CMuJava: 一个面向Java程序并发变异测试支持系统[[1]](#footnote-1)\*

孙昌爱, 耿 宁, 顾友达, 代贺鹏

(北京科技大学 计算机与通信工程学院,北京 100083)

通讯作者: 孙昌爱, E-mail: casun@ustb.edu.cn

摘 要: 并发系统由多个并发执行的流程组成,并且流程之间通常显示或隐式地共享一些存储空间.流程之间执行次序的不确定性给并发系统的测试带来了严峻的挑战.并发变异测试是一种基于并发故障的软件测试技术,依据并发变异算子改变待测程序的语法结构,生成并发变异体,通过执行并发变异体来评估测试用例集的充分性和测试技术的有效性.然而,面向Java程序的开源变异测试工具不支持生成并发变异体,人工生成并发变异体需要消耗大量的测试资源.为了提高并发变异测试的效率,本文依据Java并发变异算子设计并实现了一个并发变异体自动生成工具CMuJava,采用经验研究评估了CMuJava生成并发变异体的正确性、完备性及效率.经验研究的结果表明,该工具生成并发变异体的正确性和完备性均达到了100%,并且与人工生成并发变异体相比,生成并发变异体的效率最高提高了大约8000倍.

关键词: 并发系统;并发变异测试; 并发变异算子; 并发变异体; 测试工具

**Abstract**: Concurrent systems are composed of multiple concurrent execution flows, which usually share storage space explicitly or implicitly. Uncertainty of execution order among execution flows makes testing concurrent systems a challenging task. Concurrent mutation testing is a software testing technology based on concurrent faults, which changes the syntax of the program under test according to concurrent mutation operators, generates concurrent mutants to evaluate the adequacy of test case set and the effectiveness of test technology. However, current open source Java-oriented mutation testing tools do not support concurrent mutants generation, and manual generation of concurrent mutants requires a lot of test resources. In order to improve the efficiency of concurrent mutation testing, we designs and implements a concurrent mutation automatic generation tool CMuJava in this paper, which is based on Java concurrent mutation operators, and evaluates the correctness, completeness and efficiency of CMuJava generating concurrent mutants by empirical research. The results of the empirical study show that the correctness and completeness of the tool for generating concurrent mutants are 100%, and the efficiency of generating concurrent mutants is about 5700 times higher than that of generating concurrent mutants manually.

**Key words**: concurrent system; concurrent mutation testing; concurrent mutation operator; concurrent mutant; testing tools

随着多核计算的日益普及,并发程序已经成为人们广泛关注与重视的研究领域.并发程序中存在多个并发执行的流程,流程之间通常显示或者隐式地共享一些存储空间,并且它们的执行次序不确定[1].并发流程之间不确定的相互作用与影响的情形称为执行交错.执行交错的存在使得在完全相同的环境下运行两次同样程序的输出结果可能不同,并且程序故障难以重现,给并发程序的测试带来了严峻的挑战.如何有效地检测并发程序中潜藏的故障,提高并发程序的可靠性成为一个亟待解决的重要问题.

变异测试[2]是一种基于故障的软件测试技术,可以用来评估测试用例集的充分性与测试技术的有效性.测试人员通过分析待测程序并设计出一系列变异算子[3],然后对待测程序应用变异算子来生成大量的变异体[3]。在识别出等价变异体[3]后,若已有测试用例不能杀死所有非等价变异体,则需要设计新的测试用例来提高测试充分性.

Carver将变异测试理论和技术应用到并发程序测试中,提出了并发变异测试[29].在并发变异体生成方面[9],面向Java语言的开源变异测试工具(例如MuJava[4]、Javalance [5]和Jumble [6]等)不支持生成并发变异体,而已有的面向Java并发程序的变异体生成工具要么实现了少量并发变异算子,要么不支持开源使用.研究者在测试过程中通常依据并发变异算子人工生成并发变异体,该方式需要占用大量的测试资源.另一方面,开发面向并发程序的变异测试系统面临如下挑战:(1)提高并发变异体生成效率.一些并发变异算子定义的规则具有多样性,即针对同一并发机制可以产生大量变异体,导致并发变异系统需要多次分析源程序的结构,运用相应的规则改变源程序并生成并发变异体;(2)提高并发系统的可扩展性.随着并发理论与相关技术的不断发展,并发程序越来越复杂,可能出现新的并发变异算子来模拟并发故障.因此,实用的并发变异系统需要具备高效并发变异体生成和可扩展的特点.

MuJava是一个常用的面向Java程序的变异测试工具[34,35,36],其自身具有很好的可扩展性,例如Sun等人将非均匀分布的变异分析技术[27]应用到MuJava,扩展出支持非均匀分布的变异生成系统MuJavaX[33];Kim等人通过扩展MuJava实现了所提出的弱变异与强变异结合的方法[37].MuJava不能基于并发故障生成并发变异体,本文设计并实现了一个自动生成并发变异体的工具CMuJava,提高了面向Java程序的并发变异测试技术的自动化程度.本文的主要贡献有:

1. 基于Brabury等人提出的并发变异算子[7] 扩展了面向Java程序的变异测试工具MuJava,实现了一个在能够生成传统变异体的基础上自动生成并发变异体的工具CMuJava.
2. 采用MSG和Reflection结合的方法来提高并发变异体的生成效率及工具的可扩展性。
3. 以经验研究的方式评估了CMuJava生成并发变异体的正确性、完备性和效率.

本文第1节介绍了并发变异测试相关的背景知识.第2节讨论工具的设计与实现.第3节以经验研究的方式评估工具生成并发变异体的正确性、完备性和效率.第4节对实验结果进行了分析.第5节介绍与本文相关的研究工作.最后,第6节对本文的工作做出了总结并提出将来需要完善的方面.

# 并发变异测试

在传统变异测试中,首先,应用变异算子来改变原始程序P法结构,生成与原始程序有语法差异的变异体集合来模拟待测程序可能存在的故障.然后,在原始程序P和每个变异体上执行所有的测试用例,如果存在测试用例在P和上的执行结果不同,则称变异体为可杀死变异体,如果所有的测试用例在P和上执行的结果完全一致,则称为P的等价变异体.最后,通过式(1)计算变异得分,评估测试用例集的充分性和测试技术的有效性.

(1),

其中,MS(M,T)表示变异得分,killed(M,T)表示可杀死变异体数量,|M|表示生成的所有变异体数量,eqv(M)表示已识别的等价变异体数量.从表达式含义可以看出MS(M,T)取值范围为[0,1],取值越接近1表明测试用例集的充分性越高或测试技术越有效.

与顺序程序相比,并发程序中多个执行流之间不确定的执行交错为测试并发程序带来严峻的挑战.很多研究者从不同的角度出发,针对并发程序提出了一系列故障检测技术,例如:锁集分析[29]、happens-before[30]和并发变异测试技术[31].并发变异测试将传统变异测试技术应用到并发程序中,模拟真实的并发故障,可以有效地提高故障检测效率.

传统变异测试定义为5元组[26],其中,P是原始程序,S是规格说明,D是测试用例集,,表示P中语句的位置,,是位置处的变异算子集合.

与传统变异测试不同,并发变异测试是对程序中特有的并发机制应用变异算子.在传统变异测试模型的基础上,我们用C表示P中的并发机制所在的位置,将并发变异测试定义为6元组,其中,P是原始并发程序,S是规格说明,D是测试用例集,,表示P中语句的位置,,表示P中并发机制的位置,,是位置的变异算子集合,是变异算子的数量.

图1描述了并发变异测试的流程,主要步骤如下:

1. 根据待测并发程序所运用的并发机制选择合适的并发变异算子,然后在待测并发程序上应用并发变异算子来生成相应的并发变异体.
2. 从生成的大量并发变异体中识别出等价并发变异体并删除.
3. 在非等价并发变异体上执行测试用例,并根据并发变异体被识别的情况计算变异得分.如果变异得分满足要求则变异测试结束,否则还需要继续设计测试用例并添加到当前测试用例集中直到变异得分满足要求.



Fig.1 Concurrent Mutation Testing process

图1 并发变异测试流程

**Table 1** concurrent mutation operators for java

表1 java并发变异算子

| 变异算子类别 | 并发变异算子 |
| --- | --- |
| 修改并发方法参数 | MXT—修改方法参数 |
| MSP—修改同步代码块参数 |
| ESP—交换同步块参数 |
| MSF—修改信号量公平性 |
| MXC—修改权限许可和线程数量 |
| MBR—修改同步屏障参数 |
| 修改并发方法调用 | RTXC—删除线程方法调用 |
| RCXC—删除并发机制方法调用 |
| RNA—使用notify()替换notifyAll() |
| RJS—使用join()替换sleep() |
| ELPA—交换权限/锁获取方法 |
| EAN—将原子调用替换为非原子调用 |
| 修改关键字 | ASTK—为方法添加static关键字 |
| RSTK—删除方法static关键字 |
| ASK—为方法添加synchronized关键字 |
| RSK—删除方法synchronized关键字 |
| RSB—删除synchronized代码块 |
| RVK—删除volatile关键字 |
| RFU—删除unlock方法所在的finally代码块 |
| 交换并发对象 | RXO—替换并发机制 |
| EELO—交换锁对象 |
| 修改临界区 | SHCR—移动临界区 |
| SKCR—收缩临界区 |
| EXCR—扩大临界区 |
| SPCR—拆分临界区 |

Brabury等人[7]依据Java(J2SE 5.0)并发机制特性及Java并发程序的故障类型[32],提出了面向Java程序的25种并发变异算子,反应了真实的并发故障,具有代表性和良好的覆盖性.与传统顺序程序的变异测试技术不同,并发变异测试应用并发变异算子来改变待测程序的语法结构,生成并发变异体.并发变异算子作为并发变异测试的基础,定义了改变待测并发程序语法结构的规则,表1列出了25种并发变异算子的信息,详细的并发变异算子信息和变异体示例可以在原始论文[7]和技术报告[28]中找到.

并发变异测试执行过程主要包括并发变异体生成和并发变异体执行.针对复杂的并发程序,应用并发变异算子生成的变异体数量很大,人工生成需要消耗大量的资源,同时执行大量的并发变异体的速度缓慢.为了提高并发变异测试执行效率,需要设计自动化的并发变异分析系统.Gligori等人[23]提出变异测试优化工具MutMut,他们应用了4种技术优化了变异体包括并发变异体的执行,据报道该工具对变异测试的时间降幅达77%,然而由于缺少并发变异体生成工具,实验中应用到的并发变异体均由人工生成.Madiraju等人[24]提出了一种部分变异的方法,选择程序性中的复杂部分或方法进行变异,实现的工具Paraµ应用该方法确实减少了并发变异体的数量,但是工具仅仅实现了3个并发变异算子,不利于广泛应用.Kusao等人[21]提出的Comutation应用了选择变异的方法提高并发变异测试效率,即应用并发变异算子的子集生成并发变异体,从而减少了并发变异体的数量.以上两种方法均通过减少并发变异体数量提高了变异测试效率,但是对于并发变异体的生成,仍然采用修改原始代码的副本来单独地生成每个变异体,这种方法的低效率导致变异测试整个过程缓慢.

# CMuJava的设计与实现

基于Barbury等人提出的25种并发变异算子,我们开发了一个面向Java程序的并发变异体自动生成工具CMuJava,工具采用反射技术[38,39]分析原始程序中的并发机制,采用MSG(Mutant Schema Generation)[8]方法提高并发变异体的生成效率.接下来详细讨论CMuJava设计与实现的关键问题.

## CMuJava的基本原理

图3描述了CMuJava的基本原理.首先,工具读取原始并发程序及其依赖程序,按照Java程序的语法结构拆分并进行语法分析,得到程序中的并发机制;然后对程序中的并发机制应用并发变异算子,依据变异算子定义的规则对程序进行修改生成并发变异体;最后,对原始程序和生成的并发变异体执行测试人员编写的JUnit测试脚本,自动获取测试结果,统计最终的变异得分以及每一个变异体的“杀死”信息.



Fig.3 Basic principles of CMuJava

图3 CMuJava基本原理

生成变异体的关键步骤是按照变异算子所定义的规则对待测程序进行修改,所以首先需要对待测程序进行解析并获取待测程序的语法信息.传统的面向过程程序的变异分析系统中用一个普通的语法分析器解析原始程序,语法分析器生成抽象语法树来表示源代码,通过修改抽象语法树生成变异体.但是,传统的抽象语法树不能直接访问程序的面向对象特征(例如定义一些属性及方法)[41],而并发变异算子应用到程序并发机制中时,需要检测程序中面向对象的特征,如.因此,仅通过生成抽象语法树进行语法分析不能充分地应用并发变异算子.Chevalley[41]提出用反射解决传统的语法分析技术不适用于面向对象程序变异分析的问题.

反射技术允许程序访问其内部结构和行为,操纵该程序,从而根据另一程序提供的规则修改其行为.反射技术被应用到变异分析中的原因有:(1)允许程序员通过提供表示类定义的逻辑结构的对象来提取有关类的面向对象的相关信息,有助于实现变异分析系统的第一个阶段—解析程序.(2)提供了一个API,可以在执行过程中更改程序的行为,有助于实现变异分析系统中变异体生成的功能.(3)允许实例化对象并动态调用方法.

基于以上原因,我们将反射技术应用到并发变异分析系统CMuJava中.Java使用专用API提供内置的反射功能,允许Java程序执行诸如请求给定对象的类,查找该类中的方法以及调用这些方法之类的功能.但是,定义的Java语言不提供完整的反射功能.具体来说,Java仅支持内省,即可以修改数据结构但不支持更改程序行为的能力.已经提出了几种反射系统[42,43,44,45]来补充Java反射API的缺陷.其中,OpenJava[42]是一个编译时反射系统,其主要思想是元程序处理表示程序逻辑实体而不是标识序列或抽象语法树的类元对象.OpenJava在编译时使用元信息使得其提供了比运行时反射系统更好的性能.我们在CMuJava中使用了OpenJava对待测并发程序进行语法解析,解析完成后会将待测程序的所有信息存储到一个元对象中.每一个程序在解析之后都会得到一个对应的元对象,元对象的设计与传统的抽象语法树设计思想相似,都是通过一个树状的结构来表达程序的语法信息,不同之处在于元对象可以表示抽象语法树无法表示的程序逻辑结构信息.因此,可以有效解决解析并发机制的问题.

在生成并发变异体阶段,传统的方法只能依据并发变异算子的规则修改原始代码的副本单独地生成每个变异体,生成速度缓慢.另外,由于需要存储和编译大量变异体,对空间的需求也比较大.我们采用MSG技术降低生成并发变异体的空间开销,进而提高并发变异测试的效率.图4描述了生成和执行并发变异体的过程.



Fig.4 Mutation process for concurrent mutants

图4 并发变异体的变异分析过程

MSG技术可以将一个程序的所有变异体都编码到一个特殊的参数化程序中,这个特殊的参数化程序叫做元变异体.元变异体包含了待测程序所有变异体信息和功能,且运行过程中编译速度不变.

例子:我们以并发变异算子MSP(Modify Synchronized Parameter)为例介绍一下MSG技术在CMuJava中的使用.MSP变异算子所定义的规则就是更改synchronized代码块的同步参数,将其由this关键字替换为其他对象,或者由其他对象修改为this关键字.图4给出了三个具体的MSP变异体的例子.从MSP的变化规则可以知道,每次被更改的地方都只有synchronized的同步参数.将以上源程序的synchronized代码块进行抽象可以得到如式(4)所示的表达式:

Synchronized (getParameter(orignalArg, objList)){…} (4),

这样在程序运行的时候就可以通过getParameter方法动态的获取同步参数并生成相应的变异体.getParameter方法有两个参数,其中originalArg表示的是原来的同步参数,objList表示的是可以用于替换原来的同步参数的对象列表.在对一个并发程序应用MSP并发变异算子时,CMuJava在确定当前程序符合MSP并发变异规则之后,会自动获取程序的所有成员变量.然后将synchronized代码块的原始参数和成员变量列表传递给此方法,就可以生成待测程序的所有MSP变异体.将程序中的语句改变成这种形式后的参数化程序即为元变异体.

|  |  |
| --- | --- |
| 源程序：  ...  private Object obj1 = new Object();  private Object obj2 = new Object();  private Object obj3 = new Object();  ...  synchronized(this){...}  ... | MSP变异体1：  ...  private Object obj1 = new Object();  private Object obj2 = new Object();  private Object obj3 = new Object();  ...  synchronized(obj1){...}  ... |
| MSP变异体2：  ...  private Object obj1 = new Object();  private Object obj2 = new Object();  private Object obj3 = new Object();  ...  synchronized(obj2){...}  … | MSP变异体3：  ...  private Object obj1 = new Object();  private Object obj2 = new Object();  private Object obj3 = new Object();  ...  synchronized(obj3){...}  … |

Fig.4 Example of MSP mutant

图4 MSP变异体示例

## CMuJava的设计

目前存在多个自动化的变异测试工具,具有代表性的有CMuJava、Javalance、Jumble、Major[40]等.其中,MuJava是一个支持传统变异体生成的开源变异分析系统,具有良好的代表性和可扩展性.我们选择集成和扩展MuJava的功能设计与实现CMuJava,使其在能够生成传统变异体的基础上支持并发变异体的生成.图6展示了CMuJava的架构.

## CMuJava实现的关键问题

并发变异算子可以模拟程序中真实的并发故障,随着并发程序的发展,可能会发现新的并发故障,因此会需要新的并发变异算子.目前,CMuJava工具实现了25种并发变异算子,为了将来可以应用新的并发变异算子,我们使用了设计模式提高工具的可扩展性.

.CMuJava在完成一个Java程序解析得到一个元对象之后,需要在此对象的各个节点上运用一系列并发变异算子进行可变异点查找,然后对可变异点进行语法更改生成并发变异体.如果将所有的可变异点判断逻辑和变异体输出逻辑都放到元对象的每个节点内部,每增加一个变异算子都需要对元对象内部的一个或多个节点的操作方法进行修改以完成此变异算子的功能,这样会导致相关节点的代码变得越来越复杂,在降低了代码的可读性的同时也会增加后期的维护成本,同时不易于扩展.

Visitor设计模式[10]适用于被访问对象结构相对稳定,且访问者操作不确定的情况.CMuJava系统所涉及的元对象的数据结构非常稳定,不存在需要新加入节点或者对已有节点进行修改的情况,但是存在到对元对象的大量不确定操作,涉及到众多并发变异算子.为了提高工具的可扩展性,便于添加新的变异算子,我们在CMuJava实现过程中采用了Visitor设计模式.

图7描述了CMuJava的核心类结构.其中,CompilationUnit充当结构对象的角色,它负责存储所有的节点对象,并提供了相应的方法来遍历节点对象.访问者类Mutator和MutatorWriter均实现了ParseTreeVisitor类声明的对所有的节点的visit方法.当有新的并发变异算子出现需要添加到系统中时,只需要编写新的访问者类继承Mutator和MutatorWriter类,然后重写自己需要访问的节点的visit方法即可,极大地提高了工具的可扩展性.



Fig.7 The core class diagram of CMuJava

图7 CMuJava核心类图

## 工具演示

本文基于Java开发了工具原型,图8展示了工具的界面.界面左侧列出了所有原始待测程序,首先,测试人员在“File”栏中选择需要生成变异体的原始程序(图中以实验对象ThreadID程序为例);然后在“Java Mutation Operator”栏中选择原始程序适用的变异算子,其中最左侧为19种方法级别变异算子,中间为28种类级别变异算子,最右侧为本文实现的25种并发级别变异算子;最后,测试人员点击最下方黄色的“Generate”按钮,等待系统完成变异体生成工作,系统会依次读取选择的待测程序,为每个程序生成相应的元对象,然后在每个元对象上依次应用选中的变异算子生成相应的变异体并输出到指定目录.

变异体生成完成后,CMuJava提供了查看变异体的功能,例如,点击“Concurrent Mutants Viewer”可以查看生成的所有的并发变异体信息.图9展示了查看ThreadID程序并发变异体的界面,点击具体变异体名,界面会显示源程序信息以及变异体中被修改的位置信息.另外CMuJava还集成了MuJava执行变异体的功能,测试人员通过选择测试用例执行生成的变异体,最终得到测试结果.

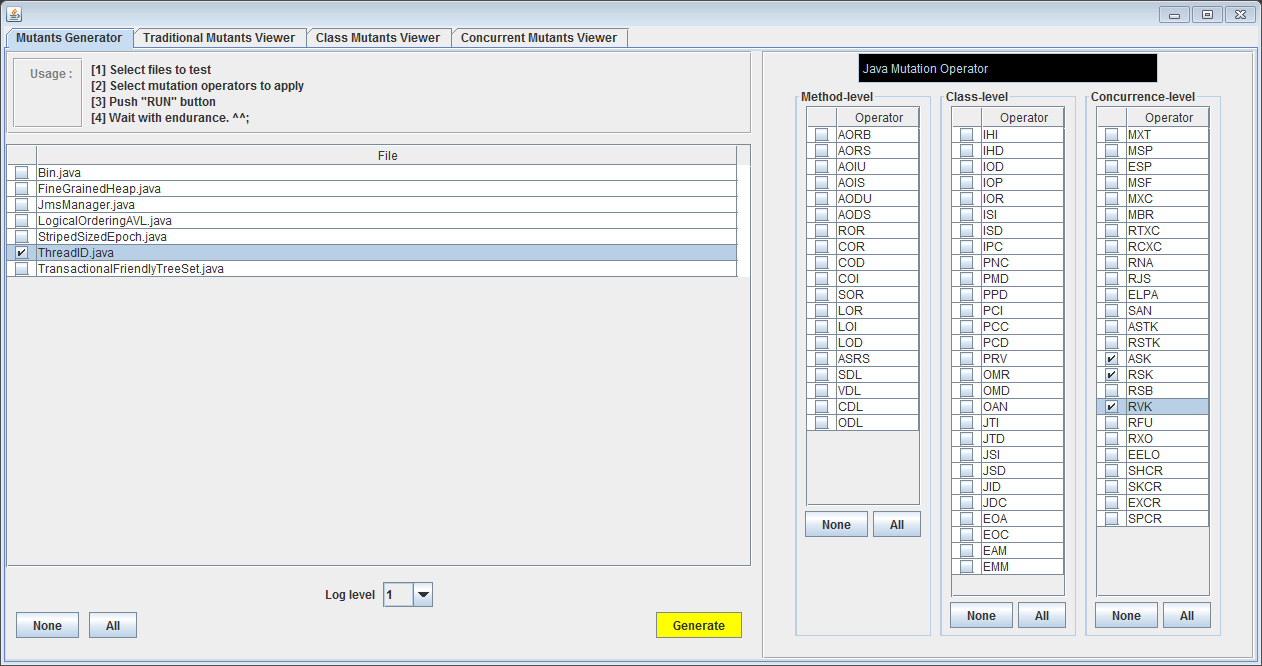


Fig.8 GUI of CMuJava

图8 CMuJava界面

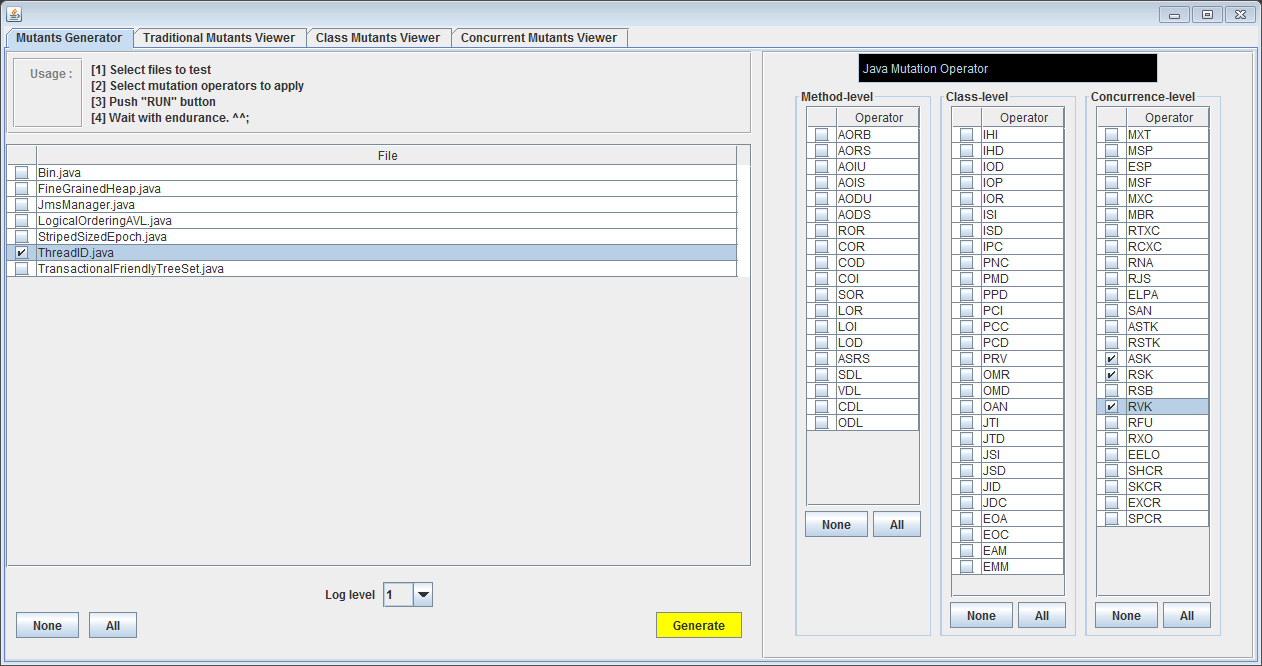


Fig.9 GUI of Concurrent Mutation Viewer

图9 并发变异体查看界面

# 经验研究

本文选取7个真实的并发测试基准程序验证所提工具CMuJava的性能,本节对实验的设置和细节进行介绍.

## 研究问题

通过实例研究,我们希望回答如下三个问题.

1. 使用CMuJava工具是否能够提高生成并发变异体的正确率?

使用一些被广泛使用的并发程序来进行实验,验证CMuJava是否能够识别程序内部所有的并发机制并生成正确的并发变异体,然后计算正确率,将实验结果和人工生成的并发变异体的正确率进行对比分析,评估工具是否提高了正确率.

1. 使用CMuJava工具生成并发变异体是否能够提高生成的并发变异体的完备性?

首先分析并发程序内部所使用的并发机制确定程序理论上应该生成的并发变异体的数量,将工具生成的变异体和人工生成的变异体数量分别与理论数量进行对比分析,评估工具生成变异体是否提高了完备性.

1. 使用CMuJava工具是否能够提高生成并发变异体的效率?

记录工具生成所有并发变异体所需要的时间,然后将实验结果和人工生成并发变异体需要的时间进行对比分析,评估工具的运行效率.

## 实验对象

为了让程序的规模尽可能的与现实情况比较接近,我们选取了覆盖了从几十行到上千行的不同复杂度的7个并发程序:Bin[12]、ThreadID(TID) [12]、FineGrainedHeap(FGH) [12]、StripedSizedEpoch(SSE)、Logi-calOrderingAVL(LOAVL)、TranscationalFriendlyTreeSet(TFTS)和JmsManager(JM).7个并发测试基准程序涵盖了Java语言的常用并发机制,不同的并发程序使用的并发机制也不完全相同.表2列出了每个实验对象的名字、代码行数以及可应用的并发变异算子.程序1~3是三个并发测试基准程序.程序4~6来自于一个专门的并发测试基准程序库Synchrobench[13],程序7来自于Apache开源库下的日志组件Log4j[[2]](#footnote-2).

**Table 2** The information of experimental object

表2 实验对象信息

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 程序名 | 代码行数 | 可应用的并发变异算子 |
| 1 | Bin | 42 | RSK |
| 2 | TID | 50 | ASK, RSK, RVK |
| 3 | FGH | 225 | ELPA, RCXC |
| 4 | SSE | 148 | MSP, RTXC, RNA, RSB, RVK |
| 5 | LOAVL | 1088 | EELO, ELPA, RCXC, RVK, RXO |
| 6 | TFTS | 617 | ASK, RJS, RSK, RVK |
| 7 | JM | 490 | ASK, RVK, MSP, SHCR, SPCR SKCR, EXCR |

## 度量指标

针对前文提出的研究问题,本次经验研究的自变量为提出的并发变异体生成工具CMuJava生成变异体和测试人员人工生成变异体.其中CMuJava为实验组,测试人员为对照组.两实验组分别对7个程序进行变异体生成实验.

本文采用正确率指标来评估生成并发变异体的正确性,其计算公式如式5所示,

(5).

采用统计生成并发变异体的数量的方法来评估生成并发变异体的完备性,根据实验结果计算CMuJava和测试人员人工生成的所有正确的并发变异体数量,并与待测程序理论上可能生成的并发变异体数量进行对比,判断两种方式是否能够正确地识别出程序里所有的可变异点并生成相应的并发变异体并进行比较.

通过记录CMuJava解析待测程序和生成并发变异体的时间,然后对比人工对相同程序进行并发变异体生成实验所需要的时间,来评估CMuJava生成并发变异体的效率.

## 实验设置

1. 实验环境

CMuJava运行在64位Windows 10操作系统,Java环境为JDK 1.8主要配置为:Intel Core i7-7700HQ CPU、16GB运行内存.测试人员手动生成并发变异体时所使用的集成开发环境为Eclipse.

1. 实验步骤

本文实验基本流程如下:

1. 从实验对象中选取一个待测程序,然后开始计时.
2. 测试人员需要先阅读程序,了解程序的基本功能和程序所使用的并发机制.CMuJava需要读取待测程序并进行解析,生成待测程序的元对象.
3. 测试人员需要理解并发变异算子所定义的规则和要求,然后根据并发变异算子寻找待测并发程序里是否存在相应的可变异点.CMuJava也需要根据并发变异算子的规则对待测程序所生成的元对象进行遍历,寻找符合并发变异算子变异规则的可变异点.如果找到相应的可变异点则进行第4步.否则继续根据下一个并发变异算子的规则进行可变异点查找直至完成所有并发变异算子的判断.
4. 按照并发变异算子所定义的规则对识别的可变异点进行语法变化生成对应的并发变异体.
5. 判断是否已生成当前程序的所有并发变异体,如果没有则继续生成下一个并发变异体.如果生成完毕则进入下一步.
6. 停止计时,计算CMuJava和测试人员完成该待测程序的并发变异体生成工作所使用的时间,同时分别统计CMuJava和测试人员生成正确并发变异体的数量.
7. 判断是否所有的待测程序都已经完成生成并发变异体的工作,如果没有完成
8. 则回到步骤(1),继续选取下一个待测程序进行实验.如果已经完成,则分别计算CMuJava和测试人员所使用的平均时间.

在完成以上所有步骤后可以得到测试人员和CMuJava生成并发变异体的实验数据,然后对实验数据进行对比分析,评估使用CMuJava对于提高生成并发变异体的正确率、完备性和效率的作用.

## 有效性威胁

采用对比实验的方法验证工具的有效性,同时为了保证数据的正确性和代表性,采用多次试验并求取平均值的方法,消除了由于实验数据不准确造成的外部有效性威胁,使实验结果更具说服力.

另一个重要的外部有效性威胁是实验对象的代表性问题.我们选取的7个常用的并发程序作为实验对象,其中,Bin、ThreadID和FineGrainedHeap三个程序,课题组前期已经应用其做了大量相关实验,能够保证程序的正确性和代表性;StripedSizedEpoch、Logi-calOrderingAVL、TranscationalFriendlyTreeSet三个程序选自专门的并发测试基准程序库Synchrobench,也可保证代表性.事实上,所有实验对象的研究结果具有一致性.

# 实验结果与分析

表3列出了测试人员和CMuJava分别对7个程序生成并发变异体的正确率数据.表4总结了两种方式生成并发变异体的数量.

**Table 3** Correct rate of concurrent mutants

表3 生成并发变异体正确率

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 生成变异体的对象 | Bin | TID | FGH | SSE | TFT | LOAVL | JM |
| tester1 | 100% | 80% | 67% | 100% | 100% | 100% | 100% |
| tester2 | 100% | 60% | 100% | 100% | 100% | 100% | 100% |
| tester3 | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| tester4 | 100% | 100% | 100% | 100% | 93% | 100% | 100% |
| tester5 | 100% | 100% | 100% | 100% | 87% | 100% | 100% |
| CMuJava | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

从表3的数据中可以看出,由于不同测试人员的能力差异和对变异算子理解的差异,所以在部分程序上正确率出现了差异.由于测试人员是在Eclipse下进行变异体生成工作,部分测试人员会生成了一些语法错误的变异体.目前几个程序产生的错误变异体都是RCXC(删除并发方法调用)变异算子定义的变异体,由于同一个程序中存在一些方法名和返回值都相同的方法调用,但是这些方法调用并不符合变异算子的规则,所以测试人员生成了这些错误的变异体.而工具每次运行产生的变异体都相同,所以表中只列出了一次的正确率.为了保证CMuJava生成的变异体不存在语法错误而导致无法运行的情况,CMuJava在设计的时候添加了语法检查功能,如果生成的变异体无法通过语法检查,则不会被输出到文件系统中.所以CMuJava生成变异体的正确率都为100%.

图9是测试人员和CMuJava生成变异体的正确率对比图.横轴表示两种变异体生成方式下的7个实验对象,纵轴表示生成变异体的正确率.从图中可以看出,使用工具对提升生成并发变异体的正确率有一定的帮助.尤其是针对一些比较特殊的并发程序的时候(如ThreadID),正确率的提高程度更为显著.

从表4中可以看出,针对第一个比较简单的并发程序所有的测试人员和CMuJava都能够正确识别里面的并发机制并生成所有的并发变异体,但是随着待测并发程序代码复杂度和可应用的并发变异算子数量的增加,人工生成方式遗漏的并发变异体数量也越来越多.图10是测试人员和CMuJava生成并发变异体的数量对比图.

**Table 4** Number of concurrent mutants

表4 生成并发变异体数量(个)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 生成变异体的对象 | Bin | TID | FGH | SSE | TFT | LOAVL | JM |
| tester1 | 3 | 4 | 10 | 14 | 22 | 35 | 46 |
| tester2 | 3 | 3 | 10 | 17 | 24 | 29 | 53 |
| tester3 | 3 | 5 | 12 | 17 | 24 | 39 | 55 |
| tester4 | 3 | 5 | 12 | 12 | 27 | 43 | 55 |
| tester5 | 3 | 5 | 12 | 16 | 27 | 40 | 59 |
| CMuJava | 3 | 5 | 12 | 17 | 27 | 43 | 61 |

由于生成并发变异体是一个比较枯燥和耗时的工作,主要就是依据并发变异算子规则识别可变异点,然后对可变异点进行微小的语法更改生成并发变异体.当一个可变异点存在多种变异方案或者待测程序存在较多可变异点时,测试人员会逐渐失去耐心,所以随着时间和可生成的并发变异体数量的增加,测试人员难免会少考虑一些情况,导致遗漏的并发变异体数量逐渐增加.但是CMuJava在生成变异体的过程中保持稳定不会出现遗漏变异体的问题.



Fig.9 Correct rate comparison chart of concurrent mutants

图9 并发变异体正确率对比图



Fig.10 Number comparison chart of concurrent mutants

图10 并发变异体的数量对比图

表5列出了5个测试人员针对不同待测程序生成并发变异体所使用的时间数据,从表5中可以看出随着程序复杂度的增加,测试人员所需要的时间也随之增加,而且即使针对同一个程序不同的测试人员所需要的时间也不完全相同.表6列出了CMuJava生成并发变异体所需要消耗的时间,由于工具所需要消耗的时间与计算机系统的资源分配有一定的关系,为了让实验数据更加合理,针对每一个待测并发程序都进行了5次实验并记录每一次实验的用时.从表中可以看出,随着待测程序的复杂度增加工具所需要的时间也逐渐增加,而且每一次所需要的时间之间也有一定的差距.

**Table 5** Time for testers to generate concurrent mutants(second)

表5测试人员生成并发变异体用时(秒)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 测试者 | 生成并发变异体所用时间 | | | | | | | |
| Bin | | TID | FGH | SSE | TFT | LOAVL | JM |
| tester1 | | 512 | 555 | 1004 | 1638 | 2903 | 4553 | 6765 |
| tester2 | | 523 | 545 | 944 | 1701 | 3251 | 4452 | 7219 |
| tester3 | | 448 | 508 | 900 | 1534 | 2953 | 4353 | 6891 |
| tester4 | | 487 | 540 | 893 | 1758 | 3361 | 4616 | 6881 |
| tester5 | | 524 | 582 | 987 | 1880 | 3097 | 4520 | 6098 |

**Table 6** Time for CMuJava to generate concurrent mutants(second)

表6 CMuJava生成并发变异体用时(秒)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 次数 | 生成并发变异体所用时间 | | | | | | |
| Bin | TID | FGH | SSE | TFT | LOAVL | JM |
| 1 | 0.063 | 0.077 | 0.140 | 0.250 | 0.581 | 0.984 | 1.187 |
| 2 | 0.070 | 0.077 | 0.156 | 0.234 | 0.503 | 0.921 | 1.321 |
| 3 | 0.054 | 0.079 | 0.098 | 0.25 | 0.519 | 0.962 | 1.191 |
| 4 | 0.052 | 0.088 | 0.120 | 0.232 | 0.499 | 0.906 | 1.201 |
| 5 | 0.062 | 0.090 | 0.109 | 0.249 | 0.523 | 0.895 | 1.296 |



Fig.11 Comparison box plot of experimental data

图11 实验数据对比盒图

从表中可以明显的看出CMuJava生成并发变异体的效率更高,针对所有的程序,人工生成并发变异体所需要的时间都是CMuJava的上千倍,而且随着程序的复杂度和可生成的并发变异体数量增加,测试人员和CMuJava的耗时差距越来越大.为了清楚地看出每次实验结果的波动情况,我们应用公式(6)对两种方式生成并发变异体的用时数据做了标准化处理,其中x是生成每个实验对象的所有并发变异体的时间值,min和max分别对应所用时间的最小值和最大值,表示标准化后的值.图11为处理后的数据对比盒图,横轴表示两种变异体生成方式下的7个实验对象,纵轴表示实验所用时间标准化后的值.从图中可以看出CMuJava生成并发变异体需要的时间相比人工生成的方式更加稳定.因此使用CMuJava可以极大地提高生成并发变异体效率,降低并发变异测试的测试资源开销.

(6).

通过使用几个实际的并发程序对人工和CMuJava生成并发变异体两种方式进行经验研究,研究结果表明,使用CMuJava可以显著地提高面向Java应用程序的并发变异测试自动化程度,降低测试资源的开销和提高生成并发变异体的效率.

在生产并发变异体的正确率方面,经验研究表明,相比于人工生成方式使用CMuJava可以将平均正确率提高约4%.对于一些容易出错的并发变异体(例如RCXC和ASK)提高的比例会更高;在生成并发变异体的完备性方面,使用CMuJava可以有效地避免人工生成并发变异体方式遗漏变异体的问题,保证能够准确识别待测程序的所有可变异点并生成相应的并发变异体,经验研究表明,相比于人工生成方式,使用CMuJava可以将生成并发变异体的数量提高约10%;在生成并发变异体效率方面,通过经验研究结果可以看出,相比于人工生成并发变异体的方式,使用CMuJava可以将生成效率平均提高约5700倍.CMuJava极大地提高了并发变异体生成的自动化程度,大幅提高了生成并发变异体的效率.

基于经验研究结果,相比于人工生成并发变异体,使用CMuJava可以显著地提高生成并发变异体的正确性、完备性和效率.

# 相关工作

自变异测试的方法被提出以来,测试人员在该领域做了大量的研究工作,涉及C/C++、C#、Java、SQL等多种编程语言[14].其中一些方面的工作为针对不同编程语言的变异算子的定义[7,11,15,16,17,18]、针对不同编程语言的变异体自动化生成工具[4,20,25],变异算子的选择,以及变异测试技术优化.本节介绍与本文提出的面向Java程序的并发变异体生成工具相关的工作.

在面向Java程序的并发变异算子研究方面,Delamaro等人根据Java语言的并发机制设计了15个并发变异算子[15],并且根据这些并发变异算子所针对的变异对象不同,将这些并发变异算子分成了针对监听对象、针对并发方法调用和针对等待集合三类.Farchi等人根据Java语言并发机制的特点并结合开发人员在实际编码中常犯的一些并发错误,系统地归纳并总结出了一套并发故障模式[16].Brabury等人[7]对Farchi等人提出的并发故障模式进行了补充,然后以这些并发故障模式为依据,设计出了包括修改并发方法参数、修改并发方法调用、修改关键字、交换并发对象和修改临界区等五类共25种并发变异算子.Wu和Kaiser认为已有的并发变异算子无法生成某些并发变异体,对已有的变异算子进行组合得到了6种新的并发变异算子,将这些并发变异算子分成了同步方法和同步代码块两类[17].Gligoric等人在实验过程中提出了3种新的并发变异算子[21],并结合Brabury等人提出的25种并发变异算子一起应用到实验中.

变异算子是变异测试的一个核心概念,依据变异算子生成变异体在变异测试过程中至关重要.考虑到程序的复杂性,人工生成变异体会耗费大量的人力物力,测试人员需要一个自动生成变异体的工具来提高测试执行效率.随着变异算子的发展,大量针对不同编程语言的自动生成变异体的工具被实现,Papadakis等人对已有的变异测试工具做出了总结[14].其中包括面向Java语言的变异体生成工具[4,5,6].MuJava[4]是一个面向Java程序的变异测试工具,实现了传统的方法级别和类级别的变异算子[4,11],但是并不支持并发变异体的生成.Javalance[5]为Java程序提供了开源的变异测试框架,但同样不支持并发变异体的生成.Jumble[6]是一个变异Java程序字节码的工具,提供了传统的方法级别的变异,并支持使用多线程程序的运行,但是文中没有提到相关的实现细节.目前已有的面向并发程序的变异工具[22,23,24]仍存在一些问题.MutMut[23]是一个面向并发程序的优化变异体执行的工具,据报道该工具对变异测试的时间降幅达77%,但MutMut并不具有优化变异体生成的功能.Paraµ[24]是一个支持并发变异体生成的工具,但是该工具的研究仍处于初级阶段,仅实现了2个类级别变异算子和3个并发变异算子.另一个面向Java程序的支持并发变异体生成的工具Comutation[21]实现了所有的28种并发变异算子,但是由于Intel没有对该工具授权发布,Comutation目前并不支持开源使用.

# 总结

变异测试是一种可以评估测试用例集的充分性与测试技术的有效性的技术.并发变异测试是基于并发故障的变异测试技术.由于缺少面向Java程序的开源的并发变异测试工具,研究者通常采用需要消耗大量测试资源的人工方式生成并发变异体.本文根据Bradbury等人提出的Java并发变异算子[7],设计并实现了一个面向Java程序的并发变异体自动生成工具CMuJava,该工具扩展了传统的变异测试工具MuJava,支持待测程序选择、变异算子选择、并发变异体生成、并发变异体查看、变异体执行和测试报告打印等功能,提高了并发变异测试的自动化程度.另外利用7个并发程序评估了CMuJava生成并发变异体的性能.实验结果表明:使用CMuJava可以显著提高生成并发变异体的正确性、完备性和效率.

未来,我们将在如下几个方面进一步完善我们在面向Java程序的并发变异体生成工具方面的研究:(1)本文使用了7个真实的并发程序进行经验研究,研究对象数量较少,未来需要使用更多的实验对象评估本文提出的支持工具的实用性.(2)完善CMuJava的功能和性能.CMuJava目前还只是一个用于并发变异测试研究的原型工具,仍存在不足之处,例如界面设计不够友好.未来需要在此原型工具的基础上进行改进,让CMuJava更好地运用到并发变异测试研究领域,提高面向Java程序的并发变异测试技术的自动化程度.

References:

1. Bianchi FA, Margara A, Pezze M. A survey of recent trends in testing concurrent software systems. IEEE Trans. on Software Engineering. 2018, 44(8): 747-783.
2. Demillo RA, Lipton RJ, Sayward FG. Hints on test data selection: Help for the practicing programmer. IEEE Computer, 1978, 11(4): 34-41.
3. Chen X, Gu Q. Mutation testing: Principal, optimization and application. Journal of Frontiers of Computer Science and Technology, 2012, 6(12): 1057-1075(in Chinese with English abstract).
4. Ma YS, Offutt AJ, Kwon YR. MuJava: An automated class mutation system. Software Testing Verification and Reliability, 2005, 15(2): 97-133.
5. Schuler D, Zeller A. Javalanche: Efficient mutation testing for java. In: Proc. of the 7th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on Foundations of Software Engineering (ESEC/FSE 2009). ACM Press, 2009: 297-298.
6. Irvine SA, Pavlinic T, Trigg L, Cleary JG, Inglis S, Utting M. Jumble java byte code to measure the effectiveness of unit tests. In: Proc. of the Testing: Academic and Industrial Conference Practice and Research Techniques – MUTATION. IEEE Computer Society. 2007: 169-175.
7. Bradbury JS, Cordy JR, Dingel J. Mutation operators for concurrent java (J2SE 5.0). In: Proc. of the 2th Workshop on Mutation Analysis, Co-located with the IEEE International Symposium on Software Reliability Engineering (ISSRE 2006). IEEE Computer Society, 2006: 83-92.
8. Untch RH, Offutt AJ, Harrold MJ. Mutation analysis using mutant schemata. In: Proc. of ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 1993). ACM Press, 1993: 139-148.
9. Wu YB, Guo JX, Li Z, Zhao RL. Mutation strategy based on concurrent program data racing fault. Joumal of Computer Application, 2016(11): 3170-3177, 3195(in Chinese with English abstract).
10. Pati T, Hill JH. A survey report of enhancements to the visitor software design pattern. Software: Practice and Experience, 2014, 44(6): 699-733.
11. Offutt J, Ma YS, Kwon YR. The class-level mutants of MuJava. In: Proc. of International Workshop on Automation of Software Test. ACM press, 2006: 78-84.
12. Herlihy M, Shavit N. The art of multiprocessor programming. Morgan Kaufmann, 2012.
13. Gramoli V. More than you ever wanted to know about synchronization synchrobench, measuring the impact of the synchronization on concurrent algorithms. In: Proc. of the 20th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming. ACM Press, 2015: 1-10.
14. Papadakis M, Kintis M, Zhang J, Jia Y, Traon YL, Harman M. Mutation testing advances: An analysis and survey. Advances In Computer, 2019: 275-378.
15. Delamaro M, Pezzè M, Vincenzi A, Maldonado JC. Mutant operators for testing concurrent java programs. In: Proc. of the Brazilian Symposium on Software Engineering (SBES 2001). IEEE Computer Society, 2001: 272-285.
16. Farchi E, Nir Y, Ur S. Concurrent bug patterns and how to test them. In: Proc. of the International Parallel and Distributed Processing Symposium (PDPS 2003). IEEE Computer Society, 2003: 286-292.
17. Leon W, Kaiser G. Constructing subtle concurrency bugs using synchronization-centric second-order mutation operators. In: Proc. of the 23th International Conference on Software Engineering and Knowledge Engineering (SEKE 2011). DBLP, 2011: 244-249.
18. Ma YS, Kwon YR, Offutt J. Inter-class mutation operators for java. In: Proc. of International Symposium on Software Reliability Engineering. IEEE, 2002: 352-366.
19. Smith B, Williams L. On guiding the augmentation of an automated test suite via mutation analysis. Empirical Software Engineering, 2009, 14(3): 341-369.
20. Delgado-Pérez P, Medina-Bulo I, Palomo-Lozano F, García-Domínguez A, Domínguez-Jiménez JJ. Assessment of class mutation operators for C++ with the mucpp mutation system. Information and Software Technology, 2017, 81: 169–184.
21. Gligoric M, Zhang L, Pereira C, Pokam G. Selective mutation testing for concurrent code. In: Proc. of the 2013 International Symposium on Software Testing and Analysis. ACM Press, 2013: 224-234.
22. Bradbury JS, Cordy JR, Dingel J. Comparative asscessment of testing and model checking using program mutation. In: Proc. of Testing: Academic & Industrial Conference Practice & Research Techniques-mutation. IEEE Computer Society, 2007: 210-219.
23. Cligoric M, Jagannath V, Marinov D. MutMut: Efficient exploration for mutation testing of multithreaded code. In: Proc. of the third International Conference on Software Testing. IEEE Press, 2010: 55-64.
24. Madiraju P, Namin AS. Paraµ: A partial and higher-order nutation tool with vonvurrency operators. In: Proc. of IEEE fourth Internation Conference on Software Testing. IEEE Press, 2011: 351-356.
25. Kusao M, Wanng C. Ccmutator: A mutation generator for concurrency constructs in multithreaded C/C++ applications. In: Proc. of the 28th IEEE/ACM International Conference on Automated Softwafre Engineering. 2013: 722-725.
26. Morell LJ. A theory of fault based testing. IEEE Trans. on Software Engineering. 1990, 16(8): 844-857.
27. Sun CA, Wang G, Cai KY, Chen TY. Distribution-Aware mutation analysis. In: Proc. of IEEE Computer Software & Applications Conference Workshops. IEEE Computer Society, 2012: 170-175.
28. Bradbury JS, Cordy JR, Dingel J. Mutation operators for concurrent java (J2SE 5.0). Technical report 2006-520, Queen’s University, 2006.
29. Lipton RJ. Reduction: A method of proving properties of parallel programs. Communications of the ACM, 1975,18(12): 717–721.
30. Lamport L. Time, clocks, and the ordering of events in a distributed system. Communications of the ACM, 1978, 21(7): 558-565.
31. Carver R. Mutation-based testing of concurrent programs. In: Proc. of International Test Conference(ITC 1993). IEEE Computer Society, 1993: 845-853.
32. Farchi E, Nir Y, Ur S. Concurrent bug patterns and how to test them. In: Proc. of 17th International Symposium on Parallel & Distributed Processing(IPDPS 2003). IEEE Computer Society, 2003: 286.2.
33. Sun CA, Wang G. MuJavaX：A distribution—aware mutation generation system for java. Journal of Computer Research and Developmen, 2014, 51(4): 874-881(in Chinese with English abstract).
34. Kintis M, Papadakis M, Papadopoulos A, Valvis E, Malevris N. Analysing and comparing the effectiveness of mutation testing tools: A manual study. In: Proc. of IEEE 16th International Working Conference on Source Code Analysis and Manipulation. IEEE, 2016, 1: 147-156.
35. Saifan AA, Alazzam I, Akour M, Hanandeh F. Regression test-tslection technique using component model based modification: Code to test traceability. International Journal of Advanced Computer Science and Applications, 2016, 7(4): 88-92.
36. Masood A, Bhatti R, Ghafoor A, Mathur A. Scalable and effective test generation for role-based access control systems, IEEE Trans. on Software Engineering, 2009, 35(5): 654-668.
37. Kim SW, Ma YS, Kwon YR. Combining weak and strong mutation for a noninterpretive java mutation system, Software Testing, Verification and Reliability. 2013, 23(8): 647-668.
38. Kiczales G, Rivieres J, Bobrow DG. The art of the metaobject protocol. MIT Press, 1991.
39. Kim S, Clark J, McDermid J. Assessing test set adequacy for object oriented programs using class mutation. In: Proc. of Symposium on Software Technology(SoST 1999). 1999, 72-83.
40. René J, Schweiggert F, Kapfhammer GM. MAJOR: An efficient and extensible tool for mutation analysis in a Java compiler. In: Proc. of 26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011). ACM, 2011, 1: 612-615.
41. Chevalley P. Applying mutation analysis for Object-oriented programs using a reflective approach. In: Proc. of 8th Asia-Pacific Software Engineering Conference (APSEC 2001). IEEE Computer Society, 2001, 1: 267-270.
42. Tatsubori M, Chiba S, Killijian M O, Itano K. OpenJava: A class-based macro system for Java. In: Proc. of the Workshop on Reflection and Software Engineering, Co-located with the Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA 1999). Springer, 1999: 117-133.
43. Chiba S. Load-time structural reflection in Java, Ecoop, 2000: 313-336.
44. Kleinoder J, Golm M. MetaJava: An efficient run-time meta architecture for Java. In: Proc. of the 5th International Workshop on Object-Orientation in Operating Systems. IEEE, 1996: 54-61.
45. Welch I, Stroud RJ. Kava: Using byte code rewriting to add behavioral reflection to Java. Technical report 704, University of Newcastle upon Tyne, 2000.

附中文参考文献:

[3] 陈翔,顾庆.变异测试:原理、优化、和应用.计算机科学与探索,2012,6(12):1057-1075.

[9] 吴俞伯,郭俊霞,李征,赵瑞莲.基于并发程序数据竞争故障的变异策略.计算机应用,2016,36(11):3170-3177,3195.

[33] 孙昌爱,王冠.MuJavaX:一个支持非均匀分布的变异生成系统.计算机研究与发展,2014,51(4):874-881.

1. \* 基金项目:

   Foundation item:

   收稿时间: 修改时间: 采用时间: jos在线出版时间:

   CNKI在线出版时间: [↑](#footnote-ref-1)
2. https://logging.apache.org/log4j/2.x/ [↑](#footnote-ref-2)