and analysis of these criteria will give a guide to applying each criterion. Axiomatic assessment of test adequacy criteria is an approach to comparing test criteria. This approach defines the intuitive requirements of ideal test adequacy criteria as some axioms, then compares the test adequacy criteria by checking if they satisfy these axioms. This paper proposes some positive properties as the intuitive requirements of ideal logic coverage criteria, and gives a generating algorithm that is used to determine whether a logic coverage criterion is complete. These properties are formally defined as an axioms system. With these formal definitions, the relations among the logic coverage criteria are described as some theorems. Finally, the common logic coverage criteria are assessed against the axioms system. From the assessing result, testers can get some conclusions that help them apply these criteria in practice.

Key words: specification-based testing; logic coverage software testing criteria; axiomatic assessment; test adequacy criteria

摘要: 由于形式规格说明采用一种精确、一致、容易被机器自动处理的符号系统来描述软件需求,因而形式规格说明为测试用例的自动生成和软件功能的验证提供了基础。在基于形式规格说明的测试过程中逻辑覆盖测试准则是一组常用的测试准则,如何选择和使用其中的每个测试准则是应用这组测试准则时面临的主要问题。因此分析和比较这组测试准则中每个测试准则的性质将为测试工程师选择测试准则提供指导和帮助。对测试充分性准则的公理化评估是一种比较测试准则的方式,这种方式将对理想的测试准则的直觉需求定义为一

* Supported by the National Natural Science Foundation of China under Grant No.60173030 (国家自然科学基金)

作者简介: 刘玲(1972—),女,安徽淮南人,博士,讲师,主要研究领域为软件形式方法,软件测试,面向对象技术;缪淮扣(1953—),男,教授,博士生导师,主要研究领域为软件形式方法,软件工程,面向对象技术。

组公理,然后通过检查测试准则是否满足该组公理来分析和比较相应的测试准则.描述了一组理想的逻辑覆盖测试准则应该具有的性质和用来确定一个测试充分性准则是否完全的生成算法.这组性质被形式化地定义为一组公理.利用这种形式化的定义,用定理的形式精确地给出了这些性质之间的关系.最后通过这组公理系统来评估现有的逻辑覆盖测试准则,评估的结果为测试人员在实际过程中选择逻辑覆盖测试准则提供了指导.

关键词: 基于规格说明的测试;逻辑覆盖测试准则;公理化评估;测试充分性准则

中图法分类号: TP311 文献标识码: A

测试充分性准则是软件测试的核心.早在1975年,Goodenough和Gerhart就指出软件测试的中心问题是“测试充分性准则是什么?”^[1].通常,测试充分性准则是指一组用来判定软件是否被充分测试了的规则.这组规则同时也是测试用例生成过程中的停机规则.根据测试需求的来源,测试充分性准则可以分为基于程序代码的测试准则和基于规格说明的测试准则两类(在软件测试的文献中,通常对测试充分性准则和测试准则不加区分.本文也遵循这一习惯).

若一个测试准则是基于规格说明的,则意味着测试需求来源于软件需求规格说明中的特征.因而,判断一个基于规格说明的测试用例集是否充分就是要判断软件需求规格说明中的一些特征是否都被验证了.常见的基于规格说明的测试准则有逻辑覆盖测试准则^[2,3]和规格说明变异测试准则^[4].其中,逻辑覆盖测试准则主要用于基于形式规格说明的测试中,通过分析规格说明中谓词和文字的真值关系来产生覆盖某些谓词和文字的真值取值的测试用例.

由于每个测试准则都有自己的优缺点,因而比较相关的一组测试准则并清楚地说明它们的性质一直是软件测试中关注的问题之一.根据软件测试充分性的直觉概念来评估软件测试准则是一种比较软件测试准则的方法^[5-9].在这方面的研究中一个重要的工作是Weyuker的公理化系统^[6]——一组关于软件测试充分性准则的基本性质.通过用这组公理评估部分现有的测试充分性准则,Weyuker对这些测试准则进行了比较.在对测试充分性准则的公理化评估中面临的一个主要问题是如何给出一些正面性质作为公理.例如,文献[6]中的反组合公理(anticomposition axiom)就不是正面性质.该公理指出,即使测试集 T 对于被测程序 P 的每个组件是充分的, T 对于 P 来说也未必是充分的.尽管这条公理真实地反映了软件测试中的一种直觉概念,但它并不能提高测试人员对某一测试用例集的信心,因而,这条公理描述的测试准则的性质是一种负面性质.这种负面性在测试准则评估中的作用是十分微弱的.针对这一问题,朱鸿提出了用于评估基于控制流的测试充分性准则的形式的、统一的公理化系统^[5].

本文对文献[5]中的公理系统针对逻辑覆盖测试准则的特点作了重新描述,并添加了1条相对性公理,从而构建了一个用于评估逻辑覆盖测试准则的形式化的公理系统.该公理系统包含了理想的逻辑覆盖测试准则希望具备的7条性质,本文还利用这7条性质对逻辑覆盖测试准则进行了评估.本文的主要贡献在于,形式化地定义和分析了逻辑覆盖测试准则的公理系统,给出了一个判断测试集是否完全的算法,并且利用给出的公理系统对这些测试准则进行了评估.这种评估有助于从整体上和直觉上区分每个逻辑覆盖测试准则,从而为测试人员选择具体的逻辑覆盖测试准则提供了指导.

本文第1节介绍常见的逻辑覆盖测试准则,这些准则的定义摘自Ammann和Offutt的文章^[3].这里,我们根据需要作了相应的改动.第2节给出一个逻辑覆盖测试准则的公理系统,并对其中每条公理进行了形式化的定义和分析.第3节利用我们的公理系统对逻辑覆盖测试准则进行评估,并对评估结果进行了总结.最后,根据第4节的评估结果得到一些有益的结论.

1 逻辑覆盖测试准则

逻辑表达式是形式规格说明中用于说明前、后置条件的主要形式,也是程序中分支控制条件的主要表达式,因而可以作为基于规格说明的测试用例的主要来源^[2,10-12]和定义测试充分性的基础.逻辑表达式的主要形式是谓词.例如, $(a > b \vee x < y) \wedge G$ 就是一个谓词表达式(简称谓词).一个谓词由若干文字(文献[3]中将不含任何逻辑

运算符的谓词定义为子句(*clause*),本文为了与经典一阶谓词逻辑的表示一致,采用了术语文字(*literal*),我们认为在这里“文字”比“子句”更合适)组成.文字是一种不包含任何逻辑运算符的原子谓词,同一个文字在一个谓词中只能出现一次.上例中的谓词包含 3 个文字: $a>b, x<y$ 和 G .在软件的形式需求规格说明中,操作的前置条件是一组谓词的集合.这里用 P 表示这组谓词的集合, L 是 P 中所有谓词的文字集.对 P 中的每个谓词 $p, p \in P, L_p$ 是 p 中所有文字构成的集合,即 $L_p = \{l | l \in p\}$.所以, $L = \bigcup_{p \in P} L_p$ [3].

在基于逻辑表达式的测试充分性准则中,一个测试集是否检测了某些谓词和文字的真值决定了该测试集的充分性.例如,对于谓词 $p = (a > b \vee x < y) \wedge G$,测试用例集 $\{a = 5 \wedge b = 4 \wedge x = 2 \wedge y = 3 \wedge G = \text{true}, a = 4 \wedge b = 5 \wedge x = 3 \wedge y = 2 \wedge G = \text{false}\}$ 就分别检测了谓词 p 取为真和为假的情况.因此,该测试用例集充分覆盖了谓词 p 的真值.在利用逻辑覆盖测试准则推导测试用例时,先要找出满足该准则的谓词和文字的真值组合,然后再通过求解约束的方式给出具体的测试数据.对于上例,如果我们使用谓词覆盖准则来推导测试用例,首先需要找出能够使得谓词 p 取值为真和为假的文字真值的组合,这里选择的是 $\{a > b = \text{true} \wedge x < y = \text{true} \wedge G = \text{true}, a > b = \text{false} \wedge x < y = \text{false} \wedge G = \text{false}\}$,然后求解其中的每一个逻辑表达式定义的约束(例如, $a > b = \text{true}$)就可以得到具体的测试数据了.通过该例可以看出,利用逻辑覆盖测试准则推导具体测试数据过程的关键是找出满足准则的文字真值取值的组合,这里将每一种文字真值取值的组合看成一个测试用例.根据这种约定,在上面的例子中,测试用例集 $\{a > b = \text{true} \wedge x < y = \text{true} \wedge G = \text{true}, a > b = \text{false} \wedge x < y = \text{false} \wedge G = \text{false}\}$ 满足谓词覆盖准则.

为了下文讨论的方便,给出一些约定.设 $TL(P)$ 是被测谓词集 P 中所有可能的文字真值取值的组合构成的集合.被测谓词集及其所有充分的测试用例集之间的关系可以用下面的二元关系 AF 来定义.

$$\begin{array}{|l} \text{[LOGCRI]} \\ A: \text{LOGCRI} \times TL(P) \rightarrow \text{BOOL} \\ \hline \text{AF: } \mathbb{P}P \rightarrow \mathbb{P}TL(P) \\ \hline \forall X: \mathbb{P}P \bullet \text{AF}(X) = \{t | t \subseteq TL(P) \wedge \exists c: \text{LOGCRI} \bullet A(c, t)\} \end{array}$$

其中,函数 A 用于判断一个测试集对某个测试准则是否是充分的,若充分则函数值为真,否则为假.给定集合 LOGCRI 表示所有逻辑覆盖准则. AF 的定义域是 $\{X | X \subseteq P\}$,对于任意的 $X \in \text{dom } AF, AF(X)$ 由所有根据某个逻辑覆盖准则对于 X 是充分的测试用例集构成.

1.1 谓词覆盖和文字覆盖

谓词覆盖和文字覆盖是两个简单的逻辑覆盖测试准则.其中,谓词覆盖准则要求一个测试集能够使得每个谓词分别为真和假.文字覆盖准则要求一个测试集能够使得每个文字分别为真和假.当被测谓词是永真(或永假)时,测试用例就是谓词本身,因而,就不存在测试用例的选择问题,下面定义的逻辑覆盖测试准则中的谓词不包括这种特殊谓词.

定义 1. 谓词覆盖(predicate coverage,简称 PC).

一组文字的真值组合构成的集合 t 满足谓词覆盖准则,当且仅当对每个 $p \in P, t$ 至少包含两个关于 p 中文字的真值组合,其中一个使得 p 为 true,另一个使得 p 为 false.

对于前面的谓词 $(a > b \vee x < y) \wedge G$,测试集 $\{a > b = \text{true} \wedge x < y = \text{true} \wedge G = \text{true}, a > b = \text{true} \wedge x < y = \text{false} \wedge G = \text{false}\}$ 满足谓词覆盖准则.

定义 2. 文字覆盖(literal coverage,简称 LC).

一组文字的真值组合构成的集合 t 满足文字覆盖准则,当且仅当对每个 $l \in L, t$ 至少包含两个关于 p 中文字的真值组合,其中一个使得 l 为 true,另一个使得 l 为 false.

对于前面的谓词 $(a > b \vee x < y) \wedge G$,测试集 $\{a > b = \text{true} \wedge x < y = \text{false} \wedge G = \text{true}, a > b = \text{false} \wedge x < y = \text{true} \wedge G = \text{false}\}$ 满足文字覆盖准则.

这里要注意的是文字覆盖并不一定包含谓词覆盖,谓词覆盖也不一定包含文字覆盖.例如,对于谓词 $A \vee B$,测试集 $\{A = \text{true} \wedge B = \text{false}, A = \text{false} \wedge B = \text{false}\}$ 满足谓词覆盖准则,但不满足文字覆盖准则.测试集 $\{A = \text{true} \wedge B = \text{false},$

$A=false \wedge B=true$ 满足文字覆盖准则,但不满足谓词覆盖准则.

1.2 活动文字覆盖

谓词覆盖和文字覆盖是两个很弱的测试准则.因为即使所有的谓词和文字的真值取值都检查了,也无法保证所有可能的文字真值的组合都能够被检查.为了检查谓词中所有文字的真值组合,出现了组合覆盖准则.

定义 3. 组合覆盖(combination coverage,简称 CoC).

一组文字的真值组合构成的集合 t 满足组合覆盖准则,当且仅当对每个 $p \in P$, t 包含 L_p 中所有可能的文字真值的组合.

例如,对于谓词 $A \vee B$,测试集 $\{A=true \wedge B=true, A=true \wedge B=false, A=false \wedge B=true, A=false \wedge B=false\}$ 满足组合覆盖准则.显然,组合覆盖准则包含谓词覆盖准则和文字覆盖准则.但该准则过于严格,以至于当一个谓词中的文字个数较多时很难在有限时间内完成测试用例的推导过程.也就是说,该准则不满足有限可应用性^[6].因为,这个性质要求一个测试准则应能够在有限的时间内被应用.为了克服组合覆盖准则的缺陷,可以从文字的真值对谓词真值的作用出发来构造有效的测试准则.术语“决定”用来定义文字的真值和谓词真值的关系.

定义 4. 决定(determination).

对于谓词 p 中的一个文字 l_i ,如果 p 中的其余的文字 $l_j, j \neq i$ 的真值使得改变 l_i 的真值会引起 p 的真值发生变化,那么称 l_i 决定谓词 p ,其中 l_i 为主要文字, l_j 为次要文字.

对于谓词 $p=A \vee B$,若 B 的真值取值为 false,那么文字 A 决定 p .这时, A 是主要文字, B 是次要文字.若 A 取 false,那么 B 决定 p .此时, B 是主要文字, A 是次要文字.利用文字和谓词之间的决定关系,可以定义如下的活动文字覆盖准则.

定义 5. 活动文字覆盖(active literal coverage,简称 ALC).

一组文字的真值组合构成的集合 t 满足活动文字覆盖准则,当且仅当对每个 $p \in P$ 和每个主要文字 $l_i \in L_p$, t 中至少含有两个 L_p 中的文字的真值组合分别使得 $l_i=true$ 和 $l_i=false$.

对于上例中的谓词 $p=A \vee B$,测试集 $\{A=true \wedge B=false, A=false \wedge B=false, A=false \wedge B=true\}$ 满足活动文字覆盖准则.其中, $A=true \wedge B=false$ 和 $A=false \wedge B=false$ 分别使得 $A=true$ 和 $A=false$, $A=false \wedge B=false$ 和 $A=false \wedge B=true$ 分别使得 $B=true$ 和 $B=false$.

在应用该准则时存在一些问题.例如,对于谓词 $p=A \leftrightarrow B$,若 A 为主要文字,无论次要文字 B 取何值, p 的值都由 A 决定.也就是说,集合 $\{A=true \wedge B=true, A=false \wedge B=true\}$ 和 $\{A=true \wedge B=true, A=false \wedge B=false\}$ 都包含分别使得 A 为 true 和 false 的文字真值组合.但在第 1 个集合的两个文字真值组合中,次要文字 B 取同样的值,而在第 2 个集合的两个文字真值组合中,次要文字 B 取不同的值.因此,在选取满足活动文字覆盖准则的测试集时,测试人员面临的一个主要问题是,当主要文字的值为 true 和 false 时,次要文字是否应该取相同的值.对于这一问题的不同回答产生了下面的 3 个逻辑覆盖测试准则——一般活动文字覆盖、相关活动文字覆盖和受限活动文字覆盖.

定义 6. 一般活动文字覆盖(general active literal coverage,简称 GALC).

一组文字的真值组合构成的集合 t 满足一般活动文字覆盖准则,当且仅当对每个 $p \in P$ 和每个主要文字 $l_i \in L_p$, t 中至少含有两个 L_p 中的文字的真值组合分别使得 $l_i=true$ 和 $l_i=false$,并且对这两个文字真值组合中次要文字的真值取值不作要求.

这里要注意的是,一般活动文字覆盖准则并不一定包含谓词覆盖准则.例如,对于上例中的谓词 $p=A \leftrightarrow B$,测试集 $\{A=true \wedge B=true, A=false \wedge B=false\}$ 满足一般活动文字覆盖准则,但是却不满足谓词覆盖准则.

定义 7. 相关活动文字覆盖(correlated active literal coverage,简称 CALC).

一组文字的真值组合构成的集合 t 满足相关活动文字覆盖准则,当且仅当对每个 $p \in P$ 和每个主要文字 $l_i \in L_p$, t 中至少含有两个 L_p 中的文字的真值组合分别使得 $l_i=true$ 和 $l_i=false$,并且这两个文字真值组合必须分别使得 p 取不同的真值.

对于上例中的谓词 $p=A \leftrightarrow B$,测试集 $\{A=true \wedge B=true, A=false \wedge B=true, A=false \wedge B=false\}$ 满足相关覆盖准则,而集合 $\{A=true \wedge B=true, A=false \wedge B=false\}$ 则不满足.这里,可以看出相关覆盖准则包含谓词覆盖准则.

定义 8. 受限活动文字覆盖(restricted active literal coverage,简称 RALC).

一组文字的真值组合构成的集合 t 满足受限活动文字覆盖准则,当且仅当对每个 $p \in P$ 和每个主要文字 $l_i \in L_p, t$ 中至少含有两个 L_p 中的文字的真值组合分别使得 $l_i = \text{true}$ 和 $l_i = \text{false}$,并且这两个文字真值组合中次要文字必须取相同的值.

对于上例中的谓词 $p = A \leftrightarrow B$,测试集 $\{A = \text{true} \wedge B = \text{true}, A = \text{false} \wedge B = \text{true}, A = \text{false} \wedge B = \text{false}\}$ 满足受限活动文字覆盖准则.其中,文字真值组合 $A = \text{true} \wedge B = \text{true}$ 和 $A = \text{false} \wedge B = \text{true}$ 对于主要文字 A 满足 RALC,文字真值组合 $A = \text{false} \wedge B = \text{true}$ 和 $A = \text{false} \wedge B = \text{false}$ 对于主要文字 B 满足 RALC.这里,可以看出,受限活动文字覆盖包含相关活动文字覆盖.

2 逻辑覆盖测试准则的公理系统

本节给出了理想的逻辑覆盖测试准则期望具有的 7 条性质,其中前 7 条性质与文献[5]的基于控制流的准则所期望的性质相对应,第 7 条是逻辑覆盖测试准则特有的性质.另外,本节还给出了这 7 条性质的公理化的描述和形式化的分析.这 7 条公理为比较和分析前一节中介绍的逻辑覆盖测试准则奠定了基础.

2.1 有限可应用性、单调性和空集不充分性

公理 1. 有限可应用性(finite applicability).

对每个被测谓词集合 P ,存在一个有限可应用的文字真值组合构成的充分测试集.该公理可以形式定义如下:

$$\forall P \exists t \bullet (t \in \text{AF}(P) \wedge \|t\| < \infty).$$

公理 2. 单调性(monotonicity).

若测试集 t 对于被测谓词集 P 是充分的,并且 $t \subseteq t'$,那么测试集 t' 对于被测谓词集 P 也是充分的.该公理的形式定义为

$$\forall P \forall t, t' \bullet (t \in \text{AF}(P) \wedge t \subseteq t' \Rightarrow t' \in \text{AF}(P)).$$

公理 3. 空集不充分(inadequate empty set).

空集对于任何被测谓词集合 P 都是不充分的.该公理的形式定义为

$$\forall P \bullet (\emptyset \notin \text{AF}(P)).$$

2.2 最小固定尺寸性质(minimum fixed-size property)

朱鸿给出了路径选择准则的最小固定尺寸性质^[5].该性质说明一个可满足的路径选择准则的最小充分测试集包含的元素的数目是固定的.本文对文献[5]中路径选择准则的最小固定尺寸性质进行了调整,用谓词代替其中的路径,并根据逻辑覆盖测试准则的特点重新进行了形式化,给出了逻辑覆盖测试准则的最小固定尺寸性质.其中,最小充分测试集是一种特殊的测试集,该集合的任何真子集对于所考虑的测试准则来说都是不充分的.因此,一个充分的测试集至少含有一个最小充分测试集.这一结论可以形式定义为

$$\forall P \bullet (\forall s \bullet (s \in \text{AF}(P) \Rightarrow \exists t \bullet (t \in \text{AF}(P) \wedge t \subseteq s \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin \text{AF}(P)))))),$$

其中 P 是任意被测谓词集, s 表示任何充分测试集, t 是最小充分测试集.

公理 4. 最小固定尺寸性质(minimum fixed-size property).

对于每个被测谓词集合 P ,若存在一些由文字真值组合构成的充分的测试集,那么这些测试集中至少含有一个最小充分测试集.该公理的形式定义为

$$\forall P \exists k: \mathbb{N} \bullet (\forall t \bullet (t \in \text{AF}(P) \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin \text{AF}(P))) \Rightarrow \|t\| = k)).$$

如果一个逻辑覆盖测试准则是可满足的,并且满足最小固定尺寸的性质,那么该准则满足有限可应用性.这一结论可以用下面的定理来描述.

定理 1. 若 LC 是一个逻辑覆盖测试准则,对每个被测谓词集合 P ,如果存在满足 LC 的充分测试集,并且 LC 满足最小固定尺寸的性质,那么 LC 是有限可应用的.该定理的形式定义为

$$\forall P \bullet (\exists s \bullet s \in \text{AF}(P) \wedge \forall P \bullet \exists k: \mathbb{N} \bullet (\forall t \bullet (t \in \text{AF}(P) \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin \text{AF}(P))) \Rightarrow \|t\| = k)) \vdash \forall P \bullet (\exists t \bullet (t \in \text{AF}(P) \wedge \|t\| < \infty))).$$

证明:设 s 是一个满足 LC 的充分测试集,那么

- | | |
|--|-------------------------------------|
| (1) $\forall P \exists s \bullet s \in AF(P)$ | 假设 |
| (2) $\exists s \bullet s \in AF(P)$ | (1), \forall -elim, P 是任意的 |
| (3) $s_0 \in AF(P)$ | (2), \exists -elim |
| (4) $\forall P \bullet (\forall s \bullet (s \in AF(P) \Rightarrow \exists t \bullet (t \in AF(P) \wedge t \subseteq s \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin AF(P))))))$ | 任意充分测试集有一个最小充分测试集 |
| (5) $\forall s \bullet (s \in AF(P) \Rightarrow \exists t \bullet (t \in AF(P) \wedge t \subseteq s \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin AF(P))))$ | (4), \forall -elim, P 是任意的 |
| (6) $s_0 \in AF(P) \Rightarrow \exists t \bullet (t \in AF(P) \wedge t \subseteq s_0 \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin AF(P))))$ | (5), \forall -elim, s_0 是某个 s |
| (7) $\exists t \bullet (t \in AF(P) \wedge t \subseteq s_0 \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin AF(P))))$ | (3), (6), \Rightarrow -elim |
| (8) $t_0 \in AF(P) \wedge t_0 \subseteq s_0 \wedge (\forall t' \bullet (t' \subseteq t_0 \wedge t' \notin AF(P)))$ | (7), \exists -elim |
| (9) $t_0 \in AF(P) \wedge (\forall t' \bullet (t' \subseteq t_0 \wedge t' \notin AF(P)))$ | (8), \wedge -elim |
| (10) $\forall P \bullet \exists k: \mathbb{N} \bullet (\forall t \bullet (t \in AF(P) \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin AF(P))) \Rightarrow \ t\ = k))$ | 最小固定尺寸性质, 公理 4 |
| (11) $\forall P \bullet (\forall t \bullet (t \in AF(P) \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin AF(P))) \Rightarrow \ t\ = k_P))$ | (10), Skolen 定理 |
| (12) $\forall t \bullet (t \in AF(P) \wedge (\forall t' \bullet (t' \subseteq t \wedge t' \notin AF(P))) \Rightarrow \ t\ = k_P)$ | (11) \forall -elim, t 是任意的 |
| (13) $t_0 \in AF(P) \wedge (\forall t' \bullet (t' \subseteq t_0 \wedge t' \notin AF(P))) \Rightarrow \ t_0\ = k_P$ | (12) \forall -elim, t_0 是某个 t |
| (14) $\ t_0\ = k_P$ | (9), (13), \Rightarrow -elim |
| (15) $t_0 \in AF(P)$ | (9), \wedge -elim |
| (16) $t_0 \in AF(P) \wedge \ t_0\ = k_P$ | (14), (15), \wedge -int |
| (17) $\exists t \bullet (t \in AF(P) \wedge \ t\ = k_P)$ | (16), \exists -int |
| (18) $\forall P \bullet (\exists t \bullet (t \in AF(P) \wedge \ t\ = k_P))$ | (17), \forall -int |
| (19) $\forall P \bullet (\exists t \bullet (t \in AF(P) \wedge \ t\ < \infty))$ | (18), $<$ -transfer |

在上述证明步骤中,右列给出的是证明证据,所用到的谓词逻辑推理规则可参见文献[13,14].

这条定理表明,一个逻辑覆盖测试准则必须同时满足可满足性和最小固定尺寸性质才蕴涵该准则也满足有限可应用性,仅仅满足一个前提条件并不能推出该准则满足有限可应用性.例如,组合覆盖准则满足最小固定尺寸性质,即,若存在某个满足该准则的测试集,那么该集合的大小是固定的.但实际上,满足该准则的测试集并不总是存在的.所以,组合覆盖准则并不满足有限可应用性.

2.3 充分测试集的完全性

通常期望测试充分性准则具有的一个性质是完全性,也就是某些满足该准则的测试集能够完全覆盖被测的点.本文针对逻辑覆盖测试准则的特点,提出对于逻辑覆盖度准则而言,一个充分测试集是完全的意味着可以根据该测试集构造出被测谓词的 DNF(disjunctive normal form).

这里判断一个充分的测试集的关键是找到一个有效的算法,使得从该测试集可以重构被测谓词集的等价的 DNF.本文给出如下从测试集构造被测谓词集的 DNF 的算法.

生成算法(generating algorithm,简称 GA).

(1) 对每个谓词 $p \in P$ 和充分测试集中的每个文字真值组合,若该真值组合能够使谓词 p 为 true,那么根据下面两条规则将该文字真值组合转换成一个谓词.

- a) 对于该真值组合中每个文字 l 的真值取值,若该取值是 $l = \text{true}$,那么得到形如 l 的文字;若该取值是 $l = \text{false}$,那么得到形如 $\neg l$ 的文字.

b) 用逻辑操作符“ \wedge ”连接所有得到的文字就形成了转换成的谓词。

(2) 用逻辑操作符“ \vee ”连接在第(1)步中得到的谓词就形成了与 p 等价的 p 的 DNF.

根据 GA 可以看出,对于任意被测谓词集合 P ,GA 作用在 P 中所有可能的文字真值取值的组合(即 $TL(P)$)上可以得到 P .这一结论可以形式地定义如下:

$$\forall P \bullet (GA(TL(P))=P).$$

对于上面提到的谓词 $p=A \vee B$,满足组合覆盖准则的充分测试集是 $\{A=\text{true} \wedge B=\text{true}, A=\text{true} \wedge B=\text{false}, A=\text{false} \wedge B=\text{true}, A=\text{false} \wedge B=\text{false}\}$.其中,使得 p 为 true 的文字真值组合是 $A=\text{true} \wedge B=\text{true}, A=\text{true} \wedge B=\text{false}$ 和 $A=\text{false} \wedge B=\text{true}$.对于这两个真值组合应用上面的生成算法,可以得到如下谓词:

$$(A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge B).$$

显然,该谓词是被测谓词 p 的等价的 DNF.因此,该充分测试集能够重构谓词 p .

公理 5 说明了这一性质,文献[5]提出了使用有效算法的概念来对该公理进行形式化,本文中使用了上面给出的生成算法 GA 对该公理进行形式化.

公理 5. 完全性(completeness property).

对每个被测谓词集 P ,将 GA 应用到 P 的任何一个充分测试集上都会产生与 P 等价的析取范式 DNF.该公理的形式定义为

$$\forall P \bullet (\forall t \bullet (t \in AF(P) \Rightarrow GA(t)=P)).$$

2.4 代表性(representative property)

文献[5]提出了路径测试准则的代表性,并使用代表关系来对代表性进行形式化,这一方法也在本文中沿用,但本文针对逻辑覆盖测试准则的特点给出了代表关系的具体定义.代表性质的产生是由于在划分测试中关注的问题之一是,是否从一个划分中选择出的元素能够代表该划分中的其余元素.当我们应用逻辑覆盖测试准则来产生测试用例时,我们通常希望对每个被测谓词,测试集中至少含有两个文字真值组合.其中,一个使得谓词取值 true,另一个使得谓词取值 false.因此,这里定义,若文字真值组合 x 和 y 能够使得其对应的同一个谓词取得相同的真值,那么 x 可以代表 y .定义 9 给出期望的逻辑覆盖测试准则的代表性.

定义 9. 代表关系(representative relation,简称 RR).

RR 是一种定义在被测谓词集 P 中文字真值组合上的等价二元关系. $RR(tc_1, tc_2)$ 表示文字真值组合 tc_1 和 tc_2 能够使得其对应的同一个谓词取得相同的真值.

有了代表关系的定义,逻辑覆盖测试准则的代表性可以用如下的公理来表示.

公理 6. 代表性(representative property).

对每个被测谓词集合 P ,根据 RR,对 P 的任何一个充分测试集 t , P 中每个可能的文字真值组合都至少能够由 t 中的一个真值组合 x 代表.该公理的形式定义为

$$\forall P \bullet (\forall t \bullet (t \in AF(P) \Rightarrow \forall tc \exists x \bullet (tc \in TL(P) \wedge (x \in t \wedge RR(x, tc))))).$$

2.5 相关性(correlative property)

当我们根据谓词来产生测试用例去测试程序时,我们通常希望知道是否在程序中原来影响谓词的文字确实如期望的那样对谓词产生了影响.也就是说,我们期望知道在一个测试集中是否包含了足够的测试用例来检查文字和谓词之间的相关性.这里,相关性是指文字和谓词之间的决定关系.下面的公理描述了该性质.

公理 7. 相关性(correlation property).

对每个被测谓词集合 P , P 中的每个谓词 p 以及 p 中的每个文字 l ,若 p 能够由 l 决定,那么必定存在一些充分的测试集,每个这样的测试集至少包含两个分别包含 $l=\text{true}$ 和 $l=\text{false}$ 的 p 的文字真值组合.其中,一个文字真值组合使得 p 为 false,另一个使得 p 为 true.该公理的形式定义为

$$\forall p, P \bullet (p \in P \wedge \forall l \bullet (l \in L_p \wedge l \text{ determine } p) \Rightarrow AF(P) \neq \emptyset \wedge \forall t \bullet (t \in AF(P) \wedge \exists tc_1, tc_2 \bullet (tc_1 \in TL(p) \wedge tc_2 \in TL(p) \wedge tc_1 \in t \wedge tc_2 \in t \wedge p(tc_1) \neq p(tc_2) \wedge (l = \text{true} \in tc_1 \wedge l = \text{false} \in tc_2 \vee l = \text{true} \in tc_2 \wedge l = \text{false} \in tc_1)))).$$

3 评 估

本节根据前一节中给出的 7 条公理对第 1 节中介绍的逻辑覆盖测试准则进行评估,以分析每个逻辑覆盖测试准则的特点,最后对比较的结果用表格形式总结出来.通过这一评估,可以从总体上对逻辑覆盖测试准则的特性有所了解,从而为应用这些逻辑覆盖测试准则提供了一个参照.这一评估结果还可以与其他比较测试准则的方法结合起来,帮助测试人员在实践中选用合适的测试准则.

3.1 有限可应用性、空集不充分性和单调性

从逻辑覆盖测试准则的定义可以看出,除了组合覆盖准则之外,其他的逻辑覆盖测试准则都满足有限可应用性.组合覆盖准则不满足有限可应用性的原因在于,当被测谓词和文字的个数非常多的时候,满足该准则的测试用例集的数目趋近于无穷.此外,所有逻辑覆盖测试准则都满足空集不充分性和单调性.

3.2 最小固定尺寸性质

谓词覆盖准则满足最小固定尺寸性质.因为对一个谓词而言,存在一个能够覆盖谓词的两种真值取值的最小测试集,该集合中元素的数目为 2.其中,一个元素使得谓词取 true,另一个使得谓词取 false.因此,若一个被测谓词集合含有 n 个谓词,那么存在一个满足谓词覆盖准则的最小充分测试集,其大小为 $2n$.

文字覆盖准则也满足最小固定尺寸的性质.因为,对每个谓词而言,存在一个满足文字覆盖的最小充分测试集,该集合中的元素数目为 2,其中,一个元素是使得所有文字取 true 的真值组合,另一个元素是使得所有文字取 false 的真值组合.因此,若一个被测谓词集合含有 n 个谓词,那么存在一个满足文字覆盖准则的最小充分测试集,其大小为 $2n$.

组合覆盖准则也满足最小固定尺寸的性质.因为,若存在一个满足该准则的充分测试集,那么这个测试集是唯一的,它含有被测谓词集中的文字的所有可能的真值组合.也就是说,该准则的最小充分测试集只有 1 个.因此,组合覆盖准则满足最小固定尺寸的性质.

一般活动文字覆盖、相关活动文字覆盖和受限活动文字覆盖都满足最小固定尺寸的性质.因为,对于一个谓词 p 而言,主要活动文字的数目就是该谓词中文字的数目,这里用 n_{L_p} 来表示该数目.这 3 个准则都要求其充分测试集中要包含每个主要文字的两个不同的真值取值.因此,对于 p 来说,存在一个最小充分测试集,该集合中应含有 $2 * n_{L_p}$ 个文字真值组合.若一个被测谓词集合含有 m 个谓词,那么对于这 3 个准则中任何一个准则,存在一个

相应的最小充分测试集,其大小为 $\sum_{i=1}^m 2 * n_{L_{p_i}}$,其中 p_i 表示一个谓词.从这个分析过程可以看出,这 3 条准则都满足最小固定尺寸的性质.

3.3 完全性

除了组合覆盖准则以外,所有其他逻辑覆盖测试准则都不满足完全性.因为,根据本文给出的生成算法 GA,要构造被测谓词集合,满足某个逻辑覆盖测试准则的充分测试集中必须包含该谓词集中所有可能的文字真值的组合.而只有满足组合覆盖准则的测试集才满足这种性质.因此,只有组合覆盖准则满足完全性.

3.4 代表性

根据定义 9 中的代表关系的定义,只要测试用例集中的文字真值组合能够分别使得谓词取真值 true 和 false,那么该测试用例集中的测试用例就可以代表所有其他的文字真值组合.因此,谓词覆盖准则、组合覆盖准则、相关活动文字覆盖准则和受限活动文字准则满足代表性.其他逻辑覆盖测试准则不满足代表性.

3.5 相关性

相关性表明,若一个谓词的真值可以由其主要文字决定,那么任何一个满足某测试准则的测试集中必须包含两个使得主要文字分别为 true 和 false 的文字真值组合,并且这两个真值组合分别使该谓词取不同的真值.根据这一描述,显然一般活动文字覆盖、谓词覆盖和文字覆盖都不满足这种性质.因为这 3 个准则中的每一个都不能保证满足该准则的充分测试集中一定包含满足上述性质的两个文字真值组合.

对于组合覆盖准则和受限活动文字覆盖准则而言,有些时候,当谓词的真值可以由其文字决定,并且谓词中的文字取值受到一定限制时,未必一定能够找到满足这两个准则的充分测试集.例如,对于谓词 $p=A \wedge (B \vee C)$,若给出限制 $\neg A \leftrightarrow B$,那么满足该限制的所有可能的文字真值组合为 $\{A=\text{true} \wedge B=\text{false} \wedge C=\text{true}, A=\text{true} \wedge B=\text{false} \wedge C=\text{false}, A=\text{false} \wedge B=\text{true} \wedge C=\text{true}, A=\text{false} \wedge B=\text{true} \wedge C=\text{false}\}$.所有的测试集只能是这个集合的子集.显然,找不到一个满足组合覆盖或受限活动文字覆盖准则的子集.所以,当在满足某些限制条件的情况下,一个谓词能够由其文字决定时,未必总能找到满足组合覆盖或受限活动文字覆盖准则的充分测试集.因此,这两个准则不满足相关性.

对于相关活动文字而言,从其定义可以看出,只要一个谓词能够由其主要文字决定,那么一定存在一个满足该准则的充分测试集,该集合中至少包含两个使得主要文字分别为 true 和 false 的文字真值组合,并且这两个真值组合分别使该谓词取不同的真值.例如,考虑上一段中的例子,测试集 $\{A=\text{true} \wedge B=\text{false} \wedge C=\text{true}, A=\text{true} \wedge B=\text{false} \wedge C=\text{false}, A=\text{false} \wedge B=\text{true} \wedge C=\text{true}, A=\text{false} \wedge B=\text{true} \wedge C=\text{false}\}$ 就满足相关活动文字覆盖准则,并且该集合也满足前面描述的相关性.因此,相关活动文字覆盖准则满足相关性.

综上所述,每个逻辑覆盖测试准则都有各自的特点,满足不同的特性.下表给出了对评估结果的一个总结.表中的“Yes”表示满足某个公理性质,“No”表示不满足某个公理性质.

Table 1 Assessment of logic coverage testing criteria

表 1 逻辑覆盖测试准则的评估结果

Property	Predicate coverage	Literal coverage	Combinational coverage	General active literal coverage	Correlated active literal coverage	Restricted active literal coverage
Finite applicability	Yes	Yes	No	Yes	Yes	Yes
Inadequate empty set	Yes	Yes	Yes	Yes	Yes	Yes
Monotonicity	Yes	Yes	Yes	Yes	Yes	Yes
Minimum fixed-size	Yes	Yes	Yes	Yes	Yes	Yes
Completeness	No	No	Yes	No	No	No
Representative	Yes	No	Yes	No	Yes	Yes
Correlative	No	No	No	No	Yes	No

4 结 论

根据上文的评估结果,我们可以得到下述结论.(1) 第 2 节给出的公理系统是可满足的.因为对每个公理而言,至少存在一个逻辑覆盖测试准则满足该公理给出的性质,也就是满足该公理.(2) 相关活动文字覆盖准则是最实用的一个逻辑覆盖测试准则.因为,该准则是有限可应用的,并且与其他有限可应用的逻辑覆盖测试准则相比,它满足的公理最多.这一结论也符合我们在应用逻辑覆盖测试准则推导测试用例时的一种直觉.(3) 该公理系统包含的都是有关逻辑覆盖测试准则的正面性质.(4) 一个逻辑覆盖测试准则包含另一个准则并不意味着该准则就比另一个准则更实用.如,受限覆盖准则包含相关覆盖准则,但相关覆盖准则却更为实用.综上所述,本文的公理化评估结果为测试人员了解每一个逻辑覆盖测试准则的性质,在实践中区别使用不同的逻辑覆盖测试准则提供了参考.

References:

- [1] Goodenough JB, Gerhart SL. Toward a theory of test data selection. IEEE Trans. on Software Engineering, 1975,1(2):156~173.
- [2] Amman P, Offutt J. Using formal methods to derive test frames in category-partition testing. In: IEEE 9th Annual Conf. on Computer Assurance. Gaithersburg: IEEE Computer Society Press, 1994. 69~80.
- [3] Amman P, Offutt J. Coverage criteria for logical expressions. In: Stephanie K, ed. Proc. of the 14th Int'l Symp. on Software Reliability Engineering. Denver: IEEE Computer Society Press, 2003. 99~107.
- [4] Black PE, Okun V, Yesha Y. Mutation operators for specifications. In: Anne R, ed. Proc. of the 5th IEEE Int'l Conf. on Automated Software Engineering. Grenoble: IEEE Computer Society Press, 2000. 81~88.

- [5] Zhu H. Axiomatic assessment of control flow-based software test adequacy criteria. *Software Engineering Journal*, 1995,10(9): 194~204.
- [6] Weyuker EJ. Axiomatizing software test data adequacy. *IEEE Trans. on Software Engineering*, 1986,12(12):1128~1138.
- [7] Gourlay J. A mathematical framework for the investigation of testing. *IEEE Trans. on Software Engineering*, 1983,9(6):686~709.
- [8] Parrish A, Aweben SH. Analysis and refinement of software test data adequacy properties. *IEEE Trans. on Software Engineering*, 1991,17(6):565~581.
- [9] Parrish A, Aweben SH. Clarifying some fundamental concepts in software testing. *IEEE Trans. on Software Engineering*, 1993,19(7):742~746.
- [10] Offutt J, Liu SY, Abdurazik A, Amman P. Generating test data from state-based specifications. *The Journal of Software Testing, Verification, and Reliability*, 2003,13(1):25~53.
- [11] Dick J, Faivre A. Automating the generation and sequencing of test cases from model-based specifications. In: *Industrial-Strength Formal Methods. Lecture Notes in Computer Science*, Odense: Springer-Verlag, 1993. 268~284.
- [12] Hierons RM. Testing from a Z specification. *The Journal of Software Testing, Verification, and Reliability*, 1997,7(1):19~33.
- [13] Miao HK, Li G, Zhu GM. *Software Engineering Language—Z*. Shanghai: Scientific and Technical Documents Publishing House, 1999 (in Chinese).
- [14] Diller AZ. *An Introduction to Formal Method*. London: John Wiley & Sons Ltd., 1990.

附中文参考文献:

- [13] 缪准扣,李刚,朱关铭. *软件工程语言——Z*.上海:上海科学技术文献出版社,1999.

第2届全国搜索引擎和网上信息挖掘学术研讨会

征文通知

为促进国内外相关领域科研人员的学术和工作交流,研讨本领域的最新技术进展和发展趋势,以推动搜索引擎和 Web 挖掘技术在中国的发展,2004年“全国搜索引擎和网上信息挖掘学术研讨会”将于2004年11月12日~13日在广州华南理工大学举行。经专家评审录用的会议论文,将在《华南理工大学学报》(自然科学版)增刊正式出版。会议还将举行“中文 Web 检索竞赛”,欢迎单位或个人组队参加。

欢迎高等院校教师、科研院所和企业的业界专家、科研人员、研究生以及业界专业人士踊跃投稿,国家和省部级资助项目支持的论文将优先录用。详情请浏览我们的网站:<http://www.scut.edu.cn/sewm2004>