APR Tools Overfitting Evaluation Framework

Robin Zhang

Motivation

Fixing bugs is time consuming and complicated

APR provides solutions without human intervention

Overfitting causes false positive solutions

⇒ Need APR with less/no overfitting

Overview

Automatic Program Repair (APR):

Feed program through test suite, generate patches and validate

Overfitting:

Validated patches (pass all tests) do not fix the problem
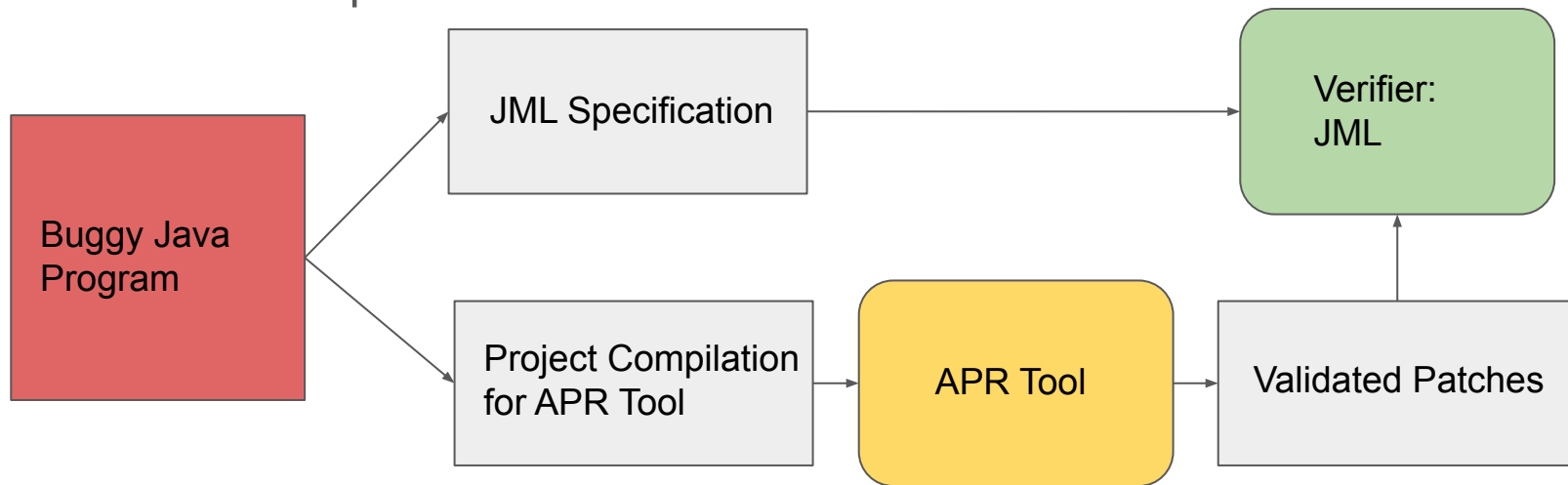
Formal Verification:

Express a program and requirements in mathematical language, and verify with completeness

Design

APR Tools: Nopol, jGenProg, SimFix, ARJA, Cardumen, jKali, and jMutRepair

Java Programs and Specifications: BuggyJavaJML

Verification Tool: OpenJML

Research Questions

1.   How efficient of the framework when assessing each APR tool? And how much time does it save comparing to the situation without using this framework?

2.   How does the program complexity affect the APR tools? How does it affect the performance of this framework?

3.   For what features or functions in a program most likely cause each APR tool overfitting problem?

Example Result

```
Program Name:Time
Test Number:74
Bug Number:58
APR Tools options:
*ARJA
*Kali-A
------
Tool no. 1:
ARJA
Validated Number:53
Verified Number:4
Overfitting Number:49
Overfitting Rate:
0.8448275862068966
Overfitting / Validated:
0.9245283018867925
Verification Rate:
0.06896551724137931
------
```

```
------
Tool no. 2:
Kali-A
Validated Number:48
Verified Number:2
Overfitting Number:46
Overfitting Rate:
0.7931034482758621
Overfitting / Validated:
0.9583333333333334
Verification Rate:
0.03448275862068966
------
Least Overfitted Tool:
ARJA
```

Example Result


```
Program Name:Time
Test Number:74
Bug Number:58
APR Tools options:
*ARJA
*Kali-A
------
Tool no. 1:
ARJA
Validated Number:53
Verified Number:4
Overfitting Number:49
Overfitting Rate:
0.8448275862068966
Overfitting / Validated:
0.9245283018867925
Verification Rate:
0.06896551724137931
------
```

```
------
Tool no. 2:
Kali-A
Validated Number:48
Verified Number:2
Overfitting Number:46
Overfitting Rate:
0.7931034482758621
Overfitting / Validated:
0.9583333333333334
Verification Rate:
0.034482758620689655
------
Least Overfitted Tool:
ARJA
```