



Università degli Studi di Salerno

Dipartimento di Studi e Ricerche Aziendali

Dottorato di Ricerca in Scienze e Tecnologia dell'Informazione,
dei Sistemi Complessi e dell'Ambiente

XII Ciclo

Abstract della tesi di dottorato
in
Software Engineering

Search-Based Software Maintenance and Testing

Annibale Panichella

Coordinatore
Prof. Roberto Scarpa

Tutor
Prof. Andrea De Lucia

Aprile 2014

Abstract

In software engineering there are many expensive tasks that are performed during development and maintenance activities. Therefore, there has been a lot of effort to try to automate these tasks in order to significantly reduce the development and maintenance cost of software, since the automation would require less human resources. One of the most used way to make such an automation is the *Search-Based Software Engineering* (SBSE), which reformulates traditional software engineering tasks as search problems. In SBSE the set of all candidate solutions to the problem defines the search space while a fitness function differentiates between candidate solutions providing a guidance to the optimization process. After the reformulation of software engineering tasks as optimization problems, search algorithms are used to solve them. Several search algorithms have been used in literature, such as genetic algorithms, genetic programming, simulated annealing, hill climbing (gradient descent), greedy algorithms, particle swarm and ant colony.

This thesis investigates and proposes the usage of search based approaches to reduce the effort of software maintenance and software testing with particular attention to four main activities: (i) program comprehension; (ii) defect prediction; (iii) test data generation and (iv) test suite optimization for regression testing. For program comprehension and defect prediction, this thesis provided their first formulations as optimization problems and then proposed the usage of genetic algorithms to solve them. More precisely, this thesis investigates the peculiarity of source code against textual documents written in natural language and proposes the usage of Genetic Algorithms (GAs) in order to calibrate and assemble IR-techniques for different software engineering tasks. This thesis also investigates and proposes the usage of Multi-Objective Genetic Algorithms (MOGAs) in order to build multi-objective defect prediction models that allows to identify defect-prone software components by taking into account multiple and practical software engineering criteria.

Test data generation and test suite optimization have been extensively investigated as search-based problems in literature . However, despite the huge body of works on search algorithms applied to software testing, both (i) automatic test data generation and (ii) test suite optimization present several limitations and not always produce satisfying results. The success of evolutionary software testing techniques in general, and GAs in particular, depends on several factors. One of these factors is the level of diversity among the individuals in the population, which directly affects the *exploration* ability of the search. For example, evolutionary test case generation techniques that employ GAs could be severely affected by *genetic drift*, i.e., a loss of diversity between solutions, which lead to a premature convergence of GAs towards some local optima. For these reasons, this thesis investigate the role played by diversity preserving mechanisms on the performance of GAs and proposed a novel diversity mechanism based on Singular Value Decomposition and linear algebra. Then, this mechanism has been integrated within the standard GAs and evaluated for evolutionary test data generation. It has been also integrated within MOGAs and empirically evaluated for regression testing.