Debugging effectiveness summary

Alessandro Valentini

06.07.2015

Summary

Intesting points

- Hypothesis
- Definition and planning
- Results
- Statistical consideration

Hypothesis

- There is no difference in the **effectiveness** (number of correctly fixed faults) of debugging between autogen e manual test cases
- There is no difference in the **efficiency** (corrected task per time) of debugging between autogen e manual test cases

Definition and planning

The experiment:

- compares manual test cases to Randoop and EvoSuite test cases
- inspects the impact of obfuscated identifiers
- subjects are divided in two groups balaced accordingly experience and ability
- two different source code are used (XML-security and JTopas), each group has to debug a specific source using either manual or autogen tests.

Results I

- **MvR:** Efficiency and effectiveness are improved in particular on experienced developer (Eclipse Debugger), autogen test are easier to understand
- **MvE:** Effectiveness shows no significant differences, efficiency is increased by manual test cases. Understandability is lower than Randoop test but do not influence debug
- **MvO:** Obfuscation strongly decrease understandability of manual test cases, however this has no impact on debug

Results II

- Autogen test cases are simpler than manual and more useful for unesperienced developers
- Experienced developer perform well with both cases and can take advantage from advanced tools
- Unexperienced developer use more time to understand test cases decreasing their performance
- In general autogen test cases are useful because can be generated quickly
- In general debug should be demanded to experienced developers

Statistical consideration

- **Confusion Matrix:** express the level of reliability of an algorithm
- **Null hypothesis:** No relationship between two events
- **Likert scale:** Express feelings (level of agreement) respect to an object
- **Statistically significant:** confidence interval etc to review
- **Threats to validity:** no reward, order of tasks, few generator (free)