Probabilistic Failure-causing Schema in Input-Domain Testing

Ziyuan Wang Yuanchao Qi Jiawei Lin

School of Computer, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

Email: wangziyuan@njupt.edu.cn

***Abstract*—To describe characteristics of failure test cases in the input-domain testing, we propose a model of probabilistic failure-causing schema. In this model, test case that contains a probabilistic failure-causing schema has a probability to be a failure test case. It may help testers to find out input characteristics that have more close relationship to the fault.**

I. Introduction

Once there are failure test cases reported in input-domain testing, input-level fault localization technique aims to find out characteristics of failure test cases.

To describe the characteristics, a model of minimal failure-causing schema was proposed [1]. Considering a boolean expression: $a \wedge (^{-}b \vee^{-} c) \wedge d \vee e$, and a clause disjunction fault (CDF) mutant: $a \wedge (^{-}b \vee^{-} c) \wedge d \vee (d \vee e)$ [2]. There are total 5 failure test cases. In all 7 failure-causing schemas, (- 1 1 1 0) and (0 - - 1 0) are minimal ones. They predict that all 5 input variables are involved in the fault. But factually, only d and e are related to this fault, where e is replaced by $d \vee e$.

	a	b	c	d	e
$test_1$	0	0	0	1	0
$test_2$	0	0	1	1	0
$test_3$	0	1	0	1	0
$test_4$	0	1	1	1	0
$test_5$	1	1	1	1	0

	a	b	c	d	e
$schema_1$	0	0	0	1	0
$schema_2$	0	0	1	1	0
$schema_3$	0	1	0	1	0
$schema_4$	0	1	1	1	0
$schema_5$	1	1	1	1	0
$schema_6$	-	1	1	1	0
$schema_7$	0	-	-	1	0

In this paper, we propose a model of probabilistic failure-causing schema to describe this phenomena.

II. Probabilistic Failure-causing Schema

Considering a program under test with n input variables, each variable has a value set V_i ($i = 1, 2, ..., n$). The input domain of program is $D = V_1 \times V_2 \times ... \times V_n$.

Definition 1 (test case). A test case is a n-tuple ($v_1 \in V_1$, $v_2 \in V_2$, ..., $v_n \in V_n$).

Definition 2 (schema). A k-value schema (or called a schema with strength k) s is a n-tuple (-, ..., -, $v_{i,1}$, -, ..., -, $v_{i,2}$, -, ..., -, $v_{i,k}$, -, ..., -) where $1 \leq k \leq n$. Where, the values of k variables have been fixes, while the values of other $n-k$ variables have not been fixed as denoted as "-".

Definition 5 (probabilisitic failure-causing schema). A k-value schema is a k-value probabilisitic failure-causing schema with a failure probability p_{fail}, if the ratio of the number of failure test cases that contain such schema to the number of test cases that contain such schema is p_{fail}.

A schema with failure probability p_{fail} means that, for arbitrary test case $t \in D = V_1 \times V_2 \times ... \times V_n$ that contains such schema, the probability that t is a failure one is p_{fail}.

Definition 6 (coverage probability). The coverage probability p_{cov} of a schema is the ratio of the number of test cases that contain such schema to the number of all test cases.

A schema with coverage probability p_{cov} means that, for arbitrary test case $t \in D = V_1 \times V_2 \times ... \times V_n$, the probability that t contains such schema is p_{cov}. There is negative correlation between the coverage probability and the strength of schema

III. Approach

People need characterize schemas with both higher failure probability and higher coverage probability, since the more p_{fail} means there is closer relationship between the schema and the fault, and the the more p_{cov} predicts less input variables that may be concerned with fault. By define a metric:

$$score(s) = p_{cov}(s) \times p_{fail}(s)$$

We can select probabilisitic failure-causing schemas with the greatest score in input-level fault localization.

For previous example, we calculate p_{fail} and p_{cov} for each schema in 5 failure test cases (see Fig. 1). Three probabilistic failure-causing schemas with the greatest scores include:

$$score(\text{- - - 1 0}) = \frac{1}{4} \times \frac{5}{8} = \frac{5}{32}$$
$$score(\text{- - - 1 -}) = \frac{1}{2} \times \frac{5}{16} = \frac{5}{32}$$
$$score(\text{- - - - 0}) = \frac{1}{2} \times \frac{5}{16} = \frac{5}{32}$$

They predict that 2 variables d and e are involved in the fault.

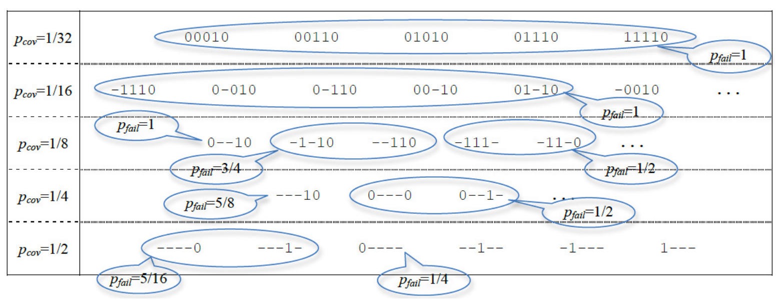

Fig. 1. Probabilistic failure-causing schemas (only some with greate p_{fail})

IV. Conclusion

A model of probabilistic failure-causing schema in input-domain testing is proposed. A simple example shows that they could reveal the source of faults more precision. More experiments are required in future works.

References

[1] C. Nie, H. Leung. The Minimal Failure-causing Schema of Combinatorial Testing. ACM Transactions on Software Engineering and Methodology (TOSEM), 2011, 20(4): 15.

[2] Z. Chen, T. Y. Chen, B. Xu. A Revisit of Fault Class Hierarchies in General Boolean Specifications. ACM Transactions on Software Engineering Methodology (TOSEM), 2011, 20(3).