Jackson Warren

CSCI 330

Prof. Seif

10/27/21

Homework #2 – CSCI 330

Debugging code is something takes up the time of every coder. Finding a way to easily find bugs and fix them is not an easy task. However, there are ways to read code such that programmers can find bugs in their code more easily. This is done in a few general key steps.

Categorizing bugs based on their cause and impact is an important first step. Causes has four subcategories (excluding other/unknown): Algorithmic, where there exist logic errors within the code; Concurrency, where there is a problem with multi-threading as a result of the code; Memory, where memory is misused/misplaced resulting in things such as memory leaks; and Programming, where there are generic error resulting in compilation errors. Furthermore, for impact, there are three subcategories: Security, where the program runs but can be exploited; Performance, where there is a delayed response to the program; and Failure, where the impact results in the crashing of the program.

To classify these bugs, classification is divided into two phases. The first phase, Keyword Search, is done so by taking a random 10% of the error-messages and categorizing them accordingly to create a restrictive set of key-words to define both the cause and impact categories. Then, using the first phase as training set, the second phase, Supervised Classification, is classified using these same keywords for the remainder of the bugs.

Using these techniques, Baishakhi Ray et al. [1] found an average accuracy of 84% for 180 selected bug fixes for both concurrency and recall. They only received 1.04% unclassifiable bug fixes, meaning their techniques were capable of classifying almost all the data at a high percentage of accuracy. With encouraging numbers, this study shows that potential for classifying bugs in code, and reducing the time spent searching for these bugs.

References:

[1] Ray, B., Posnett, D., Devanbu, P., & Filkov, V. (2021). *Communications of the ACM Volume 60, 2017*. Association for Computing Machinery.