A Replication and Reproduction RoSE 6 ICSE 2020 A Replication Study: Just-In-Time Defect Prediction with Deep Learning

Steven Young[1], Tamer Abdou[2], and Ayse Bener[3]

[1] Github ID:
[2] Github ID: @Tamer-Abdou
[3] Github ID:
{steven.young, tamer.abdou, ayse.bener}@ryerson.ca

ABSTRACT

Links: A link to the **original** paper[4]:
https://www.sciencedirect.com/science/article/abs/pii/S0950584917302501
A link to the **replicated** paper[5]:
https://ieeexplore.ieee.org/document/8452881

What: The original study proposed a methodology that employs a two-layer ensemble called TLEL. The performance of TLEL is compared against previously proposed methodologies for just-in-time defect prediction. The replication study was performed externally without involvement from the original researchers. Data from the same six open source projects and the same feature set as the original study was used. The replication of the experiment, research questions, and design are similar to the original experiment.

Why: Just-in-time defect prediction, which is also known as change-level defect prediction, can be used to efficiently allocate resources and manage project schedules in the software testing and debugging process. The replication study is an empirical validation of hybrid ensemble methodologies to improve the performance in defect prediction relative to previously tested methodologies. The aim of the replication study was to analyze findings and draw cross study conclusions for both the original and replication studies.

How: We applied a set of guidelines on how to replicate a study in software engineering domain. We analyzed and highlighted conclusions about using multiple base learners, optimized weightings, and additional layers with respect to their relationship to classification performance on tested datasets.

Where: The used datasets come from six open source projects: Bugzilla, Columba, JDT, Platform, Mozilla, and PostgreSQL. Combined, they contain data on 137,417 changes. Changes are labeled as either buggy or not buggy and classes are imbalanced with the portion of buggy changes ranging from 5% to 36% within the datasets.

Discussion: Although the study was a replication, we presented a new technique with strong generalization abilities that can be used in future research. This generalization ability enables superior performance on larger datasets where defects may have different characteristics that are better identified by different classifiers. The replication of earlier just-in-time defect prediction approaches enabled us to confirm the external validity of our new proposed prediction approach.

[4] Xinli Yang, David Lo, Xin Xia, and Jianling Sun. 2017. TLEL: A Two-layer Ensemble Learning Approach for Just-in-time Defect Prediction. In Information and Software Technology, Vol. 87. 206–220.

[5] S. Young, T. Abdou and A. Bener, "A Replication Study: Just-in-Time Defect Prediction with Ensemble Learning," 2018 IEEE/ACM 6th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE@ICSE 2018), Gothenburg, 2018, pp. 42-47