A Review on Automated Test Data Generation (ATDG)

Prof. Sambhaji J. Nawale¹, Prof. Amit A. Kadam², Prof. Jitendra C. Musale³, Prof. Anil T. Lohar⁴

Assistant Professor,

¹ Department of Electronics & Telecommunication Engineering, Sanjivani Pratishthan's Institute Of Technology, Kurund, Ahmednagar, India.

^{2,3,4} Department of Computer Engineering AnantraoPawar College of Engineering & Research, Parvati, Pune, India.

sam.nawale90@gmail.com, kadamamit1811@gmail.com, jitendra.musale99@gmail.com, loharanil04@gmail.com

Article Info	Abstract
Page Number: 780-787	In software engineering, software testing plays a vital role in development
Publication Issue:	of software. In software testing, Test data generation is a show up amongst
Vol. 70 No. 1 (2021)	the most significant and crucial phases. Software testing is not possible
	without adequate test data. Software testing can be performed by using
	different test cases like, unit testing, integration testing, or system level
	testing. The first phase of testing is single module and we called as a Unit
	Testing, But by combining many unit testing module one other test type
	we can use i.e. Integration Testing. Integration testing tests the
	connections of different mechanism, when they are integrated together in
	precise application, for the smooth functionality of software arrangement.
	Many automated and manual test data generation techniques have been
	proposed for software testing. Coupling based testing is an integration
	testing which is based upon coupling relationships that exist among
	different variables within many other call sites are present. Up until now.
	test data generation approaches deal only with unit level testing. There is
	no work for software test information generation for coupling based
	testing. So, in this Paper, we proposed a novel methodology for automated
Article History	test data generation based testing type, classification and different
Article Received: 26 January 2021	algorithms that mostly used
Revised: 07 February 2021	agonanis and mostly used.
Accented: 25 February 2021	Keywords: -Test data generation Software testing, classification,
Publication: 15 March 2021	automation, integration testing, Genetic Algorithm

Introduction

Software testing is a research going to give stakeholders with data concerning the excellence of the result or tune under test. Software test can be perform by using different test cases like, unit testing, integration testing, or system level testing. The first phase of testing is single module and we called as a Unit Testing, Same thing for other test cases but not only the single module but also every modules like integration testing. Integration testing tests the connections of different mechanism, when they are integrated together in precise application, for the smooth functionality of software arrangement [1].Integration stage testing, tests the interaction of different components after integration with other mechanism. System stage testing treats the system as black carton and checks the functionality of the framework as a intact. Integration testing is an vital level of testing which verifies the different components interactions and message passing during interfaces. Unit level testing is a base for integration testing, if the units work properly then different units are included together using different interfaces covered by different components. Integration level testing verifies with the intention of the interfaces are correctly integrated and message transient through interfaces is right. Integration testing is apprehensive with the connections among works. Does a component call other components appropriately? Are the right parameter with right types and ranges arepassed? Does the called method return the proper type and the charge is in the correct range? These questions are focus of the combination testing. unfortunately, very little research has been done in area of coupling based integration testing test data invention using evolutionary approaches [1]. So, now mechanization has been achieved by various means of computers, usually in grouping of others. Test data are an important part of test case, exclusive of test data test case execution is not possible. Research has explored several techniques for test data invention using evolutionary approaches [3-4]. A number of test data generation approaches have been urban and automated. Random test data generation, generates test data based on choosy random inputs form some circulation. Path- oriented and structural approaches use the program's organize flow graph for test data generation; they select a path, and use a technique such as symbolic execution for making of test data. Goaloriented test-data invention approaches select inputs to execute the selected goal, such as statement, condition coverage, decision coverage, irrespective of the path taken. Evolutionary test approaches use evolutionary algorithms i.e. genetic algorithm, for selection and generation of test data by applying evolutionary operators, i.e., crossover and mutation. Most of the work on test data generation has been done at unit level. Unit level test data generation involves the test data that executes the assessment case for unit level testing [3-4].

1.0 Basic Concepts

Concept Related To Software Testing

Software testing is the process of ensuring right software product to achieve full customer satisfaction. The following terms are mostly used for automated test data generation research. Test data: Test Data are data which have been specifically identified for use in testing computer program.

- Test case: Test case is a collection of conditions as well as variables under which a tester will conclude whether an application or software system is working properly or not.
- Test oracle: The method for determining whether a software program or system has passed.
- Test suite: A set of test cases is called test suite. Test plan: is a document which contains all the information about the testing of all stages.
- Test Automation: Developing software for testing a software product.
- Path: sequence of nodes and edges. If we start from entry node and end at exit node then we

call complete path [5].

Automated Test Data Generation

In General here are different reasons to computerize test data generation task in software testing. Some of the most vital reasons are as follows. Reducing number of test cases: Generation efficient test cases are the essential identification for simplifying the test work and improving the test efficiency. The test work is inefficient because of the great number of the initial test cases, so some Automation algorithms are needed to optimize the test cases [3].

Reducing software testing cost :During testing phase the cost can increase more than the expected value due to inappropriate test data. These inappropriate test data cause wastage of organizational resources as well as time. There is a need to minimize the cost for getting an satisfactory product [6].

Reducing human errors: In order to find out how a test case is valid there is no definite mechanism. It basically depends on the testers understanding of the requirement. In this process there are lots of human errors and testers basic knowledge taken into concern. This leads to the addition of bugs in the system after testing. To overcome this problem, automated test data generation phase should be measured [7].

Increasing quality of software products: In general physical testing is becoming a holdup and is a frequent cause of project delays especially for large programs. Therefore, automated test data design has become important to ensure the quality of present day large software products [8].

2.0 Classificatioin Based On Testing

There are different types of software testing approaches are available. With respect to the fact that test data generation is an important phase in software testing, test data generation approaches directly depends on software testing types. There are different approaches available that represents a classification for different automated test data generation approach based on testing type as shown in figure 1.



Fig. 1 Search based techniques

1. Structural Based Testing

In this approach test data are generated using system source code or control flow graph of the program. The main aim of structural testing is to cover a test adequacy criterion.

2. Functional Based Testing

In this approach test data are generated by using system specification. The goal of functional testing is to test the functionality of the software under test.

3. Gray-Box Based Testing

In this move toward both the structural and functional in order to used for generation of test data.It takes the benefits of both structural and functional basedtesting.

4. Non-Functional Based Testing

Non-functional testing is a trying of "how" the framework functi levels of testing. The expression non- functional trying to describe the tests required to measured uniqueness of the systems and software that can be qualified on the verifying scalability of the system.

3.0 CLASSIFICATION BASED ON ALGORITHM

The classification for automated test data generation approaches based on the variety of algorithm they used.

1. Random Based Testing

The test cases and events sequences are generated randomly. This algorithm is mostly used to analyze the efficiency of other algorithms in comparison with it.

Advantages

- It is easy to implement
- It require lesscost
- It work well for simpleprogram
- It is good at exercising systems when the source code and the specification of the
- program under test are not available orincomplete.
- Simply generating a huge number of test cases automatically.

Disadvantages

- It doesn't use available information for test case generation
- May give up a large number of event sequences that are not label and hence not executable, wasting valuabl resources.
- The test designer has no control over choice of eventsequences
- They may not have acceptable testcoverage[11] Application
- It mostly used for analyze the efficiency of other algorithm in comparison with thisalgorithm

2. Search Based Testing

In these approach the trouble of generating the test data are considered as an optimization dilemma, and try to find optimize solution including best test put for problem under test.

Advantage of search based testing result shows efficiency of this approaches and disadvantage is that it requires large spaces. Application for search-based method is used in web/desktop applications. In search based automated test data generation a diversity of search algorithm to use for the purpose of test data generation. The Figure 2 shows these search based methods[11].



Fig. 2 Search based techniques

3. Hill Climbing

Hill climbing is a family of local search which is mathematical optimization technique. It is an iterative calculation that begins with an arbitrary answer for an topic, and then endeavors to determine a better arrangement by incrementally varying a single factor of the result. In the event that the change delivers superior arrangement, an incremental change is made to the new comeabout, rehashing until no further enhancements can be make. For instance, hill climbing can be applied to the TSP. It is not difficult to discover a beginning arrangement that visits all the cities however will be exceptionally poor contrasted with the ideal result. The calculation begins with such a come about and makes little upgrades to it, for example, exchanging the request in which two cities are visited. At the end, greatly shorter route is to beobtained

4. Memetic Algorithm

The memetic algorithms (MAs) are meta- heuristics that use both global search and local search for instance GA with a hill climbing. It is motivated by the cultural growth. A meme is a unit of reproduction in educational transmission. The plan is to mimic the process of the growth of these memes. With the point of view of an optimization, a MA can be described as a population based meta-heuristic in which, every time an issue or offspring is generated, unless it reaches local optimum until a local search is applied toit.

5. Scatter Search

Scatter search is a search based evolutionary means that mechanism on a set of solutions, called the reference set, which stores the best solutions that have been generated secluded. The solutions in this set are pooled in sort to achieve new ones, so it trying to generate each time better solutions

6. Tabu Search

Tabu search is a global search technique that solves combinatorial optimization problems by

using meta-heuristic algorithm, such as the travelling salesman problem (TSP). Tabu search is system that uses a local or district search method to iteratively move from a explanation s to a solution s' in the region of s, until some stopping condition has been pleased. To search area of the search space that would be left strange by the local search procedure, tabu search modifies the neighborhood structure of each result as the search progresses [6].

7. Genetic Algorithm

In the long-ago, evolutionary algorithms have been applied in many real life problems. Genetic algorithm is one such evolutionary algorithm. Genetic algorithm has emerged as a practical, robust optimization technique and search method. A Genetic algorithm is a search algorithm thatis motivated by the way scenery evolves type using natural selection of the fittest persons. The possible solutions to problem can be solved which is represented by a population of chromosomes. A DNA is a string of binary digits and each digit that makes up a chromosome is called a gene. This primary population can be entirely random or can be created by hand using processes such as greedy algorithm. The pseudo code of a basic algorithm for Generalized Genetic algorithm is as follows [13]–

Initialize (population) Evaluate (population) While (stopping state not satisfied) { Selection (population) Crossover (population) Mutate (population) Evaluate (population)

8. Simulated Annealing

It is one type of global search technique which is simulated annealing. It samples the whole part and improve the result by recombination in some form. In simulated annealing a value y1, is selected for the product y, and the effect which has the minimal cost or objective function, E, is chosen. Cost functions define the comparative and desirability of particular solutions. Minimizing the objective role is usually referred to as a cost function

4.0 Conclusion

This paper provides an overview of the automated test data generation by using different testing type classification and based on algorithms which are an important area of research

for reducing cost of software development. Test data generation is used to satisfy functional, non functional and business requirements. Some non functional requirement testing can be done only by automation; where manually it is not possible. The paper emphasizes the basic concepts of automated test data generation. It also focuses much on the test data generation classification techniques and methodology. We feel that these concepts are mandatory to perform research in the area of automated test data generation whether it is conventional programming or modern programming.

5.0 Acknowledgement

We would like to express our sincere gratitude Mr. Atul Marathe (Adjunct Professor), Dr. Ashok Saraf (Innovation Club Member) for their invaluable guidance and support throughout the research process. We also wish to thank Dr. Sunil B. Thakare (Principal, APCOER, Pune) for their support. Finally, we are grateful to all of the research participants who generously gave their time and effort to this project.

6.0 References

- [1] Shaukat Ali Khan and AamerNadeem, Automated Test Data Generation for Coupling Based Integration Testing of Object Oriented Programs using Evolutionary Approaches, *IEEE 10th International Conference on Information Technology*,**2013**.
- [2] Ankur Pachauri and Gursaran, Test data generation for Unit testing of Java programs Using JML and Genetic Algorithm, *International Journal of Software Engineering Research and Practice*,2011.
- [3] Dhabliya, D. (2021a). AODV Routing Protocol Implementation: Implications for Cybersecurity. In Intelligent and Reliable Engineering Systems (pp. 144–148). CRC Press.
- [4] Lee Copeland, A Practitioner's Guide to Software Test Design, STQE Publishing, 2004
- [5] Boris Beizer, *Software TestingTechniques*, International Thomson Computer Press, **1990**
- [6] Dhabliya, D. (2021c). Designing a Routing Protocol towards Enhancing System Network Lifetime. In Intelligent and Reliable Engineering Systems (pp. 160–163). CRC Press.
- [7] Hitesh Tahbildar and BichitraKalita, Automated Software Test Data Generation: Direction of Research, *International Journal of Computer Scienc & engineering Survey*, 2011, Vol 2.
- [8] Y A Sharma, A Jadhav, P R Srivastava and R Goyal, Test Cost Optimization using Tabu Search, *J Soft EngAppl*, **2010**, vol 3, pp 105-156.
- [9] Dhabliya, D. (2021d). Examine Several Time Stamping Systems and Analyse their Advantages and Disadvantages. International Journal of Engineering Research, 1(2), 01–05.
- [10] V Rajappa, A Biradar and S Panda, Efficient Software Test Case Generation using Genetic Algorithm Based Graph Theory, *Proceedings of the First International Conference on Emerging Trends in Engineering and Technology*, 2008, pp298-303.
- [11] S K Swain, D P Mohapatra and R Mall, Test Case Generation Based on State and

Activity Models, *Journal of Object Technology*, **2010**, vol 9, pp1-27.

- [12] W T Tsai, D Volovik, T F Keefe and M EFayad, Automatic Test Case Generation from Relational Algebra Queries, *IEEE Transactions on Software Engineering*, **1990**, vol 16.
- [13] A Alhroob, K Dahal, A Hossain, Automatic Test Cases Generation from Software Specifications, *e- Informatica Software Engineering Journal*, **2010**, vol4.
- [14] P McMinn, Search-Based Software TestDataGeneration: A Survey, Software Testing, Verification & Reliability, 2004, vol 14, pp 105–156.
- [15] Raquel Blanco, Javier Tuya, BelarminoAdenso- Díaz, Automated Test Data Generation using a Scatter Search Approach, Information and Software Technology, 2009, pp708-720.
- [16] Dhabliya, D., & Others. (2021). An Integrated Optimization Model for Plant Diseases Prediction with Machine Learning Model. Machine Learning Applications in Engineering Education and Management, 1(2), 21–26.
- [17] Chayanika Sharma, SangeetaSabharwal, RituSibal, A Survey on Software Testing Techniques using GeneticAlgorithm, *International Journal of Computer Science Issues*, 2013, Vol 10,Issue1.
- [18] [14] Amit A. Kadam, Anant M. Bagade, A Survey on Generation of Automated Test Data for Coupling Based Integration Testing, European Journal of Advances in Engineering and Technology, 2014, 1(2):69-74
- [19] Amit A. Kadam, Anant M. Bagade, —AutomatedTest DataGenerationforCouplingBased IntegrationTesting, Spvryan's International Journal of Engineering Sciences & Technology (SEST), 2015.