

Test case prioritization: a systematic mapping study

Catal, C (Catal, Cagatay)^[1]; Mishra, D (Mishra, Deepti)^[2]

Abstract

Test case prioritization techniques, which are used to improve the cost-effectiveness of regression testing, order test cases in such a way that those cases that are expected to outperform others in detecting software faults are run earlier in the testing phase. The objective of this study is to examine what kind of techniques have been widely used in papers on this subject, determine which aspects of test case prioritization have been studied, provide a basis for the improvement of test case prioritization research, and evaluate the current trends of this research area. We searched for papers in the following five electronic databases: IEEE Explorer, ACM Digital Library, Science Direct, Springer, and Wiley. Initially, the search string retrieved 202 studies, but upon further examination of titles and abstracts, 120 papers were identified as related to test case prioritization. There exists a large variety of prioritization techniques in the literature, with coverage-based prioritization techniques (i.e., prioritization in terms of the number of statements, basic blocks, or methods test cases cover) dominating the field. The proportion of papers on model-based techniques is on the rise, yet the growth rate is still slow. The proportion of papers that use datasets from industrial projects is found to be 64 %, while those that utilize public datasets for validation are only 38 %. On the basis of this study, the following recommendations are provided for researchers: (1) Give preference to public datasets rather than proprietary datasets; (2) develop more model-based prioritization methods; (3) conduct more studies on the comparison of prioritization methods; (4) always evaluate the effectiveness of the proposed technique with well-known evaluation metrics and compare the performance with the existing methods; (5) publish surveys and systematic review papers on test case prioritization; and (6) use datasets from industrial projects that represent real industrial problems.