Relationship Between Test Effectiveness and Coverage

Ran Ye and Yashwant K. Malaiya
Computer Science Dept.
Colorado State University
Fort Collins, CO 80523
+1 970 491 7031
ye|malaiya@cs.colostate.edu

ABSTRACT

A model has recently been developed that relates test coverage and defect coverage. It has been shown that this model can be used to estimate the residual defect density. Defect finding capability of sets of tests with different coverage values has recently been studied by Frankl and Iakounenko. They have defined measure *test effectiveness*, and have evaluated it for programs containing a single fault. Here we explore the applicability of Malaiya et al's model to the test data reported by Frankl and Iakounenko.

1 INTRODUCTION

Several test coverage metrics can be automatically evaluated using coverage tools. A coverage metric evaluates thoroughness of testing and thus can be expected to be related to probability of detecting defects. We can define defect coverage as the fraction of defects encountered during testing. A model relating test coverage and defects found has been presented by Malaiya et al. [1,2] that can been used to estimate number of residual defects [4,5].

Coverage enumerable can include structural elements like statements or decisions, or data-flow elements like define-use associations (dua coverage). The fault detection ability of decision coverage and dua coverage has recently been studied by Frankl and Iakounenko [3]. They have defined a measure called *test effectiveness* and have evaluated it experimentally. In this paper we examine the relationship between test effectiveness and defect coverage. We obtain a model for test effectiveness using Malaiya et al.'s model, and fit it to the data experimentally obtained by Frankl and Iakounenko.

The model by Malaiya et al. is based on the Logarithmic Poisson reliability growth model, which assumes that the number defects, found grows logarithmically with the number of tests applied. They also assumed that the number of coverage enumerables (like decisions) exercised also grows according to a similar logarithmic expression. If we eliminate the number of tests and express defect coverage C^0 directly as a function of test coverage C^i, we can write, using suitable parameters [1],

$$C^0 = a_0^i \ln[1 + a_1^i(\exp(a_2^i C^i) - 1]$$
$$\approx a_0^i \ln[a_1^i(\exp(a_2^i C^i)] = -A + BC^i \quad C^i > C^i_{knee}$$

It has been found that for the values of C^i greater than a knee-value C^i_{knee}, this model can be approximated by a linear function [2,4]. We will refer to this as the logarithmic-exponential (LE) model.

2 TEST EFFECTIVENESS EXPERIMENTS

An empirical evaluation of the fault-detecting ability of two white-box software testing techniques, decision coverage and the all-use data flow testing criterion was reported by Frankl and Iakounenko [3]. They examined a C program with about 11,640 source lines. They created 33 faulty versions by reintroducing one defect in each of them. Of these, 8 versions were selected because of their lower failure rates. They tested each of them by applying a fixed number of randomly chosen tests drawn from a universe of 10,00 test cases. Here we consider versions V7 and V8. Each of them was tested by applying many randomly selected sets of 20 tests. For each set of 20 tests, they evaluated the decision coverage and dua coverage in addition to Test Effectiveness (defined below). They also estimated the error bounds of those estimates.

3 ANALYSIS OF EMPIRICAL DATA

Frankl and Iakounenko define Test Effectiveness (TE) as an estimate of the proportion of the tests in the set (of a specific size) of coverage of at least c_i that detects the single fault. We can rewrite this definition as:

TE = {Sum of all the probability of detect the fault at a given coverage c_x, when test coverage ≥ c_x}

$$TE = \frac{\int_{c=c_x}^{1} \Pr\{detection \mid cov = c\} \cdot \Pr ob\{cov = c\} \cdot dc}{\int_{c=c_x}^{1} \Pr ob\{coverage = c\} \cdot dc}$$

Since the test vectors were selected from a large universe, we can make a preliminary assumption that coverage of test vectors obey to the normal distribution, with different mean value (μ) and standard deviation (σ) respectively. If we assume that Malaiya et al.'s LE model yields the probability of detecting the fault, then the expression of test effectiveness can be rewritten as:

TE=

$$\frac{\int_{c=c_x}^{1} a_0 \cdot \ln[1 + a_1 \cdot (e^{a_2 c} - 1)] \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp[-\frac{1}{2}(\frac{c-\mu}{\sigma})^2] \cdot dc}{\int_{c=ci}^{1} \frac{1}{\sqrt{2\pi\sigma}} \exp[-\frac{1}{2}(\frac{c-\mu}{\sigma})^2] \cdot dc}$$

This expression, referred below as the TE-coverage model serves a model to correlate TE with coverage using Malaiya et al.'s LE model.

Table 1: Parameters of the TE-Coverage Model

Data	a_0	a_1	a_2
V7 DUA cov	10.65	$3.79*10^{-3}$	3.00
V7 Deci cov	10.64	$3.79*10^{-3}$	2.99
V8 DUA cov	4.93	$1.68*10^{-9}$	18.42
V8 Deci civ	4.93	$1.68*10^{-9}$	18.42

Numerical computation was used to fit the data (Figure 1) to the TE coverage model. Table 1 gives the values of parameters a_0, a_1 and a_2 for two of the versions, V7 and V8. Version 7 and Version 18 have the same program size but very different failure exposure rates for the two defects, which result in the large difference for the a_1^i value.

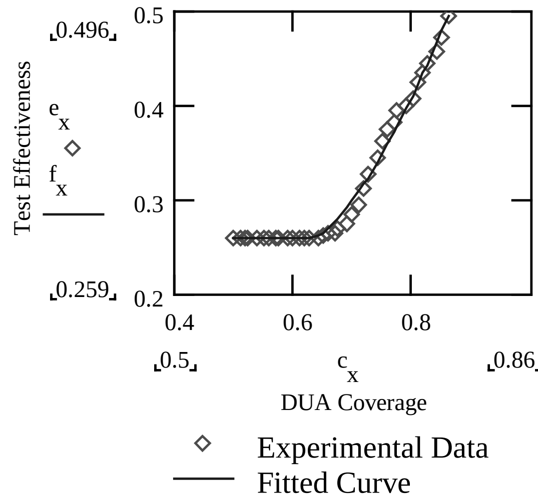

Figure 1: Experimental data (V7 DUA) and fit

CONCLUSIONS

The plots for defect coverage vs. test coverage [1,2,4] and TE vs. test coverage [3] show a knee after which the plot rises sharply. The TE plots show a flat region before the knee. We have presented an approach to characterize dependence of TE on test coverage in terms of Malaiya et al.'s LE model. The plots presented in [3] provide further insight into dependence of defect detection probability and test coverage. The fit of the data to the TE coverage model suggests that the model proposed in [1] is consistent with the data obtained in [3].

REFERENCES

1. Y.K. Malaiya, N. Li, J. Bieman, R. Karcich and B. Skibbe, "The Relationship between Test Coverage and Reliability" *Proc. ISSRE*, Nov. 1994, pp.186-195.
2. Y. Malaiya, N. Li, J. Bieman, R. Karcich, Software Reliability Growth with Test Coverage" to appear in IEEE Trans. Reliability, Dec. 2002.
3. P. Frankl and O. Iakounenko, "Further Empirical Studies of Test Effectiveness", *SIGSOFT '98*, Nov, 1998
4. Y.K. Malaiya and J. Denton, "Estimating the Number of Residual Defects" *Proc. IEEE High Assurance Systems Engineering Symposium*, November 1998, pp. 98-105.
5. H. Buczilowski, "Applying Malaiya et al.'s Method for Estimating the Number of Defects Based on Coverage of Large Datasets", *Proc. ISSRE Fast Abs & Ind. Practices*, Nov. 1999, pp 141-149.