A Comprehensive Study of Various Regressions Testing Techniques: A Survey

Ms. Sneha Thakkar

M.phil Student, Department of Computer Science, Gujarat Vidyapith, Gujarat

Prof. Dr. Dhiren B. Patel,
Professor, Department of Computer Science, Gujarat Vidyapith, Gujarat

Abstract - In software development life cycle software maintenance phase is most crucial part. Software maintenance means error corrections, enhancement of capabilities, optimization, new requirement given by client and many more. So the changed or modified software or application required testing which is known as regression testing. Regression testing for any Software or Application is very essential process to make any software development correct, complete and qualitative. At the same time it time regression testing is consuming and expensive process. In regression testing, a small piece of change in coding, changes and improvements in an application required testing for whole application each time. It required to process to verify the application under positive and negative conditions in manual and automated manner to check Specifications, Functionality, and Performance. Many researches are done in regression testing for test case selection, prioritization by using the history based test cases, in this paper we are trying to present the various types of regression testing techniques and their classifications and also will try to analyze and study on history base regression testing and minimize the testing criteria in selection, and priority which are presented by various researchers.

Keyword - *Regression testing, Software testing, Software maintenance, Test case selection, Software development life cycle, Software testing Life Cycle,*

1. INTRODUCTION

An activity of Software maintenance which is mostly includes enhancements of application, error corrections during development or testing, bug fixing, optimization of product, and removing of obsolete functionality. A small piece of change in coding, changes and improvements in an application required testing for whole application each time. Sometimes, due to these types of modifications in the software or application can cause of any error or bug in the software and possible that software might not work correctly and might also affect the other parts of the software, so to prevent this Regression testing is performed. Regression testing is always focuses on retesting after any changes are made. Because of continuous modifications in any software or application,

Regression testing is used to revalidate again and again Regression testing can be classified in two ways:

a. One is traditional regression testing and;

b. Another one is risk-oriented regression testing.

In traditional regression testing, we can reuse the same test cases and whether risk-oriented regression testing is opposites of traditional regression testing in this we tests the same areas as before, but we have to use different test cases which are increasingly more complex. Techniques of test cases Optimization can reduce the cost of the regression testing by selecting test cases from a subset of an existing whole test-suit to use in retesting for a modified program. Regression testing is crucial and often important software maintenance activity for any software testing life cycle. Regression testing is a process in which test case and test suits are executed to make sure that no new errors or bug have been occurred into previously tested code. Several regression testing techniques have been studied and described in the literature. In this paper, we are trying to represent the results of a comparative empirical study of different regression testing techniques.

Regression testing attempt to mitigate two risks:

- A change that was intended to fix a bug failed,

- Some change had a side effect, unfixing an old bug or introducing a new bug [8]

In section 2 of this paper we have to show the various types of regression testing techniques broadly and as well as we have discussed the classifications of types given by various authors, after that discussion moving into the details of various techniques such as selection, prioritizing test cases for regression testing.

In Section 3, we have discussed some different approaches which might be used to compare various Regression Testing techniques.
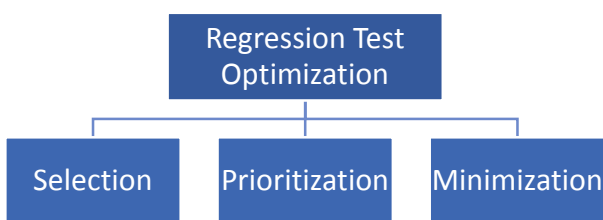
2. REGRESSION TESTING

Regression testing can be defined as "the process of the retesting the modified parts of the software and to ensure that no new errors have been introduced into previously tested code."[8]

Let P be a procedure or program, let P' be a modified version of P and let T be a test suite for P. Regression testing consists of reusing T on P′, and determining where the new test cases are needed to effectively test code or functionality added to or changed in producing P′. A typical regression test proceeds as follows [3]:

- Select T' ⊆ T, a set of test cases to execute on P'.

- Test P' with T'', Establish P'' correctness with respect to T''.

- If necessary, create T'', a set of new functional or structural test cases for P'.

- Test P' with T'', establishing P'''''s correctness with respect to T''

- Create T''', a new test suite and test history for P', from T, T', and T''.

When developing T'', three problems arise:

1. which test cases in T should be used to test P'' (regression test selection);

2. Which new test cases must be developed (excess of test-suite); and

3. Which order should be used to run the test cases (prioritization of test-suite) [5].

There are various regression testing techniques: [10]

1. Re-test All techniques;

2. Regression Test Selection;

3. Test Case Prioritization;

4. Hybrid Approach. Figure 1 shows various regression testing techniques.
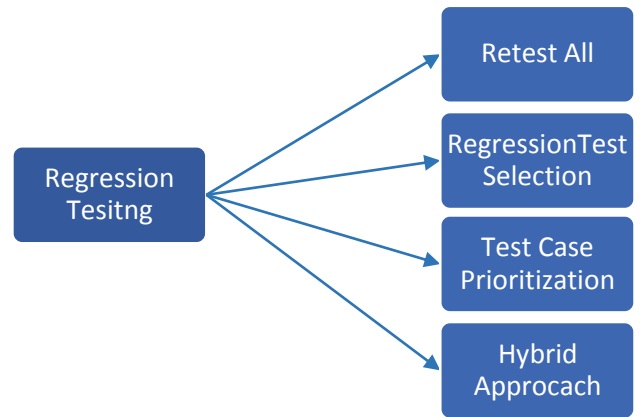
3. REGRESSIONTESTING TECHNIQUES

3.1 Re-test All technique:

Re-test all method is one of the traditional methods for regression testing in which all the tests in the existing test suits are rerun. In this method not a single test case is left to test. So this method is very time consuming and costly to execute in full because it required too much time and it is always high budget process. This method may be used when effectiveness of test is the extreme priority and cost is less priority.

3.2 Regression test selection technique (RTS):

Because of re-testing all technique is time consuming and as well as most expensive technique, regression test selection method is presented. Rather than selecting the whole test suite, we select the part of the test suite which will examine that what is less costly that either to run a part of tests suits or to run a whole test suite. Mostly RTS allow us to skip the test cases which are non usable from the test suits and it will allow us to rerun the test cases which are affected to the application. RTS make an effort to reduce the cost of regression testing by selecting the part of the test suits that was used during the development and also to test the modified program by using that part of test cases.[8] RTS divides the existing test suites into as following:

- Reusable test cases;

- Re-testable test cases;

- Obsolete test cases [8].

Additionally, the classifications of RTS may create new test cases that for modified area which is not covered by existing test cases.

RTS techniques are broadly divided into following categories:

1. Code Coverage/ Dataflow techniques: This technique generally focuses on modified part only and chooses the test coverage criteria and selects only those test cases which affecting on modified part.

2. Random/Ad-Hoc Technique: In this technique, the testers rely on their previous experiences and knowledge and

base on that experience and information testers usually select the test cases and decide which test cases need to be rerun. This can include selecting a percentage of test cases randomly. [8]

3.   Minimization technique: This technique is equivalent to code coverage/dataflow technique but in this technique testers always focus to select test suite with minimum test cases.

4.   Safe technique: This technique is totally different from above mentioned techniques; it is an alternative technique to selects those test cases which are not related to the modified part and generally this technique produce different output with a modified program as compare to original version of software of application.

First three types can be classified under checking positive condition category of regression testing and the last type can be classified under checking negative condition category of regression testing.

Rothermel [11] has identified the various categories in regression test selection technique can be evaluated and compared. These categories are

- Inclusiveness;
- Precision;
- Efficiency;
- Generality;
- Accountability,[12]

a)   Inclusiveness: is the way to measure of extent up towhich a technique chooses test cases which willcause the change to program to produce differentoutput than the original program.[10]

b)   Precision: is the way to measure the ability oftechnique to prevent choosing the test case that willnot make change program to produce different outputthan the original program. [10]

c)   Efficiency: measure the practicality of a technique. [10]

d)   Generality: is the way to measure the ability of atechnique to handle the complex modification,realistic language construct and realistic testingapplication. [10]

e)   Accountability:Accountability refers the extent to which a selective retest strategy promotes the use of structural coverage criteria as it increase the effectiveness of testing. If a program is initially tested with such a criterion, then after modifications it is desirable to confirm that the criterion remains satisfied. [12]

3.3 Test case Prioritization:

After selecting the test cases and test-suit the tester usually focus on priority of the selected test cases that in which order those test cases can be rerun. So, on the bases of some important factors such as efforts of execution of testing, cost-effective part, time of the execution, deploying time etc., tester prioritize the test cases in this technique. This will useful to increase a test - suit's rate of fault detection. The prioritize technique for reorders test cases, scheduling test cases with the highest priority according to some specific criterion earlier in the testing process using Least Recently Used or some random technique.[12]

This technique can be classified as:

- Generic Prioritization
- Version Specific Prioritization

In General Prioritization technique, it try to select an order of test case which will effect on average successor version of application and in Version specific Prioritization technique, it concerned with particular version of the software or application.

At high level test cases prioritization can be work as apply RTS to test-suit existing T, yielding T' then assign a selection probability to each test case in T' then draw a test case from T' by using the probabilities which had assigned previously and run it and repeat this step again and again until testing is executed.

3.3.1. Test Case Prioritization problem

Rothermel et al., as we have studied define the test caseprioritization problem as follows:

Given:T, a test suite; PT, the set of permutations of T; f, a function from PT to the real numbers. [13]

Problem: Find $T' \in PT$ such that $(\forall T'')$ $(T'' \in PT)$

$(T'' \neq T')$ $[f(T') \geq f(T'')]$.

Here, PT represents the set of all possibleprioritizations (orderings) of T and f is a function that, applied to any such ordering, yields an award value for that ordering.[9]
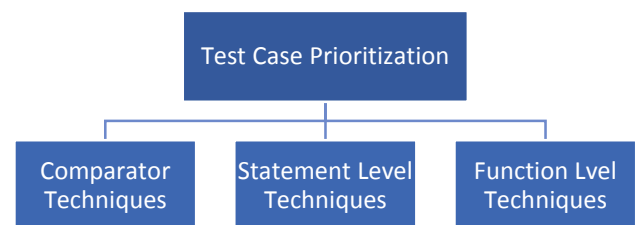
3.3.2. Test Case Prioritization Techniques

There are 18 different test case prioritizations techniques numbered P1-P18 which are divided into three groups as shown in figure.

Comparator Technique

P1: Random Ordering: Test cases are randomlyprioritized;

P2: Optimal Ordering: Test cases are poetized tooptimize the fault detection rate. [14]

Statement level techniques: (Fine Granularity)

P3: Total Statement Coverage prioritization: Testcases are prioritized in terms of total number ofstatement by sorting them in order on coverageachieved;

P4: Additional statement coverage prioritization: It is similar to total coverage prioritization but differ in one sense that it depends on the feedback about coverage attained to focus on areas which are not yetcovered. [14]

P5: Total FEP prioritization: Prioritization isdone onthe probability of the exposure of faults by test case and to approximate Fault Exposure Potential (FEP), mutation analysis is used. [14]

P6: Additional FEP Prioritization: Total FEP isextended to additional FEP when the total statementcoverage prioritization is extended to additional statement coverage prioritization. [14]

Function level Techniques (Coarse Granularity)

P7: Total function coverage prioritization: It is sameas total statement coverage but in place of usingstatements it used functions. It has got coarsegranularity so the process of collecting statement level trace in the total statement coverage. [14]

P8: Additional Function coverage prioritization: Thisis similar to Additional statement coverage but only difference is that it is considering function levelcoverage. [14]

P9: Total FEP prioritization: It is analogous to totalFEP prioritization but differ in one aspect that it isusing functions. [14]

P10: Additional FEP Prioritization: Similar to

Additional FEP with only difference that it usesfunctions instead of statements. [14]

P11: Total Fault index (FI) Prioritization: Faultproneness is measurable software attribute whichused for this technique. [14]

P12: Additional fault index (FI) prioritization: Totalfault index (FI) prioritization is extension to Additional Fault index (FI) prioritization. [14]

P13: Total FI with FEP coverage prioritization: It is the combination of Total FI and FEP coverage prioritization to achieve good rate of fault detection. [14]

P14: Additional FI with FEP coverage prioritization: Total FI with FEP coverage prioritization is extended to Additional FI with FEP coverage prioritization. [14]

P15: Total Diff Prioritization: It is same as Total Fault Index prioritization with difference that Total FI prioritization requires collection of metrics and total difference prioritization requires calculation of syntactic difference between program and modifiedprogram. [14]

P16: Additional Diff Prioritization: Additional Diff prioritization is extension of Total Diff Prioritization. [14]

P17: Total diff with FEP prioritization: Same asTotal FI with FEP coverage prioritization but it isdependent on data derived from different. [14]

P18: Additional Diff with FEP Prioritization: This isTotal diff with FEP prioritization extension. [14]

3.3 Hybrid Approach:

The last regression technique is the Hybrid Approach of both Regression Test Selection and Test Case Prioritization. Hybrid Approach is the combination of the both Regression techniques – Regression Test Cases Selection and Regression Test Cases Prioritization. Hybrid approach is purely based on the selection and prioritization of test cases for internal procedural program. This technique can be classifies as a version specific technique as in this technique tester apply test cases for the old as well as modified part also. This technique requires the test cases from original test suit with each test case's identification, with expected input and output and variable also through will be checked by test suit and module for updated part. Till now so many researches are working on this approach and have been many algorithms has been proposed for this approach by many researcher. Following are some examples of proposed algorithm by various researchers:

- Test Selection Algorithm: proposed by Aggarwal et al. Implementation of algorithm [15]: (a) Input (b) Test Selection algorithm: Adjust module and Reduce module (c) output.
- Hybrid technique proposed by Wong et al which combines minimization, modification and prioritization based selection using test history [16].
- Hybrid technique proposed by Yogesh Singh et al is based on Regression Test Selection and Test Case Prioritization. The proposed algorithm in detail can be studied in [17].

4. A VERY NOTIFIABLE DIFFERENCE BETWEEN REGRESSION TESTING AND RE-TESTING[10]

Retesting denote to retesting the functionality for modified part of any software or application or bug again to validate that the code is fixed properly with predefined criteria. If the bug is not fixed, then defect is reopened and if the bug is fixed then it is closed. Basically, Regression Testing is executing to done to make sure that bug is resolved and previously failed functionality is working or not. Retesting is always classified as a planned testing and Regression Testing is a generic testing for any modifications in application. Regression Testing denotes to test the software or application when it undertakes any code change for adding new functionality or enhancement of software that are made by the developer, to certify that no new bug or error is occurred in the application due to which other working functionality are not affected. Re-testing has always high priority as compared to Regression Testing.

5. APPROACHES AND CHALLENGES

Regression testing is always become a big challenge for any IT professionals as regression testing is very time consuming cost effective process. To implement a single line code change in existing application required a whole application testing. So, to make successful regression test the test-suit always become extremely large. Due to time constraint it becomes next to impossible to execute the entire test-suite every time. Even selection, minimization and prioritization of test case are also big issue. Even frequency of Regression test need to be also categorizes after every modification or bug fixing. Some possible approaches which can be performed to compare and analyze above all discussed techniques which are: [10]

- Controlled Experiments
- Case Studies

Controlled Experiments:this technique is performed on objects that are pinched the code and are further created and modified in controlled environment. The advantage for this technique is that independent variable can be changed to find their effect on dependent variable. Disadvantage of the same can be that it is threat to external validity shaped by manufacturing of errors, test cases and manipulations. [10]

Case Studies: this technique is performed on objects taken "from field", that have error data and existing test suites. The advantage of this technique is that this approach reduces cost and disadvantage is that some factors are not controlled.

Therefore, Case studies and Controlled Experiments have some advantages and disadvantages. Finding object to study such as modified program, test suits, test cases, fault data, selection of the appropriate approach these are big challenges.

Conclusion:

Regression testing is performing under the maintenance phase of the software development life cycle to retest the modified software or application for the enhancement it has undergone. Approximately 50% of the software cost is involved in the maintenance phase of any software development life cycle. So researchers are working aggressively to find out the concrete output and come up with best results by developing new Regression Testing techniques so that the costs and overhead occupied in this phase can be reduced to some level. Lots of Researches are done in these area and the motive of all researcher is seems general to reduce the time and cost to execute the Regression testing.

Many foundations like [9] US National Science Foundation (NSF), Galileo Research Group at Oregon State University, Mapsext Group at University of Nebraska Lincoln, Aristotle Research Group at Georgia Institute of Technology, NCSU Software Engineering Realsearch Group, National Natural Science Foundation of China are few of the many names who are working on development of new regression testing techniques and improving existing techniques on different aspects.[9]

In this paper we have tried to, explain the complete structure of Regression Testing, Illuminate the possible areas of Regression Testing to give proper understanding and importance to researchers and create a scope and motivate new researchers who are planning to start their research "to work on it". Through this paper we have discussed Various Regression Testing techniques and have additionally classified each technique respectively as per explained by various author. Also, we have explaining Regression Test Selection and Test Case Prioritization techniques in detail.

There are several Prioritization techniques algorithm are there such as Genetic Algorithm, Greedy Algorithm, Hi Climbing and many more are applied on various examples to get the effective results. In future we can work on the mentioned prioritization techniques to get the effective and optimal results.

References:

[1] J. E. Bentley, "Software Testing Fundamentals—Concepts, Roles, and Terminology," SAS Proceedings, Wachovia Bank, Charlotte NC, 2004.

[2] K.K.Aggarwal&Yogesh Singh, "Software Engineering Programs Documentation, Operating Procedures," New Age International Publishers, Revised Second Edition – 2005.

[3]J.-M. Kim and A. Porter, "A History-Based Test Prioritization Technique for Regression Testing in Resource Constrained Environments," ACM Proceedings of the 24th International Conference on Software Engineering, pp. 119-129 , 2002.

[4] T. L. Graves, M. J. Harrold, J.-M. Kim, A. Porter and G. Rothermel, "An Empirical Study of Regression Test Selection Techniques," ACM Transactions on Software Engineering and Methodology, Vol. 10, No. 2, pp. 184-208, 2001.

[5]A. Orso, T. Apiwattanapong and M. J. Harrold, "Leveraging Feild Data for Impact Analysis and Regression Testing," ACM SIGSOFT Software Engineering, vol. 28, no. 5, pp. 128-137, 2003.

[6]R. Rothermel, "Efficient Effective Regression Testing Using Safe Test Selection Techniques," Ph.D Thesis, Clemson University, 1996.

[7]L. C. Briand, Y. Labiche, K. Buist and G. Soccar, "Automating Impact Analysis and Regression Test Selection Based on UML Designs," TR SCE-02-04, Version 2, 2003.

[8] Jyoti, Kamna Solanki "A Comparative Study of Five Regression Testing Techniques : A Survey", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 3, ISSUE 8, August 2014

[9]Gaurav Duggal, Mrs. Bharti Suri, "UNDERSTANDING REGRESSION TESTING TECHNIQUES"

[10]Anjali Choudhary, TarunDalal, "A Review on Regression Testing techniques", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 4, Issue 3, May-June 2015

[11] G. Rothermel, R.H Untch, C. Chu and M.J Harrold,"Priortizing test cases for Regression testing." IEEE Trans Software Engg., Vol 27, no. 10, pages 929-948, Oct 2001

[12] Rakesh Garg, Department of Computer Science & Engineering, Amity University, Noida, Uttar Pradesh,"A Comprehensive Review on Regression Testing Techniques." International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 4, Issue 3, Sept 2017, pp. 44-49

[13] S. Kadry, "A New Proposed Technique to Improve Software Regression Testing Cost," International Journal of Security and Its Applications vol.5 no. 3, 2011.

[14] G. Rothermel, R.H. Untch, C. Chu, and M.J. Harrold, "Prioritizing Test Cases for Regression Testing," IEEE Trans. Software Eng., vol. 27, no. 10, pages 929-948, Oct. 2001.

[15] K. K. Aggrawal, Yogesh Singh, A. Kaur, " Code coverage based technique for prioritizing test cases for regression testing," ACM SIGSOFT Software Engineering Notes, vol 29 Issue 5 September 2004.

[16] W. E.Wong, J. R. Horgan, S. London and H.Agrawal, "A study of effective regression testing in practice," In Proceedings of the 8th IEEE International Symposium on Software Reliability Engineering (ISSRE' 97), pages 264-274, November 1997.

[17] Yogesh Singh, Arvinder Kaur, Bharti Suri, "A new technique for version-specific test case selection and prioritization for regression testing," Journal of the CSI ,Vol. 36 No.4, pages 23-32, October-December 2006.