Preliminary Study on the Effectiveness of ChatGPT at Identifying, Explaining, and Revealing Semantic Conflicts in Merge Commits

Duarte Guedes Sardão { up201905497@fe.up.pt }

Abstract. In collaborative software development, parallel work done by several different branches often has to be merged. Due to the differences of work done in the difference branches, often conflicts arise. Some conflicts are easy to detect and rectify, such as textual merge conflicts, where different developers have altered the same line in different branches: most version control systems can detect that the same line has been changed and urge the merger to decide on a solution, for example, by keeping one change and discarding the other. Among merge conflicts, semantic merges arise as particularly difficult to resolve, as they avoid detection by version control systems and lacking syntactic errors, compile successfully. An example of a semantic conflict can be seen in the integration of parallel changes in a class such as Point, containing a method to calculate the distance to another Point: one developer in branch A changes the distance calculation (from euclidean to manhattan, for example), while in B another introduces a move() method, which uses the distance calculation as the value to move by. We find that upon a merge, while no errors are raised, the code exhibits altered behaviour from the original branches: specifically, the movement will behave differently. We sought to explore how the emerging field of Large Language Models can provide a breakthrough in our ability to test for the altered, introduced and lost behaviours arising from semantic conflicts. Specifically, we analysed the ability of ChatGPT to describe the conflict present in a merge, as well as its capability to generate appropriate unit tests that highlight the presence of semantic conflicts, with the description provided. We concluded that, in most cases, ChatGPT demonstrates capacity for both tasks when presented with low complexity code. Another notable obstacle is the association between "conflict" and active software fault, which leads to difficulty identifying simpler conflicts.

Supervisor: José Campos; **Second Supervisor:** Antónia Lopes.

Keywords: Large-language Models (LLMs). ChatGPT. Semantic Conflicts. Unit Test Generation.

ACM Computing Classification System:

- Software and its engineering → Software creation and management → Software verification and validation
- Computing Methodologies → Machine learning