

Finding Higher Order Mutants Using Variational Execution*

Extended Abstract

Serena Chen

Franklin W. Olin College of Engineering
Needham, Massachusetts

ACM Reference Format:

Serena Chen. 2018. Finding Higher Order Mutants Using Variational Execution: Extended Abstract. In *Proceedings of ACM SIGPLAN International Conference on Systems, Programming, Languages, and Applications: Software for Humanity (SPLASH'18)*. ACM, New York, NY, USA, Article 4, 3 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Mutation testing is an effective but time consuming method for gauging the quality of a test suite. It functions by repeatedly making changes, called mutants, to the source code and checking whether the test suite fails (i.e., whether the mutant is *killed*) [1]. Recent work [4] has shown cases in which applying multiple changes, called a *higher order mutation*, is more difficult to kill than a single change, called a *first order mutation*. Specifically, a special kind of higher order mutation, called a *strongly subsuming higher order mutation (SSHOM)*, can enable equivalent accuracy in assessing the quality of the test suite with fewer executions of tests. Little is known about these SSHOMs, as they are difficult to find. Jia et al [4] used a genetic search to find SSHOMs; however, since it runs over many generations and has an element of randomness, the algorithm is time consuming and the results are often incomplete.

Our goal in this research is to identify a faster, more reliable method for finding SSHOMs in order to characterize them in the future. We propose an approach based on variational execution to find SSHOMs. Preliminary results indicate that variational execution performs better than the existing genetic algorithm in terms of speed and completeness of results. Out of a set of 33 first order mutations, our variational execution approach finds all 38 SSHOMs in 4.5 seconds, whereas the genetic algorithm only finds 36 of the 38 SSHOMs in 50 seconds.

2 APPROACH

In this section, we discuss our approach to finding SSHOMs using variational execution.

2.1 Variational Execution

Variational execution is a dynamic analysis that explores the behavior of a program under all inputs by sharing redundant executions.

* Accepted to SPLASH 2018 Student Research Competition

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SPLASH'18, November 2018, Boston, Massachusetts USA

© 2018 Copyright held by the owner/author(s).

ACM ISBN 123-4567-24-567/08/06.

https://doi.org/10.475/123_4

```

1 boolean equalsThree(int a, int b) {
2   - int c = a + b;
3   + int c = m0 ? a - b : a + b;
4   - return c == 3;
5   + return m1 ? c != 3 : c == 3;
6 }
```

Listing 1: Diff of mutations encoded as runtime variability.

It was originally designed to examine the interactions of configuration options in a program [5]. The output of variational execution is an expression for the input conditions for which certain outputs are achieved.

We encode all possible first order mutants as runtime variables. An example is shown in Listing 1. Then, we can use variational execution to efficiently and exhaustively explore the behavior of the test suite under all combinations of mutants. Variational execution will then output the mutant configurations in which a test case fails as a *variability context*, or a propositional formula over the input options. For example, let's say we have two tests for the code example in Listing 1:

- (1) `assertTrue(equalsThree(1, 2));`
- (2) `assertFalse(equalsThree(3, 4));`

Variational execution would output two variability contexts:

- (1) `{test1 failed} : (m0 ∧ ¬m1) ∨ (¬m0 ∧ m1)`
- (2) `{test2 failed} : m1`

From these variability contexts, we derive SSHOMs using logical reasoning, which we will discuss next.

2.2 Finding Strongly Subsuming Higher Order Mutants

Current approaches to finding SSHOMs, such as genetic algorithms or greedy algorithms, are search based [4]. In contrast, our approach directly generates SSHOMs. This approach uses the output of variational execution to construct a formula that encodes the definition of a SSHOM and evaluates to true for all SSHOMs. Since this formula is a satisfiability problem, we can use a BDD or SAT solver to cheaply find SSHOMs.

To derive the formula, we outline the criteria for identifying SSHOMs as defined by Jia et al [4] and construct a logical expression for each criterion.

Let T be the set of all tests, M be the set of all first order mutants, and $f(t)$ be the propositional formula for the mutant configurations in which test t fails, as demonstrated in the variability contexts in the previous section. Let $\Gamma(m, t)$ be the result of evaluating $f(t)$ with first order mutant m ; in other words, whether or not test t fails with first order mutant m . The criteria for SSHOMs follow:

- (1) The SSHOM must fail at least one test (must not be an equivalent mutant)

$$\bigvee_{t \in T} f(t) \quad (1)$$

In order to find the mutant configurations in which at least one test fails, it is sufficient to take the disjunction of the conditions in which each test fails. Thus, if a mutant configuration successfully kills a test, the whole expression is true.

- (2) Every test that fails the SSHOM must fail each constituent first order mutant

$$\bigwedge_{t \in T} (f(t) \Rightarrow \bigwedge_{m \in M} (\neg m \vee \Gamma(m, t))) \quad (2)$$

Either m is not included or is killed by t

A certain mutant configuration (i.e. higher order mutant) killing a test implies one of two things for each first order mutant in the program: either the first order mutant is not in the higher order mutant ($\neg m$) or the test fails for the first order mutant ($\Gamma(m, t)$). This must hold over all tests and first order mutants.

In addition, we can optimize for SSHOMs that are harder to kill than the constituent first order mutants, excluding those that are equally difficult to kill [4]. We call these *strictly subsuming higher order mutants* and add a third criteria to find them.

- (3) The set of tests that kill the higher order mutant must be a strict subset of the tests that kill all the first order mutants

The selected mutants do not kill the test

$$\bigvee_{t \in T} (\neg f(t) \wedge \bigwedge_{m \in M} (\neg m \vee \Gamma(m, t))) \quad (3)$$

Either m is not included or is killed by t

If we replace all the constituent first order mutants with this strictly subsuming higher order mutant, then we will have a smaller set of tests that can kill the set of mutants. As such, there must exist at least one test that does not fail for the included mutants but does fail for each first order mutant of the higher order mutant.

To compute the SSHOMs, we take the conjunction of Equations 1 and 2 (and Equation 3 for strictly subsuming higher order mutants) and use a BDD or SAT solver to iteratively get all satisfiable solutions. Thus, this formula is guaranteed to generate all of the possible SSHOMs for the given set of first order mutants, the given test set.

3 PRELIMINARY EVALUATION

We compared three approaches to finding SSHOMs: a naive exhaustive search method, the state of the art genetic algorithm outlined in the work of Jia et. al [3, 4], and our variational execution approach.

The exhaustive search serves as a baseline. We iterate through all pairs of first order mutants, then all triples, and so on. For each candidate higher order mutant, we run the tests and check whether the results indicate that the higher order mutant is strongly subsuming. This algorithm has an exponential runtime.

As we could not find an existing implementation of the genetic algorithm, we reimplemented the algorithm presented by Jia et. al [3, 4] with a few modifications optimizing for the current limited number of mutation operators.

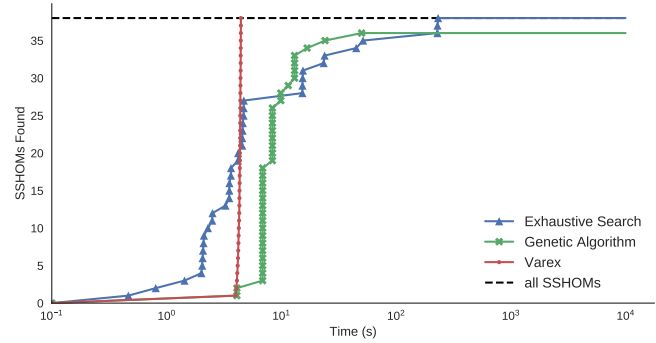


Figure 1: The number of mutants found over time in Triangle. Note, the x axis is in log scale.

We tested the three approaches on the program *Triangle*. Triangle is a simple program that takes three integers representing three side lengths of a triangle and outputs whether the triangle is equilateral, scalene, isosceles, or invalid. This program is used in mutation testing literature and was used to test the original genetic algorithm for finding SSHOMs [4]. Triangle contains about 40 lines of code and 33 first order mutations. The test suite we used was generated by EvoSuite [2], with a few test cases manually added for completeness.

After running the three algorithms on Triangle, we plotted the number of SSHOMs found over time, as shown in Figure 1. Since the variational execution and exhaustive search always find all the SSHOMs, we establish that from the 33 first order mutations, there are 38 total SSHOMs. The exhaustive search finds all the SSHOMs in just under 4 minutes, but does not terminate as it continues testing more higher order mutants. The genetic algorithm finds its last SSHOM after 50 seconds but it only finds 36 of the 38 SSHOMs. Due to the random nature of the genetic algorithm, all four of our trials resulted in the genetic algorithm eventually hanging and no longer able to find more SSHOMs. The variational execution approach finds all the SSHOMs 50 times faster than the exhaustive search and 10 times faster than the genetic algorithm approach. The variational execution approach takes 4 seconds to run variational execution and 0.5 seconds to find all the SSHOMs, a total of 4.5 seconds. Although variational execution takes a while, finding the SSHOMs is very quick as satisfiability checks are inexpensive and fast compared to running a test suite, as happens in the exhaustive search and genetic algorithm.

4 CONCLUSIONS

We propose a new way of finding SSHOMs using variational execution and compared it to the state of the art genetic algorithm and a baseline exhaustive search method. The significant improvement in performance in a small program like Triangle leads us to believe that the variational execution approach is promising. In addition to extending to larger programs, future work includes analyzing the SSHOMs and finding any defining characteristics.

REFERENCES

- [1] R. A. DeMillo, R. J. Lipton, and F. G. Sayward. Hints on test data selection: Help for the practicing programmer. *Computer*, 11(4):34–41, 1978.

- [2] G. Fraser and A. Arcuri. Evosuite: Automatic test suite generation for object-oriented software. In *Proceedings of the 19th ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering, ESEC/FSE '11*, pages 416–419. ACM, 2011.
- [3] M. Harman, Y. Jia, P. Reales Mateo, and M. Polo. Angels and monsters: An empirical investigation of potential test effectiveness and efficiency improvement from strongly subsuming higher order mutation. In *Proceedings of the 29th ACM/IEEE international conference on Automated software engineering*, pages 397–408. ACM, 2014.
- [4] Y. Jia and M. Harman. Higher Order Mutation Testing. *Information and Software Technology*, 51(10):1379–1393, 2009.
- [5] J. Meinicke, C. P. Wong, C. Kästner, T. Thüm, and G. Saake. On Essential Configuration Complexity : Measuring Interactions in Highly-Configurable Systems. In *Proceedings of the International Conference on Automated Software Engineering (ASE)*, pages 483–494, 2016.