Name of the Scholar : **Sapna Varshney**

Name of the Supervisor : **Dr. Monica Mehrotra**

Department : **Department of Computer Science,**
Faculty of Natural Sciences

Month/Year : **November, 2016**

Title : Use of Metaheuristic Optimization Techniques for Automatic Generation of Effective Test Cases

Keywords : Automated test Data Generation, Data-Flow Coverage, Search Based Software Testing, Genetic Algorithm, Particle Swarm Optimization, Differential Evolution

---

Abstract

Software testing is the process of assessing the quality of the developed software by detecting as many faults as possible. Test data generation according to a test adequacy criterion is an important aspect of software testing. Manual software testing (test data generation, in particular) is expensive and a labor-intensive process. Therefore, automated fault revealing test strategies need to be defined that are cost effective to improve the quality of the modern complex software systems. Structural testing is the most widely practiced form of testing. Several approaches that have been proposed to generate test data for dynamic structural testing include random, path-oriented, goal-oriented and intelligent approaches.

In the past few decades, the researchers have widely employed evolutionary search-based algorithms such as Genetic Algorithm (GA) to optimize the generation of test data with minimum human effort - research area that came to be known as Search Based Software Testing (SBST). Test data is generated according to a test adequacy criterion encoded as a fitness function to guide the search process. The measures that have been used to assess the effectiveness of a search-based technique for structural testing are control flow coverage (statement coverage, branch coverage, path coverage) and data-flow coverage. Control-flow based coverage criteria, branch coverage in particular, are the most often used measures. Data flow coverage have received relatively little attention. More recently, the focus is now on the use of other highly adaptive metaheuristic search-based techniques such as Particle Swarm Optimization (PSO) for test data generation problem. Also, there has been little work on the application of hybrid techniques to structural test data generation problem.

In this thesis, application of an elitist GA, an adaptive PSO, Differential Evolution (DE) algorithm, and finally a hybrid (adaptive PSO and DE) algorithm has been studied for optimal test data generation in accordance to the all-uses data-flow coverage test adequacy criterion. For the hybrid algorithm, a local neighbourhood search strategy is also proposed to handle the

boundary constraints effectively accelerating the search for optimal test data. The performance of the proposed approaches have been experimentally evaluated on standard benchmark programs with respect to the measures 'mean percentage coverage' and 'mean number of generations'. The experiments have been repeated a number of times for different population sizes to ensure randomness.

A novel fitness function is proposed for data-flow dependencies of a program based on the concepts of dominance relations and branch distance. The novel fitness function is validated by proposing an elitist GA-based approach for data-flow coverage of a program. Gray encoding is used to encode test data. Experimental results indicate that the proposed elitist GA outperformed the other GA-based studies for data-flow coverage and random search with respect to the measures collected for all the subject programs. The performance gains for the proposed approach can be attributed to elitism, gray encoding and the novel fitness function.

A PSO-based approach with adaptive inertia weight strategy guided by the novel fitness function is proposed for data-flow coverage of a program; adaptive inertia weight is used to balance the exploitation and exploration capabilities of the PSO technique. The adaptive PSO-based approach outperformed the elitist GA and random search with respect to the measures collected for all the subject programs.

In this research work, DE has been applied for the first time for structural test data generation according to data-flow dependencies of a program. Experimentally, performance of the proposed DE-based approach is shown to be comparable to that of the adaptive PSO and better than the elitist GA and random search with respect to the measures collected for all the subject programs.

Finally, a hybrid algorithm is proposed that incorporated the best of the evolution schemes of both PSO and DE - adaptive inertia weight strategy and knowledge about the previous solutions in PSO along with differential operator and greedy selection in DE. A local neighbourhood strategy is applied to handle the boundary constraints effectively. Experimental results and the statistical analysis reveal that the proposed hybrid (adaptive PSO and DE) algorithm is the best performing approach; it outperformed DE, PSO, GA and random search with respect to the measure 'mean number of generations' for all the population sizes that are considered. Performance of the proposed hybrid algorithm is comparable to that of DE, PSO and GA with respect to the coverage achieved for smaller population sizes; however, only the proposed hybrid algorithm achieved full data-flow coverage as the population size is increased to 20 and 25 for complex subject programs. Performance of random search is worst. Here, a promising hybrid global optimization algorithm is presented for data-flow test data generation.