



Università degli Studi di Salerno
Dipartimento di Informatica

Tesi di Laurea di I livello in
Informatica

Template tesi ISISLab

Relatore

Giuseppe Scanniello

Correlatore

Dott. Nome Cognome

Candidato

Michelangelo Esposito

Academic Year 2021-2022

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

List of Figures

3.1 NSGA-II algorithm main loop 10

Contents

1	Introduction	1
2	Problem formulation	2
2.1	Coverage testing	2
2.2	An overview on search problems	3
2.3	Search-based test case generation	4
2.4	Iterative single-target search	4
2.5	Whole test suite approach	5
2.6	Many-objective search	6
3	Literature	8
3.1	EvoSuite	8
3.2	NSGA-II	9
3.3	SPEA2	11
3.4	MOSA	12
3.5	DynaMOSA	17
3.6	OCELOT and object-oriented versus procedural test case generation	20
4	Conclusions	21
5	Math reference	22

List of Acronyms and Abbreviations

yo

Chapter 1

Introduction

Things to add:

- where different types of coverage are useful
- Add formulas for coverage criteria

Chapter 2

Problem formulation

2.1 Coverage testing

Coverage is one of the metrics employed during testing to assess what portion of the source code is "covered" by the test suite i.e., what portion of the code is executed when the tests run; it is essential to extract information about the general quality of a test suite and helps determine how comprehensively the software is being verified. As a result, coverage can be classified as a white-box testing technique.

Source code coverage can be expressed according to different criteria:

- Statement coverage aims executing every single statement in the code.
- Branch coverage, also known as decision coverage, measures how many decision structure have been fully explored by the test cases. Out of all the branches in a program, some may be unreachable/infeasible [1], i.e. there is no input configuration that will result in their execution, due to a developer error; as a result, when performing coverage testing, such branches may pose a threat since, if undetected, may drain test budget.
- Mutation coverage, also known as fault-based testing, aims at purposefully introducing faults in the program in order to check whether the test suite is able to identify them. If the fault is correctly detected, the mutant is "killed". One issue of mutation is scalability, since generating and compiling the mutants, before running the test cases, can be time-consuming and quickly exhaust testing resources. Additionally, the introduced mutations can be classified as weak or strong: with strong

mutation, the artificial fault is propagated to an observable behavior, while weak mutants are confined to more specific environments.

- Function coverage measures how many functions have been called by the test cases.
- Condition coverage determines the number of boolean conditions/expressions executed in the conditional statements.

To reach statement coverage, it is sufficient to execute a branch in which the statement is control dependent.

A high coverage can sometimes be deceiving, however: in the case of Machine Learning Systems (MLS), for example, where typically the source code is made up of a sequence of library functions and API invocations [2], thus resulting in very high statement and branch coverage with relatively modest test suites. Additionally, the effectiveness of such systems is highly determined by the dataset employed for model training and validation, which cannot be covered by tradition test cases.

Coverage can also be measured at any testing levels; while at the unit test-level we focus mostly on the coverage of statements and branches, at the system-testing level, the coverage targets shift towards more complex elements, such as menu items, business transactions or other operations that require multiple components of the system to work properly.

A coverage goal is a particular target that we want to cover in respect to the chosen criterion, i.e. a particular branch of an if statement.

2.2 An overview on search problems

As humans, everyday we perform search: from looking for our car keys in the house to searching for a new book to purchase, this action is engraved in our daily life. In the same way, search is one of the most fundamental operations on which computer science has focused since its early days: performing efficient search is the core operation of many classes of algorithms, such as the ones responsible for route planning, computer vision, robotics, automated software testing, puzzle solving, and many others.

Any search problem is typically defined by:

- A search space: the set of elements in which we search for our solution. Examples include paths in a graph, the numbers to insert in a sorted list, or the web pages accesses by a web index.

- A condition that defines the characteristics of candidate solutions.

formal problem definition with graphs, distance functions, ...

A search algorithm will therefore examine the search space according to some criteria and attempt to find a suitable solution. Finding any candidate solution is just one of the objective of search problems however, often times we are interested in finding the best possible solution, the optimum; such problems are referred to as optimization problems.

The exploration of the search space can happen in many different ways. The most basic approach consists of exploring the elements one by one, till a solution/the optimum is found, in a brute-force manner. Applying this simple solution for problems whose search space is somewhat limited can be done without too many repercussions on execution time, however, given the exponential size of the search spaces of most practical problems, a brute-force approach is infeasible.

For problems in which we have no choice but to employ brute-force search, there may be some room for improvement by using heuristics. Heuristics are estimates of the distance function to reach a goal state from the current node [3].

Therefore, what it is preferred to obtain a solution that may not be the optimum, but rather it is "good enough" for our objective. This approach is called local search.

Hill climbing, simulated annealing, GAs... Genetic Algorithms (GAs) are an example of an evolutionary search approach for test case generation; starting from an initial, often randomly generated, population of test cases, the algorithm keeps evolving the individuals according to simulated natural evolution theory principles. In this context, a typical fitness function of a GA would measure the distance between the execution trace of the generated test cases and the coverage targets.

2.3 Search-based test case generation

Search-based approaches for test case generation use optimization algorithms to attempt to find the best candidate test case with the objective to maximize fault detection.

2.4 Iterative single-target search

The simplest way of approaching the evolutionary search problem for test case generation is by iteratively determining a coverage goal in the source

code (i.e. a particular branch), and executing the GA to find a test case that achieves this coverage. A strategy for iterative search typically includes:

- Enumerating the targets to cover, i.e. the individual branches.
- Performing the single-objective search for each target, until all the targets are covered, or the budget has expired.
- Building the final test suite by combining all the generated test cases.

Formally, the search problem for branch coverage can be formulated as follows:

Problem 1. Let $B = \{b_1, \dots, b_k\}$ be the set of branches in a program; find a test suite $T = \{t_1, \dots, t_n\}$ of n test cases that covers all feasible branches, minimizing the fitness function:

$$\min f_B(T) = |M| - |M_T| + \sum_{b \in B} d(b, T) \quad (2.1)$$

where $|M|$ is the total number of methods, $|M_T|$ is the number of executed methods by all test cases in T , and $d(b, T)$ indicates the minimal normalized branch distance for branch $b \in B$

Add infeasible branches

Focusing on one coverage goal at a time can be a poor strategy, however. Foremost, a search algorithms could get "trapped" while attempting to cover an expensive branch and waste a large portion of the testing budget [4]. Secondly, the order by which coverage targets are selected may end up largely impacting the final performance. Additionally, this approach assumes that all coverage goal are equally important and independent of each other; this is often not the case as, for example, covering the true branch of an if statement may be easier than covering the true branch of another if statement that requires a complex chain of operations to be properly satisfied. Finally, covering a particular branch may have collateral coverage over other branches in the test case's path.

2.5 Whole test suite approach

The issues with iterative search suggest that multi-target evolutionary approaches may reveal more effective and reliable. These are known as whole test suites approaches (WS) [5] and their goal is to evolve the entire test suite

simultaneously, rather than iteratively covering the single branches/statements; this eliminates the two issues of the iterative approach: the algorithms can't get stuck on an expensive branch, since the all coverage targets are being searched simultaneously and, for the same reason, the order of test case generation becomes irrelevant, preventing adversary influence in the final test suite.

In the context of the WS approach, the fitness function used is still one for the entire test suite and is computed as an aggregated value from the fitness values measured for the single test cases, in order to take into consideration all coverage targets simultaneously for the chosen criterion.

In fact, each test case in a test suite is associated with the target closest to its execution trace. The sum over all test cases of such minimum distances from the targets provides the overall, test-suite-level fitness. The additive combination of multiple targets into a single, scalar objective function is known as sum scalarisation in the theory of optimization [6].

Branch coverage is typically the most used coverage criterion for test case generation. In this context, a fitness function is based on two parameters:

- Approach level: represents how far the execution path of a given test case is from covering the target branch.
- Branch distance: represents how far the input data is from changing the conditional value of the branch.

Given that the branch distance can be arbitrarily greater than the approach level, usually this value is normalized [7].

While being more effective than the iterative approach, algorithms based on WS principles suffer from the problems of sum-scalarization in many-objective optimization, among which the inefficient convergence occurring in the non-convex regions of the search space

2.6 Many-objective search

The final approach consists in performing a many-objective search; here different coverage targets are considered different objectives to be optimized. Studies in literature [8] have demonstrated that when applying many-objective search to a complex problem previously approached with

single-objective search may bring improvements, since the probability of being trapped in a local maxima is reduced, also leading to a faster convergence rate [9].

The new problem can then be formulated as follows [10]:

As a result of the nature of this problem, the fitness function doesn't stem from the aggregated fitness scores of the individual test cases anymore; rather, it is measured according to the multi-objective nature of optimality, with the goals of minimizing each of the fitness functions of the generated test case population.

Chapter 3

Literature

In this chapter we provide a general overview of the literature concerning evolutionary test case generation, in both the Object-Oriented and procedural contexts. The main techniques and algorithms employed for the problem will be examined and compared, to act as the base for our study.

3.1 EvoSuite

EvoSuite is an example of an evolutionary algorithm that optimizes the whole test suite towards just one coverage criterion, rather than generating test cases directed towards multiple coverage criteria. With EvoSuite, any collateral coverage isn't a concern since all coverage is intentional, given that the ultimate goal is to generate the whole test suite. The algorithm starts with a randomly generated population of test suites. The fitness function rewards better coverage of the source code; if two suites have the same coverage, the one with fewer statements is chosen. For each test suite, its fitness is measured by executing all of its test cases and keeping track of the executed methods and of the minimal branch distance for each branch.

- Expand on bloat in EvoSuite
- algorithm pseudocode

3.2 NSGA-II

A popular algorithm for many-objective search problems is the Non-dominated Sorting Genetic Algorithm II (NSGA-II). This algorithm is based on three principles:

- It uses elitism when evolving the population: the most fit individuals are carried over along the offsprings.
- It uses an explicit diversity-preserving mechanism, the Crowding distance.
- It emphasizes the non-dominated solutions, as its name suggests.

In the context of test cases, domination can be expressed by the following relation:

Definition 1. A test case x dominates another test case y , also written $x \prec y$ if and only if ...

Algorithm 1: NSGA-II

input : $U = \{u_1, \dots, u_m\}$ the set of coverage targets of a program
Population size M
output: A test suite T

```

1 begin
2    $t \leftarrow 0$ 
3    $P_t \leftarrow \text{RANDOM-POPULATION}(M)$ 
4   while not(search_budget_consumed) do
5      $Q_t \leftarrow \text{GENERATE-OFFSPRING}(P_t)$ 
6      $R_t \leftarrow P_t \cup Q_t$ 
7      $F \leftarrow \text{FAST-NONDOMINATED-SORT}(R_t)$ 
8      $P_{t+1} \leftarrow \emptyset$ 
9      $d \leftarrow 1$ 
10    while  $(|P_{t+1}| + |F_d| \leq M)$  do
11       $\text{CROWDING-DISTANCE-ASSIGNMENT}(F_d)$ 
12       $P_{t+1} \leftarrow P_{t+1} \cup F_d$ 
13       $d \leftarrow d + 1$ 
14     $\text{Sort}(F_d)$  // according to the crowding distance
15     $P_{t+1} \leftarrow P_{t+1} \cup F_d[1 : (M - |P_{t+1}|)]$ 
16     $t \leftarrow t + 1$ 
17   $S \leftarrow t + 1$ 

```

The NSGA-II algorithm works as follows:

- Starting from an initial population of individuals P_t , generate an offspring population Q_t of equal size and merge the two together, obtaining the population R_t .
- Perform non-dominated sorting of the individuals in R_t based on target indicators and classify them by fronts, i.e. they are sorted according to an ascending level of non-domination. This ensures that the top Pareto-optimal individuals will survive to the next generation.
- If one of the fronts in the sorted sequence doesn't fit in terms of population size, crowding distance sorting is performed.
- Create the new population based on crowded tournament selection, then perform crossover and mutation.

Figure 2.2 summarizes the main loop of the algorithm:

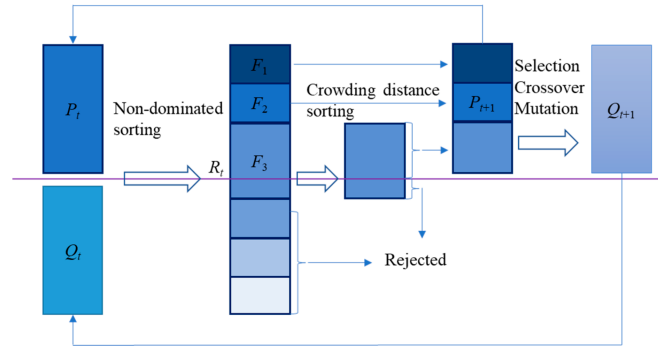


Figure 3.1: NSGA-II algorithm main loop

In the context of software engineering, NSGA-II has been applied to problems such as software refactoring and test case prioritization, with two or three objectives. If the number of objectives begins to grow, however, the performance of the algorithm doesn't scale up well [11]. To overcome these limitations, there have been various adjustments and optimization of this algorithm...

3.3 SPEA2

Strength Pareto Evolutionary Algorithm (SPEA) is another multi-objective evolutionary search algorithm based on the concept of Pareto domination for fitness and selection, originally proposed by Zitzler and Thiele in 1999 [12]. SPEA2 was presented as an evolution of the algorithm by the same authors [13]. A peculiarity of SPEA2 is that it uses an archive to store non-dominated solutions generated in each iteration; as we will see later, the archive approach is also employed by more modern solutions.

During the evolution, a strength value is calculated for each individual in each generation and is used for the selection of the fittest individuals. For any solution i , this strength value is measured based on the numbers of j individuals, belonging to the archive and the population, dominated by i .

Algorithm 2: SPEA2

input : $U = \{u_1, \dots, u_m\}$ the set of coverage targets of a program
 An archive A
 Population size M

1 begin
2 $P_t \leftarrow \text{RANDOM-POPULATION}(M)$
3 **while** *not*(*search_budget_consumed*) **do**
4 $\text{CALCULATE-STRENGTH}(P_t, A)$
5 $\text{UPDATE-ARCHIVE}(P_t, A)$
6 $\text{BINARY-TOURNAMENT-SELECTION}(P_t)$
7 $\text{CROSOVER}(P_t)$
8 $\text{MUTATION}(P_t)$

If the size of the archive doesn't reach the population size, dominated individuals are used to fill up the remaining spaces. Similarly to NSGA-II isn't suitable for problems with more than three objectives[14].

3.4 MOSA

The Multi-Objective Sorting Algorithm, MOSA [9] is a GA proposed as an improvement over NSGA-II and SPEA2 for many-objective test case generation, with the main goal of allowing for a higher number of objectives to optimize. With this algorithm, the coverage of the different branches makes up the set of objectives to be optimized. The main characteristic about MOSA is that instead of ranking candidates for selection based on their Pareto optimality, it uses a preference criterion; this criterion selects the test case with the lowest objective score for each uncovered target; these selected individuals are given a higher chance of survival, while other test cases are ranked with the traditional NSGA-II approach.

As the first step, MOSA starts from a randomly generated initial population of test cases. To create the next generation, MOSA generates offsprings by implementing the classic operations of selection, crossover and mutation. Before selection occurs however, for each uncovered branch, the test case with the lowest objective score(branch distance + approach level) is determined; these test cases will have assigned the rank 0 and will make up the first front. The remainder of the test cases will be sorted according to the traditional NSGA-II approach.

After the rank assignment step is complete, the crowding distance is calculated to determine which individuals to select: the individuals with higher distance from the rest are given a higher chance of being selected. The algorithm attempts to select as many test cases as possible, starting from front 0, until the population size is reached.

Algorithm 3: MOSA

input : $U = \{u_1, \dots, u_m\}$ the set of coverage targets of a program
Population size M
output: A test suite T

```
1 begin
2    $t \leftarrow 0$ 
3    $P_t \leftarrow \text{RANDOM-POPULATION}(M)$ 
4    $archive \leftarrow \text{UPDATE-ARCHIVE}(P_t, \emptyset)$ 
5   while  $\text{not}(\text{search\_budget\_consumed})$  do
6      $Q_t \leftarrow \text{GENERATE-OFFSPRING}(P_t)$ 
7      $archive \leftarrow \text{UPDATE-ARCHIVE}(Q_t, archive)$ 
8      $R_t \leftarrow P_t \cup Q_t$ 
9      $F \leftarrow \text{PREFERENCE-SORTING}(R_t)$ 
10     $P_{t+1} \leftarrow \emptyset$ 
11     $d \leftarrow 0$ 
12    while  $(|P_{t+1}| + |F_d| \leq M)$  do
13       $\text{CROWDING-DISTANCE-ASSIGNMENT}(F_d)$ 
14       $P_{t+1} \leftarrow P_{t+1} \cup F_d$ 
15       $d \leftarrow d + 1$ 
16     $\text{Sort}(F_d)$  // according to the crowding distance
17     $P_{t+1} \leftarrow P_{t+1} \cup F_d[1 : (M - |P_{t+1}|)]$ 
18     $t \leftarrow t + 1$ 
19   $T \leftarrow archive$ 
```

Algorithm 4: PREFERENCE-SORTING

input : $T = \{t_1, \dots, t_n\}$, a set of candidate test cases

Population size M

output: A non-dominated ranking assignment F

```
1 begin
2    $F_0$  // first front
3   for  $u_i \in U$  and  $u_i$  is uncovered do
4     // for each uncovered test target, the best test case according
5     // to the preference criterion is selected
6      $t_{best} \leftarrow$  test case in  $T$  with minimum objective score for  $u_i$ 
7      $F_0 \leftarrow F_0 \cup \{t_{best}\}$ 
8    $T \leftarrow T - F_0$ 
9   if  $|F_0| > M$  then
10     $F_1 \leftarrow T$ 
11  else
12     $U^* \leftarrow \{u \in U : u \text{ is uncovered}\}$ 
13     $E \leftarrow$  FAST-NONDOMINATED-SORT( $T$ ,  $\{u \in U^*\}$ )
14     $d \leftarrow 0$ 
15    for All non-dominated fronts in  $E$  do
16       $F_{d+1} \leftarrow E_d$ 
```

Algorithm 5: DOMINANCE-COMPARATOR

input : Two test cases to compare, t_1 and t_2
 $U = \{u_1, \dots, u_m\}$ the set of coverage targets of a program

```
1 begin
2    $dominates1 \leftarrow false$ 
3    $dominates2 \leftarrow false$ 
4   for  $u_i \in U$  and  $u_i$  is uncovered do
5      $f_i^1 \leftarrow$  values of  $u_i$  for  $t_1$ 
6      $f_i^2 \leftarrow$  values of  $u_i$  for  $t_2$ 
7     if  $f_i^1 < f_i^2$  then
8        $dominates1 \leftarrow true$ 
9     if  $f_i^2 < f_i^1$  then
10       $dominates2 \leftarrow true$ 
11     if  $dominates1 == true$  and  $dominates2 == true$  then
12       break
13   if  $dominates1 == dominates2$  then
14     //  $t_1$  and  $t_2$  don't dominate each other
15   else
16     if  $dominates1 == true$  then
17       //  $t_1$  dominates  $t_2$ 
18     else
19       //  $t_2$  dominates  $t_1$ 
```

Finally, MOSA uses an archived population that keeps track of the best performing test cases, in order to form the final test suite. The archive is continuously updated as follows:

Algorithm 6: UPDATE-ARCHIVE

input : A set of candidate test cases T

An archive A

output: An updated archive A

```

1 begin
2   for  $u_i \in U$  do
3      $t_{best} \leftarrow \emptyset$ 
4      $best\_length \leftarrow \infty$ 
5     if  $u_i$  already covered then
6        $t_{best} \leftarrow$  test case in  $A$  covering  $u_i$ 
7        $best\_length \leftarrow$  number of statements in  $t_{best}$ 
8     for  $t_j \in T$  do
9        $score \leftarrow$  objective score of  $t_j$  for target  $u_i$ 
10       $length \leftarrow$  number of statements in  $t_j$ 
11      if  $score == 0$  and  $length \leq best\_length$  then
12        replace  $t_{best}$  with  $t_j$  in  $A$ 
13         $t_{best} \leftarrow t_j$ 
14         $best\_length \leftarrow length$ 
15   return  $A$ 

```

3.5 DynaMOSA

One of the limitations of MOSA is its inability to consider the implicit dependencies between targets, which simply appear as independent objectives to optimize. Such dependencies can be quite common in practice: most commonly, there exist branches which may only be satisfied if and only if other branches in the outer scope have been already covered.

DynaMOSA, Dynamic Many-Objective Sorting Algorithm [10] is an algorithm that focuses on these dynamic dependencies, and has been proposed as an evolution of MOSA. Before introducing the algorithm, a few definitions are needed [10]:

Definition 2. (Dominator): A statement $s1$ dominates another statement $s2$ if every execution path to $s2$ passes through $s1$.

Definition 3. (Post-dominator): A statement $s1$ post-dominates another statement $s2$ if every execution path from $s2$ to the exit point passes through $s1$.

Definition 4. (Control dependency): There is a control dependency between program statement $s1$ and program statement $s2$ iff: (1) $s2$ is not a postdominator of $s1$, and (2) there exist a path in the control flow graph between $s1$ and $s2$ whose nodes are postdominated by $s2$.

Definition 5. (Control dependency graph): The graph $G = \langle N, E, s \rangle$, consisting of nodes $n \in N$ that represent program statements, and edges $e \in E \subseteq N \times N$ that represent control dependencies between program statements, is called control dependency graph. Node $s \in N$ represents the entry node, which is connected to all nodes that are not under the control dependency of another node.

This definition can be extended to other coverage criteria. For example, given two branches $b1$ and $b2$, we say that $b1$ holds a control dependency on $b2$ if $b1$ is postdominated by a statement $s1$ which holds a control dependency on a node $s2$ that postdominates $b2$.

DynaMOSA uses the control dependency graph to identify which targets are independent from each other and which ones can be covered only after satisfying previous targets in the graph. Algorithm 6 highlights the test case evolution in DynaMOSA.

Algorithm 7: DynaMOSA

input : $U = \{u_1, \dots, u_m\}$ the set of coverage targets of a program
Population size M
 $G = \langle N, E, s \rangle$: control dependency graph of the program
 $\phi : E \rightarrow U$: partial mapping between edges and targets
output: A test suite T

```
1 begin
2    $U^* \leftarrow$  targets in  $U$  with not control dependencies
3    $t \leftarrow 0$ 
4    $P_t \leftarrow$  RANDOM-POPULATION( $M$ )
5    $archive \leftarrow$  UPDATE-ARCHIVE( $P_t, \emptyset$ )
6    $U^* \leftarrow$  UPDATE-TARGETS( $U^*, G, \phi$ )
7   while not( $search\_budget\_consumed$ ) do
8      $Q_t \leftarrow$  GENERATE-OFFSPRING( $P_t$ )
9      $archive \leftarrow$  UPDATE-ARCHIVE( $Q_t, archive$ )
10     $U^* \leftarrow$  UPDATE-TARGETS( $U^*, G, \phi$ )
11     $R_t \leftarrow P_t \cup Q_t$ 
12     $F \leftarrow$  PREFERENCE-SORTING( $R_t, U^*$ )
13     $P_{t+1} \leftarrow \emptyset$ 
14     $d \leftarrow 0$ 
15    while ( $|P_{t+1}| + |F_d| \leq M$ ) do
16      CROWDING-DISTANCE-ASSIGNMENT( $F_d, U^*$ )
17       $P_{t+1} \leftarrow P_{t+1} \cup F_d$ 
18       $d \leftarrow d + 1$ 
19    Sort( $F_d$ ) // according to the crowding distance
20     $P_{t+1} \leftarrow P_{t+1} \cup F_d[1 : (M - |P_{t+1}|)]$ 
21     $t \leftarrow t + 1$ 
22   $T \leftarrow archive$ 
```

The main difference with MOSA is how the initial target population is selected; with DynaMOSA, only the targets that are free from dependencies are part of this set. Then, in each iteration, the set of non-dependent targets is updated by using the following procedure:

Algorithm 8: UPDATE-TARGETS

input : $G = \langle N, E, s \rangle$: control dependency graph of the program
 $U^* \subseteq U$: current set of targets
 $\phi : E \rightarrow U$: partial mapping between edges and targets
output: U^* : updated set of targets to optimize

```

1 begin
2   for  $u \in U$  do
3     if  $u$  is covered then
4        $U^* \leftarrow U^* - \{u\}$ 
5        $e_u \leftarrow$  edge in  $G$  for the target  $u$ 
6       visit( $e_u$ )
7 function visit( $e_u \in E$ ):
8   for each unvisited  $e_n \in E$  control dependent on  $e_u$  do
9     if  $\phi(e_n)$  is not covered then
10       $U^* \leftarrow U^* \cup \{\phi(e_n)\}$ 
11      set  $e_n$  as visited
12   else
13     visit( $e_m$ )

```

This routine update the sets of selected targets U^* in order to include any uncovered targets that are control dependent on the newly covered target. In the case of newly covered targets, the procedure iterates over the control graph to find all control dependent targets.

3.6 OCELOT and object-oriented versus procedural test case generation

Generally speaking, both EvoSuite and MOSA/DynaMOSA have been designed with the OO paradigm in mind

Optimal Coverage sEarch-based tooL for sOftware Testing, OCELOT [15] is a test case generation tool for C programs that implements both a multi-objective approach based on MOSA, and a new iterative single-target approach named LIPS, Linear Independent Path-based Search. OCELOT can automatically detect the input types of a C function, without requiring any specification of parameters; Additionally it can handle the different data types of C, including pointers and structs, and produces test cases based on the Check framework [16]. Similarly to other tools, OCELOT is not able to generate oracles. From an implementation point of view, the GA implementation in OCELOT was realized with the JMetal, a Java framework for multi-objective optimization with meta-heuristics [17]. Furthermore, JNI is used as interface with the target program to try different combinations of test data. SBX Crossover, Polynomial Mutation and Binary Tournament Selection, for both the multi-objective and iterative approaches.

Running OCELOT consists of two main phases: build and run [18]. Focusing on the latter,

Chapter 4

Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 5

Math reference

1.

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = \lim_{n \rightarrow \infty} \frac{n}{\sqrt[n]{n!}}$$

2.

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

3.

$$e = 2 + \frac{1}{1 + \frac{1}{2 + \frac{1}{3 + \frac{1}{4 + \ddots}}}}$$

4.

$$\int_a^b f(x) dx$$

5.

$$x_1, x_2, \dots, x_n$$

Bibliography

- [1] Derek F. Yates and Nicos Malevris. “Reducing the Effects of Infeasible Paths in Branch Testing”. In: *Proceedings of the ACM SIGSOFT '89 Third Symposium on Testing, Analysis, and Verification, TAV 1989, Key West, Florida, USA, December 13-15, 1989*. Ed. by Richard A. Kemmerer. ACM, 1989, pp. 48–54. DOI: 10.1145/75308.75315. URL: <https://doi.org/10.1145/75308.75315>.
- [2] Vincenzo Riccio et al. “Testing machine learning based systems: a systematic mapping”. In: vol. 25. 6. 2020, pp. 5193–5254. DOI: 10.1007/s10664-020-09881-0. URL: <https://doi.org/10.1007/s10664-020-09881-0>.
- [3] Stefan Edelkamp and Stefan Schrodl. *Heuristic Search: Theory and Applications*. Morgan Kaufmann, 2008, pp. 403–449.
- [4] Gordon Fraser and Andrea Arcuri. “Whole Test Suite Generation”. In: vol. 39. 2. 2013, pp. 276–291. DOI: 10.1109/TSE.2012.14. URL: <https://doi.org/10.1109/TSE.2012.14>.
- [5] Gordon Fraser and Andrea Arcuri. “Evolutionary Generation of Whole Test Suites”. In: *International Conference On Quality Software (QSIC)*. Los Alamitos, CA, USA: IEEE Computer Society, 2011, pp. 31–40. DOI: <http://doi.ieeecomputersociety.org/10.1109/QSIC.2011.19>.
- [6] Edmund K. Burke and Graham Kendall. *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*. Springer US, 2014, pp. 403–449.
- [7] Andrea Arcuri. “It Does Matter How You Normalise the Branch Distance in Search Based Software Testing”. In: *Third International Conference on Software Testing, Verification and Validation, ICST 2010, Paris, France, April 7-9, 2010*. IEEE Computer Society, 2010, pp. 205–214. DOI: 10.1109/ICST.2010.17. URL: <https://doi.org/10.1109/ICST.2010.17>.

- [8] Julia Handl, Simon C. Lovell, and Joshua D. Knowles. “Multiobjectivization by Decomposition of Scalar Cost Functions”. In: *Parallel Problem Solving from Nature - PPSN X, 10th International Conference Dortmund, Germany, September 13-17, 2008, Proceedings*. Ed. by Günter Rudolph et al. Vol. 5199. Lecture Notes in Computer Science. Springer, 2008, pp. 31–40. DOI: 10.1007/978-3-540-87700-4_4. URL: https://doi.org/10.1007/978-3-540-87700-4%5C_4.
- [9] Annibale Panichella, Fitsum Meshesha Kifetew, and Paolo Tonella. “Reformulating Branch Coverage as a Many-Objective Optimization Problem”. In: *8th IEEE International Conference on Software Testing, Verification and Validation, ICST 2015, Graz, Austria, April 13-17, 2015*. IEEE Computer Society, 2015, pp. 1–10. DOI: 10.1109/ICST.2015.7102604. URL: <https://doi.org/10.1109/ICST.2015.7102604>.
- [10] Annibale Panichella, Fitsum Meshesha Kifetew, and Paolo Tonella. “Automated Test Case Generation as a Many-Objective Optimisation Problem with Dynamic Selection of the Targets”. In: vol. 44. 2. 2018, pp. 122–158. DOI: 10.1109/TSE.2017.2663435. URL: <https://doi.org/10.1109/TSE.2017.2663435>.
- [11] Bingdong Li et al. “Many-Objective Evolutionary Algorithms: A Survey”. In: vol. 48. 1. 2015, 13:1–13:35. DOI: 10.1145/2792984. URL: <https://doi.org/10.1145/2792984>.
- [12] Eckart Zitzler and Lothar Thiele. “Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach”. In: vol. 3. 4. 1999, pp. 257–271. DOI: 10.1109/4235.797969. URL: <https://doi.org/10.1109/4235.797969>.
- [13] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. “SPEA2: Improving the Strength Pareto Evolutionary Algorithm”. In: (2001).
- [14] Xi Liu and Dan Zhang. “Multi-objective Investment Decision Making Based on an Improved SPEA2 Algorithm”. In: *Artificial Intelligence and Security - 5th International Conference, ICAIS 2019, New York, NY, USA, July 26-28, 2019, Proceedings, Part I*. Ed. by Xingming Sun, Zhaoqing Pan, and Elisa Bertino. Vol. 11632. Lecture Notes in Computer Science. Springer, 2019, pp. 434–443. DOI: 10.1007/978-3-030-24274-9_39. URL: https://doi.org/10.1007/978-3-030-24274-9%5C_39.

- [15] Simone Scalabrino et al. “Search-Based Testing of Procedural Programs: Iterative Single-Target or Multi-target Approach?” In: *Search Based Software Engineering - 8th International Symposium, SSBSE 2016, Raleigh, NC, USA, October 8-10, 2016, Proceedings*. Ed. by Federica Sarro and Kalyanmoy Deb. Vol. 9962. Lecture Notes in Computer Science. 2016, pp. 64–79. DOI: 10.1007/978-3-319-47106-8_5. URL: https://doi.org/10.1007/978-3-319-47106-8%5C_5.
- [16] *Check: framework for unit testing in C*. URL: <https://libcheck.github.io/check/>.
- [17] *JMetal Java framework*. URL: <http://jmetal.sourceforge.net/>.
- [18] Simone Scalabrino et al. “OCELOT: a search-based test-data generation tool for C”. In: *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ASE 2018, Montpellier, France, September 3-7, 2018*. Ed. by Marianne Huchard, Christian Kästner, and Gordon Fraser. ACM, 2018, pp. 868–871. DOI: 10.1145/3238147.3240477. URL: <https://doi.org/10.1145/3238147.3240477>.
- [19] A. Author and A. Author. *Book reference example*. Publisher, 2099.
- [20] A. Author. “Article title”. In: *Journal name* (2099).
- [21] *Example*. URL: <https://www.isislab.it>.
- [22] A. Author. “Tesi di esempio ISISLab”. 2099.