

Paper Six Summary

Matthew Neal, Joseph Sankar, and Alexander Sobran

October 22, 2015

Reference

Rahman et al. [1] listed below.

Important Keywords

Cross-project defect prediction Using data from one project to predict defects in another.

Within-project defect prediction Using data from previous releases of a project to predict defects in future releases. For new projects, the lack of historical defect data makes this kind of defect prediction almost impossible.

F-measure The harmonic mean of precision and recall. A unified score used to balance the trade-off between precision and recall.

Over-fitting When a model has completely learned all the variances of the training data and has lost significant generality leading to significantly worse predictive performance on unseen test data. This property usually manifests as very low training error and very high test error. Over-fitting becomes more probable as a model gains greater complexity.

Feature Extraction

Motivational Statements While within-project defect prediction can be very effective, new projects don't have the volume of data needed to create these models. Cross-project defect prediction models aim to help with this issue, but so far the results have largely been disappointing. The authors hope to show that cross-project defect prediction can be roughly as effective as traditional defect prediction by using a different set of measures, namely those based on a variety of tradeoffs of time-and-cost vs. quality.

Possible Improvements

Connection to Other Papers

References

- [1] Foyzur Rahman, Daryl Posnett, and Premkumar Devanbu. Recalling the "imprecision" of cross-project defect prediction. In *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering*, FSE '12, pages 61:1–61:11, New York, NY, USA, 2012. ACM.