A Search Based Approach Towards Robust Optimization in Software Product Line Scoping

Reza Karimpour and Guenther Ruhe
Software Engineering Decision Support Lab
Department of Computer Science, University of Calgary
Calgary, Canada
{reza.karimpour, ruhe}@ucalgary.ca

ABSTRACT
Software product line (SPL) scoping is important for planning upfront investment. One challenge with scoping comes from inaccuracies in estimated parameters and uncertainty in environment. In this paper, a method to incorporate uncertainty in SPL scoping optimization and its application to generate robust solutions is proposed. We model scoping optimization as a *multi-objective* problem with *profit* and *stability* as heuristics. To evaluate our proposal, a number of experiments are conducted. Analysis of results show that both performance stability and feasibility stability were improved providing the product line manager enhanced decision-making support.

Categories and Subject Descriptors
• Software and its engineering ~ Search-based software engineering;

Keywords
Software product line portfolio scoping; Robust optimization; Uncertainty; Multi-objective

1. INTRODUCTION
SPL Scoping is an activity concentrated on deciding the boundaries of problem domain and other scope dependent decisions [5]. Although, once adopted, the SPL may help reduce risks associated with software production [5], yet the decision of which product(s) to include in product line portfolio requires careful evaluation. One reason is that most scoping approaches depend on some input data that are generally estimated. Depending on the type of the project, this estimation may come from stakeholders, marketing, development team, or management [3]. Most studies that model SPL scoping using profit as the main directing measure do not address inaccuracies in estimates [1, 3]. In case of SPL scoping, estimated measures are, among others, customer's willingness-to-pay, production cost, and offerings from competitors. Being estimates mean that these measurements may be inaccurate and also may change after measurement.

In this study, we propose a method based on search-based optimization that performs SPL scoping while considering uncertainty in input data. This is done by adding extra objectives that target stability and measure the quality of a solution with regard to uncertainty. For implementation, we use a multi-objective genetic algorithm and offer new crossover and mutation operators in order to achieve reasonable performance. As for evaluation, we use Monte Carlo simulation to check the stability of solutions.

GECCO'15 Companion, July 11-15, 2015, Madrid, Spain
ACM 978-1-4503-3488-4/15/07.
http://dx.doi.org/10.1145/2739482.2764650.

2. ROBUST SCOPING OPTIMIZATION
Lin and Branke [4] report two groups of approaches for robust optimization: Single-objective and multi-objective. The first group represents uncertainty as part of the fitness function while the second group captures uncertainty as additional objective(s). We adopt the second approach. To support stability, we model scoping as a multi-objective optimization problem that tries to create stable solutions considering fluctuations of input parameters. Our model is based on Muller's work [6] with some modification. For example, to make the model simpler, we assume that each segment has a single customer and the cost of deployment is negligible. Additionally, we see asset scoping as an integral part of whole scoping picture and include it in the modeling. In this paper, we will only consider the uncertainty in environment and try to find solutions that are resistant to changes in assumptions/estimates. In addition, to represent uncertainty, we will be using a probabilistic measure to model the behavior of uncertain part of the model.

First objective function is to model the profit maximization, which is equal to the difference between revenue and cost. Two other objective functions represent *Profit stability* and *Feasibility stability*. To measure these additional objectives, for each individual, we perform a small random sampling on the environment variables and measure the standard deviation of profit and the number of samples with constraint violation respectively. For optimization, the original objective is maximized while the other two are minimized.

3. STUDY DESIGN
We would like to compare regular profit driven approach to the approach proposed in this study. We created two different setups: *Single-Objective* (SO) and *Multi-Objective* (MO). In SO, only the profit objective function is calculated and others are simply set to zero. In MO, other objectives are also evaluated.

To be able to investigate the quality of results, we define profit variance (PVAR**)** and percentage of solutions with violation **(**PSWV**)** measures. PVAR is the variance of profits while PSWV is defined as the percentage of simulations that had at least one constraint violation. To evaluate these measures, Monte Carlo simulation was used. To simulate uncertainty, we use an approach similar to the one introduced by Cantor [3]. For each individual, we create 200 variations of the original environment and evaluate the above measures in these environments. Triangular distribution is used for representing uncertainty. We assume δ is the percentage of deviation from original estimate and lower limit, mode and upper limit are equal to $-\delta$, 0 and δ respectively. To model different level of uncertainty, we define three values $\delta \epsilon \{5,15,30\}$ for representing *low*, *medium*, and *high* uncertainty.

For this study, we use jMetal's implementation of MOCell [7] algorithm. Population size, crossover and mutation probability are set to 100, 0.9 and 0.1 respectively. A single run is defined as 10000 objective function evaluations. We run each setting 10 times to reduce the effect of randomness. For SO, we pick the best resulting solution from each run while for the MO we collect the Pareto front from each run. Finally, for MO, random sampling size of 25 was selected as to be small relative to simulation size. As for dataset, the home automation system [2] is used. This dataset has 36 features, 5 customer segments, and 5 products.

4. RESULTS AND CONCLUSION

When targeting maximum profit, SO appears to attain solutions with higher profits compared to MO. This is because MO considers stability objectives and refrains from very high profit values to avoid solutions that are very likely to be infeasible or may result in profits that are far from original prediction.

To make more direct comparison, first objective function was modified so that it would minimize the distance to a specific profit. This way we would be able to directly compare simulation measures as the profit would be relatively similar for compared solutions. Different values for target profit where chosen to cover a range of attained profits. Finally, to see if two approaches are actually different, we perform Wilcoxon rank sum test and also report the Vargha-Delaney A-measure effect size.

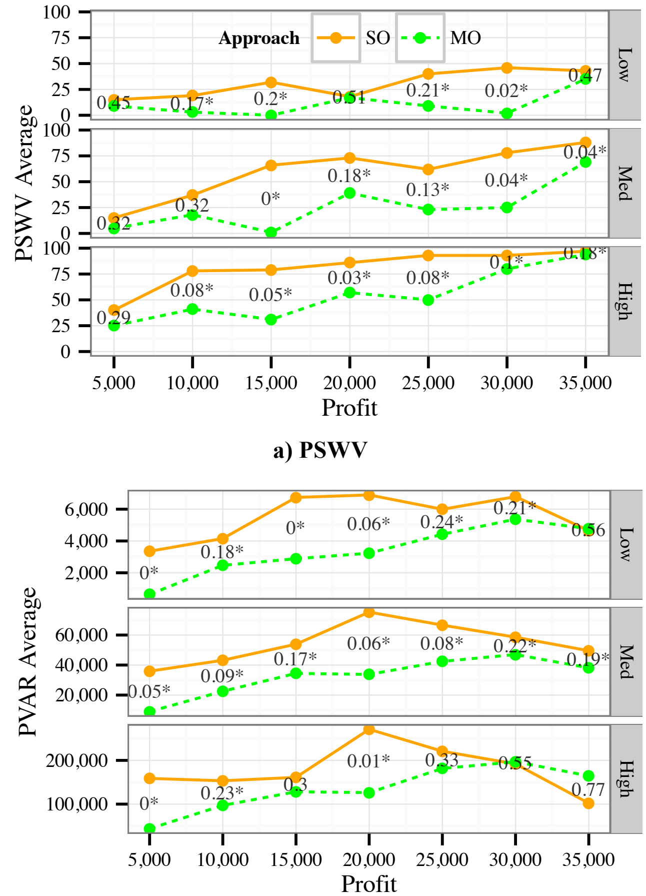
Results are summarized in Figure 1. Solutions included in each approach are a subset of all solutions which had a profit within ±1000 from the target profit. On Figure 1, numbers indicated between two approaches represent the effect size and * is an indication that the test p-value was less than 0.05. Results show that in most settings, MO achieved lower values on both measures. This indicates that MO was able to find equally performing solutions with better stability. For the highest target profit of 40000, MO could not find any solutions.

Within each uncertainty level, for MO, higher target profits generally decrease the stability of solutions. This is more inconsistent for SO. Since we are reporting average measures here, this random behavior may be an indication that there are few near-optimum solutions among feasible ones that are more stable comparatively and are missed by SO. With very high profits (above 30000), MO converges to SO as the number of feasible solutions starts to decrease. As expected, across uncertainty levels, higher uncertainty results in solutions with lower stability. For detailed results of this study refer to http://pages.cpsc.ucalgary.ca/~rkarimpo/g15.html.

One key benefit of Multi-objective robust optimization is that it provides rich decision-making support for product line manager. By presenting the results ranging from low-stability to high-stability, they can evaluate different solutions and perform risk analysis against the profit of each solution.

5. ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant 250343-12 and Alberta Innovates Technology Futures award.



a) PSWV

b) PVAR

Figure 1. Comparison of SO and MO approaches across different target profits.

6. REFERENCES

[1] Ali, M.S., Babar, M.A. and Schmid, K. 2009. A Comparative Survey of Economic Models for Software Product Lines. *35th Euromicro Conference on Software Engineering and Advanced Applications, SEAA* (2009), 275–278.

[2] Alsawalqah, H.I., Kang, S. and Lee, J. 2014. A method to optimize the scope of a software product platform based on end-user features. *Journal of Systems and Software*. 98, (2014), 79–106.

[3] Cantor, M. 2011. Calculating and Improving ROI in Software and System Programs. *Communications of the ACM*. 54, 9 (2011), 121–130.

[4] Jin, Y. and Branke, J. 2005. Evolutionary optimization in uncertain environments-a survey. *IEEE Transactions on Evolutionary Computation*. 9, 3 (Jun. 2005), 303–317.

[5] John, I. and Eisenbarth, M. 2009. A decade of scoping: a survey. *Proceedings of the 13th International Software Product Line Conference* (2009), 31–40.

[6] Müller, J. 2011. Value-Based Portfolio Optimization for Software Product Lines. *15th International Software Product Line Conference, SPLC* (2011), 15–24.

[7] Nebro, A.J., Durillo, J.J., Luna, F., Dorronsoro, B. and Alba, E. 2009. MOCell: A cellular genetic algorithm for multiobjective optimization. *International Journal of Intelligent Systems*. 24, 7 (2009), 726–746.