Using Genetic Algorithms to Improve Test Suite Prioritizations

Alexander Conrad

Regression Testing

- Test Case: A small piece of code designed to test a part of a larger program.
- Test Suite: A collection of test cases.
- Regression Test Suite: All of the test cases written for a single program.
- Regression testing is an important technique for ensuring that modifications do not create new defects in other parts of the program.

Test Suite Prioritization

- Regression testing can be extremely expensive - commercial test suites are often enormous and can take weeks to run.
- Test Suite Prioritization: Reordering the test cases in a test suite so that requirements are covered as quickly as is possible.
- Prioritization does not reduce the amount of time necessary for the test suite to run, but allows the suite to reveal defects earlier than an unprioritized suite, enabling engineers to begin addressing them sooner.
- Coverage Effectiveness (CE): A metric that reflects how rapidly a test suite covers a set of requirements, such as segments of a program's code.

Genetic Algorithms

- A genetic algorithm is an evolutionary approach to solving optimization problems, imitating the biological processes of mutation and recombination.

Parts of a Genetic Algorithm

- Chromosome: Represents the smallest indivisible unit of data.
- Individual: A collection of chromosomes representing one potential solution to the problem.
- Population: A collection of different individuals.
- Mutation Operator: Randomly moves the chromosomes around within a small number of individuals.
- Crossover Operator: Creates new individuals by recombining patterns of chromosomes in existing individuals.

Experiment Design

- A genetic algorithm-based prioritization system was implemented in the Java programming language.

Experimental Results

- The genetic algorithm produces prioritizations of significantly higher coverage effectiveness than the control prioritization method (a random search through the domain of potential prioritizations).
- The genetic prioritizer's execution time is slightly greater than that of the control method.
- Some crossover operators, particularly the POS (position based) and OX2 (order based) operators, showed promise because they rapidly produced prioritizations of noteworthy CE values.
- This research demonstrates the viability of a genetic algorithm as a method for test suite prioritization.

Contributions and Future Work

- This research is one of the first that has been done on applying genetic algorithms to the problem of regression test suite prioritization.
- The source code for the genetic algorithm-based test suite prioritization system will be released in the near future.
- Senior Thesis Research:
  - A detailed empirical study comparing genetic algorithm-based prioritization and greedy-based prioritization (the current industry standard).
  - Continue development of the genetic algorithm-based prioritizers, producing a releasable research tool which can be used by others.
  - Create a comprehensive suite of visualization and statistical analysis functions for the analysis of regression test suite prioritization techniques.