Gregor Engels, Regina Hebig, Matthias Tichy (Hrsg.): SE 2023,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2023 121

A comprehensive empirical evaluation of generating test suites for mobile applications with diversity – Summary

Thomas Vogel[1] Chinh Tran[1] Lars Grunske[1]

Abstract: In this extended abstract, we summarize our work on analyzing the fitness landscape of the search-based app testing problem and building on that, improving and evaluating a specific solution for this problem. This work has been published under the title of "A comprehensive empirical evaluation of generating test suites for mobile applications with diversity" in the journal Information and Software Technology (IST) in 2021 [VTG21].

Keywords: Mobile apps; Search-based testing; Fitness landscape analysis

Context. In search-based software testing, we often use popular heuristics with default or best-practice configurations to automatically generate tests. Such an out-of-the-box use typically leads to suboptimal results, for instance, in terms of achieved coverage of the software under test. To yield better results, costly trial-and-error experiments are performed to find suitable configurations of a heuristic for a given search problem [AF13]. One example in this context is SAPIENZ [MHJ16] that uses a default NSGA-II heuristic to generate test suites for mobile applications (apps) without adapting this heurisitc to this specific testing problem. Consequently, a promising way to improve the effectiveness of SAPIENZ could be the identification and use of suitable configurations of NSGA-II for this specific problem.

Objective. Focusing on app testing, our objective was to improve the effectivenss of SAPIENZ while avoiding costly trial-and-error experiments to identify suitable configurations of the used NSGA-II. Particularly, we wanted to analytically understand the search problem of SAPIENZ and use this understanding to identify suitable configurations in an informed way. To achieve this objective, we performed a *fitness landscape analysis* [PA12, ME13] of SAPIENZ and used the obtained results to sytematically adapt the NSGA-II heuristic of SAPIENZ. While the analysis of SAPIENZ has been conducted earlier [VTG19], a comprehensive evaluation of our adaptation of SAPIENZ has been presented more recently [VTG21]. In the context of search-based testing, our work is novel as it targets the testing of mobile apps while others have analyzed the fitness landscape for the problem of unit testing (e.g., [AMG17]).

Method. Our fitness landscape analysis focused on the genotypic diversity and evolvability, that is, how the evolved test suites are spread in the search space and evolve over time regarding their fitness (achieved coverage, detected faults, and length of the test cases). We particularly selected diversity as a major aspect for our analysis since it is considered

[1] Humboldt-Universität zu Berlin, Software Engineering Group, Unter den Linden 6, 10099 Berlin, Germany.
{thomas.vogel, grunske}@informatik.hu-berlin.de

important for the performance of evolutionary algorithms, while the performance is analyzed by the evolvability. To perform the fitness landscape analysis, we implemented 11 metrics from the literature to characterize the search space of SAPIENZ regarding diversity and evolvability throughout the search process. Executing SAPIENZ with these metrics on five selected apps, we obtained data that we analyzed regarding diversity and evolvability.

Results. Our analsis indicated that the diversity of the evolved test suites decreases during the first 25 generations of search to a low level. At the same time, SAPIENZ loses its ability to produce better test suites (evolvability)—the search stagnates after 25 generations. Given these results, we adapted the heurisitc of SAPIENZ to preserve the diversity of the test suites during the search by four techniques: initializing the search with diverse test suites, dynamically controlling the diversity during search, eliminating duplicate test suites, and incorporating diversity into the selection. We evaluate the resulting SAPIENZ^{div} in a head-to-head comparison with SAPIENZ on 34 apps. SAPIENZ^{div} significantly outperformed SAPIENZ for coverage on 9/34 and for fault revelation on 20/34 apps while performing similarly on the remaining apps and tending to produce longer test cases than SAPIENZ.

Conclusion. Our work has shown that the understanding of the search problem obtained by the fitness landscape analysis helped us to find a more suitable configuration of SAPIENZ without the need of trial-and-error experiments.

Data Availability. SAPIENZ^{div} is available on GitHub: `https://github.com/thomas-vogel/sapienzdiv-ssbse19`

Bibliography

[AF13] Arcuri, Andrea; Fraser, Gordon: Parameter tuning or default values? An empirical investigation in search-based software engineering. Empirical Software Engineering, 18(3):594–623, 2013.

[AMG17] Aleti, Aldeida; Moser, I.; Grunske, Lars: Analysing the Fitness Landscape of Search-Based Software Testing Problems. Automated Software Engg., 24(3):603–621, 2017.

[ME13] Malan, Katherine M.; Engelbrecht, Andries P.: A survey of techniques for characterising fitness landscapes and some possible ways forward. Information Sciences, 241(Supplement C):148–163, 2013.

[MHJ16] Mao, Ke; Harman, Mark; Jia, Yue: Sapienz: Multi-objective Automated Testing for Android Applications. In: ISSTA'16. ACM, pp. 94–105, 2016.

[PA12] Pitzer, Erik; Affenzeller, Michael: A Comprehensive Survey on Fitness Landscape Analysis. In: Recent Advances in Intelligent Engineering Systems. Springer, pp. 161–191, 2012.

[VTG19] Vogel, Thomas; Tran, Chinh; Grunske, Lars: Does Diversity Improve the Test Suite Generation for Mobile Applications? In: SSBSE '19. Springer, pp. 58–74, 2019.

[VTG21] Vogel, Thomas; Tran, Chinh; Grunske, Lars: A comprehensive empirical evaluation of generating test suites for mobile applications with diversity. Information and Software Technology, 130:106436, 2021.