Título: The Role of Bug Report Evolution in Reliable Fixing Estimation

Data: 26/05/2021

Horário: 10h

Local: Videoconferência

Resumo:

Context: Bug reports contain information that can be used by researchers and practitioners to better understand the bug fixing process and to enable the estimation of the effort necessary to fix bugs. In general, estimation models are built using the data (e.g., fixing time, severity, number of comments, number of attachments, and number of patches) present in the reports of fixed bugs (i.e., the report final state). However, we claim that this approach is not reliable in a real setting. Effort estimation is necessary for bug fix scheduling and team allocation tasks, which happens closer to the bug report opening than its closing. At that moment, the data available in the bug report is less complete than the data used to build the model, which may lead to an unrealistic estimation. Objective: We propose a new approach to estimate bug-fixing time, i.e., the time span between the moment the bug was first reported until the bug is considered fixed. Different from previous studies, we consider not only the final state of the bug report to create our estimation model but all the previous available states. The concept of bug report evolution is used to create a dataset containing all investigated report states. Method:

First, we verify how often the bug reports and their fields are updated. Next, we evaluate our approach using different machine learning methods with distinct output configurations, class balancing techniques and both standard classification and ordinal regression settings for the bug reports at different states of evolution. The experimental analysis is performed with data from the JIRA issue tracking system of five open-source projects. By leveraging the best models (considering all possible configurations) for the different states of evolution of a bug report, we can assess whether there are significantdifferences in the models' estimation ability due to the report's state. Results: We gather edevidence that the reports' fields are updated often, which characterizes the reports' evolu-tion, impacting the building of bug-fixing estimation models. The models' evaluation shows promising results for Gaussian Process and Logistic Regression when predicting whether abug will be fixed in less or more than five days. We verify higher performance when classifying the initial reports' versions, which is suitable for real-world scenarios, with f-measure values from 0.62 up to 0.74, and AUC from 0.62 up to 0.70. Conclusions: Our experiments show that field updates have a meaningful impact on the models' performance. Furthermore, we present a new approach to deal with the bug report evolution by considering each report version as an independent report. Finally, we also make available our dataset to the community.

Banca:

- Prof. Dr. João Paulo Pordeus Gomes (MDCC/UFC - Orientador)
- Prof. Dr. Lincoln Souza Rocha (MDCC/UFC - Coorientador)
- Prof. Dr. César Lincoln Cavalcante Mattos (MDCC/UFC)
- Prof. Dr. João Bosco Ferreira Filho (MDCC/UFC)