

## 1 假设

为了在数学上易于表示，我们做了如下假设：

1. 测试用例集的失效率为  $\theta$ ，它表示从测试用例集中随机选取一个能够揭示故障的测试用例的概率。将输入域划分分区之后，每一个分区对应的失效率为  $\theta_i (i = 1, 2, \dots, m)$ 。当不能确定每一个分区对应的失效率时，可以通过估计的方式得到。
2. 将待测软件的输入域划分分区之后总存在一个分区  $c_M$ ，它的失效率最大  $\theta_M > \theta_i (i \neq M)$ 。
3. 在测试的过程中每一个分区对应的失效率  $\theta_i (i = 1, 2, \dots, m)$  保持不变，即在测试的过程中不移除检测到的故障。
4. 测试用例放回，即当一个测试用例被选中并执行之后仍有机会被选中。

将待测软件的输入域划分分区之后至少存在一个失效率最大的分区  $c_M$ ，它的失效率为  $\theta_M$ 。DRT策略最突出的特点就是根据以往的测试信息使得具有更高失效率的分区更有机会被选中。一方面参数  $\varepsilon$  在测试的过程中更新测试剖面。另一方面，一般将均等概率分布作为初始测试剖面，分区数目越多每一个分区被选择的概率越低， $\varepsilon$  的大小决定了具有更高失效率的分区比其它分区具有更高的被选择概率的速度。因此在一定的分区数目下，设置合理的  $\varepsilon$  可以使 DRT 策略更快地找到失效率最高地分区  $\theta_M$  并且尽可能多地选中该分区。

在测试的过程中，为了保证具有最大失效率的分区  $c_M$  更有可能被选中，我们希望  $c_M$  分区被选择的概率越来越大，数学表示如(1),其中  $n$  表示已经执行过的测试用例。

$$E(p_M(n+1)) > p_M(n) \quad (1)$$

## 2 理论分析

假设初始测试剖面为  $p_1(0), p_2(0), \dots, p_m(0)$ ， $n$  个测试用例已经被执行了，目前的测试剖面为  $p_1(n), p_2(n), \dots, p_m(n)$ 。在实际的测试过程中  $p_i < \varepsilon / (m-1)$  或者  $p_i < \varepsilon$  的可能性非常小，因此在下面的讨论中不考虑这些情况。执行第  $n+1$  个测试用例  $c_i$  分区的期望如下所示：

$$\begin{aligned} E(p_i(n+1)) &= p_i(n)\theta_i(p_i(n) + \varepsilon) + p_i(n)(1 - \theta_i)(p_i(n) - \varepsilon) \\ &\quad + \sum_{j \neq i} p_j(n)\theta_j(p_i(n) - \frac{\varepsilon}{m-1}) \\ &\quad + \sum_{j \neq i} p_j(n)(1 - \theta_j)(p_i(n) + \frac{\varepsilon}{m-1}) \\ &= p_i(n) + Y_i(n) \end{aligned} \quad (2)$$

其中:

$$Y_i(n) = \frac{\varepsilon}{m-1}(2mP_i\theta_i - p_im - 2p_i\theta_i + 1) - \frac{2\varepsilon}{m-1} \sum_{j \neq i} p_j\theta_j \quad (3)$$

由于  $\sum_{i=1}^m E(p_i(n+1)) = 1$ , 所以  $\sum_{i=1}^m Y_i(n) = 0$ 。并且对于  $\forall i \in \{1, 2, \dots, m\} \& i \neq M$  有:

$$Y_M(n) - Y_i(n) = \frac{2\varepsilon}{m-1}(m(p_M\theta_M - p_i\theta_i) - \frac{m}{2}(p_M - p_i)) \quad (4)$$

对  $\forall i \in \{1, 2, \dots, m\} \& i \neq M$  如果都有  $Y_M(n) > Y_i(n)$  则可以保证  $Y_M(n) > 0$ , 即:

$$Y_M(n) - Y_i(n) > 0 \iff (p_i(n) - p_M(n)) > 2(p_i(n)\theta_i - p_M(n)\theta_M) \quad (5)$$

要使公式(5)成立有以下3中情况:

1. 当  $p_i = p_M$  时, 公式(5)恒成立。
2. 当  $p_i > p_M$  时, 由于  $p_i(n)\theta_i - p_M(n)\theta_M < (p_i(n) - p_M(n))\theta_i < (p_i(n) - p_M(n))\theta_i \frac{m+k\varepsilon}{m}$ , 只要保证  $p_i(n) - p_M(n) > 2(p_i(n) - p_M(n))\theta_i \frac{m+k\varepsilon}{m}$  成立, 则(5)式成立, 可得:

$$\frac{k\varepsilon}{m} < \frac{1}{2\theta_i} - 1 \quad (6)$$

3. 当  $p_i < p_M$  时, 公式(5)可以转化为  $(p_M(n) - p_i(n)) < 2(p_M(n)\theta_M - p_i(n)\theta_i)$ , 由于  $p_M(n)\theta_M - p_i(n)\theta_i > (p_M(n) - p_i(n))\theta_M > (p_M(n) - p_i(n))\theta_M \frac{m-1+g\varepsilon}{m}$ , 只要保证  $p_M(n) - p_i(n) < 2(p_M(n) - p_i(n))\theta_M \frac{m-1+g\varepsilon}{m}$  则(5)式成立, 可得:

$$\frac{1}{2\theta_M} - \frac{m-1}{m} < \frac{g\varepsilon}{m} \quad (7)$$

综上所述, 只要

$$\frac{1}{2\theta_M g} - \frac{m-1}{mg} < \frac{\varepsilon}{m} < \frac{1}{2k\theta_\Delta} - \frac{1}{k} \quad (8)$$

成立就可以保证  $E(p_M(n+1)) > p_M(n)$ 。其中  $\theta_\Delta = \max\{\theta_i | i = 1, 2, \dots, m, i \neq M\}$ ,  $\theta_\Delta < \theta_M$ 。