Particle Swarm Optimization with Cross-Over Operator for Prioritization in Regression Testing

Dr. Arvinder Kaur
University School of Information Technology,
Guru Gobind Singh Indraprastha University

Divya Bhatt
University School of Information Technology,
Guru Gobind Singh Indraprastha University

ABSTRACT

Software Testing is continuous process of development and maintenance in life of software. In maintenance phase, regression testing gets exercised with additional resources/time for performance. The prioritization of test cases helps to reduce the cost-time of regression testing. Hence, completing Regression Testing effectively and on schedule is challenge for software tester. In this research paper, the Particle Swarm Optimization (PSO) technology has been studied and used with the blend of Genetic Algorithm (GA) and the hybrid prioritized algorithm has been proposed. The Particle Swarm Optimization is an optimization algorithm based on heuristic search which can be used to solve time-constraint environment of Test Case Prioritization and the concept of Genetic Algorithm will further help in diversifying the solution within whole search space. For finding the effectiveness of hybrid prioritization algorithm: the efficiency %, saving %, reduction % and APFD/APCC has been calculated.

Keywords: Regression Testing, Particle Swarm Optimization, Genetic Algorithms

1. INTRODUCTION

During maintenance phase, modifications/defects in software are corrected. Hence, it is very difficult to decide which test case should be executed, or which test case to mark as “effective” in order to detect the cause of modification in software. Regression Testing is done within maintenance phase. Regression Testing assures that modification done in software has not over-rulled user specification and expectations. Hence, to make Regression Testing more scheduled many criteria’s can be followed: reduction of test case, test case selection, prioritization of test cases. In this research paper, the concentration is in the area of “prioritization of test cases”. The time constrained prioritization of test cases can be reduced to 0/1 knapsack problem [1] which is a NP-hard [2] and solved using Particle Swarm Optimization [3] and Genetic Algorithm [4] techniques. Thus, combination of two optimization techniques has been used here, to make effective solution on-schedule.

The techniques described are: Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) operator. The PSO is a global optimization algorithm based on heuristic search [5]. The idea was given by John Kennedy and Eberhert, in 1995 [6], after observing the group of animals, flock of birds and fishes, where each individual follows the path of “global best” particle with in its population, e.g. in ocean, while searching for good food source (plankton), a school of fish, travels everywhere making cylindrical shape. The inside story: every fish is observing its neighbor’s position and velocity then compare it with global best position and velocity. The “best” position and velocity is chosen and updates are made by individuals in their position and velocity. Thence, each fish converges towards the best position after modification in its velocity, which helps to move towards the food faster.

In this paper, we have blended the PSO with crossover operator of GA [5], to avoid group of population from converging to local best. The GA process consists of [7]: selection, crossover and mutation. In GA, from general population, a set of best fitted population is selected on the basis of fitness function and crossover, mutation operators are applied to get optimal solution. The crossover operator helps in PSO to make each individual a widened look of search space, by crossing over them with other individuals. This resulted in hybrid prioritization algorithm of PSO & GA.

The PSO has been applied to NP-hard combinatorial problems Traveling Salesman Problem [8], and NP Knapsack [3] whereas GA is known for optimizing scheduling problems [9]. Here, we have tried to use combination of PSO and GA to solve, time-constrained prioritization and code based prioritization. To bring clarity between prioritized and non-prioritized test-suite Average percentage fault detection (APFD) [10] and Average percentage Condition Coverage (APCC) [11] has been calculated and represented using graphs. The APFD has been used to evaluate average percentage of faults detected as per average percentage of test-suites executed whereas APCC has been evaluated average number of condition covered as per average percentage of test suite executed.

2. RELATED WORK

The process of Regression testing is quite expensive and time consuming even for small systems. To address problem of regression testing many test case prioritization techniques have been explored by researchers such as: hybrid approach by Wong [12], Test Selection Algorithm by Aggarwal [13], version specific technique [14] where test cases are prioritized for specified version of software, hybrid technique based on variable method for regression testing [15], optimization algorithm techniques [16] by observing natural behavior of ants and swarms [17] for test case prioritization, code-coverage based technique [18] where internal structure is analyzed and therevalidation strategy [19] based on extension of Fischer
algorithm decides which Test case to re-run for Regression Testing, CBSE (component based software engineering) for regression testing where features associated with test cases are exercised [20], SBSE (search based software engineering) for regression test optimization [21, 22] MORTO (multi objective regression testing optimization) for selection and prioritization usage in real world [23]. The regression testing techniques use greedy approach [24] taking feedback from previous selection is fast to achieve.

Many researchers have explored area of PSO and on being an optimization approach it has been used to solve routing optimization [25] problems where the optimized path for traffic routing, packet routing is search. The scheduling problems such as: job-scheduling [26], task scheduling problems in distributed systems [27] where task or job is divided into subsets to make performance cost efficient. The PSO has been applied in the field of cryptography and crypt-analysis [28] for military applications such as code encryption and effective code security. The traveling salesman problems is to find optimal path to reach from source to destination with less time and cost, that make it NP-hard combinatorial problem has been solved using PSO technique [29] and other combinatorial problems such as: packing and knapsack [30]. The applications of detection of fault and recovering from them includes field: test pattern generation for circuits [31], software faults detection [32] have been applied by PSO technique.

The GA is favorable field of many scientists, explorers and researchers; this makes application of GA in the diverse fields. The field of regression testing and its techniques has been explored using GA and tools are available to effectively implement regression testing [33]. The GA allows data to evolve naturally and it makes logical patterns that help in information processing for trend, financial analysis such as: data mining [34]. The GA is efficiently been used in the field of encryption and code breaking [35]. The GA approach can be used for pattern detection for the learning accuracy improvement and noise filtering [36] of image.

3. HYBRID PRIORITIZED ALGORITHM

In PSO, while searching for best solution in the search space, the particle can converge towards point between particle best and local best (neighbor best), this results particle to converge towards single point, ignoring other aspects of search spaces (ignoring global best conditions). To overcome this limitation, GA factor has been used along with PSO algorithm to propose hybrid prioritized algorithm. The reason to apply crossover [5] in PSO is to increase population diversity and ability of PSO to avoid local maxima [37] and make searching process fast.

3.1 Assumptions

a) **Test cases** of given problem is equal to particles in population.

b) Randomly generate initial population.

c) The position, velocity and stopping criteria of particles depend on problem.

d) In our case, two lists are maintained:

   (1) Position= rank of test case based on faults or conditional node covered.

(2) Velocity= execution time or independent paths covered.

e) **Stopping Criteria** is total number of faults covered in minimum time and independent path covered.

**Input:** ‘n’ number of Test Case is selected.

**Output:** Test suite with minimum number of test cases and fulfillign stopping criteria.

3.2 Algorithm

Proposed Algorithm

1. **Step I**—Generate Population

   ‘n’ number of particles [p1…pn] generated.

2. **Step II**—Initialization

   (Here, a particle represents a test case)

   For each ‘n’ particle
   
   Initialize position.

   Initialize velocity

   Set pos_velo_list(n); //for each particle.

   For each ‘n’ particle
   
   Identify local best.

   Identify global_best particle (within search space).

3. **Step III**—Proposed Algorithm

   For each ‘n’ particle
   
   sort_list(n); //read

   pos_velo_list(n); //read

   for each ‘n’ particle
   
   sort_list(n); //read

   For each ‘n’ particle

   pos_velo_list(n); //read

   sort_list(n); //read

   sort_list(n); //read

   sort_list(n); //read
update_position_velocity(n);
}  //combine with either local or global best
} //end of for

If(updated_position & velocity > old_position & velocity) {
    Keep the new position & velocity
} //end of if

Else { revert back to old position and velocity }
Generate 'New_Population';
} //end of for

Step-IV—Crossover Operator
Divide 'New_Population' into two set.
Apply crossover operator on both parents.
//two new off-springs generated.
For each 'n'
    Remove duplicity. /// comparison /////with in both
    n/2 + n/2 off-springs
} //end for

Step-V—Examine Test Suite
Evaluate new off-springs
    // if loop runs for both off-springs of length n/2 each
    If (any off-spring meets stopping criteria)
    {
        Set as final solution.
    } //end of if
Else
    
    Compare both off-springs
    Select best //near to stopping criteria
    Apply PSO on that off-spring
    (Go to Step-III).
} //end of else

} //end of Proposed Algorithm

3.3 Explanation of Hybrid Prioritized Algorithm
Initially, particles are generated randomly. The particles are initialized with position and velocity. Each particle contains information regarding its own position and velocity. The position in our problem is either number of faults covered or number of conditional node covered and the velocity is either execution time or independent paths covered. Now, set fitness for each particle. The particle covering maximum number of faults/conditions in less time or maximum independent path covered is set as best particle. Each particle does comparison: its own position and velocity with best. On getting new position and velocity each particle compares it with old position and velocity respectively. The crossover operator is applied on new generated population such that: the off springs generated full fills the stopping criteria. The proposed Hybrid Prioritized algorithm has been automated for analyzing the test case prioritization. The algorithm has been implemented using JAVA in appropriate IDE.

3.4 Analysis of Hybrid Prioritized Algorithm
The analysis is based on complexity and correctness of algorithm [38]. The complexity depends on execution time and cost taken by algorithm whereas correctness of algorithm depends on modification-traversing and modification-revealing property of algorithm. The algorithm proved to select modification-revealing (fault-revealing) test cases of any program P, is said to be safe which means if any modification done to P, the faults will be detected by algorithm. The algorithm proposed in this research paper has been tested for code coverage testing where CFG (control flow graph) are used to test the program. The complexity of algorithm depends on number of calls made in algorithm and on data available. In proposed algorithm, the population is generated of size 'n' that takes O (n) operations. The initialization step three lists made that required n*(O (n) + n (2O (n))) operations. Thereafter, in updation and comparison of each particle w.r.t local best and global best comparison between particles takes O(n*logn) operation and updation for each particle is done. The crossover operator does n*O (n) operations for 'n' population. The last step executed for two off-springs of length n/2 each. In best case, the proposed algorithm's run time is O (n'). But in worst case, the recursion operation is applied (in last step) and run maximum 2^n times. This makes complexity to reach O (2^n, n') which can make it NP complete. This algorithm has been executed on two approaches: fault-revealing and code coverage and it is analyzed that this algorithm works on best case prominently.

3.5 Fault based Testing Experimentation
To test perfectly, a program is tested for every single input, whether inputs are valid or invalid. The fault based testing can be used as quantifying measure for test cases. Test cases, which are not able to find mutant or “kill mutant” can be removed from test suite, that is, quality of test suite is tested and raised using mutation testing. Nevertheless, to execute mutation testing is quite expensive task, particularly on the larger applications.

For analyzing the hybrid prioritized algorithm, all examples are executed with implemented code (Triangle, Hotel, Student, Railway, and Quadratic). The examples are executed by crossing over each other and analyzing faults one-by-one. The algorithm is executed 25 times on the each problem and the best prioritized test suite is noted with its execution time. In hybrid prioritized algorithm, the result of each 25 runs is same; therefore, final optimized suite is represented in analysis table. It is declared above that best test suite should cover all faults with minimum time. The table shown below briefly explains the result of each run. The best test suite of each run is shown in Table-1.
With the table shown above it can be generalized that the priority test suite is the shortest time taken test suite. The student example takes T8T6 in '18.33' with average build of 3sec; hotel example takes T5T6 in '20.64' with average build of 3sec, triangle example takes T5T2 in '7.0' with average build of 4sec; quadratic example takes T1T7T3 in '10.0' with average build of 4sec and railway example takes T2T3 in '30.0' with average build of 5sec. The effectiveness of run will be 100% as the best optimal test suite is retrieved with one run only. The saving % of crossover hybrid particle swarm optimization is same as of mutated hybrid particle swarm optimization as the number of test cases prioritized are same as of mutated one. Hence, the crossover hybrid particle swarm optimization algorithm is more efficient in getting optimal result.

3.6 Code Coverage Testing (Path Based) Experimentation

The code coverage based testing consists of; branch coverage, flow coverage, condition coverage. The path coverage testing is also a form of code coverage where internal structure of program is analyzed. In path coverage testing, the numbers of independent paths are searched (McCabe, 1976) based on control flow graph (CFG) that helps in finding the paths and flow of code. Each independent path assures covering at the least one condition.

According to McCabe, the basic path testing technique consists of for steps:

\[ V(G) = e - n + 2p \]

where, \( e = 23, n = 18, p = 1 \)

Test Cases are as follows:

- T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15, T16, T17, T18, T19.

On the basis of DD-graph shown above, the Independent Paths are as follows: ABCDEGHJLMQR, ABCDEGHKMQ, ABCDEGHNQR, ABCDEGNQPR, ABCDEGHNQPR, ABCDEGHJLMQR, and ABFGNQR

3.6.1 APPLYING HYBRID TECHNIQUE ON ABOVE EXAMPLE

According to PSO technique, each particle will look for global best and local best and will update its position and velocity for maximum conditional node and independent path coverage. The crossover operator is applied for swapping test cases to diversify search process.

STEP-I & II:

First, each particle will maintain its own position and velocity. Here, T2, T3, T5 T9, T10, T11, T15, T16 are global best particles as they cover maximum number of conditions. Then, each particle compares itself with best particle. Here, test case T1 has compared itself with T2 (local best) and global best (T2, T3, T5, T9, T10, T11, T15, T16) and since its local best is also one of test case in global best. It had updated its position and velocity w.r.t T2.

STEP-III:

Now, each particle will examine the updated position and velocity with respect to old position and velocity. The new values are retained if it improves position and velocity of particle otherwise revert back to old position.

STEP-IV:

Now, apply crossover operation on updated generation of population. In crossover operation, the population is divided in set of two parents. Then, two-point crossover is applied to search problem space more thoroughly. The one-point crossover is not suited for this problem as it was not diversifying the search space.

![Figure-1: Representing Crossover function](image-url)
STEP-V:
Now, the generated off-springs are compared to get best among them. They are compared on the basis of maximum conditional node coverage and independent path covered. Hence, the second table covers all independent paths and conditions. On removing duplicity: TEST-SUITE A: T10, T11, T3, T4, T5, T6, T7, T8, T9
TEST-SUITE B: T1, T2, T12, T13, T16, T17, T18
TEST-SUITE B covers all independent paths. In this example, the total code coverage has been achieved using path coverage (for identifying independent paths of code).

4. COMPARISION
4.1 Fault Coverage Analysis
The fault detection examples are being compared on the parameter of: random order, reverse order, optimized order and prioritized order for better clarity in results. The comparison has been shown with the help of Average Percentage of Fault Detection (APFD) formula as defined by Rothermal [8]. This formula gives better vision towards final result. In below Sections the APFD of various orders are been calculated and represented with the help of bar graphs respectively.

4.1.1 Example-I
Table-2: Represents % of various orders of Triangle

<table>
<thead>
<tr>
<th>Technique</th>
<th>APFD %</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Order</td>
<td>83.4</td>
</tr>
<tr>
<td>Random Order</td>
<td>84.3</td>
</tr>
<tr>
<td>Reverse Order</td>
<td>78.4</td>
</tr>
<tr>
<td>Optimal Order</td>
<td>90.3</td>
</tr>
<tr>
<td>Proposed Order</td>
<td>90.3</td>
</tr>
</tbody>
</table>

4.1.2 Example-II

Table-3: Represents % of various orders of Hotel

<table>
<thead>
<tr>
<th>Ordering of Test Cases</th>
<th>APFD %</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Order</td>
<td>46%</td>
</tr>
<tr>
<td>Random Order</td>
<td>66%</td>
</tr>
<tr>
<td>Reverse Order</td>
<td>62%</td>
</tr>
<tr>
<td>Optimum Order</td>
<td>66%</td>
</tr>
<tr>
<td>Prioritized Order</td>
<td>66%</td>
</tr>
</tbody>
</table>

Fig-4: Represents APFD for prioritized hotel suite with other Orders

4.1.3 Example-III

Table-4: Represents % of various orders of student

<table>
<thead>
<tr>
<th>Ordering of Test Cases</th>
<th>APFD %</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Order</td>
<td>76.8%</td>
</tr>
<tr>
<td>Random Order</td>
<td>76.8%</td>
</tr>
<tr>
<td>Reverse Order</td>
<td>70%</td>
</tr>
<tr>
<td>Optimum Order</td>
<td>78.9%</td>
</tr>
<tr>
<td>Prioritized Order</td>
<td>78.9%</td>
</tr>
</tbody>
</table>

Fig-5: Represents APFD for prioritized student suite with other orders
4.1.4 EXAMPLE-IV

Table-5: Represents % of various orders of Railway

<table>
<thead>
<tr>
<th>Ordering of Test Cases</th>
<th>APFD %</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Order</td>
<td>81.1</td>
</tr>
<tr>
<td>Random Order</td>
<td>71.6</td>
</tr>
<tr>
<td>Reverse Order</td>
<td>44.2</td>
</tr>
<tr>
<td>Optimal Order</td>
<td>87.9</td>
</tr>
<tr>
<td>Proposed Order</td>
<td>87.9</td>
</tr>
</tbody>
</table>

Fig-6: Represents APFD for railway prioritized suite with other orders

4.1.5 EXAMPLE-V

Table-6: Represents % of various orders for Quadratic

<table>
<thead>
<tr>
<th>Ordering of Test Cases</th>
<th>APFD %</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Order</td>
<td>62.9</td>
</tr>
<tr>
<td>Random Order</td>
<td>82.8</td>
</tr>
<tr>
<td>Reverse Order</td>
<td>83.4</td>
</tr>
<tr>
<td>Optimal Order</td>
<td>90.3</td>
</tr>
<tr>
<td>Proposed Order</td>
<td>90.3</td>
</tr>
</tbody>
</table>

Fig-7: Represents APFD for prioritized quadratic suite with other orders

4.1.6 FINAL ANALYSIS OF FAULT COVERAGE:

The saving % achieved by proposed algorithm has been calculated:

\[ S = \frac{\text{frequency of optimal test cases}}{\text{total number of runs}} \times 100 \]  

(1)

Table-7: Represents summarized form of fault coverage examples

<table>
<thead>
<tr>
<th>Programs</th>
<th>Fault Seed</th>
<th>Effectiveness %</th>
<th>Saving %</th>
<th>Optimal APFD</th>
<th>Prioritized APFD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>5</td>
<td>28%</td>
<td>77.77%</td>
<td>78.9</td>
<td>78.9</td>
</tr>
<tr>
<td>Hotel</td>
<td>5</td>
<td>60%</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Tri</td>
<td>6</td>
<td>16%</td>
<td>88.235</td>
<td>90.3</td>
<td>90.3</td>
</tr>
<tr>
<td>Quad</td>
<td>9</td>
<td>20%</td>
<td>80</td>
<td>90.3</td>
<td>90.3</td>
</tr>
<tr>
<td>Railway</td>
<td>9</td>
<td>36%</td>
<td>89.473</td>
<td>87.9</td>
<td>87.9</td>
</tr>
</tbody>
</table>

Fig-8: Representing the saving of test cases for execution

The Table-7 represents the summarized qualities of proposed algorithm’s solutions for all examples. The table gives information regarding the number of faults seeded in each program respectively. The effectiveness of 25 runs for each program with the percentage of saving in test cases, then the optimal order and prioritized order (proposed order) are compared with each other for all examples respectively.

4.2 CODE COVERAGE ANALYSIS:

To analyze code coverage effectively the Average Percentage of Condition Coverage (APCC) approach has been used [99].

4.2.1 EXAMPLE-I

Table-8: Represents % of various orders of Triangle

<table>
<thead>
<tr>
<th>Technique</th>
<th>APCC %</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Order</td>
<td>89.6</td>
</tr>
<tr>
<td>Random Order</td>
<td>88.7</td>
</tr>
<tr>
<td>Reverse Order</td>
<td>87.8</td>
</tr>
<tr>
<td>Optimal Order</td>
<td>93.1</td>
</tr>
<tr>
<td>Proposed Order</td>
<td>87.8</td>
</tr>
</tbody>
</table>
5. CONCLUSION:
The algorithm has been proposed to prioritize test cases using PSO technique along with crossover operator used in Genetic Algorithm. The technique is used to prioritize test cases on the basis of two different selection criteria namely: (i) total fault coverage with in time constrained environment, (ii) amount of code coverage. The results show that the proposed algorithm can do prioritization of test cases on different selection criteria other than used in this paper, as the algorithm uses the phenomena of convergence (PSO) while diversifying search space (GA operator) for regression testing. In proposed algorithm, PSO technique initializes particles to get fitness which further used for comparison within population to get local candidate solution. Thereafter, crossover operator has been used for diversifying the localization of solution by crossing over two parent strings with each other for getting two new off-springs which are compared according to desired stopping criteria. The effectiveness of proposed algorithm has been shown with the help of APFD and APCC values respectively. The APFD has been calculated for all examples for fault based testing and APCC has been calculated for code coverage testing example. The APFD and APCC calculation helps in evaluate usefulness of proposed algorithm. The APFD & APCC values are comparable w.r.t. optimal result, that proves algorithm prioritizes efficiently. In this paper, test cases have been selected from large set of generated test data. The automation of prioritization has simplified the process. Further improvement is going on.

6. REFERENCES
[8] Lope, H.S., Colelho, L.S., 2005 Particle Swarm Optimization with fast local search for the blind traveling salesman problem, Proceedings of Fifth International Conference on hybrid intelligent systems (HIS05), Brazil, pp.: 245-250.


