

Evaluate the Effectiveness of Test Suite Prioritization Techniques Using APFD Metric

Mr. Anil Mor

M.Tech. (Computer Science & Engineering) University Institute of Engineering & Technology, M.D.U. Rohtak

Abstract: Regression testing is a testing activity that is performed to provide confidence that changes do not harm the existing behavior of the software. Test suites tend to grow in size as software evolves often making it too costly to execute entire test suites. To reduce the cost of regression testing, test cases are prioritized. One goal of test case prioritization technique is to increase a test suite's rate of fault detection and to reduce the cost of regression testing. We have provided a metric Average Percentage of Fault Detected (APFD), for measuring rate of fault detection during prioritization. APFD is also being used in its mutant form as APFDc, APFDp, ASFD, WPFd, TSFD, APBC, APDC, APSC, NAPFD, APMC, TPFd, APRC and BPFg. This paper presents a review of all metrics used for assessing the rate of fault detection of prioritized test cases based on coverage criteria for the various prioritization techniques.

Keywords: Regression Testing, Test case prioritization, Average Percentage of Fault Detection (APFD), APFD Metric

I. Introduction

Regression Testing: - (Onoma, Tsai, Poonawala, & Suganuma, 1988)[9] Regression testing can be defined as follows:-

Let P be a program, let P' be a modified version of P and let T be a test suite developed by P. Regression testing is concerned with validating P'. To facilitate regression testing, engineers typically reuse T, but new test cases may also be required to test new functionality. Both reuse of T and creation of new test cases are important, however, it is test case reuse that is of concern here, as such reuse typically forms a part of regression testing processes.

Test Case Prioritization: - Test case prioritization techniques (G, R, C., & M.J., 2001)[3] schedule test cases so that those with the highest priority, according to some criteria, are executed earlier in the regression testing process than lower priority test cases. A potential advantage of these techniques is that unlike test case reduction and non-safe regression test selection techniques, they don't discard tests. Many different prioritization techniques have been proposed (Kim & Porter, 2002)[6] but the techniques utilize simple code coverage information like statement and method coverage. For example, proper test case prioritization techniques increase the fault detection rate of a test suite and the chance of executing test cases with higher rate of fault detection earlier (Elbaum, Malishevsky, & Rothermel, 2002)[1].

We adopt the test case permutation problem from as follows (G, R, C., & M.J., 2001)[3]:-
Given:- T, a test suite; PT, the set of permutations of set T; and f, a function from PT to real numbers. (For example, f may calculate the fault detection rate of a permutation of T).

Problem:- To find T' ∈ PT such that, $\forall T'' \in PT, T'' \neq T' \Rightarrow f(T') \geq f(T'')$

Metrics: - To properly understand the effects and the outcomes of any case study or experiment, one needs to quantify the results or analyze them with respects to the measures, well known in the testing field as metrics (Yogesh & Arvinder, 2012)[11]. Metrics can be classified as shown in Figure 1.

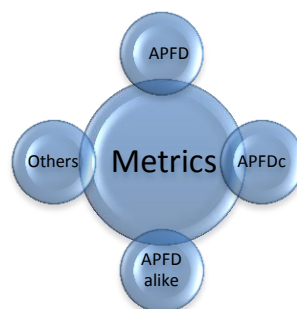


Figure 1:- Classification of metrics

There are some types of metrics as explained below:-

- Average Percentage of Fault Detected (APFD):- APFD metric denotes the weighted average of the faults detected. The metric of APFD is widely adopted in evaluating test case prioritization techniques.
- Average Percentage of Faults Detected with Cost (APFDc):- APFDc is the modified APFD to include the costs of faults.
- Average Percentage of Fault Detected alike (APFD alike):- Depending on the coverage criteria (Li, Harman, & Hierons, 2007)[7] (statement, branch etc.) considered, the following metrics were computed:-
 - Average Percentage of Statement Coverage (APSC):- this measures the rate at which a prioritized test suite covers the statements (Li, Harman, & Hierons, 2007)[7].
 - Average Percentage of Branch Coverage (APBC):- This measures the rate at which a prioritized test suite covers the branches (Li, Harman, & Hierons, 2007)[7]. This metric is also represented as Average Percentage of Block Coverage.
 - Average Percentage of Loop Coverage (APLC):- This measures the rate at which a prioritized test suite covers the loops.
 - Average Percentage of Condition Coverage (APCC):- This measures the rate at which a prioritized test suite covers the condition.
- Others;- Various other metrics, whether available or self-developed, such as PTR, RFFT, ATEL, FDD, LOC count, size reduction, Kruskal Wallie Test and distance etc. have also been taken.

II. Measuring Effectiveness Using Various Metrics

Average Percentage of Fault Detected (APFD):- Depending upon the fault criterion considered, APFD metric was computed (H & G, 2006)[4] to measure the rate of fault detection of coverage based prioritization techniques. APFD measures the weighted average of the percentage of faults detected over the life of a test suite. APFD values ranges from 0 to 100, higher number simply faster fault detection rates.

Let T be a test suite containing n test cases, and let F be a set of m faults revealed by T. Let TF_i be the first test case in ordering T_0 of T which reveals fault i. The APFD for test suite T_0 is given by the equation as:-

$$APFD = 1 - \frac{TF_1 + TF_2 + \dots + TF_m}{nm} + \frac{1}{2n}$$

Average Percentage of Fault Detected with Cost (APFDc):- The cost of the test case is related to the resources required to execute and validate it. Various measures are possible like:-

- When the primary required resource is machine or human time, test cost can be measured in terms of the actual time needed to execute a test case.
- Another measures considers the monetary costs of test case execution and validation, this may reflect hardware cost, wages, cost of materials required for testing and so on.

Depending on the cost criterion we have considered, APFDc metric is computed (Malishevsky, Ruthru, Gregg, & Elbaummy, 2006)[8]. Let T be a test suite containing n test cases with costs t_1, t_2, \dots, t_n . Let F be a set of m faults revealed by T_i and let f_1, f_2, \dots, f_m be the severities of those faults. Let TF_i be the first test case in an ordering T' of T that reveals fault i. the (cost-cognizant) weighted average percentage of faults detected during the execution of test suite T' is given by an equation as:-

$$APFDc = \frac{\sum_{i=1}^m (f_i \times (\sum_{j=TF_i}^n t_j - \frac{1}{2} t_{TF_i}))}{\sum_{i=1}^n t_i \times \sum_{i=1}^m f_i}$$

APFD alike: - “APFD alike” are basically the metrics which are calculating average percentage of faults detected with some variations in calculation method. These variations can be statements, branches, loops and conditions etc. (Yogesh & Arvinder, 2012)[11]

- APBC (Average Percentage Branch Coverage):- In calculating APBC metric, a test suite containing n test cases that covers a set of m branches is taken. Let TF_i be the first test case in ordering T' of T which covers branch i. The APBC for test suite T' , is given by the equation as:-

$$APBC = 1 - \frac{TB_1 + TB_2 + \dots + TB_m}{nm} + \frac{1}{2n}$$

APBC measures the weighted average of the percentage of branches covered over the life of a test suite. APBC values ranges from 0 to 100.

- APSC (Average Percentage of Statement Coverage):- Similar to the APBC, in APSC the statements are covered in place of branches. On the basis of the value of APSC metric, the test case prioritization takes place. The value of the APSC can be calculated as:-

$$APSC = 1 - \frac{TS1 + TS2 + \dots + TS_m}{nm} + \frac{1}{2n}$$

APSC measures the weighted average of the percentage of statements are covered over the life of a test suite. Similarly APLC and APCC are calculated on the basis of the coverage of loops and conditions respectively.

Other Metrics: - There are some other metrics which are available to measure the effectiveness of Test Suite Prioritization e.g. PTR metric (Raju & Uma, 2012)[10].

- PTR Metric:- The PTR Metric is another way that the effectiveness of a test prioritization may be analyzed (Raju & Uma, 2012)[10]. Recall that an effective prioritization technique would place test cases that are most likely to detect faults at the beginning of test sequence. It would be beneficial to calculate the percentage of test cases that must be run before all faults have been revealed. PTR is calculated as follows:-

Let T be the test suite under evaluation, n be the total number of test cases in T and n_d the total number of test cases needed to detect all faults in the program under test P.

$$PTR(T, P) = \frac{n_d}{n}$$

There are some other metrics in the category of other metrics such that RFFT, FDD, ATEL etc. which are also used for the analysis of effectiveness of test prioritization.

III. Experimentation And Analysis

To quantify the goal of increasing a subset of the test suite's rate of fault detection, I use a metric called APFD developed by Elbaum et. al.[1,8] that measures the average rate of fault detection per percentage of test suite execution. The APFD is calculated by taking the weighted average of the number of faults detected during the run of the test suite. APFD can be calculated using a notation:

Let T -> The test suite under evaluation

m -> the number of faults contained in the program under test P

n -> the total number of test cases and

TF_i -> the position of the first test in T that exposes fault i

$$APFD = 1 - \frac{TF_1 + TF_2 + \dots + TF_m}{nm} + \frac{1}{2n}$$

So as the formula for APFD shows that calculating APFD is only possible when prior knowledge of faults is available.

Suppose for my study,

m (number of faults) = 19

n (number of test cases) = 97

T1 be the test suite (Non-Prioritized) is like:

TC-001, TC-002, TC-003TC-081, TC-082, TC-101, TC-102, TC-103 TC-115

T2 be the test suite (Prioritized) is like:

TC-001, TC-002, TC-113, TC114, TC-003----TC-023, TC-026, TC-025, TC-024, TC-102---TC-107, TC-028, TC-029, TC-027, TC-101, TC-108----TC-112, TC-030----TC-045, TC-055, TC-056, TC-058, TC-057, TC-046---TC-054, TC-059----TC-082, TC-115

Test Cases which contain faults are as:

TC-026, TC-029, TC-030, TC- 055, TC-058, TC-101, TC-102, TC-103, TC-104, TC-105, TC-106, TC-107, TC-108, TC-109, TC-110, TC-111, TC-112, TC-113, TC-114

APFD value for non-prioritizes test cases i.e. for test suite T1

$$APFD = 1 - \frac{26 + 29 + 30 + 55 + 58 + 101 + 102 + 103 + 104 + 105 + 106 + 107 + 108 + 109 + 110 + 111 + 112 + 113 + 114}{19 \times 97} + \frac{1}{2 \times 97}$$

$$APFD = 1 - \frac{1703}{1843} + \frac{1}{194}$$

$$APFD = 1 - 0.924 + 0.005$$

$$APFD = 1.005 - 0.924$$

$$APFD = 0.081$$

APFD value for prioritized test cases i.e. for test suite T2

$$APFD = 1 - \frac{26 + 36 + 44 + 60 + 62 + 38 + 29 + 30 + 31 + 32 + 33 + 34 + 39 + 40 + 41 + 42 + 43 + 3 + 4}{19 \times 97} + \frac{1}{2 \times 97}$$

$$APFD = 1 - \frac{667}{1843} + \frac{1}{194}$$

$$APFD = 1 - 0.361 + 0.005$$

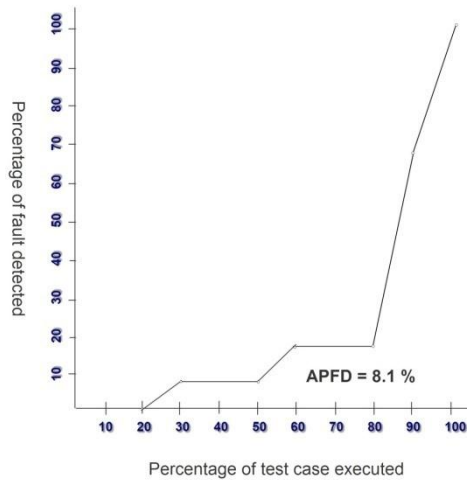
$$APFD = 1.005 - 0.361$$

$$APFD = 0.64$$

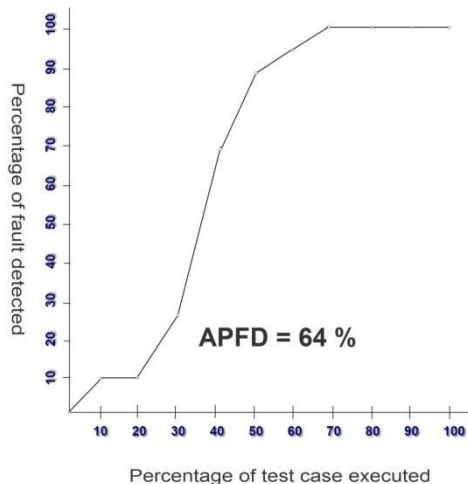
Analysis of APFD

The comparison is drawn between prioritized and non-prioritized case, which shows that value obtained for prioritized case is more than non-prioritized case, hence more effective of prioritized case. Below two graphs showing the results for prioritized (T2) and non-prioritized case (T1)

APFD graph for non-prioritized case (T1)



APFD graph for prioritized case (T2)



IV. Conclusion

To improve the rate of fault detection we have computed a number of metrics that predicts the quality and effectiveness of the various prioritization techniques required for regression testing. In this paper we found how to measure the effectiveness using APFD (Average Percentage of Fault Detected). By using the information or values given by the metric, test prioritization can be done using the appropriate technique. Higher the value of APFD metric gives the better rate detection.

References

- [1] Elbaum, S. G., Malishevsky, A. G., & Rothermel, G. (2002). Test Case Prioritization: A Family of Empirical Studies. IEEE Transaction on Software Engineering .
- [2] Elbaum, S., Gable, D., & Rothermel, G. (2001). Understanding and Measuring the Sources of Variation in the Prioritization of Regression Test Suites. International Software Metrics Symposium.
- [3] G, R., R, U., C., C., & M.J., H. (2001). Prioritization for Regression Testing. IEEE Transaction for Software Engineering .
- [4] H, D., & G, R. (2006). On the Use of Mutation Faults in Empirical Assessment of Test Case Prioritization Techniques. IEEE Transaction of Software Engineering .
- [5] Hou, S. S., Zhang, L., Xie, T., & Sun, J. S. (2008). Quota-Constrained Test Case Prioritization for Regression Testing of Service Centric Systems. IEEE International Conference on Software Maintenance (ICSM 2008).
- [6] Kim, J., & Porter, A. (2002). A History Based Test Prioritization for Regression Testing in Resource Constrained Environments. Proceeding International Conference Software Engineering.
- [7] Li, Z., Harman, M., & Hierons, R. M. (2007). Search Algorithm for Regression Test Case Prioritization. IEEE Transaction on Software Engineering .
- [8] Malishevsky, A. G., Ruthru, J. R., Gregg, Y., & Elbaum, R. S. (2006). Cost-Cognizant Test Case Prioritization . Technical Report TR-UNL-CSE-2006-0004, Dept of Computer Science, Lincoln, USA .
- [9] Onoma, K., Tsai, W.-T., Poonawala, M., & Suganuma, H. (1988). Regression Testing in an Industrial Environment. Comm.ACM .
- [10] Raju, S., & Uma, G. V. (2012). Factors Oriented Test Case Prioritization Techniques in Regression Testing using Genetic Algorithm. European Journal of Scientific Research .
- [11] Yogesh, S., & Arvinder, K. (2012). Systematic Literature Review on Regression Test Prioritization Techniques. Informatica 36 .
- [12] Yoo, S., & Harman, M. (2007). Regression Testing Minimisation, Selection and Prioritization : A Survey. Wiley InterScience .