1. 实验评估

本章选择三个java程序进行实例研究: （1）航空行李托运计费程序；（2）中国联通计费程序；（3）停车计费程序。验证本文提出方法的可行性和有效性。三个实例程序均根据真实的规格说明，由课题组同学开发。

* 1. 研究问题

本章的实例研究围绕三个问题展开讨论。

1. **验证基于符号执行的原始测试用例生成技术的可行性与适用性。**
2. **验证基于符号执行的原始测试用例生成技术的有效性。**实验中将针对相同的蜕变关系，分别用上述方法和随机值法生成原始测试用例进行蜕变测试，比较两种原始测试用例生成方法的故障检测能力。
3. **验证基于路径距离的原始测试用例优先级排序方法的有效性。**实验中分别从按优先级排序的原始测试用例集和未排序的原始测试用例集分别选取相同比例的原始测试用例，在相同蜕变关系集的条件下，比较两组选取的原始测试用例集进行蜕变测试的故障检测能力。选取比例设定为10%，20%，30%，40%，50%，60%，70%，80%，90%。
   1. 实验步骤

针对上面的三个研究问题，设计两组对比实验分别进行研究。

为了验证基于符号执行的原始测试用例生成技术的可行性和有效性，设计一组对比实验。在获取相同的蜕变关系，生成相同变异体的条件下，比较上述方法和随机值法生成原始测试用例的故障检测能力。具体的实验步骤如下：

1. **获取蜕变关系**：3个实例均采用了基于数据变异的蜕变关系获取技术[24],并结合程序的属性设计蜕变关系。
2. **生成变异体：**实验中采用变异分析工具MuJava为待测程序生成变异体，其中包含所有典型的变异算子，并已去除等价变异体。
3. **生成原始测试用例：**实验中测试用例的生成是对比实验的关键问题，采用了两种不同的测试用例生成策略：(1)基于符号执行的原始测试用例生成方法——实验中根据程序的源文件，解析程序的所有路径，根据路径约束条件，生成满足路径约束条件的原始测试用例。(2)随机值法——根据程序的输入规格随机生成测试用例，实验中设定TC=n、TC=2n、TC=5n、TC=10n四组原始测试用例集(其中n代表方法(1)生成的原始测试用例数量)。
4. **判断原始测试用例是否满足蜕变域**：针对步骤3)生成原始测试用例集，分别判断是否满足步骤1)获取的蜕变关系集，如果满足，根据蜕变关系为原始测试用例生成相应的衍生测试用例，并添加到待测测试用例集中。
5. **执行测试：**将原始测试用例集和对应的衍生测试用例集分别在步骤2)生成的非等价变异体上执行，比较原始测试用例的输出结果和衍生测试用例的输出结果是否满足蜕变关系，得到测试结果。
6. **结果分析**：实验结果采用变异分析技术进行分析，度量指标选用变异得分(Mutation Score)。变异得分指的是用测试用例集TC能够杀死程序P的变异体数量占生成非等价变异的数量的百分比(测试用例集能够检测到变异体的故障称为变异体被杀死)。计算公式如下:



其中表示被杀死的变异数量，Nm表示总的变异体数量，表示等价变异体数量。变异得分能有效的反应测试用例集的故障检测能力。

为了验证基于路径距离的原始测试用例优先级排序方法的有效性，同样设计一组对比实验。实验步骤与第一组对比实验除步骤3)之外都相同，在本实验中，需要先设定测试用例选取比例，分别为10%，20%，30%，40%，50%，60%，70%，80%，90%。采用基于符号执行的方法生成原始测试用例集并根据路径距离进行排序，从按优先级排序的原始测试用例集和未排序的原始测试用例集中分别选取相同比例的原始测试用例进行蜕变测试。

1. 实例研究

为了验证基于场景的BPEL测试用例生成技术的有效性及TSTG工具的实用性，使用几个真实的BPEL例子作为待测程序与TSTG工具进行实例研究。

* 1. 度量指标

在本实验中使用变异分析技术评估本文提出的测试用例生成技术的有效性及工具的实用性。根据变体在测试用例集上的故障检测结果，我们使用如下两个指标度量蜕变测试方法的有效性。

第一个指标是变异得分[40]。变异得分定义了测试用例集杀死的非等价变体的比例，评估了测试用例集的有效性。显然，变异得分越高测试用例集越有效。

第二个指标是故障检测率(Fault Discovery Rate, FDR)[40]，定义如下：



(6-1)

其中，*Nf*是能够杀死变体m的测试用例数量，*Nts*是测试用例集ts中测试用例总数量，*Ni*是测试用例集中的非法测试用例数量。故障检测率定义了某条蜕变关系生成的测试用例能够有多少比例杀死变体m。故障检测率越高说明某测试用例杀死变体m的几率越大。

* 1. 实验步骤

按照如下步骤使用TSTG对待测程序进行实例研究：

1. **生成测试路径**

利用工具TSTG生成满足一定覆盖准则下的测试路径，这里包括弱覆盖及一般覆盖准则下的测试路径集。

1. **生成测试数据**

利用工具TSTG可以针对每条测试路径生成满足每条测试路径的测试数据，也可以通过解析WSDL生成一组测试数据集，然后再人工分析每组数据覆盖的测试路径。为了最大化的实现自动化，我们选择第一种直接针对每条测试路径生成测试数据集。

1. **生成测试用例**

针对每条测试路径生成测试数据之后，由于测试数据覆盖掉所有的测试路径，这些测试数据便组合成一个新的集合，即测试用例集。

1. **植入错误**

为了采用变异分析技术，需要向BPEL流程中植入一些错误。目前，虽然有研究涉及BPEL流程的变异测试，也有工具实现自动化的植入错误，但该工具并没有供人使用。因此，这里采用手工植入一些错误的形式进行试验。

根据目前研究中提出的针对BPEL流程的变异算子(Mutation Operator)，例如文献[24]，文献[45]等，并结合例子的具体情况，可以手工植入错误，具体操作时，可根据上述的每一种变异算子所表明的常见错误类型，对例子以可能的错误形式替换，例如：ERR代表关系运算符的错误类型，可通过替换同种操作符(<，>，<=，>=，=，!=)的方式得到多个变异体。由于不同的BPEL流程未必能够完全反映出上述文献中总结的所有变异算子，因此针对具体的例子而言，可能的变异体的错误类型将少于总结的变异算子数量。

1. **执行测试**

利用TSTG分别在原始程序和错误版本上执行测试用例并截取它们的输出，对比输出结果，查看变体是否被杀死。

1. **测试结果评估**

在原始程序与植入错误的程序上执行所有的测试用例，统计变异得分与故障检测率，分析统计结果。

1. 经验研究

本章对课题组前期的经验研究进行扩充，以增强实验结果的可靠性；采用经验研究评估所提出的优化技术的有效性。

* 1. 实验对象与度量指标

1. **实验对象**

本文实验对象共包括6个BPEL程序实例。SupplyChain[10]是一个供销链管理的实例。客户输入商品的名字和数量，零售商将会根据库存的状况反馈信息。SmartShelf[10]是一个商品货架管理的实例。用户输入货物的相关信息，例如货物名称、数量，系统处理这些信息并输出商品库存数量、货架位置和库存状态三个信息。SupplyCustomer[2]是项目订单管理的实例。用户输入订单的名称和地址，系统处理并反馈订单的查询结果。LoanApproval[2]是贷款审批的实例。用户输入个人信息及贷款金额，系统进行处理并反馈贷款审核结果。CarEstimate[41]是用于汽车评估的实例。用户发出评估请求，系统提供初步、简单、复杂评估方式，并返回最终评估结果。TravelAgency[42]是一个旅行社预订实例，组合旅店预订、订票、旅行社预订、和银行结算服务。用户提供预订旅客的信息和人数，系统进行处理并反馈结果。表5-1给出这些实例的具体信息，包括实例的基本功能描述（Basic functionality）、组装的服务数目（No. of services composed）和实例的规模（Size）。文中，使用P1至P6依次代表SupplyChain、SmartShelf、SupplyCustomer、LoanApproval、CarEstimate和TravelAgency实例程序。

表5-1 BPEL的实例程序基本信息

|  |  |  |  |
| --- | --- | --- | --- |
| **Program** | **Basic functionality** | **No. of services composed** | **Size (LOC)** |
| P1 | Management of supply chains | 2 | 50 |
| P2 | Management of commodity shelves | 14 | 194 |
| P3 | Management of project orders | 5 | 122 |
| P4 | Examination of loan applications | 3 | 120 |
| P5 | Assessment of car repairs | 7 | 121 |
| P6 | Booking of travels | 9 | 543 |