A Replication and Reproduction
RoSE 6 ICSE 2020
A Replication Study: Mining A Proprietary Temporal Defect Dataset

Tamer Abdou[1], Atakan Erdem[2], Ayse Bener[3], and Adam Neal[4]

[1] Github ID: @Tamer-Abdou
[2] Github ID:
[3] Github ID:
[4] Github ID:
{tamer.abdou, atakan.erdem, ayse.bener}@ryerson.ca
adam_neal@persistent.com

ABSTRACT

Links: A link to the **original** paper[5]:
https://ieeexplore.ieee.org/document/7180127
A link to the **replicated** paper[6]:
https://dl.acm.org/doi/abs/10.1145/2989238.2989246

What: The original study is based on mining Firefox Bug Repository and stores data related to the temporal alignment of developers' interactions. We had a high level of interaction with the authors of the original study. We had full access to their code and data extraction scripts. However, we have developed our own code and data extraction scripts in this study as the datasets were different.

Why: Comprehending temporal activity sequences assist software quality teams to optimize the allocation of their human resources as well on manage project schedules more efficiently. One of the main values of a replication is the comparison of its results with the results of the original study. Through comparison, the consistency between original study and the replication can be identified. We replicated that study to check whether its findings are supported in the context of proprietary projects.

How: We applied a set of guidelines on how to replicate a study in software engineering domain. We met the goals of our study by (i) Providing a descriptive analysis of a proprietary temporal dataset, (ii) Examining the temporal relationship between activities, and (iii) Studying the distribution of activities grouped by the bug status.

Where: The original dataset contains 86,444 bug reports covering a period of eight years, from January 1, 2006, to December 31, 2014. The replicated study employed datasets from an issue tracking system repository of a proprietary and enterprise level software lifecycle management tool. We defined the temporal patterns of activity sequences and compared them with the original Firefox dataset.

Discussion: On the basis of these analyses, we observed that some activities of the sequences are more critical than the others in the context of proprietary projects. Similarities and differences of the relevant activities were highlighted and explained. Integrating results from various empirical studies helped us in gradually generalizing evidences observed in the original study, and identifying the consistencies between open-source projects and proprietary ones.

---

[5] Habayeb, M., Miranskyy, A., Murtaza, S. S., Buchanan, L., and Bener, A. The Firefox Temporal Defect Dataset. In 2015 IEEE/ACM 12th Working Conference on Mining Software Repositories (may 2015), pp. 498–501

[6] Abdou, T., Erdem, A., Bener, A., & Neal, A. (2016). A replication study: mining a proprietary temporal defect dataset. Proceedings of the 2nd International Workshop on Software Analytics, 15–21.