MAPi DOCTORAL PROGRAMME IN COMPUTER SCIENCE

Thesis Proposal

Automated Software Process Performance Analysis and Improvement Recommendation

Student: Mushtaq Raza
Supervisor: João Pascoal Faria
Hosting research unit: INESC Porto

March 4, 2012

Background

High-maturity software development processes and process improvement frameworks, making use of metrics and quantitative methods, such as the Personal Software Process and Team Software Process from the SEI / CMU, generate a significant amount of data that can be periodically analysed, to identify performance problems, determine their root causes, and identify improvement actions.

For example, it is known that projects delays have as typical causes: effort underestimation, resource allocation inferior to estimation, too many open tasks. In the case of effort underestimation, there are several actions that can be made to improve estimates. Similar analyses can be undertaken in relation to other performance indicators (quality, productivity, etc.).

Currently, there are several tools that automate data collection and produce performance charts for manual analysis in the context of processes such as the PSP/TSP, but practically no tools exists for automating the data analysis and the recommendation of improvement actions.

Manual analysis of this performance data has important problems: there is too much data to analyse, the effort needed is significant, and expert knowledge is required to do the analyses.

So the goal of this research work is to automate this kind of analyses.

Research question

Main question:

Is it possible to automatically identify performance problems, root causes and recommend improvement actions that actually lead to performance improvements (or at least that experts recognize as good recommendations), in industrial software development practice, namely medium to high-maturity organizations?

Some sub-questions:

- is it possible to build or derive a performance model that allows one to automatically identify performance problems (e.g., through control limits) and root causes (e.g., through cause-effect relationships)?

- is there a way to automatically evaluate (predict the cost and impact, possibly based on some kind of model) and rank potential improvement actions in a meaningful way?

Goals

- Construct and validate one or more performance models (defining performance indicators, cause effect relationships between them, and recommended ranges) that can be used to automatically identify performance problems and root causes in software development organizations using medium to high-maturity processes such as PSP/TSP;
- Construct and validate a catalogue of possible improvement actions for performance problems and root causes identified with the previous performance model(s);
- Develop and validate algorithms and processes for automatically evaluating and ranking improvement actions as mentioned above;
- Develop and validate a tool for automating the performance analysis and improvement recommendation;
- If possible, apply the above models, catalogues and tools in a real world context, to show that the recommended improvement actions actually lead to performance improvements.

Methodology

Initially, a state of the art analysis on all the topics relevant to the PhD work (mainly in the scope of the Thesis Planning unit) will be performed. Some initial preparation work will also be done in the scope of the "Free Option" and "Scientific Activities" (see appendix).

The performance model will be constructed based on a combination of expert knowledge (collected through bibliographic search and interviews), and, possibly, automatic discovery of correlations between performance variables in existing (training) data sets.

The constructed performance model (hypothesis) will be validated on existing (training) data sets (possibly, from PSP/TSP).

A tool will be implemented to automate the performance analysis (performance problems and root causes identification) and experiment on other data sets (possibly different from the training data sets). The results produced by the tool will be compared to the results obtained by manual expert analysis for possible additional validation.

Regarding the automatic recommendation of improvement actions, a catalogue of possible improvement actions for each root cause will be constructed based on a combination of expert knowledge (collected through bibliographic search and interviews), and, possibly, automatic means. For evaluating and ranking the improvement actions, will take advantage of the state of the art in

recommendation systems and preference ranking techniques. The above performance model should also be useful to predict the impact of each action and rank the identified actions. The above mentioned tool will be extended to automate the recommendation part, and validate the approach on real data.

For practical reasons - given the existing partnership with the SEI/CMU, that has already agreed on giving access to PSP data, and we expect will also give access to TSP data – our current plan is to focus the work on the PSP/TSP processes. Nevertheless, we'll also seek specific companies that might be interested in this work, in order to validate the work on their data.

Relevant conferences and journals

- ICSSP – International Conference on Software and Systems Processes
- Euro SPI European System & Software Process Improvement and Innovation
- ICSE – International Conference on Software Engineering
- IEEE Transactions on Software Engineering

References

- PSP: A Self-Improvement Process for Software Engineers, Watts S. Humphrey, 2005
- TSP: Coaching Development Teams, Watts S. Humphrey, Addison-Wesley, 2006
- D. Burton and W. S. Humphrey, "Mining PSP Data," in TSP Symposium", 2006
- "Quantitative Project Management Framework via Integrating Six Sigma and PSP/TSP", Sejun Kim, BISTel, Okjoo Choi, KAIST, Jongmoon Baik, KAIS, CrossTalk Journal, July/August 2011
- "A fuzzy logic model for predicting the development effort of short scale programs based upon two independent variables", Cuauhtemoc Lopez-Martin, Applied Soft Computing 11 (2011) 724–732

The PhD Student	The PhD Supervisor
_____	_____
(Mushtaq Raza)	(João Pascoal Faria)

Appendix: Selection of Free Option and Cultural Option

Free Option / Supervised Study

With the agreement of the supervisor, and also the teacher of the curricular unit (because of the teaching language), the student chose the curricular unit "Data Analysis", of the Master Programme in Information Science, 2nd Semester, taught by João Mendes Moreira (that agreed to teach in English).

Approximately half of the curricular unit covers topics on statistics (central limit theorem, confidence intervals, hypothesis testing, analysis of variance, linear regression), whilst the other half covers more advanced topics on data mining (self-learning, prediction algorithms, analysis of clusters, association rules). Given that the student already has some background on statistics, the teacher agreed to replace the attendance of some of the initial classes by the realization of a practical work applying the concepts learned to the analysis of performance data collected in the PSP training courses, e.g., validation of a performance model proposed in a previous work (hypothesis testing), and discovery of interesting correlations (association rules?) between performance variables. This practical work is fully aligned with the PhD goals.

Cultural Option / Scientific Activities

With the agreement of the supervisor, the student chose the alternative "Scientific Activities", corresponding to about 135H of work, with the following goals and tasks:

1) [1 March – 15 May] Conception of a Performance Analysis Component of TSP performance data, that can be incorporated and explored in the context of the AIMS platform (a SaaS platform being developed by FEUP and Strongstep, supporting the Accelerated Improvement Method, combining CMMI, TSP and Six Sigma):
    - 1.1) Formalization of base and derived performance measures, at the personal, team and organizational levels
    - 1.2) Design OLAP cube(s)
    - 1.3) Six Sigma Analysis Charts
    - 1.4) Six Sigma Control Charts

2) [16 May- 30 June] Conception of workflows and processes for handling Process Improvement Proposals (PIPs), that can be incorporated and explored in the context of the same AIMS platform:
    - 2.1) Collecting PIPs connected with process assets (process scripts, templates, role definitions, metrics definitions, checklists, etc.);
    - 2.2) Evaluation of PIPs – popularity, cost/benefit, ranking;
    - 2.3) Implementation of PIPs (possibly generating new versions of process assets, training, change management, etc.).
    - 2.4) Results assessment and decision (accept, backtrack).

This work is very relevant for the proposed PhD work, because the goal of the PhD work is basically to perform automatically the kind of performance analysis and process improvement

recommendations described above, and so the student will get a very good insight on the problem addressed.