Grading programming assignments using Big Data techniques
My research

Programming Languages

Software Engineering

Big Data
Roadmap

1. Problem we want to solve
2. Unraveling keywords
3. Our research
4. Conclusions
Assignments

Assignment → Solution → Grading

Solution → Feedback
CS/BS enrollment in North America
CS1 at RIT

One instructor

50 students
Assignments in CS1@RIT

• Ten assignments in total, 50 submissions to grade every week.
• In total, 500 submissions.
• Instructors usually have additional 2-3 sections.
• Students demand personalized feedback as soon as possible.
Are you feeling like this?
## CS1 enrollments in Fall’15

<table>
<thead>
<tr>
<th>University</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rochester Inst. of Tech.</td>
<td>900</td>
</tr>
<tr>
<td>Univ. of Toronto</td>
<td>1,200</td>
</tr>
<tr>
<td>UC Berkeley</td>
<td>1,400</td>
</tr>
<tr>
<td>Univ. of Michigan</td>
<td>1,600</td>
</tr>
<tr>
<td>Univ. of Washington</td>
<td>2,800</td>
</tr>
</tbody>
</table>
How are you feeling now?
Massive Online Open Courses (MOOCs)
Some numbers

• The Introduction to Computer Science and Programming MOOC offered by MIT reported 35,000+ submissions to 8 assignments in Fall’12.

• The Stanford’s Machine Learning MOOC reported 1,000,000+ submissions to 42 assignments in Fall’11.
How are you feeling now?
Wouldn’t it be nice...?

Assessing (Instructors)

Feedback (Students)
Roadmap

1. Problem we want to solve
2. Unraveling keywords
3. Our research
4. Conclusions
MOOCs

**Massive**
Classes may consist of up to 100,000+ students

**Online**
The course is taken completely online

**Open**
Registration is open to anyone around the world

**Course**
The course is similar to college courses
Database

Organizing data

Efficient retrieval
# Relational databases

<table>
<thead>
<tr>
<th>Patient</th>
<th>ssn</th>
<th>firstName</th>
<th>middleName</th>
<th>lastName</th>
<th>primaryDoctor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>235-14-7854</td>
<td>Sandra</td>
<td>null</td>
<td>Smith</td>
<td>943-23-9874</td>
</tr>
<tr>
<td></td>
<td>192-48-0924</td>
<td>John</td>
<td>Richard</td>
<td>Moore</td>
<td>862-74-3611</td>
</tr>
</tbody>
</table>
Big Data
Finding subgraphs
So, how is that related to assignments?
Roadmap

1. Problem we want to solve
2. Unraveling keywords
3. Our research
4. Conclusions
Looking for patterns in the code
void foo(int[] a) {
    int o = 0;
    for (int i = 0; i <= a.length; i++)
        if (i % 2 == 1)
            o += a[i];
    System.out.println("My result:" + o);
}
Patterns as subgraphs

- \( u_0, \text{Assign} \)
  - \( c = 1 \mid \text{Digit} \)

- \( u_1, \text{Cond} \)
  - _

- \( u_2, \text{Cond} \)
  - _

- \( u_3, \text{Assign} \)
  - \( c \ast= _ \)
void foo(int[] a) {
    int o = 0;
    for (int i = 0; i <= a.length; i++)
        if (i % 2 == 1)
            o += a[i];
    System.out.println("My result:"+o);
}

• You are cumulatively adding variable \( o \) correctly.
• You should not go beyond \( a.length \).
• You are printing \( o \) to console correctly.
Roadmap

1. Problem we want to solve
2. Unraveling keywords
3. Our research
4. Conclusions
Using subgraphs to detect patterns
Would you like to help?

CHECK THIS OUT!

https://crivero.gitlab.io/personalized-feedback/