A Hybrid Adaptive Regression Testing Strategy

Priyanka
Research scholar, YMCAUST Faridabad

Harish Kumar
Assistant Professor, YMCAUST Faridabad

Naresh Chauhan
Professor, YMCAUST Faridabad

Abstract – All software systems need modifications with time, these modifications involve different types or amounts of code modifications in different versions. To validate these modifications many regression testing sessions are needed. But researchers do not have a single regression testing technique that can be used on every version. To date, no work has been done on the problem of choosing the most cost effective testing technique to use on every version. To address this critical problem, a hybrid adaptive regression testing (HART) strategy has been proposed that attempts to identify the regression testing techniques that will be the best for each regression testing session considering organization's situations and testing environment. The result showed that prioritization techniques selected by proposed technique is more cost-effective than those used by the previous approaches.

Index Terms – Hybrid adaptive regression testing strategy, Multiple criteria decision making, Regression testing, Test case prioritization techniques.

1. INTRODUCTION

Regression testing is an important and necessary activity that can be performed on modified software to provide confidence that the changes do not have adverse affect on working of software and behave as expected and also that the unchanged parts of the code are not adversely affected by the modifications. To date, researchers have provided many regression testing techniques. For instance, regression test selection techniques [1,2,3] reduce testing costs by selecting a subset of the existing test cases to rerun on the modified software in order to reduce the costs of regression testing. Test case prioritization [4,5,6] reorder test cases, schedule test cases for execution in an order that attempts to increase their effectiveness at meeting some performance goal.

While this research has made significant progress in regression testing areas, one critical problem has been overlooked. As all software systems require modifications and these modifications involved different types or amounts of code modifications and these changes can affect the costs and benefits of regression testing techniques in different ways. . But researchers do not have a single regression testing technique that can be used on every version.

To address this critical problem a strategy has been proposed known as *hybrid adaptive regression testing* (HART) strategy. HART strategy is a approach that apply across system lifetimes, and attempt to identify the regression testing techniques that will be the best for each regression testing session in term of cost and benefit. HART strategy evaluates regression testing techniques in terms of decision criteria such as cost and benefit factors and select the best alternative among techniques considering organization's situations and feedback from prior regression testing sessions.

The problem of finding the best alternative is known as the "multiple criteria decision making" (MCDM) problem, and MCDM approaches have been used in many science, engineering, business areas an many real life problem[7,8].

To date, there are many MCDM approaches including the Weighted Sum Model (WSM), the Weighted Product Model (WPM), the Analytical Hierarchy Process (AHP), the Technique for Order Preference by Similarity of Ideal Solution (TOPSIS) and other variants. Among these MCDM methods, AHP and TOPSIS have been the more popular methods, having been used by researchers and practitioners in various areas including software engineering.

In this work two step methods has been used. In first step AHP is used for calculating the weights of the attributes or

criteria as well as the overall weights of the candidates in each attribute. In second step these weights are considered and used in TOPSIS process. Then TOPSIS is applied for the evaluation problem and the result shows the preference order of alternatives. Hence result showed that the prioritization techniques selected by this hybrid approach is more cost-effective than those used by the control approaches.

2. RELATED WORK

Regression testing attempts to validate modifications in the programs to see whether changes have produced adverse effects. Depending on various factors, such as the size and complexity of the program and its test suite, regression testing session can be very costly. Thus, there are numerous cost-effective regression testing techniques including regression test selection, test suite reduction/minimization, and test case prioritization, but in this paper work is limited to test case prioritization, which is directly related to work.

Test case prioritization techniques [6,10] reorder test cases, schedule test cases for execution in an order that attempts to increase their effectiveness at meeting some performance goal. There are many test case prioritization techniques, for example *total statement coverage* orders the test cases based on the total number of statements covered by them. It counts the number of statements covered by the test cases and orders them in descending order of this number. One variation of this technique, *additional block coverage prioritization* iteratively selects a test case T1,that yields the greatest statement coverage, and then selects a test case which covers a statement uncovered by the T1. Repeat this process until all statements covered by at least one test case have been covered.

Yoo and Harman [11] in their recent paper provide a comprehensive overview of these techniques. The goal of the proposed techniques is to improve the effectiveness of regression testing. *Alessandro Orso and Mary Jean Harrold* in their paper presents a new regression-tests election technique for Java programs that is safe, precise, and yet scales to large systems.

These studies have allowed researchers to understand factors that affect the use of techniques and to compare techniques in terms of costs and benefits relative to actual software systems. However, no considerable research has been made for selecting appropriate techniques under particular circumstances as systems evolve. Only few studies [12,13] have done on the problem of helping practitioners choose appropriate techniques under particular system.

Selecting appropriate techniques for each version can be a multiple criteria decision making (MCDM) problem. Analytic Hierarchy Process (AHP) and the Technique for Order Preference by Similarity of Ideal Solution (TOPSIS) are two widely used MCDM methods. For instance, *Kamal and Al-Harbi* [14] use AHP in project management to determine the contractors' competence or ability to participate in the project bid. AHP has also been used in determining the best manufacturing system [15], layout design [16], and the evaluation of technology investment decisions [17].

Md. Junaid Arafeen and Hyunsook Do[18] in their paper present a adaptive regression testing strategy which use AHP method that attempt to identify the regression testing techniques that will be the best for each regression testing session in term of cost and benefit. *Pema Wangchen Bhutia*[19] in his paper use both AHP and TOPSIS method in determining the suitable suppliers who are able to provide the buyer with the right quality products and/or services at the right price, at the right time and in the right quantities.

In this paper, AHP and TOPSIS method has been used to develop adaptive regressions testing strategy, which identify the best test case prioritization technique.

3. PORPOSED MODELLING

This section, describes how AHP and TOPSIS methods are used to achieve the goal.

3.1. Ahp method

To use AHP method, a hierarchy has been defined that describes the problem. An AHP hierarchy consists of goal has to be achieve, alternative techniques that are available to reach to goal, criteria that are factor that may be used in decision making about these alternative techniques. Criteria can be further partitioned into sub-partition if required.

Once an AHP hierarchy has been defined as shown in figure 1, a comparison is performed: between pairs of criteria as shown in Table 1. When comparing pairs of criteria relative importance weights has been assigned to criteria; C1 is given importance 4 relative to C4. After completing the matrix, local priority of each criterion is calculated using (1).

$$LP_i = \frac{\sum_{j=1}^{N} RW_{ij}}{\sum_{i=1}^{N} \sum_{J=1}^{N} RW_{ij}} \qquad (1).$$

where LP_i is a local priority of criterion i, RW_{ij} is a relative weight of criterion i over criterion j, and N is the number of criteria.

To apply AHP to prioritization strategy, the following steps are required:

1. Set a goal.
2. Identify alternatives that are available to reach goal.
3. Identify evaluation criteria for alternatives techniques.
4. Comparisons: between pairs of criteria.
5. Obtain local priorities of criteria.

In proposed approach AHP method is used to calculate local priority of each criteria and these local priority can use as a input in TOPSIS method to rank the alternative techniques.
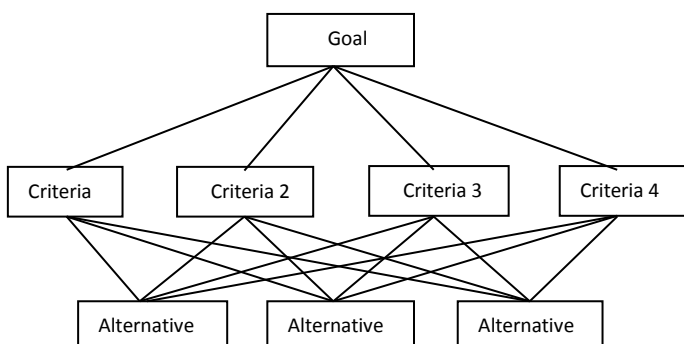

Figure 1 An AHP Hierarchy

	C1	C2	C3	C4	Local priority
C1	1	1/3	2	4	0.1
C2	3	1	5	3	0.4
C3	½	1/5	1	1/3	0.3
C4	¼	1/3	3	1	0.2

Table 1 Example of AHP Method

3.2. TOPSIS Method

TOPSIS is a decision making technique. It is a goal based approach for finding the alternative that is closest to the ideal solution. TOPSIS is based on the idea that the chosen alternative should have the shortest distance from the Positive Ideal Solution (PIS) and on the other side the farthest distance of the Negative Ideal Solution (NIS). The Positive Ideal Solution maximizes the benefit criteria and minimizes the cost criteria, whereas the Negative Ideal Solution maximizes the cost criteria and minimizes the benefit criteria.

So basically in this method two artificial alternatives are hypothesized:

Ideal alternative: the one which has the best level for all attributes considered.

Negative ideal alternative: the one which has the worst attribute values.

1) *Input to TOPSIS Method:* TOPSIS assumes that there are m alternatives (options) and n attributes/criteria. The weight of each option with respect to each criterion and weight of each criterion is also provided to TOPSIS method as a input. Let x_{ij} is score of option i with respect to criterion j. This form a matrix $X = (x_{ij})$ known as normalized matrix as shown in Table 2.These weights has been assigned according to Satty's rule[7]. Let J be the set of benefit attributes or criteria (more is better) and J' be the set of negative attributes or criteria (less is better).

Weight	0.1	0.4	0.3	0.2
	C1	C2	C3	C4
A1	7	9	9	8
A2	8	7	8	7
A3	9	6	8	9
A4	6	7	8	6

Table 2 Normalized Matrix

2) *Steps Of TOPSIS Method:* General TOPSIS process with 5 steps is listed below:

Step1: Construct normalized decision matrix.

This step transforms various attribute dimensions into non-dimensional attributes, which allows comparisons across criteria. Normalize scores or data are calculated as in (2):

$$r_{ij} = X_{ij} \Big/ X^2{}_{ij} \qquad (2).$$

for i = 1, …, m; j = 1, …, n

Step 2: Construct the weighted normalized decision matrix. Multiply each column of the normalized decision matrix by its associated weight.

Element of the new matrix are calculated using (3).

$$v_{ij} = w_j \times r_{ij} \qquad (3).$$

Step 3: Determine the positive ideal solution and negative ideal solution.

Positive ideal solution: A* = {(max vij │jϵJ), (min vij │jϵJ')}

Negative ideal solution: A' = {(min vij │jϵJ),(max vij │jϵJ')}

J = 1,2,3,….,n where J is associated with the benefit criteria

J' = 1,2,3,….,n where J' is associated with the cost criteria.

Step4: Calculate the separation measure

The separation (S^*_i) of each alternative from the positive ideal one is given by (4).

$$S^*_i = \sqrt{\sum_{J=1}^{N}(V_{ij} - V^*_j)^2} \qquad (4).$$

Similarly, the separation of each alternative from the negative ideal one is given by (5).

$$S'_i = \sqrt{\sum_{J=1}^{N}(V_{ij} - V'_j)^2} \qquad (5).$$

Step5. Calculate the relative closeness (C^*_i) of ideal solution using (6).

$$C_i^* = S_i' \big/ (S_i^* + S_i') \qquad (6).$$

Rank the preference order according to the relative closeness of ideal solution.

4.   RESULTS AND DISCUSSIONS

To analyze and validate the proposed method a C program[9] and four test case prioritization techniques[9] and four criteria which have influence on test case prioritization techniques have been taken.. Four test case prioritization techniques are:

4.1. Control structure weighted test case prioritization (A1)

In this technique test cases which covering higher number of non-dc paths are executed first. In case two test cases cover same number of non-dc paths then the test case which covers higher number of dc paths are executed first out of those two test cases. Further, if a condition arises that two test cases cover same number of dc paths then the test case which covers the higher number of lines of code is executed first.

4.2. Total Statement Coverage Prioritization (A2)

This technique orders the test cases based on the total number of statements covered by them. Count the number of statements covered by the test cases and orders them in descending order of this number. If multiple test cases cover the same number of statements, then a random order may be used.

4.3. Additional statement coverage prioritization (A3)

This test case prioritization technique iteratively selects a test case T1,that yields the greatest statement coverage, and then selects a test case which covers a statement uncovered by the T1. Repeat this process until all statements covered by at least one test case have been covered.

4.4. Random prioritization (A4)

This technique orders the test case randomly.

Four criteria are

C1:Earlier fault detection.

C2: Feedback.

C3: Reliability.

C4: Cost.

C1,C2,C3 are the benefit criteria and C4 is negative criteria.

Table 3 shows the final result of experiment. It shows that out of four alternative techniques one technique which has highest value of C_i^* is the most cost effective technique that can be used on every version. Table 3 shows that alternative technique A1 has highest value of C_i^*. So from Table 3 it is concluded that technique A1 (Control structure weighted test case prioritization) is the best prioritization technique among others.

	$S_i' /(S_i' + S_i^*)$	C_i^*
A1	0.083/0.112	0.74
A2	0.040/0.097	0.41
A3	0.019/0.109	0.17
A4	0.047/0.105	0.45

Table 3 Relative Closeness Of Ideal Solution

To validate this result above four test case prioritization techniques has been applied on a C program[9] and calculated the APFD (Average Percentage of Fault Detection) for each test case prioritization technique. The test case prioritization technique which has highest value of APFD is best for particular system. The APFD values for all four prioritization techniques according to our experiment are shown in Table 4. According to Table 4 A1 (Control structure weighted test case prioritization) has highest APFD value. Hence it validates the result shown in Table 3.

Techniques	APFD
A1	0.79166
A2	0.745
A3	0.745
A4	0.45833

Table 4 APFD OF Prioritization Techniques

5. CONCLUSION

In paper, a hybrid adaptive regression testing (HART) strategy has been proposed that utilizes two multiple criteria decision making (MCDM) approaches, Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity of Ideal Solution (TOPSIS). Result showed that HART strategy can help researchers and practitioners in selecting most cost-effective techniques across system lifetime.

This study also has several limitations. These limitations can be addressed only through further studies of additional artifacts and regression testing techniques.

First, in this study, the AHP and TOPSIS method has been chosen to implement an HART strategy, but there are many other MCDM approaches available including Weighted Sum Model, Weighted Product Model ,modified AHP methods etc.

Second, in this study, only 4 evaluation criteria and 4 test case prioritization techniques are used, it may be required to investigate HART strategies considering other types of evaluation criteria and other test case prioritization techniques.

REFERENCES

[1] A. Orso, N. Shi, and M. J. Harrold, "Scaling regression testing to large software systems," in Proceedings of the International Symposium on Foundations of Software Engineering, Nov.2004.

[2] X. Ren, F. Shah, F. Tip, B. G. Ryder, and O. Chesley, "Chianti: A tool for change impact analysis of Java programs," in Proceedings of the International Conference on Object- Oriented Programming Systems, Languages, and Applications, Oct. 2004, pp. 432–448.

[3] G. Rothermel and M. J. Harrold, "A safe, efficient regression test selection technique," ACM Transactions on Software Engineering and Methodologies, vol. 6, no. 2, pp. 173–210, Apr. 1997.

[4] S. Elbaum, A. Malishevsky, and G. Rothermel, "Prioritizing test cases for regression testing," in Proceedings of the International Symposium on Software Testing and Analysis, Aug. 2000, pp. 102–112.

[5] A. Srivastava and J. Thiagarajan, " Effectively prioritizing tests in development environment," in Proceedings of the International Symposium on Software Testing and Analysis, Jul. 2002, pp. 97–106.

[6] W. E. Wong, J. R. Horgan, S. London, and H. Agrawal, "A study of effective regression testing in practice," in Proceedings of the International Symposium on Software Reliability Engineering, Nov. 1997, pp. 230–238.

[7] T. L. Saaty, The Analytic Hierarchy Process. McGraw-Hill,1980.

[8] E. Triantaphyllou and K. Baig, "The impact of aggregating benefit and cost criteria in four MCDA methods," IEEE Transactions on Engineering Management, vol. 25, no. 2, pp. 213–226, Feb. 2005.

[9] Dr. Naresh Chauhan ,"Software Testing – Principle and Practice" ,Oxford university press,   2010

[10] G. Rothermel, R. Untch, C. Chu, and M. J. Harrold, "Prioritizing test cases for regression testing," IEEE Transactions on Software Engineering, vol. 27, no. 10, pp. 929–948, Oct. 2001.

[11] S. Yoo and M. Harman, " Regression testing minimisation, selection and prioritisation : A survey," Software Testing, Verification, and Reliability, Mar. 2010.

[12] S. Elbaum, G. Rothermel, S. Kanduri, and A. G. Malishevsky, "Selecting a cost-effective test case prioritization technique," Software Quality Journal, vol. 12, no. 3, 2004.

[13] M. J. Harrold, D. Rosenblum, G. Rothermel, and E. Weyuker, "Empirical studies of a prediction model for regression test selection," IEEE Transactions on Software Engineering, vol. 27, no. 3, pp. 248–263, Mar. 2001.

[14] K. M. A.-S. Al-Harbi, "Application of the AHP in project management," International Journal of Project Management, vol. 19, pp. 19–27, Jan. 2001.

[15] R. N. Wabalickis, "Justification of fms with the analytic hierarchy process," IEEE Transactions on Software Engineering, vol. 7, pp. 175–182, 1988.

[16] K. E. Cambron and G. W. Evans, "Layout design using the analytic hierarchy process," Computers and Industrial Engineering, vol. 20, pp. 221–229, 1991.

[17] T. O. Boucher and E. L. MacStravic, "Multiattribute evaluation within a present value framework and its relation to the analytic hierarchy process," The Engineering Economist, pp.55–71, 1990.

[18] Md. Junaid Arafeen, "Adaptive Regression Testing Strategy: An Empirical Study" 2011 22nd IEEE International Symposium on Software Reliability Engineering.

[19] Pema Wangchen Bhutia, Ruben Phipon2 "Appication of ahp and topsis method for supplier selection" IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, p-ISSN: 2278-8719, www.iosrjen.org Volume 2, Issue 10 (October 2012).