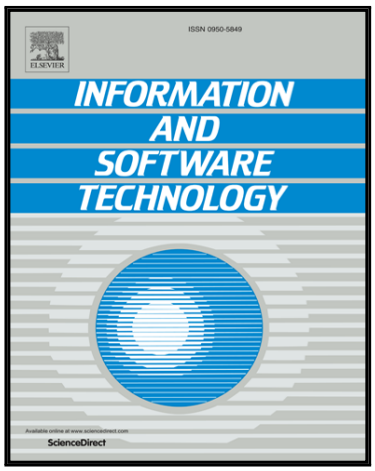

Software Metrics Thresholds Calculation Techniques to Predict Fault-Proneness: An empirical comparison

Alexandre Boucher, Mourad Badri

PII: S0950-5849(16)30313-5
DOI: 10.1016/j.infsof.2017.11.005
Reference: INFSOF 5911

To appear in: *Information and Software Technology*

Received date: 4 November 2016
Revised date: 9 October 2017
Accepted date: 8 November 2017

Please cite this article as: Alexandre Boucher, Mourad Badri, Software Metrics Thresholds Calculation Techniques to Predict Fault-Proneness: An empirical comparison, *Information and Software Technology* (2017), doi: 10.1016/j.infsof.2017.11.005

Software Metrics Thresholds Calculation Techniques to Predict Fault-Proneness: An empirical comparison

Alexandre Boucher[1], Mourad Badri[2,*]

University of Quebec, Trois-Rivières, Canada

Abstract

Context: Nowadays, fault-proneness prediction is an important field of software engineering. It can be used by developers and testers to prioritize tests. This would allow a better allocation of resources, reducing testing time and costs, and improving the effectiveness of software testing. Non-supervised fault-proneness prediction models, especially thresholds-based models, can easily be automated and give valuable insights to developers and testers on the classification performed.

Objective: In this paper, we investigated three thresholds calculation techniques that can be used for fault-proneness prediction: ROC Curves, VARL (Value of an Acceptable Risk Level) and Alves rankings. We compared the performance of these techniques with the performance of four machine learning and two clustering based models.

Method: Threshold values were calculated on a total of twelve different public datasets: eleven from the PROMISE Repository and another based on the Eclipse project. Thresholds-based models were then constructed using each thresholds calculation technique investigated. For comparison, results were also computed for supervised machine learning and clustering based models. Inter-dataset experimentation between different systems and versions of a same system was performed.

Results: Results show that ROC Curves is the best performing method among the three thresholds calculation methods investigated, closely followed by Alves Rankings. VARL method didn't give valuable results for most of the datasets investigated and was easily outperformed by the two other methods. Results also show that thresholds-based models using ROC Curves outperformed machine learning and clustering based models.

Conclusion: The best of the three thresholds calculation techniques for fault-proneness prediction is ROC Curves, but Alves Rankings is a good choice too. In fact, the advantage of Alves Rankings over ROC Curves technique is that it is completely unsupervised and can therefore give pertinent threshold values when fault data is not available.

Keywords: Metrics Thresholds, Class-Level Metrics, Object-Oriented Metrics, Faults, Fault-Proneness Prediction, Machine Learning, Clustering, Cross-validation, Code Quality, Object-Oriented Programming

*Corresponding author

Email addresses: Alexandre.Boucher2@uqtr.ca (Alexandre Boucher), Mourad.Badri@uqtr.ca (Mourad Badri)

[1]Master Student, Software Engineering Laboratory, Department of Mathematics and Computer Science

[2]Professor, Software Engineering Laboratory, Department of Mathematics and Computer Science