Lecture

Mutation Testing - II

1

Mutation Testing

- To measure the effectiveness
- To measure the adequacy
- To generate test cases

2

Outline

- Killing mutants
- Mutation scores
- Effectiveness metric
- Adequacy criterion

3

Killing Mutants

Given a program P, let M be its mutant

M is said to be killed by a test case (t), if
the output of executing P with t
is different to
the output of executing M with t

4

Example 1: Killing Mutants

Given program (P)	Mutant (M)
Input A, B	Input A, B
C= A + B	C= A * B
Output C	Output C

Test case (A=3, B=2) kills M, because

- Executing P with (A=3, B=2), output is 5
- Executing M with (A=3, B=2), output is 6

5

Example 2: Killing Mutants

Given program (P)	Mutant (M)
Input A, B	Input A, B
C= A + B	C= A * B
Output C	Output C

Test case (A=2, B=2) does not kill M, because

- Executing P with (A=3, B=2), output is 4
- Executing M with (A=3, B=2), output is 4

6

Killing Mutants (continued)

Given a program P, let M be its mutant.
M is said to be killed by a test case (t), if
the output of executing P with t
is different to
the output of executing M with t

Note: P acts as a test oracle for M

7

Killing Mutants (continued)

P acts as a test oracle of M

Two possible scenarios:

- P has a test oracle
- P does not have a test oracle

8

Mutation Score

Given

- a set of k mutants, SM={M1, M2, …Mk}
- a set of n test cases, TC={t1, t2, …, tn}

A mutant Mj (1 <= j <=k) is said to be killed by TC, if it is killed by some elements of TC

9

Mutation Score (continued)

The mutation score (MS) for TC and SM is defined as the ratio of the number of killed mutants and the total number of mutants

That is, MS = m / k

where m denotes the number of mutants killed by TC, and k denotes the total number of mutants

10

Example

Suppose

- SM={M1, M2, M3, M4, M5, M6}
- TC={t1, t2, t3, t4}

Assume further that

- t1 kills M2, M3 and M5
- t2 kills M3
- t3 kills M1, M5
- t4 kills M1, M2

11

Example (continued)

Then

- Set of killed mutants = {M1, M2, M3, M5}
- MS = (4 / 6) = 2/3 (or 0.67)

12

Effectiveness Metric

How to compare the failure detection effectiveness of two sets of test cases (test suites)?

Suppose we have TC1 and TC2, which are designed by different testers

How to know whose test suite is better?

13

Effectiveness Metric (continued)

- Define a set of mutants (SM)
- Execute mutants against TC1 and TC2
- Compute TC1's mutation score (MS1)
- Compute TC2's mutation score (MS2)
- Compare MS1and MS2 to determine which test suite (TC1 or TC2) is more effective

14

Example

Suppose

- SM={M1, M2, M3, M4, M5, M6}
- TC1={t1, t2, t3, t4}
- TC2={t5, t6, t7}

15

Example (continued)

Assume further that

- t1 kills M2, M3 and M5
- t2 kills M3
- t3 kills M1 and M5
- t4 kills M1, M2, M5

Then, TC1 kills {M1, M2, M3, M5}

MS of TC1 = (4/6) = 0.67

16

Example (continued)

Assume further that

- t5 kills M2, M3 and M5
- t6 kills M3, M5 and M6
- t7 kills M1 and M5

Then, TC2 kills {M1, M2, M3, M5, M6}

MS of TC2 = (5/6) = 0.83

17

Example (continued)

TC2 is more effective than TC1

because

MS of TC2 (=0.83) > MS of TC1 (=0.67)

18

Mutation Score

Is mutation score a good effectiveness metric?

19

Example (continued)

Assume that another TC3={t8, t9, t10, t11}

- t8 kills M2 and M3
- t9 kills M2
- t10 kills M5
- t11 kills M1and M5

Then, TC3 kills {M1, M2, M3, M5}

MS of TC3 = (4/6) = 0.67

20

Example (continued)

TC1	TC3
t1 kills M2, M3 and M5	t8 kills M2 and M3
t2 kills M3	t9 kills M2
t3 kills M1 and M5	t10 kills M5
t4 kills M1, M2 and M5	t11 kills M1 and M5

Same MS for TC1 and TC3

Any preference: TC1 or TC3?

21

Example (continued)

TC1	TC3
t1 kills M2, M3 and M5	t8 kills M2 and M3
t2 kills M3	t9 kills M2
t3 kills M1 and M5	t10 kills M5
t4 kills M1, M2 and M5	t11 kills M1 and M5

Same MS for TC1 and TC3

Any preference: TC1 or TC3? TC1

22

Example (continued)

TC1	TC4
t1 kills M2, M3 and M5	t12 kills M2 and M6
t2 kills M3	t13 kills M4
t3 kills M1 and M5	t14 kills M5
t4 kills M1and M2	t15 kills M2 and M4

Same MS for TC1 and TC4

Any preference: TC1 or TC4?

23

Example (continued)

TC1	TC4
t1 kills M2, M3 and M5	t12 kills M2 and M6
t2 kills M3	t13 kills M4
t3 kills M1 and M5	t14 kills M5
t4 kills M1and M2	t15 kills M2 and M4

Same MS for TC1 and TC4

Any preference: TC1 or TC4? ???

24

Some faults are more difficult to detect

25

Some faults are more difficult to detect

Example:

Correct statement: if (A+B >= C)

Incorrect statement: if (A+B > C)

26

Some faults are more difficult to detect

Example:

Correct statement: if (A+B >= C)

Incorrect statement: if (A+B > C)

A+B = C (A=12, B= 31, C=42)

27

Test Adequacy

Testing only demonstrates the presence of faults but not the absence of faults

Question

When to stop testing?

28

Test Adequacy (continued)

- Define a test adequacy criterion
- Stop testing upon the satisfaction of the test adequacy criterion

29

Test Adequacy (continued)

- Use a set of mutants as a test adequacy criterion
  - killing all mutants

30

Test Adequacy (continued)

- Use a set of mutants as a test adequacy criterion
  - killing all mutants

Question???

31

Summary

32

References

- Y. Jia and M. Harman, “An Analysis and Survey of the Development of Mutation Testing”, IEEE Transactions on Software Engineering, Vol. 37(5), 649-678, 2011.

33