Review 16

3/22/12

CS6V81.502

Seungtack Baek

**Perracotta: Mining Temporal API Rules**

**from Imperfect Traces**

In this Paper, Yang et al. studied how to infer rules from a program by running the program itself. In order to achieve this,

1. Instrument the code so it can be traced.
2. Ran the program with test inputs and obtain the execution traces.
3. From trace, infer the program rules.
4. Refine program rules with applied heuristics (refinement/optimization)

To evaluate this tool, they used this tool to three projects: Daisy, a toy file system, JBoss, an application server and Windows Kernel. As a result, they could obtain quite a few of interesting rules that were not properly documented – especially from Microsoft kernel (but, didn’t you say that Microsoft has strict rules for using their product?).

For their contributions, they claim that this study reveals the problem of imperfect trace, develops an inference algorithm based on both context-neutral and context-sensitive and introduce new heuristics on finding “interesting” properties. Yet, they did not explicitly discuss how serious problem imperfect trace is.

As they mentioned in the paper, they can only inference rules on *Alternating* properties, since it will be little more complex to build such tool to find various features. Still, they came up with clever way of winnowing rules that are likely to be interesting. Their heuristics use reachability information and naming convention.

I do not heavily agree with Naming convention – even if I do use conventions, ‘similar’ names might not come out to be too similar in their system. For an example, I could have two variables “ldapSecurityProvider” and “ldapInitialContext”. Although they might be related to each other, this tool can rule it out to be “not enough similarity”. However, I do think their making use of reachability to determine interesting rules. It is nice to see how two components without call relationship could infer a rule.

One mistake I noticed was, they forgot to label figure 4 (or did they do that on purpose?) I had hard time trying to see what the figure was trying to convey.

**Question:**

1. As with AMNESIA, this tool also takes “hybrid” approach – both static and dynamic. Would it be better if this tool was all dynamic?