

Variance of Source Code Quality Change Caused by Version Control Operations

Csaba Faragó

Motivation

Maintainability is one of the six sub-characteristics of software quality, as defined by the ISO/IEC 9126 standard [5]. Software maintenance consumes huge efforts: based on experiences, about half of the total amount of software development costs are spent on this activity. As maintainability is in direct connection with maintenance costs [2], we investigated the effect of a particular development process task – performing version control operations – on maintainability. Our goal was to explore typical patterns causing similar results in software quality change, which could either help to avoid software erosion, or provide information how to better allocate efforts spent on improving software quality.

Rich data about the developer actions can be found in the version control systems. Number of operations is the most obvious information we can gain for every commit, therefore we decided that we first examine the impact of these data on maintainability. We already performed research in this topic. First we found that there is a strong connection between the version control operations and the maintainability change of the source code [4]. Afterwards, we investigated the impact of the version control operations on maintainability change [3]. We showed that file additions have rather positive, file updates have rather negative effect on maintainability, while a clear effect of file deletions was not identified.

In this current paper we examine the variance of the maintainability changes caused by version control operations. We decided to check this for several reasons. First of all, if the net effect of one commit set is similar to the other's one, the difference in amplitudes could be important. The limited amount of efforts which could be spent on source code quality improvements could be better allocated to those commits which statistically cause higher amplitude. Eliminating the drastic maintainability decreases will result in net maintainability increase (similarly to the greenhouse effect). Second, discovering other dimension of the connections between version control operation and quality change could help in fine-tuning the results of the long-term research. We were especially interested in the variance caused by file deletions, as we did not identify clear impact of this operation on the maintainability previously. Finally, by discovering new connections other questions may raise. These potential new questions result in new research which might bring us closer and closer to the final goal: to create the formula of the developer interactions' impact on the quality of the source code.

Methodology

We took all available revisions of the source code of one industrial and three open-source projects. For each revision we recorded the following values: *number of each version control operation* (Add, Update, Delete), and *maintainability change* caused by that commit.

We estimated the maintainability of each revision by employing the ColumbusQM probabilistic software quality model [1]. Then we calculated the absolute maintainability change based on these values with transformations.

The variance tests are generally executed on two sets of numbers, with the null hypothesis that their variances are the same. Therefore we created two disjoint subsets of commits based on the number of version control operations in several ways. We examined all three operations one-by-one, and defined 7 combinations of divisions for every operation based on the existence, and absolute and relative medians. Then we considered the maintainability change values in

both subsets of every division as the input of variance tests. We executed the test itself with help of `var.test()` function of the R statistical program [6].

In the study we asked the following research question: *What is the impact of each operation (Add, Update, Delete) on the variance of maintainability change?*

Results

We executed the variance tests on one industrial (Gremon) and three open-source projects (Ant, Struts2, Tomcat). As result, we found clear connection between version control operations and the variance of the maintainability change. File Additions and Deletions caused significantly higher variance of maintainability change, compared to file Updates. Commits containing higher number of operations – regardless which operation it was – caused higher variance of maintainability change than those commits containing lower number of operations.

These results help us for better allocating the efforts spent on improving code quality. It is recommended to pay special attention on operations which could cause drastic change on maintainability. These are file additions and file deletions. There is nothing to do with deletions: if something needs to be removed, then it should be removed. On the other hand, it is recommended to pay special attention on new code. For example, it is recommended to mandate code review at least in case of new code development; or, if the code review is mandatory anyway, then it is recommended to do this in these cases with more strict rules. That the number of commits containing file Additions is relatively low considering all the commits, and this is especially true for commits containing file Additions exclusively.

Acknowledgements

This work was supported by Lufthansa Systems Hungary. The author would like to thank the help provided by Rudolf Ferenc and Péter Hegedűs in providing advices to this article.

References

- [1] Bakota, T., Hegedűs, P., Körtvélyesi, P., Ferenc, R., and Gyimóthy, T. A Probabilistic Software Quality Model. In *Proceedings of the 27th IEEE International Conference on Software Maintenance (ICSM 2011)*, pages 368–377, Williamsburg, VA, USA, 2011. IEEE Computer Society.
- [2] Bakota, T., Hegedűs, P., Ladányi, G., Körtvélyesi, P., Ferenc, R., and Gyimóthy, T. A Cost Model Based on Software Maintainability. In *Proceedings of the 28th IEEE International Conference on Software Maintenance (ICSM 2012)*, pages 316–325, Riva del Garda, Italy, 2012. IEEE Computer Society.
- [3] Faragó, Cs., Hegedűs, P., and Ferenc, R. The impact of version control operations on the quality change of the source code. In *14th International Conference on Computational Science and Applications (ICCSA 2014)*. Springer Verlag (accepted, to appear), June 2014.
- [4] Faragó, Cs., Hegedűs, P., Végh, Á. Z., and Ferenc, R. Connection Between Version Control Operations and Quality Change of the Source Code. *Acta Cybernetica (submitted)*.
- [5] ISO/IEC. *ISO/IEC 9126. Software Engineering – Product quality 6.5*. ISO/IEC, 2001.
- [6] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.