Extracting Concise Bug-Fixing Patches from Human-Written Patches in Version Control Systems

High-quality and large-scale repositories of real bugs and their concise patches collected from real-world applications are critical for research in software engineering community. A number of such repositories, e.g.,Defects4J, have been proposed. However, such repositories are rather small because their construction involves expensive human intervention. In this paper, we propose an automatic approach, called BugBuilder.

BugBuilder can extract complete and concise bug-fixing patches from human-written patches in version control systems. It excludes refactorings by detected refactorings involved in bug-fixing commits, and reapplying detected refactorings on the faulty version. It enumerates all subsets of the remaining part and validates them on test cases. If none of the subsets has the potential to be a complete patch, the remaining part as a whole is taken as a complete and concise patch. Evaluation results on 809 real bug-fixing commits in Defects4J suggest that BugBuilder successfully generated complete and concise patches for forty percent of the bug-fixing commits, and its precision(99%) was even higher than human experts. This tool and more information can be found at https://github.com/jiangyanjie/BugBuilder

We plan to apply for badge reusable or available. BugBuilder is functional and is very carefully documented and well-structured to the extent that reuse and repurposing is facilitated. Furthermore, it is an interesting tool that it can extract complete and concise bug-fixing patches from human-written patches in version control systems. Based on BugBuilder, we can build large-scale and high-quality repositories of real bugs automatically in the future.