Automatic Test Data Generation using Metaheuristic Cuckoo Search Algorithm

Madhumita Panda, Seemanta Engineering College, Baripada, India
Partha Pratim Sarangi, Seemanta Engineering College, Baripada, India
Sujata Dash, North Orissa University, Baripada, India

ABSTRACT

The proposed work emphasizes on the automated process of test data generation for unit testing of structured programs, targeting complete path coverage of the software under test. In recent years, Cuckoo Search (CS) has been successfully applied in many engineering applications because of its high convergence rate to the global solution. The authors motivated with the performance of Cuckoo search, utilized it to generate test suits for the standard benchmark problems, covering entire search space of the input data in less iterations. The experimental results reveal that the proposed approach covers entire search space generating test data for all feasible paths of the problem in few number of generations. It is observed that proposed approach gives promising results and outperforms other reported algorithms and it can be an alternative approach in the field of test data generation.

Keywords: Cuckoo Search (CS), Metaheuristic Algorithm, Path Testing, Search Based Software Engineering (SBSE), Search Based Software Testing (SBST)

INTRODUCTION

The automated test data generation is an undecidable problem and the main challenge in this field is the size and complexity of the software. Nature inspired algorithms are good at seeking good solutions for combinatorial optimization problems within confined time at a reasonable computational cost. A good number of Nature based algorithms coming under metaheuristic have been used by researchers to generate test data for white box, black box as well as grey box testing (Harman, 2015). Most of the research work has been done for structural testing, functional testing, mutation testing etc., but very few papers are available related to grey box and quality testing (Harman, 2009). The first workshop on software engineering using metaheuristic algorithms was conducted in 2001 (Harman, 2001). At that time only limited work has been done by

DOI: 10.4018/IJKDB.2015070102
the researchers in the area of software testing using search algorithms like Simulated Annealing (SA), Tabu search and Genetic algorithms (Harman, 2001; McMinn, 2004; Sthamer, 2002).

Search based software engineering (SBSE) is a reformulation of the software engineering problems (Harman, 2004; de Oliveira, 2015) coming under the category of un-decidable, NP hard or NP complete problems. Here the specific problem is designed as a search problem and then the target is to select best solutions out of a large domain of solution space using guided heuristics mimicked from nature (Harman, 2015).

Search based techniques are perfect for the software engineering problems as this is the area where the engineers have no idea regarding what are the good solutions but very clear idea about what qualities the good solutions must have, and they have to choose the best one out of a number of competing and conflicting solutions (Harman, 2001; McMinn, 2004). To redesign any problem of software engineering as a search problem, three components are necessary to be specified clearly; these include first an intermediate representation of the problem, second a fitness function on the basis of the targeted objective and finally a set of operators to deploy (Harman, 2001).

Out of these three the most crucial role is played by the fitness function. The fitness function captures the vital information and guides the search process by selecting the best solutions from a number of suitable solutions on the basis of the testing strategy adopted (Sthamer, 2002; Baresel, 2002). For example, in white box testing if path based testing approach is followed then the suitable test cases are those test cases that are able to show maximum path coverage.

Many metaheuristic algorithms including, Genetic algorithm (GA), differential evolution (DE) algorithm, Artificial bee colony algorithm (ABC), Particle swarm optimization (PSO), Bat algorithm, Cuckoo search algorithm (CS), are used for different problems of software engineering but still research is going on using different nature inspired algorithms. Researchers are trying to develop a deep empathetic study of the search algorithms features and their suitability for different engineering approaches.

Out of the number of papers published more than sixty percent (Orso & Rothermel, 2014) of the papers are related to the area of software testing, therefore search based software testing (SBST) is considered as a sub domain of search based software engineering (SBSE).

This search based software testing is a reformulation of software testing as a search problem, where the correct test data are selected from a large solution space. The term SBSE was coined by Harman and Jones in 2001, however work on the application of search techniques to software engineering problems set in motion as early as 1992 (Harman, 2001).

Here in this paper a brief analysis of some of the most popular bio-inspired metaheuristic have been presented and their capabilities for test data generation for bench mark triangle classification problem has been analyzed.

The rest sections of the paper are organized as follows, section 2 includes background study including the concepts of the widely used metaheuristic algorithm Cuckoo Search (CS), section 3 includes basic concepts of Structural testing, section 4 includes the special characteristics of the triangle classification problem that makes it a bench mark problem in search based testing domain, and section 5 includes the experimental settings and results. Finally, Section 6 includes conclusion with a discussion of the future perspectives of using bio-inspired algorithms in the field of search based test data generation.

Related Work

The principle behind using Bio-inspired algorithms in software testing is the conversion of the test goal into an optimization problem (Harman, 2001) and then solving this problem with the
Related Content

Digital Model for Diagnosis of Postoperative Complications in Medicine Using Bioinformatics: Evaluate and Predict the Patient's Condition
[www.igi-global.com/article/digital-model-for-diagnosis-of-postoperative-complications-in-medicine-using-bioinformatics/237197?camid=4v1a](www.igi-global.com/article/digital-model-for-diagnosis-of-postoperative-complications-in-medicine-using-bioinformatics/237197?camid=4v1a)

ROC Curve, Area under it, other Classification Performance Characteristics and Statistical Tests
[www.igi-global.com/chapter/roc-curve-area-under-other/53913?camid=4v1a](www.igi-global.com/chapter/roc-curve-area-under-other/53913?camid=4v1a)
Genome Sequence Analysis in Distributed Computing using Spark
[www.igi-global.com/article/genome-sequence-analysis-in-distributed-computing-using-spark/167708?camid=4v1a](www.igi-global.com/article/genome-sequence-analysis-in-distributed-computing-using-spark/167708?camid=4v1a)

New Technologies in Personalized Decision Support to Enhance Patient Choice: Applications and Challenges
[www.igi-global.com/chapter/new-technologies-personalized-decision-support/64074?camid=4v1a](www.igi-global.com/chapter/new-technologies-personalized-decision-support/64074?camid=4v1a)