TSE Second Progress Report

Changes

We don't have any changes in approach , novelty , dataset , comparison or value to user community

Research Questions

The research questions we are hoping to answer are:

- RQ1: How many unique patches are generated by the system and what effect does removing certain approaches have on the running of the system?
    - The number of patches by the system as well as an ablation study
- RQ2: How much time does the system take on average to produce a fix for a bug
    - An average of time taken per bug plotted against the various tools (bar chart/table)
- RQ3: How many of the patches generated can be counted as correct and have been correctly classified as such?
    - Will be looking into a subset of patches and making this evaluation based on manual effort
- RQ4: How much time does it take for a human user to go through incorrect patches?
    - Based on the manual effort in hours taken in RQ3 to run through the patches
- RQ5: How much time does the ranking mechanism save in terms of providing the user the best possible patch?
    - Based on a rough estimate from RQ4 and evaluating if the patch that was generated is the best one and if not what was the ranking of the best one for a subset of patches chosen in RQ3

Demo

For the demo we plan to give a brief overview of the approach and use an example to show how the system first finds a patch , ranks the patches and then finally generates a classification of correct/incorrect and gives examples of each stage of the tool.

Then we would go through the # of bugs fixed, time taken on average and the details of the other research questions

The github libraries that we are planning of using for now:

- https://github.com/SpoonLabs/astor
- https://github.com/yyxhdy/arja
- https://github.com/SerVal-DTF/TBar
- https://github.com/rjust/defects4j
- https://github.com/Ultimanecat/DefectRepairing