Título: Contributions to Bug-Fixing Time Estimation

Data: 26/05/2022

Horário: 10h00

Local: Videoconferência

Resumo:

Fixing bugs is a crucial aspect of software maintenance. In practice, maintainers have to manage many bug reports that need immediate attention, despite limited resources and tight deadlines. Bug reports aggregate a plethora of information regarding the bug fixing process under different perspectives. Generally, software projects use issue tracking systems to report and monitor bug-fixing tasks. In recent years, several researchers have been conducting bug tracking analysis to understand the problem better and thus provide means to reduce costs and improve the efficiency of the bug-fixing task. This thesis presents three contributions to the bug fixing process, especially bug fix-time estimation. The first contribution is a curated dataset with its mining script, alongside a series of characterizing analyses and visualizations. We describe the acquisition process, the necessity to mine a new dataset, its content, and provide a deeper analysis of selected features that we use in the following contributions. The second contribution is a new approach to estimate bug-fixing time, i.e., the period between the moment the bug was first reported until the bug considered fixed. We consider not only the final state of the bug report to create our estimation model but all the previous available states, different from previous studies that sometimes do not consider the reports' updates. We use the concept of bug report evolution to create a dataset containing all investigated report states. First, we verify how often the bug reports and their fields are updated. Next, we evaluate our approach using different machine learning methods as a classification problem, with distinct output configurations and class balancing techniques. By leveraging the best models (considering all possible designs) for the different stages of the evolution of a bug report, we can assess whether there are significant differences in the models' estimation ability due to the report's state. We gathered evidence that the reports' fields are updated often, which characterizes the reports' evolution, impacting the building of bug-fixing estimation models. The models' evaluation shows promising results when predicting whether a bug will be fixed in less or more than five days, with metrics values similar or better than comparable approaches. The third and final contribution is a study regarding the relationship between bug fixing time and a selected group of features: priority, links (the connection between reports), and code-churn (related to the committed code associated with the bug report). We test five different models - one 'simple' for each project and four hierarchical models, using all projects at once, under different assumptions of how the data could be generated. After describing the modeling process and the preliminary results, we gathered evidence that bug reports with links and a higher value of code churn tend to demand more time to be fixed. On the other hand, the priority level of the reports seems to not influence the bug fixing time.

Banca examinadora:

- 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. Pedro de Alcântara dos Santos Neto (UFPI)
- Prof. Dr. Ajalmar Rego Rocha Neto (IFCE)