Early Applicability of the Coverage/Defect Model

Yashwant K. Malaiya, Conrad Braganza and Chintan Sutaria
Computer Science Dept.
Colorado State University
Fort Collins, CO 80523
malaiya@cs.colostate.edu

Abstract

Test coverage is closely related to the defect detection efficiency. A model to relate test coverage and defects found has recently been proposed and tested using actual data sets. Most of the data sets used so far have been obtained by testing software that has already been debugged to some extent. Here we examine the question of the applicability of the MLBKS model very early when little or no prior debugging has been performed. Data collected from an actual project is analyzed. The results demonstrate that the model successfully describes the early relationship between the test coverage and the defects found. The results confirm earlier analysis that suggests that the position of the knee of the curve describing the model should depend on the initial defect density.

1. COVERAGE & DEFECTS

The test coverage tools can automatically evaluate test coverage metrics. A coverage metric evaluates thoroughness of testing and is thus related to the probability of detecting defects [1,2]. Coverage enumerable can include structural elements like statements or decisions, or data-flow elements like P-uses or C-Uses. A model relating test coverage and defects found was proposed by Malaiya et al. [1,3]. This model, termed the MLBKS model here, can be used to estimate number of residual defects [4,5]. The model is consistent with the experimental evaluation of the fault detection ability of the coverage metrics using a measure called "test effectiveness" by Frankl and Iakounenko [6,7]. Another model relating defects and coverage was recently proposed by Bishop [8].

Frequently a developer would like to use a model in the initial phases of a development project. This makes it desirable that a model should be applicable in the early phases of the project. However the data sets often represent the later phases of the project, when the initial defect density is already low. The initial defect densities represented by Vouk's data sets are about 2-4/KOLC [1,3] and about 6-7/KLOC for Pasquini et al's data set [4]. The final defect densities for these data sets can be estimated to be about 0.5-1/KLOC and 2-3/KLOC respectively. Here we examine the question of the validity of the MLBKS model during the early phases when the defect density is significantly higher.

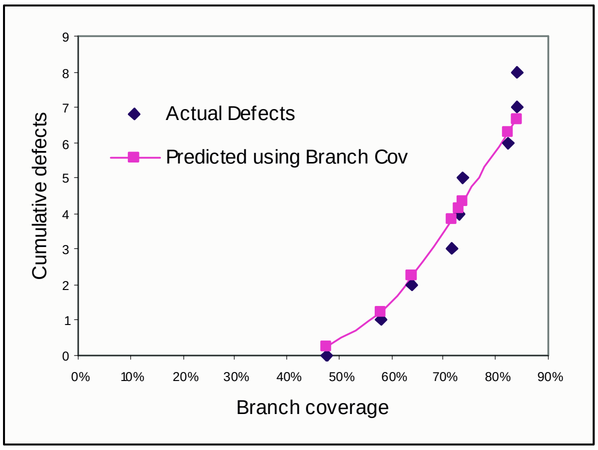

Figure 1: Vouk's Data set 1

The MLBKS model assumes that both the number of defects found and the coverage enumerables exercised, grow logarithmically with the number of tests applied. The model eliminates the number of tests and expresses defect coverage C^0 directly as a function of a test coverage metric C^i. We can write, using suitable parameters [5],

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

For the values of C^i greater than a knee-value C^i_{knee}, this model can be approximated by a linear function [4], as given above in terms of parameters A and B. The data sets examined in the past show a linear trend as suggested by the model. It can also be shown that the knee (where the linear trend would intersect the x-axis) can be approximated by [4]

$$C^{i}_{knee} = 1 - FD_0$$

where F is a parameter and D_0 is the initial defect density. Thus with higher defect density the knee should occur at lower coverage values.

The next section describes an experiment conducted to explore the relationship between coverage and defect found early during the development cycle.

2. EARLY RELATIONSHIP EXPERIMENT

A team of two programmers developed two programs, one with no initial inspection or debugging, which was found to have 20 defects. The other program was carefully inspected and only three defects were found in it. Data from the first program `stringfunction.c` is examined here. During testing of the 250 LOC program, ten test cases were applied, and with the final 96.7% block coverage and 83.8% decision coverage, twenty defects were found. A plot of the decision coverage along with the fitted MLBKS model is shown in Figure 2.

A chi-square test found that the MLBKS model fits the data well. The first two test cases, which resulted in coverage of 1.0% and 1.9%, did not uncover any defects. The third test case detected eight defects. The data points with coverage exceeding 40% form a straight line which would intersect the x-axis at 12.1% coverage. If we assume that `stringfunction.c` does not have any dead (or unreachable) code, then at 100% decision coverage, about 24-25 defects would have been found. This suggests that about 4-5 undiscovered defects are likely to be present [4,5].

Table 1 gives the parameters values for branch coverage for `stringfunction.c`, Vouk's Dataset 1 and Pasquini's dataset. As we would expect, with much higher initial defect density, the knee for `stringfunction.c` occurs much earlier. A similar result is obtained when block coverage data is used.

Table 1: The parameters for the three data sets

Data Set (branch)	A	B	C^{i}_{knee}
Vouk's Data set 1	-1.201	2.201	54.6%
Pasquini's data set	-0.801	1.693	47.3%
stringfunction.c	-0.138	1.138	12.1%

3. CONCLUSIONS

The results demonstrate the applicability of the MLBKS model early in the development life-cycle. The results conform the expectation [4] that when prior testing is minimal and the initial defect density is higher, the knee should occur earlier. In such a case, the defects with high detection probability are still present, to be found with relative ease. The model can be used for early determination of the potential number of faults for optimum resource allocation, and for estimation of the number of potential residual defects.

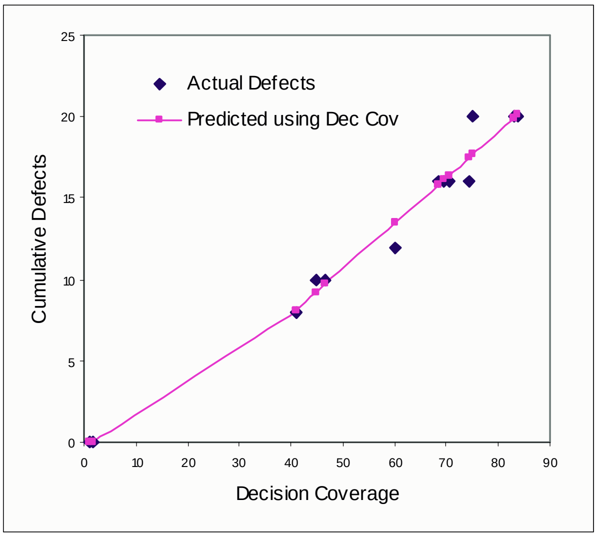

Figure 2: `stringfunction.c` data

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. M.-H. Chen, M. R. Lyu and W. E. Wong, "An empirical study of the correlation between code coverage and reliability estimation," METRICS 1996, pp. 133-141.
3. Y. Malaiya, N. Li, J. Bieman, R. Karcich, "Software Reliability Growth with Test Coverage" IEEE Trans. Reliability, Dec. 2002, pp 420-426.
4. Y.K. Malaiya and J. Denton, "Estimating the Number of Residual Defects" Proc. HASE, 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
6. P. Frankl and O. Iakounenko, "Further Empirical Studies of Test Effectiveness", SIGSOFT '98, Nov, 1998, pp. 153-162.
7. R.Ye and Y.K. Malaiya, "Relationship between test effectiveness and Coverage," in Supp. Proc. ISSRE, Nov 2002, pp. 159-160.
8. P.G. Bishop, "Estimating Residual Faults from Code Coverage," Proc. SAFECOMP, Sept. 2002, pp. 163-174.