

# 1. 实验对象

所有实验对象的变异体在 RT 策略下均需要至少 20 个测试用例才能揭示，具体变异情况如表 1 所示。由于航空托运服务的“较难”杀死的变异体的变异位置均处于同一行，因此不存在故障集中度。

表 1 实验对象

Web services	代码行数	故障集中度	变异体数目
航空托运	116	-	12
联通计费	131	5-5	50
		4-6	42
		3-7	22
		2-8	19
		1-9	10
停车计费	129	5-5	106
		4-6	69
		3-7	26
		2-8	25
		1-9	9

# 2. 变量

## 2.1 自变量

测试技术是我们研究的自变量。DRT 是其中一个变量，另外 RPT 在 RT 的基础上增加了分区概念，而 DRT 又在分区的基础上增加了控制理论。因此本文选择 RPT 和 RT 作为与 DRT 策略的比较对象。

## 2.2 因变量

因变量用来度量 DRT 策略在 Web services 中测试效率。有很多度量标准判断测试技术是否有效：P-measure(在当前测试用例集中至少检测出一个故障的概率)、E-measure(在当前测试用例集中期望检测的故障数目)、F-measure(当前测试用例集中检测一个故障需要的测试用例数目)、T-measure(当前测试用例集中检测出所有故障需要的测试用例数目)和 NF-measure(当前测试用例集中，揭示第一个故障之后再揭示一个故障需要的测试用例数目)。在这些度量标准中我们认为 F-measure 和 NF-measure 是最合适评估 DRT 测试有效性的度量标准。

# 3. 实验设置

## 3.1 分区设置

三个实验的分区情况如表 2 所示。

表 2 每一个实验对象的分区数目

Web services	分区方式 1	分区方式 2
航空托运	24	7
联通计费	20	3
停车计费	18	3

航空托运服务有两种分区方式：一，根据规格说明书将所有的输入进行组合形成原始决策表，决策表中的每一条规则对应一个分区；二，在原始决策表的基础上将具有相同动作的规则进行合并，合并后的决策表每一个规则对应一个分区。

联通计费有两种分区方式：一，根据规格说明书将所有的输入进行组合形成决策表，决策表中的每一条规则对应一个分区；二，仅考虑一个参数：套餐类型 A、B、C，按照套餐类型行进分区。

停车计费有两种分区方式：一，根据规格说明书将所有的输入进行组合形成决策表，决策表中的每一条规则对应一个分区；二，仅考虑一个参数：交通工具的类型：摩托车、2 门的汽车和 4 门的汽车。

### 3.2 测试剖面

由于测试用例是在测试的过程中随机产生的，因此用均等的概率分布作为初始测试剖面是一个保守、可行的方案。另外测试人员也可以根据以往的测试经验将某一个或某些分区被选择的概率增大，并且减少某些分区被选择的概率也是可行的初始剖面的设置方案。

### 3.3 参数的设置

参数的取值为：

$$\varepsilon = \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.3, 0.5, 0.75\}$$

$\varepsilon = 0.75$  已经是一个很大的数，此时很容易产生这样的情况：当某一个分区内的测试用例没有揭示软件中的故障时，并且此时该分区被选择的概率比较低，由于 DRT 策略的算法原因，此时很有可能导致该分区被选择的概率变为 0。此时明显是不合理的。例如在停车计费服务中当在分区方式 2 下进行测试时，此时  $\varepsilon = 0.75$  初始测试剖面为均等概率分布，即  $p_i = 1/3$ 。假如第一次测试在第一个分区选中一个测试用例并执行后没有揭示软件中的故障，由于  $p_i < \varepsilon/(3 - 1)$ ，所以  $p_1 = 0$ 。即便是失效率很高的分区也不可能每一个测试用例都能揭示软件中的故障，但是

当 $\epsilon$ 取值很大时，往往会因为一次的测试结果将某一个分区被选择的概率直接降为 0，这显然是不合理的。因此 $\epsilon$ 取很大的值 ( $\epsilon \leq 1$  始终成立) 没有意义。