Review of Genetic Algorithm and Application in Software Testing

Patrick Kwaku Kudjo
Jiangsu University
China Zhenjiang
P.R.China, 212013

Elias Nii Noi Ocquaye
Jiangsu University
China Zhenjiang
P.R.China, 212013

Wolali Ametepe
Jiangsu University
China Zhenjiang
P.R.China, 212013

ABSTRACT
The recent technological advancement and complexity of software makes it very difficult to maintain and improve the quality of software. Software testing is a technique used in validating and verifying that a develop software meets its requirements and specification hence the need to properly generate suitable test case to test programs. Software testing consumes a lot of time, effort and money, therefore the focus of testing largely depends on test case generation, execution and evaluation. Automated testing is a technique used to maximize test coverage, detect more errors, increase test execution, decrease cost as well as improving the quality of software. This paper reviews the use of Genetic algorithm to automatically generate test case using the random method.

Keywords
Automated Testing; Genetic Algorithm; Evolutionary Algorithms; Random Testing; Test Case

1. INTRODUCTION
Software testing remains an integral part of the software development life cycle (SDLC); it is an important phase in the SDLC as it helps remove errors in different forms ranging from minor to major errors that can seriously affect the quality of product when delivered to the intended users. Previous research shows that, software testing phase takes a lot of effort, time and cost hence the need to find appropriate test case generation technique to effectively improve the testing process. Different test case generation techniques have been proposed in several research work with an objective of enhancing the testing process, these include random test data generators [1-3], path oriented test data generators[4-6], goal oriented test data technique[7], intelligent test data generator technique[8], other automatic test generation technique for java programs [9-15]. In this paper, we review the use of Genetic Algorithm to automatically generate test case using the simple random testing technique. Random testing is a simple and effective testing technique that allows a test case to be selected randomly from the input domain [16-18]. The uniform technique of random testing helps ensures that all inputs have the same probability of being selected from the input domain. The application of Genetic Algorithm (GA) in software testing is a young and active area of scientific research since much study has not been conducted in the field.

The paper review the application of Evolutionary algorithm (EA) specifically genetic algorithm based on test case generation using the simple random testing technique. An evolutionary algorithm (EA) is a stochastic search for an optimal solution to a given problem [19]. It uses a population of individuals; where each individual is classified as a chromosome. Chromosome represents the characteristics of individuals in the population; these characteristics are thus referred as gene[20, 21].

1.1 Genetic Algorithm
Genetic algorithm is one of the evolutionary algorithms that implements optimization strategies by simulating evolution of species through the process of natural selection, survival of the fittest and reproduction [22-24]. Genetic Algorithm is a subset of a branch of Evolutionary algorithms that include Evolutionary programing[25], evolutionary strategies[26, 27], differential evolution[28], cultural evolution and coevolution [29, 30]. This paper reviews the application of genetic algorithm in software testing specifically the generation of test case using the simple random testing technique. The structure of the paper is as follows, Section 2 will briefly give a background to genetic algorithm, application of genetic algorithm for the test case generation using the random technique is given in Section 3, Section 4 gives a detail analysis of Genetic Algorithm with other test case generation techniques, the related work is presented in Section 5 and Section 6 concludes and gives recommendations for future works.

2. GENETIC ALGORITHM
Genetic algorithm was proposed by John Holland and his fellow colleagues in Michigan University [31]. However the initial concept was first investigated by J.D. Bagley’s in 1967 “The behavior of Adaptive systems which employ Genetic and Correlative Algorithms”[32]. Other independent study of Evolutionary algorithms include [33-38]. Genetic algorithm is a search based optimization technique [39-41] that evolve through a search space of candidate population to identify the best pool of population. To help understand the concept and its application in software testing, we will briefly explain some of the related concepts used in Genetic Algorithm.

A. Related Terms
i. Chromosome: are the strings or numerical values which represent the candidate solution to the search problem[42-44]

ii. Gene: the elements of the solution are known as the gene.

iii. Population: refers to the total number of candidate solutions to be used in the optimization problem.

iv. Allele: the possible values of a gene are known as allele.

v. Genotype refers to the set of genes contained in a genome.

vi. Phenotype: Are the behavioral traits or characteristics of an individual[45, 46].
B. Pseudocode For Genetic Algorithm
The general pseudocode is shown below

```
Start
 Initialize population;  
 Randomly generate t number of individuals;
 Evaluate fitness of individuals;
 While (fitness value! = termination criteria) do
   Selection;
   Crossover;
   Mutation;
   Calculate fitness function;
 End
```

Fig 1: Flow chart of GA

2.1 Genetic Operators
Genetic algorithm uses three main operators which are selection, crossover and mutation.

2.1.1 Selection
The selection process determines which solutions are to be preserved and allowed to reproduce and which ones deserves to die out. The process normally involves identifying the good solutions in a population, making multiple copies of the good solutions, eliminating bad solutions from the population.[45, 47, 48]. The final stage is applying the fitness function to evaluate the solutions. A fitness function represents the quality of the solution[49, 50]. There are different types of selection methods used by GA, this include rank selection[51], proportionate selection[51, 52], roulette wheel selection [49], tournament selection [53], and steady selection.

2.1.1.1 Crossover Operator
The crossover operator is used to create new solutions from the existing solutions by recombination of the information from both parents.[54]. It thus involves the exchange of genes randomly between the chromosomes of two parents to create new offspring [55, 56]. Crossover can either be one point or two points.

2.1.1.2 Mutation
The final genetic operator is mutation which introduces new features to a given population or problem to obtain and maintain diversity in the population. It thus involves randomly changing individuals to allow diversity among the population. The various types of mutation are Flip mutation, Uniform mutation inversion mutation, Scramble mutation, Swap mutation, interchanging mutation and Creep mutation [57-60]

3. TEST CASE GENERATION USING GENETIC ALGORITHM
Manual testing and automated testing are the two main testing techniques that have been employed over the past decades[61, 62]; however automated testing appears to be the most efficient and effective method based on previous research[63], due partly to its efficiency and effectiveness in maximizing test coverage, detecting more errors, increasing test execution, decreasing cost and improving the quality of the product. The application of GA in test case generation usually starts with an initial population. Previous studies on the population size[64] indicate that, the size of the population used in any optimization problem could affect an optimization process, C. R. Reeves investigated the use GA with small population [65] and other studies that involve the use of large population are[66, 67] based on this, P. A. Diaz-Gomez and D. F. Hougen[68] proposed three types of metrics to help evaluate diversity of fixed length population of chromosomes. Based on the randomly generated test case from the population or input domain, a fitness function is used to evaluate the test cases to determine if it can be used for future generation of more test cases; the fitness function must be defined in relation to the testing goals. The test case is executed and feedback information is collected, all test cases with high fitness value are selected and added to the best test pool. Test cases with low fitness value are discarded and not used for future generation or can be recombine and mutated to produce new test cases, the newly created offspring are evaluated against the fitness value, this process continues until a stopping condition is achieved.

4. ANALYSIS OF GENETIC ALGORITHM WITH OTHER TEST CASE GENERATION TECHNIQUES
Based on previous studies conducted, genetic algorithm proves to be an effective technique for the generation of test case for detecting errors in complex systems[69] as compared to other test case generation techniques investigated by [10, 70] N. Kosinrdecha and J. Daengdej presented a number of test case generation process and techniques such as random testing technique, goal oriented techniques, specification-based techniques and sketch diagram based techniques, the findings shows a number of weakness related to these test data generation techniques. Some of the problems identified include the following:

i. Existing methods lacks the ability to identify and reserve the critical domain requirements in the test case generation process
ii. Inefficient automated test case generation techniques

iii. Failure to use most of the generated test cases

iv. Inefficient test case generation techniques with limited resources.

Although genetic algorithm has been considered an efficient and efficient method, D. J. Berndt and A. Watkins proposed two new strategies to help improve the performance of genetic algorithm:

i. Neural Network Based oracles: this method involves replacing the real target system execution to avoid expensive execution costs for evaluating test cases.

ii. The second method involves improving and enhancing the major genetic operators such as selection, crossover and mutation.

5. RELATED WORK

Genetic algorithm is applied in several fields of studies such as machine learning,[71, 72] business [40], Ecological[73] as well as medical data mining[74, 75]

Jarmo T. et al. [76] study program testing automation by using optimization via genetic algorithm. The findings of their investigation show that, genetic optimization performs better than simple random testing.

Rakesh Kumar et al. [77] showed how genetic algorithm can used to Automate test case generation. The findings shows that the use of genetic algorithm for full automation of test case generation enhance the overall quality of the software as compared with random testing.

Amit Kumar Sharma [26] also investigated how Genetic algorithm can be used to improve the quality and reliability of the software by generating optimized test cases, the results show that GA is a good technique in searching the input domain for the required test cases and as well improving the test case optimization process.

R.P. Pargas et.al [15] presented a GenerateData, an algorithm for automatic generation of test data using genetic algorithm directed by the control dependence graph of the program, the approach adopted in the paper shows a potentially useful way of generating test data using the Genetic algorithm.

Kuvinder Singh et.al [78] generated test case using the genetic algorithm with anti-random population, experimental results of the works shows a significant average of 20% and maximum 40% improvement over random test case generation technique.

F. S. Babamir et al. [79] investigated the application of genetic algorithm based tester using different parameters to automate the structural oriented test data generation on the principiple of internal program structure.

S. P. Tripathy and D. Kanhar proposed a heuristic algorithm with sampling techniques to optimize test suite. The genetic algorithm was used as an element to optimize the testsuite [80].

J. McCart et al. [81] used genetic algorithm to improve the testing technique process by reducing the execution time as well as improving the ability of the technique in detecting more error

N. K. Gupta and M. K. Rohil conducted a research by using the genetic algorithm to generate test cases for testing object oriented programs(OOP). The method uses a tree representation of statements in test cases. The approach adopted facilitates the automatic generation of object oriented programs using genetic algorithm[82].

S. Wappler and F. Lammermann [83] in their study proposed an evolutionary algorithm for the automatic generation of test cases for white- box testing of object-oriented programs. The experimental result of their approach outperforms random testing.

P. Tonella proposed a genetic algorithm to automatically generate test case to test classes in Java programs. The approach employed effectively achieve good coverage. [84]

A. Rauf et al. [85] proposed a graphical User Interface (GUI) testing and coverage analysis technique using genetic algorithm to help generate the best possible test parameter based on a predefined testing goals. They used some of the modern tools and techniques of automating GUI testing such as control- flow graph, even-flow graph and event-flow model etc. to help reduce the challenges associated with testing of such systems.

B. F. Jones [86] also investigated structural testing using genetic algorithm.

6. CONCLUSION

The work review various application of Genetic Algorithm in software testing specifically automated test generation. The ultimate aim of testing is to produce quality programs that meet the needs of customer’s hence the need to select appropriate test generation techniques to help improve the testing process. Based on the study conduct we can conclude that, the application of Genetic Algorithm as an optimization technique to automatically generate test case helps detect more errors in program as it uses the best possible test case. The findings further shows that, Genetic algorithm performs better than other automated test generation techniques such as random testing and also increases the number of test cases thereby improving efficiency. Future study will be focus on investigating the application of GA and Particle Swarm Optimization (PSO) in test case automation. Another area of study that can be investigated is the use of Genetic algorithm in regression testing as well as test case prioritization. Finally a more robust way of improving upon the fitness function can be investigated.

7. ACKNOWLEDGMENTS

Our sincere thanks go to all the Professors and PhD students in the Computer Science and Communication Engineering Department of Jiangsu University China.

8. REFERENCES


