More Course Overview:
Models, Tests, Bugs, and Symbols

But first
• Homework 1 and 2
• Yi and Jingbo’s lectures
• What’s coming up
• Idea proposal assignment
• Overview of the final topics in this class

Homework 1
• Posted on class website
• Due Tuesday March 2, 9 AM on gradescope
  – Everyone in the class should have gotten a notification from gradescope about being added to the class. If you didn’t get it, let me know!

Homework 2
• A little different.
• In-person, 2.5-hour session.
• Sign up for a slot soon.
• Slots will take place between March 6 and March 28
• More info soon!

Yi Ding
A Holistic View on Machine Learning for Systems

Minimize Energy under Performance Constraints
Tuesday, 2/21, noon in CS151

Trustworthy Software Enabled by Program Analysis and Synthesis

Jingbo Wang

Coming Up: 3 guest lectures and a day off

- Tuesday, February 28: guest lecture
- Thursday, March 2: Use class slot to discuss project ideas and form teams. No lecture.
- Tuesday, March 7: guest lecture
- Thursday, March 9: guest lecture
- ... no more guest lectures after!

Tuesday, 2/28, noon CS 151

From Barriers to Bridges: Designing Processes and Tools for Inclusive Open-Source Communities.

Idea Proposal Assignment
On to semester overview

Static analysis
- Using the source code to improve a program
- Manual code reviews and inspections
- Automatic inference of properties

Improve the software quality

Dynamic analysis
- Using the program executions to improve the program
- Manual with debuggers, etc.
- Automatic inference over logged behavior
- Does not need source code or even binaries

Improve the software quality

Areas we will cover in this course
- Static analysis
- Dynamic analysis
- Model checking
- Mutation testing
- Bug localization
- Symbolic execution

areas for your projects

As we go over each topic...
- Think whether this sounds interesting
- Think about what kind of a tool you could make that uses this
- You are all programmers: think about things you’ve done while programming that were hard, and how these kinds of analysis might make it easier

Model checking
- I actually meant:
  - Model checking
  - Model inference
  - Model simulation
Model inference

**problem:**
I have a system (or a log of executions).
I want a small, descriptive model of what the system does.

Model can be used to understand the system, debug, detect anomalies, document.

Logs are hard to read

- First, parse out the executions
  - check-out → valid-coupon → check-out → reduce-price → get-credit-card
  - check-out → invalid-coupon → check-out → reduce-price → get-credit-card
  - check-out → get-credit-card
  - ...hard to understand

So what’s the magic?

- Lots of ways to do it:
  - Try merging the executions into a small model
  - Mine properties then build a model from the properties alone
  - Use static or dynamic analysis to determine what events can legally take place after others

K-Tails

- let’s use k=1 as an example
- merge two states if their name is the same

- (k=2 means merge two states if their name, and all the states to which they have transitions are “the same”)
- and so on for larger k
Model checking
• Given a property and a model, check if the model satisfies that property
  • Reduce-price always followed by get-credit-card?

Model simulation
• Given a model, you can generate new executions that have not been observed before!

Mutation testing
• Evaluate the tests
  – not the program!
  – not a type of testing!
  – does not improve a program directly; improves tests!

Mutation
• Take a program
• Create a mutant with one or a few small changes:
  – change a + to a –
  – add/subtract 1 somewhere
  – increment/decrement a loop counter
  – delete a line
  – insert a line from one place in another
• Repeat to create many mutants

Why create mutants?
• Suppose you have a program and a test suite
• All the tests pass
• What does that mean about your program?
  1. Program is correct
  2. Tests only test parts of the program that are correct and the rest, who knows
  3. Tests and program may be written by the same person, using the same implicit assumptions

Let’s write some tests
// returns the factorial of the input n
int factorial (int n) {
  if (n <= 0)
    return 1;
  if (n == 1)
    return 1;
  else
    return n * factorial(n-1);
}
OK, so how do we test the tests?

• Run the tests on the main program
• Run the tests on the mutants — what if the tests pass?

Mutation testing evaluates the tests

• If a test “kills a mutant” then that’s a good test
• If some mutants aren’t killed, the test suite is lacking
• Solution: write more tests!

• Is it OK to write more tests until all mutants are killed and then stop?

Consider this mutant

// returns the factorial of the input n
int factorial(int n) {
    if (n <= 2)
        return 1;
    if (n == 1)
        return 1;
    else
        return n * factorial(n-1);
}

Consider this mutant

// returns the factorial of the input n
int factorial(int n) {
    if (n == 0)
        return 1;
    if (n == 1)
        return 1;
    else
        return n * factorial(n+1);
}

Consider this mutant

// returns the factorial of the input n
int factorial(int n) {
    if (n == 0)
        return 1;
    if (n == 1)
        return 1;
    else
        return n * factorial(n-1);
}

Bug localization

• Narrowing down the most likely place to contain a bug
Failure-inducing input

- This HTML input makes Mozilla crash (segmentation fault).
- Which portion is the failure-inducing one?

Delta Debugging: Try half the input

- Will the program still crash?

Minimizing via binary search

- 57 test to simplify the 896 line HTML input to the "<SELECT>" tag that causes the crash
- Each character is relevant (as shown from line 20 to 26)
- Only removes deltas from the failing test

Impact analysis

- Run the code on passing test cases
- Run the code on failing test cases
- Keep track of which lines execute
- Lines that executes only on passing test cases are OK. So are lines that execute on both.
- Lines that only execute on failing test cases are suspicious.

What else can you do to localize a bug?

- Regressions: suppose a test used to pass and now fails.
  - consider the latest changes
  - do delta debugging on the changes

Can we automatically fix bugs?

- Take a program that passes most test cases and fails one or two, and tweak it
  - write (tweak) a very similar program (with minimal change) that passes all the tests
  [see Weimer et al., Automatically Finding Patches Using Genetic Programming, ICSE 2009]
Symbolic execution

- “Think” about the code, rather than execute it, but execute it anyway. But don’t use numbers. Just think about the numbers.

- Clear, right?

Why symbolic execution?

- A different way to reasoning about the code
- Can determine what parts are reachable and under what conditions
- Can be compared to developers’ expectations about those conditions
- Can be used to document
  - For example, “this method can only be called if x>0” or “this method throws an exception is pts == null”