

UNIVERSITÀ DEGLI STUDI DI SALERNO

Dottorato di Ricerca in Scienze Matematiche, Fisiche ed
Informatiche

Curriculum in Sistemi e Tecnologie del Software

XI Ciclo

Abstract

Search-Based Approaches for Software
Development Effort Estimation

Doctoral Dissertation of

Federica Sarro

Coordinator: Prof. Patrizia Longobardi

Advisor: Prof. Filomena Ferrucci

A.A. 2011/2012

Abstract

Effort estimation is a critical activity for planning and monitoring software project development and for delivering the product on time and within budget. Significant over or under-estimates expose a software project to several risks. As a matter of fact under-estimates could lead to addition of manpower to a late software project, making the project later (Brooks's Law), or to the cancellation of activities, such as documentation and testing, negatively impacting on software quality and maintainability. Thus, the competitiveness of a software company heavily depends on the ability of its project managers to accurately predict in advance the effort required to develop software system. However, several challenges exist in making accurate estimates, e.g., the estimation is needed early in the software lifecycle, when few information about the project are available, or several factors can impact on project effort and these factors are usually specific for different production contexts.

Several techniques have been proposed in the literature to support project manager in estimating software project development effort.

In the last years the use of Search-Based (SB) approaches has been suggested to be employed as an effort estimation technique. These approaches include a variety of meta-heuristics, such as local search techniques (e.g., Hill Climbing, Tabu Search, Simulated Annealing) or Evolutionary Algorithms (e.g., Genetic Algorithms, Genetic Programming).

The idea underlying the use of such techniques is based on the reformulation of software engineering problems as search or optimization problems whose goal is to find the most appropriate solutions which conform to some adequacy criteria (i.e., problem goals). In particular, the use of SB approaches in the context of effort estimation is twofold: they can be exploited to build effort estimation models or to enhance the use of existing effort estimation techniques. The usage reported in the literature of SB approaches for effort estimation have provided promising results that encourage further investigations. However, they can be considered preliminary studies. As a matter of fact, the capabilities of these approaches were not fully exploited, either the employed empirical analyses did not consider the more recent recommendations on how to carry out this kind of empirical assessment in the effort estimation and in the SBSE contexts. The main aim of the PhD dissertation is to provide an insight on the use of SB

techniques for the effort estimation trying to highlight strengths and weaknesses of these approaches for both the uses above mentioned.

In particular, on the basis of the weakness highlighted in the state of the art, the research has been carried out aiming at answer the following questions:

- How the design choices characterizing the use of SB approaches impact on the performance of these techniques?
- Are there any differences in the use of different SB techniques?
- Are SB techniques as effective as widely used effort estimation methods?
- Are SB techniques effective to improve the accuracy of other data-driven effort estimation techniques?

Outline

The thesis is structured as follows. Chapter 1 provides background on software development effort estimation and search-based approaches. Chapter 2 discusses the work carried out so far on the use of search-based approaches for software development effort estimation. Chapter 3 focuses on the definition and assessment of three search-based approaches (i.e., Hill Climbing, Tabu Search and Genetic Programming) to build effort estimation models reporting results relative to their setting, comparison and effectiveness. In Chapter 4 we assess the impact of using different objective functions with both Tabu Search and Genetic Programming. In Chapter 5 we propose the use of a search-based approach, namely Tabu Search, to improve the use of a machine learning technique for effort estimation, namely Support Vector Regression. Final remarks conclude the thesis.