Regression test selection and product line system testing

Engström, Emelie

Published in:
[Host publication title missing]

2010

Link to publication

Total number of authors:
1

Regression Test Selection and Product Line System Testing

Emelie Engström

Dept. of Computer Science
Lund University
Lund, Sweden
emelie.engstrom@cs.lth.se

Abstract— Context: **Software product lines (SPL) are used in industry to achieve more efficient software development. To test a SPL is complex and costly and often becomes a bottleneck in the product line organization.** *Objective*: **This research aims to develop and evaluate strategies for improving system test selection in a SPL.** *Method*: **Initially industrial practices and research in both SPL testing and traditional regression test selection have been surveyed. Two systematic literature reviews, two industrial exploratory surveys and one industrial evaluation of a pragmatic test selection approach have been conducted.** *Results*: **There is a lack of industrial evaluations as well as of useful solutions, both regarding regression test selection and SPL testing. Test selection is an activity of varying scope and preconditions, strongly dependent on the context in which it is applied.** *Conclusions*: **Continued research will be done in close cooperation with industry with the goal to define a tool for visualizing system test coverage in a product line and the delta between a product and the covered part of the product line.**

Keywords- software product line testing; regression testing; industrial practices, literature review, test case selection; test coverage

I. INTRODUCTION

Software product line engineering is a means for organizations to customize large numbers of software products instead of developing one-off solutions to each customer or end product. This is achieved through systematic reuse of artefacts throughout the development process. Commonality and variability are identified in an early stage and the software process is divided into domain (platform) engineering and application (product) engineering.

Efficient testing strategies are important for any organization with a large share of their cost in software development. In an organization using software product lines (SPL) it is even more crucial since the share of testing costs increases as the development costs for each product decreases.

Testing a software product line is a complex and costly task since the variety of products derived from the product platform is huge. The major challenge in testing of a software product line regards the large number of required tests and consequently costs. In order to fully test a product line, all possible uses of each generic component, and preferably even all possible product configurations, need to be tested. The close relationship between the developed products and the fact that they are derived from the same specifications indicates an option to reduce the number of tests, due to redundancy.

The overall goal of my research is to: *develop and evaluate strategies for improved system test selection in large scale software product line organisations by minimizing the amount of redundant testing.*

This problem is closely related to the problem of regression testing of evolving software. The goal of regression testing is to verify that previously working software still works after a change. The test scope for regression testing is often set by selecting test cases from an existing test pool, based on knowledge about changes between the system under test and previously tested versions of the system. This could be compared with the testing of a new product configuration in a software product line were the previously tested product line is the older stable version of the system. My starting point has been regression test selection since this activity is researched and practiced to a greater extent. Based on existing knowledge on regression test selection I expect to find and evaluate strategies for system test selection in a product line context

II. PRESENTATION OF RESEARCH

Initially my research has mainly been exploratory, seeking answers to the following questions:

1. What do we know about regression test selection?
2. How is regression testing applied in practice?
3. How can we improve current regression testing practices?
4. What do we know about software product line testing?
5. How is product line testing applied

A. Completed studies

In order to address the first and fourth questions two different kinds of systematic literature reviews have been conducted: a systematic review on regression test selection

[6] and a systematic mapping on software product line testing [5].

The use of systematic reviews, literature reviews in which evidence of a specific question is systematically searched for, assessed, and summarized according to a predetermined criterion, in the software engineering domain has been subject to a growing interest in the last years. In our application of the methodology we followed the, to the specific characteristics of software engineering research adapted, guidelines proposed by Kitchenham *et al.* [8] Contributions of the systematic literature review on regression test selection were:

- A classification scheme for regression test selection techniques intended to make research results more accessible to practitioners within the field.
- Overview and classification of regression test selection techniques evaluated in literature: most of the proposed regression test selection techniques are not feasible to scale up to testing of large complex real time systems.
- Overview and qualitative analysis of reported evidence on regression test selection techniques: Most of the presented techniques are not evaluated sufficiently for a practitioner to make decisions based on research alone. In many studies, only one aspect of the problem is evaluated (e.g. only test suite reduction and not fault detection ability or analysis cost) and the context is too specific to be easily applied directly by software developers.
- Overview of metrics and strategies used for evaluation of regression test selection strategies. Standards for conducting empirical studies, and which measures to evaluate, differ greatly across the studies.

Another observation made was that few studies are replicated, and thus the possibility to draw conclusions based on variations in test context is limited. In order for a practitioner to make use of results of a single study, the study context must be considered and compared to the actual environment into which a technique is supposed to be applied. This is discussed further in our paper [16] which includes proposals for test technique benchmarks.

A mapping study is a variant of a systematic review and could be used if the amount of empirical evidence is too little, which was the case for the software product line testing research, or if the topic is too broad, for a systematic review to be feasible. Both methods are systematic in that a well defined protocol for study selection and analysis is followed but the goal and use differs. A mapping study is performed at a higher granularity level than a systematic review with the aim to identify research gaps and clusters of evidence in order to direct future research, while the goal of a systematic review is to analyse and aggregate the base of empirical evidence [14].

By conducting a systematic mapping of existing research on Software Product Line Testing, the following was observed:

- Most research effort is spent on high level test case derivation. The most frequent goal is automatic generation of test cases from requirements.
- Most solutions are only applicable in a specific context where for example requirements and variability is modelled in a specific way
- No solution is based on traditional regression test selection techniques.
- Research contributions are mainly method proposals.
- There is a need for more empirical evaluations.
- There is a need for tool support.

In addition to the exploratory part an evaluating part in the form of a case study has been conducted [7], which partly answers question 3. We have evaluated a pragmatic strategy, Fix-cache based regression testing, for improving effectiveness of regression testing in a large industrial setting. This case study was applied post hoc. Two different selection strategies were applied to a large scale industrial project and data collection and analysis was made afterwards.

B. On-going studies

In the second part of this work I have focused on industry practice of regression testing and product line testing which is often based on experience alone, and not on systematic approaches.

In order to address the second question, a qualitative survey of industry practice of regression testing is conducted, by means of focus group discussions in a software process improvement network (SPIN) and a questionnaire to validate the results [4]. Issues discussed in the focus group were definitions and practices of regression testing in industry as well as challenges and improvement suggestions. A total of 46 software engineers from 40 different companies participated in the survey. Results are qualitative and of great value in that they highlight relevant and possible directions for future research.

The fifth question is partly addressed in an extensive interview study on the alignment between requirements and test. 15 software engineers representing different parts of a large product line organisation are interviewed by the means of a semi-structured interview-approach. Interviews cover many different aspects of the connections between requirements work and test work e.g. organizational, process related, communication related, configuration management, traceability, artefacts on different stages of the development etc.

C. Future plans

Focus on future work will be on test selection in the product line context. Two aspects of the test selection will be researched.

- Product Line Coverage - How should test cases be related to the system and what is the coverage of a test case? How should the total system test coverage of the product line be monitored?
- Test Scope Selection - How many and which test cases should be executed for a particular

configuration of the product line? How should the delta between the product under test and the product line be expressed?

The overall goal is to define a tool for visualizing system test coverage in a product line and the delta between a product and the covered part of the product line and, by extension, develop and evaluate a tool-based strategy for test selection in a product line context.

1) Product Line Coverage.

A method for expressing and visualizing the completed test coverage (not limited to code coverage but could also refer to coverage at any level such as requirements coverage or variability coverage) of the product line would be a great help in setting the test scope for each new product in the product line. In order to achieve that, a number of questions have to be answered: At which granularity level should test coverage be expressed? – a cost benefit trade off. How should test cases be connected to these smallest pieces? When is a piece covered, is it useful to have several levels of coverage?

2) System Test Selection in a Product Line.

Given this information about to what extent the product line is covered by previous testing, the issue is to determine the test scope for a particular product in the product line. The delta between two products in a product line could be considered a well specified case of evolution between versions in any software system. Thus it is relevant to compare the product line test selection with regression test selection. The major difference between the two situations is the variability model from which the different configurations are derived in a product line. In the case of evolution of software systems, changes may not be well specified and may vary widely in type, size, importance, where in the development process and why they are introduced, while the delta between products in the product line is the result of a well planned and specified product strategy.

All regression test selection strategies evaluated in literature are based on the assumption that test cases not covering changes in the system are not likely to detect new faults [4]. If this assumption holds or not depends on how test coverage of a system is expressed. Questions to answer are: How should the delta between the product under test and the tested product line be expressed? Which factors should be considered when setting the test scope? How to select/prioritize test cases based on this information?

Most of the suggested coverage measures evaluated in the literature used for regression test selection are code based and not feasible to apply in large scale, real time systems [4]. Literature on product line testing propose many idealistic strategies for deriving test cases covering variabilities and commonalities on different levels in a product line (e.g. [1] [12]). Few have been evaluated in a real life setting.

To complete the picture from literature reviews I intend to conduct a case study at a product line company, seeking answers to what the challenges, current practices and possible improvements are regarding system test coverage and system test scope selection for configured products.

The next step is then to evaluate the effects of different levels of granularity of coverage and different ways of relating test cases to parts of the system. We have previously reported on an empirical evaluation of a pragmatic approach for regression test selection where test cases were related to the system based on information from the error reporting system and configuration management system in a large real life setting [7]. Our results were promising but further adaption and evaluation of the strategy is desirable.

3) Additional questions.

The following are examples of sub questions which would be interesting to study but are not the main focus of my continuous work:

- Regression in the tested product line – How is that handled?
- How should the testing be balanced between domain and application testing?
- Which factors in the testing context affect the efficiency of a test strategy?

III. RELATED WORK

Software product line testing is a relatively new research area. The first papers were published in 2001[10]][11], and most papers have been published in workshops and conferences. There is a well established understanding about challenges [9]. However, when looking for solutions to these challenges, we mostly find proposals, and empirical evaluations are sparse. A thorough and basic report on techniques and activities for meeting the challenges is [11]. This work is the starting point for many researchers within the field.

McGregor discusses the possibility of product line organizations to retrieve a high level of structural coverage by aggregating the test executions of each product in the product line [11]. Cohen *et al.* [2] define a family of cumulative coverage criteria based on a relational model capturing variability in the feasible product line instances, e.g. the orthogonal variability model [15]. They further propose application of interaction testing and connect this to the combinatorial coverage criteria [2]. Muccini and van der Hoek [12] propose regression testing, based on comparison of code execution with the architectural design.

However, research on regression testing has been going on for a while, empirical studies are reported on since 1980 [4] and the field is one of the more mature in software engineering. An overview of research on regression test selection is given in our paper [6]. Most of the research is conducted as experiments or small scale case studies and one of the challenges is to scale up solutions and apply them in different industrial contexts. A few large scale case studies have been undertaken e.g. [17]. Another challenge, which is common for all research in software engineering, is how to generalize results and benchmark solutions [16]

REFERENCES

[1] S. Bashardoust-Tajali and J-P. Corriveau, "On extracting tests from a testable model in the context of domain engineering," *Proc. 13th IEEE Int. Conf. Engineering of Complex Computer Systems (ICECCS 06),* pp.98-107, (2008).

[2] M.B. Cohen, M. B. Dwyer and J. Shi, "Coverage and Adequacy in Software Product Line Testing," *Proc. ACM ISSTA Workshop on Role of Software Architecture for Testing and Analysis (ROSATEA 06),* pp 53-63,. New York (2006)

[3] S. Easterbrook, J. Singer, M-A. Storey and D. Damian, "Selecting empirical methods for software engineering research," In F. Shull, J. Singer, and D. Sjøberg, (Eds.), *Guide to advanced empirical software engineering*, pp.285-311, Springer. (2008)

[4] E. Engström and P. Runeson, "A qualitative survey of regression testing practices," unpublished.

[5] E. Engström and P. Runeson, "Software product line testing - a systematic mapping study," unpublished.

[6] E. Engström, P. Runeson and M. Skoglund, "A systematic review on regression test selection techniques." *Information and Software Technology*, Vol. 52, Issue 1. (2010)

[7] E. Engström, P. Runeson and G. Wikstrand, "An empirical evaluation of regression testing based on fix-cache recommendations," *Proc. Third IEEE International Conference on Software Testing, Verification and Validation (ICST 2010)*, in press.

[8] B. Kitchenham, "Procedures for Performing Systematic Reviews," Technical Report TR/SE0401, Keele University, and Technical Report 0400011T.1, National ICT Australia, July 2004.

[9] R. Kolb. and D. Muthig, "Challenges in Testing Software Product Lines.," *Proc. Conf. Quality Engineering in Software Quality (CONQUEST 03),* pp. 81—95. (2003)

[10] J. D. McGregor, "Structuring Test Assets in a Product Line Effort," *Proc. Second ICSE Workshop on Software Product Lines: Economics, Architectures, and Implications*, pp. 89--92, (2001).

[11] J. D. McGregor, "Testing a Software Product Line," Technical Report, CMU/SEI-2001-TR-022, ESC-TR-2001-022. (2001)

[12] H. Muccini and A. van der Hoek, "Towards Testing Product Line Architectures," *Electronic Notes in Theoretical Computer Science 82* No. 6. (2003)

[13] C. Nebut, Y. Le Traon and J. M. Jézéquel, "System Testing of Product Lines: From Requirements to Test Cases," in *Software Product Lines, Research Issues in Engineering and Management* pp. 447–477, Springer. (2006)

[14] K. Petersen, R. Feldt, M. Shahid and M. Mattsson, "Systematic Mapping Studies in Software Engineering", *Proc. 12th Int. Conf. Evaluation and Assessment in Software Engineering (EASE 08).* (2008)

[15] K. Pohl, G. Böckle, and F. van der Linden*, Software Product Line Engineering: Foundations, Principles,and Techniques,* Springer, Heidelberg. (2005)

[16] P. Runeson, M. Skoglund and E. Engström, "Test Benchmarks -- what is the question?," *Proc. First IEEE ICST Workshop Software Testing Benchmark (TESTBENCH 08),* pp.368-371. (2008)

[17] M. Skoglund and P. Runeson, "A case study of the class firewall regression test selection technique on a large scale distributed software system," *Proc. IEEE Int. Symp. Empirical Software Engineering (ISESE 05) .* (2005)

[18] A. Tevanlinna, J. Taina and R. Kauppinen, "Product family testing: a survey," *ACM SIGSOFT Softw. Eng. Notes* 29(2): 12-17. (2004)