Visual Analytics for
Personalized Programming Learning

Dr. Sharon, I-Han Hsiao
School of Computing, Informatics & Decision Systems Engineering
Arizona State University

Sharon.Hsiao@asu.edu
http://www.public.asu.edu/~ihsiao1/
@shoha99
CSI (Computing Systems & Informatics) Lab

https://sites.google.com/a/asu.edu/csi/

• LAK, AIED/EDM, SIGCSE, CSCL, ICALT, ECTEL
• IJAIED in CS associate editor
• 2017 AIED, EDM, ECTEL publicity chair
Experience the Joy of Programming
GIRL...

I LOVE CODING TOO
Visual Analytics & Educational Technology in Computing Education

1. Problem Solving for Personalized Learning
   - Open Student Modeling (+adaptive navigation support)
   - Distributed practices (AIED’17...stay tuned)
   - Semantic topic modeling + visualization (LAK’15 & ’16)

2. Behavior/Learning Activity Modeling
   - Programming information seeking (FIE’15, EDM’16, ICALT’16)

3. Feedback Analytics
   - Augmented automated assessments (ECTEL’16, SIGCSE’17)
   - Reviewing & reflecting behavioral analytics (LAK’17)
Data-Driven Engineering to Improve Computing Education
Programming Learning Literature

• **Technology:**
  - ITS (Barnes et al., 2008; Boyer et al., 2011);
  - **Adaptive technology** (Brusilovsky et al., 2009; Hsiao et al., 2010; Yudelson et al., 2005; Sosnovsky & Brusilovsky, 2015);
  - Social computing (Denny et al., 2011; Hsiao & Brusilovsky, 2011);
  - Automated services (SIGCSE) ...

• **Pedagogy:**
  - Peer instruction (Porter et al., 2013);
  - Media computation (Simon et al., 2010; Guzdial et al., 2013);
  - Flipped classroom (Sarawagi, 2013);
  - Classroom orchestration (Martinez-Maldonado et al., 2013; Slotta, et al, 2013) ...

• **Community:**
  - Software engineering (Ko. et al., 2006; Stylos & Myers, 2006; Brandt, 2010);
  - Online communities (Glassman et al., 2015; Rivers & Koedinger, 2015; Hsiao & Naveed, 2015)...

  **Learning to program is NOT easy!**
Figure 1. The collaboration management cycle, showing points at which the responsibility for analyzing and guiding the interaction might shift from the collaborators to the system.

Questions that we’re trying to answer:

• How much do novices learn programming in *Just-In-Time* learning paradigm?

• Will integrate *formal* and *informal* learning analytics bridge the gap between cyber and physical programming classes?

• How to design *next generation* of automated assessment tools to assist computing education? (i.e. examples in math)
Research-in-progress in “orchestration technology”

- Dillenbourg: smart classroom
- Roberto Martinez-Maldonado: MTClassroom
- Kurt VanLehn: FACT project
- ...

11
The reality is...

• Majority of lower division CS courses are still Blended!
  – face-to-face instruction in the classrooms supported by online tools

• Weak links between assessment and instruction!
#1. Strengthen the link between Assessment and Instruction
Make students improve from feedback and keep working!

• The Power of Feedback: feed up, feed back, and feed forward (Hattie & Timperley, 2007)
  – To reduce discrepancies between current understandings and performance and a goal.

• Success from ITS:
  – Inner loop feedback: next step
  – Outer loop feedback: problem selection (VanLehn, 2006)
Today’s blended programming classes:

• Formal:
  – Automated submission assignments
  – Paper-based exams

• Informal
  – Optional readings, tools, exercises...
Problems & Goals:

• Problems:
  – *how do students receive partial credits?*
  – *was it a single concept or multiple concepts mistake?*
  – *a careless mistake or a long-term misconception?*
  – *Limited class time (no personalized feedback)*
  – *Graders’ training or inconsistency issue*
  – ...

• Goals: using **visual learning analytics** to capture & reinforce
  – identification of strength and weakness
  – characterization of the nature of their errors or any recurring patterns if any
  – assessment of appropriateness of their study strategies and preparation
Studies & Results

• 3 Lab studies
  – 20 graders, sample 30 exams

• 1200 student users in 5 classroom studies w/ Augmented Grading Tools
Graders can be VERY inconsistent!
1. Write the following for loop as a **while** loop: (5pt)

```java
for(int i = 0; i <= 10; i++)
{
    System.out.println(i);
}
```

// full score must be exactly the same codes
// as the correct answer (except variable names)
int i = 0;  // -1 incorrect initialization
while (i<=10){  // -3 incorrect while statement
    System.out.println(i);
    i++;  // -1 incorrect increment statement
}
// -0.5 other errors

2. Instantiate an ArrayList that contains decimal numbers and assign it to an appropriate variable. Write an **enhanced for loop** (for-each loop) that iterates through your ArrayList of decimal numbers and displays their sum to the console: (7pt)

```java
ArrayList<Double> numList = new ArrayList<Double>();
double sum = 0;
```

// -2 incorrect ArrayList

```java
for(Double d: numList){
    sum += d;
}
System.out.println(sum);  // -0.5 other errors
```

---

15. Instantiate an ArrayList that contains decimal numbers and assign it to an appropriate variable. Write an enhanced for loop (for-each loop) that iterates through your ArrayList of decimal numbers and displays their sum to the console: (10pt)

```java
double sum = 0;
for (Double element: decimalNumbers)
    sum += element;
System.out.println(sum);
```
Feedback Type Distribution

<table>
<thead>
<tr>
<th>Feedback Type</th>
<th>Easy</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue</td>
<td>52.93%</td>
<td>56.32%</td>
</tr>
<tr>
<td>Diagnostictic</td>
<td>11.06%</td>
<td>15.19%</td>
</tr>
<tr>
<td>Corrective</td>
<td>9.06%</td>
<td>27.92%</td>
</tr>
<tr>
<td>Reinforcement</td>
<td>3.70%</td>
<td>3.01%</td>
</tr>
<tr>
<td>None</td>
<td>18.17%</td>
<td>0.38%</td>
</tr>
<tr>
<td>Motivational</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Negative</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

SIGCSE’17
Approaches to address above issues:

• Consistent *feedback* to students
• Grader support
Augmented Grading Tools: PGA & WPGA

• (Web-based) Programming Grading Assistant
2. Instantiate an `ArrayList` that contains decimal numbers and assign it to an appropriately named variable. Write an enhanced `for` loop (for-each loop) that iterates over this list of decimal numbers and displays their sum to the console. (7pts)

```java
ArrayList num = new ArrayList();
for (double d : num)
{
    // Your code here
}

ArrayList<Double> num = new ArrayList<Double>();
double sum = 0;
for (Double d : num)
{
    sum += d;
}
```

5.25 3
Auto-calculate grades based on semantic partial credit algorithm

- **Concept Similarity**: reward conceptual soundness
- **Concept Saliency**: highlight key concept and demote peripheral ones
- **Miscellaneous**: non-conceptual mistakes, i.e. careless

(ECTEL’16, TLT under review)
Innovative educational technology supports grading paper-based programming exams

- Automatic semantic partial credit approach combines grading schemes and conceptual feedback
- Improves grading consistency and coherence
- Default provides conceptual feedback
- Grading efficiency?
- How do students receive the feedback and impacts?
Cross Physical & Digital Learning Space Orchestration Technology Stakeholders

• **School**
  – Digitize paper-based formal assessments.
  – Enable systematic & longitudinal comprehensive learning analytics data mining.

• **Teacher**
  – Manage grading progress, including grader/exam/question assignment, grader synchronization/communication, etc.
  – Retrieve student's assessment item via searchable interface.
  – Meaningful analytics on student's performance.

• **Grader**
  – Leverage grading consistency (by grading the same question at once; grading scheme and feedback are tied and preconfigured)
  – Comments can be reused.
  – Grading progress is visible.
  – Fast grading turnaround rate.

• **Student**
  – Immediate and organized fashion on the received feedback and graded items.
  – Accessibility to graded assessment anywhere anytime.
  – Reflective prompts to facilitate review & learning.
  – Navigation support to guide students focus on attention-needed exams and questions.
  – Learning analytics to support self-regulated learning.
Instructor/grader UI

Student UI
Review & Reflect

- **Review**: exam, correct/incorrect question, filter
- **Reflect**: Keep notes; Bookmark question; Tick a checkbox to acknowledge.

Assumption:
R(E)->R(C)->R(C)->R(I)->Reflect
R(E)->R(I)->Reflect
$\pi$: start state
$E$: exam
$C$: correct answer
$I$: incorrect answer
$R$: reflection
$F$: filter
Findings (LAK’17)

• Students review quizzes regardless of if they count towards academic performances or not.
• Hard working reviewers perform better than the others.
• Both A & B students perform on overview first, detail on demand.
• A students review & reflect strategically: they strive to get the wrong right.
• B students review persistently, but fail to engage in deeper reflection.
Educational Implications

Lack-of-deep-reflection behaviors can be dangerous.
Strengthen the link between Assessment and Instruction:

Capture and push for meaningful “reflection”.
Quiz of the Day (QuizIT)
Learning Sciences Principles

• **Distributed Retrieval Practice** (Benjamin & Tullis, 2010; Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Rohrer, 2015)

• **Reflection** (Bjork, Dunlosky, & Kornell, 2013; Butler & Winne, 1995; Hacker et al., 1998; Schraw, Crippen, & Hartley, 2006; Zimmerman & Schunk, 2012)

• **Peer Interaction and Feedback** (Butler & Winne, 1995; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Shute, 2008; Falchikov & Goldfinch, 2000; Patchan & Schunn, 2015; Patchan, Schunn, & Correnti, 2016; Topping, 1998)
#2. Semantic topic modeling + visualization
Theoretical Background

- Accountable Talk (Resnick, L.)
- ICAP (Interactive, Constructive, Active, Passive) learning activity framework (Chi, M.T.H.)

Vialogues in EdLab@Columbia Univ.

- Computational Modeling
- Social Communicative Factor
- Visualization
- Informal Learning Activity
How to facilitate novices learn programming informally?
StackOverflow 3 years data: Topic: Java
What are the students reading? How do those content contribute learning?

(c) Behavior Modeling

{ Reading, Searching, Locating, Evaluating, ...

Procedural / Declarative Knowledge

(b) Programming Knowledge Mapping

Syntactic
Semantics
Pragmatics

(a) Semantic Content Modeling

Posts’ Semantics

Topic Facet Modeling
a. **Label “constructive” content** *(CSCL’15, ICALT’15, FIE’15)*

   – *inferring, creating, integrating, elaborating, comparing, contrasting, analogizing, generalizing, including, reflecting on conditions, explaining why something works...*

   – **MPQA (Multiple Perspective Question Answering) opinion corpus developed by U Pitt**
     
     *Emphasis: “...this is why...”*
     
     *Causation: “...as a result...”*
     
     *Generalization: “...everything...”*
     
     *Conditionals: “...it would be...”*

   – **Positive correlation between user favorites and the constructiveness in Accepted Answers (r= 0.0781, p< 0.01)**

   – **User bookmarked and up voted more in Answers when the content is more constructive.**

b. **Label content associated concepts**

c. **Collect students’ learning activities**
a. Label “constructive” content

b. Label content associated concepts (LAK’15,16, Computers in Human Interaction: Technology Behaviors in Education Innovation)

- 2 teachers & MTurk
  - 103/149/406 MTurkers passed & completed the qualifier survey ($0.5 dollars for 2.5 weeks)
    - Each of question in the corpus is indexed by at least 2 to 5 experts from the crowd
  - Semi-auto indexing interfaces: Natural + Programming language parser (topic facet modeling)
    - Can extract sig. more and diverse concepts.

c. Collect students’ learning activities
EduAnalysis: Visual Learning Analytics for both teachers & students

- Supports indexing, authoring, delivery

**EXAM1**

1. What is the final value of sum displayed to the console:
   ```java
   for(int i = 0; i < 5; i++)
   {
       int sum = 0;
       sum = sum + i;
       System.out.println(sum);
   }
   ```
   a. 5
   b. 10
   c. 15
   d. Compiler error
   e. None of the above

   Answer: e

   Marks: 5 / 8

   (separate multiple answers with |)
• Persistent traces of learning analytics in semantic level.

• Conceptual feedback for students.

• Semantic Partial Credit Assignment
  – “award the logic soundness instead of code completeness”
1. Label “constructive” content
2. Label content associated concepts
3. Collect students’ learning activities (EDM’16)
   – Novices are not developers!
   – The more they read, the more they learned (r=0.418, p<0.01)
Search on StackOverflow

**Personalized Information Seeking Assistant**

Tag: Words

2015/9/26

**P.I.S.A.**

javascript remove multiple whitespaces

---

Q: Remove multiple (but keep single) whitespaces
   I need to remove multiple whitespaces in a javascript string entirely but keep single ones the way they are. I saw string replace(/s+/g, "") and this worked as expected but it removed all whitespaces.

-3 votes
1 answer

---

Q: Find word with certain character in javascript and replace whole ...
   http://stackoverflow.com/questions/4910272/replace-a-b-in-string
   I want to search for the char $ in a string. However, this char can be joined to multiple things, even whitespaces. Few examples " $ " "Sanother" "San other" or $more "What I want to do is remo...

0 votes
2 answers

---

Q: How to compress PHP output before sending it to browser without J... 
   http://stackoverflow.com/questions/1611626/how-to-compress...
   I have a function that compress the php output but it gives me problems with the inline javascript. I found a page with related topic but the sample from there it's not work: http://jeromejaglaie.co...

1 vote
0 answers

---

Q: Convert a input string to GET variable
   http://stackoverflow.com/questions/10525337/encoded-e-args
   I'm trying to create an ajax search page, which shows live results similar to google. Right now I am struggling with how to convert a search query such as small dog to a GET parameter, in this case ...

0 votes
3 answers

---
Summary & Future On-going Work

• Data-driven engineering to strengthen learning between formal/informal programming assessments & instructions
  • Semantic indexing methods (to label programming-related content to programming topic)
  • Practical solutions (minimum intrusion for class orchestration) to facilitate (large-scaled) blended programming classes

• Conduct more exhaustive studies to examine effects on learning
  • How to provide creative way to excite students in using feedback?
  • How do students prepare exams? cheat sheets? How can I help?
  • How does distributed (bite-size) practice strategy help in programming learning?
#CS4All is real.
It is happening.
We take it very seriously.
We do not take today's programming classes for granted.

Thank you, Questions?
Sharon.Hsiao@asu.edu
http://www.public.asu.edu/~ihsiao1/
EDM 2017

VENUE
Crimson Resort and Spa, Cebu, Philippines
June 23 — 25, 2017

CALL FOR PAPERS
The EDM conference is a leading international forum for high-quality research that mines large data sets in order to answer educational research questions that shed light on the learning process. These data sets may come from the traces that students leave when they interact, either individually or collaboratively, with learning management systems, interactive learning environments, intelligent tutoring systems, educational games or when they participate in a data-rich learning context. The types of data therefore range from raw log files to eye-tracking devices and other sensor data.

AIED 2017

VENUE
Crimson Resort and Spa, Cebu, Philippines
June 26 — 30, 2017

CALL FOR PAPERS
The 18th International Conference on Artificial Intelligence in Education (AIED 2017) is the next in a longstanding series of biennial international conferences for high quality research in intelligent systems and cognitive science for educational computing applications. Empirical and theoretical papers are solicited for submission.