Research on Bug-fix Distance and Development Robustness

Group 04

1. Data Acquisition

We are interested in the relationship between bug-fix distance and development robustness. To find the distance, we need data of commit with fix tags and the commit history of the function. To judge whether the subsystem development is robust, we need data including new files over time, new features over time and bug density.

We will not see much in Linux-stable as this is not where new features are actually developed. They are developed on the subsystems dedicated git tree and then during the merge window enter. Since we want to use the development speed of features, we will try and extract that from many subsystems at file level in Linux-next.

2. Data Preprocessing

3. EDA

4. Assumption and Hypothesis

4.1 Assumption

If there are few changes happened between introduction of a bug and the fix of it, then it would mean that the developers are good because they can write readable code

and/or understand the code they are working on, thus there would be a good development.

4.2 Hypothesis

A system whose bugs are fixed in closer commit has a more robust development.

5. Model Development

We assume that the development of a system is robust if new features and new files are added evenly over time, and the system has a low bug density. We will use regression analysis to find out the relationship between the distance and the robustness of the subsystem development.

6. Evaluation and Interpretation

7. Verification