Discovering and removing barriers to learning

Ken Koedinger
Human-Computer Interaction & Psychology
Director of LearnLab
Carnegie Mellon University

NIPS Workshop on Data Driven Education
Dec 10, 2013
**General Argument**

- Most of what we learn is outside our conscious awareness
- Thus, instructor intuition produces *flawed course materials*
- Instead: *Use data-driven models of learners to design instruction*
  - Machine discoveries predict data better than human-built models
- Instructional experiments demonstrate improved learning
Cognitive Tutors: Interactive Support for Learning by Doing

My current cell phone company charges me $14.95 per month for service and $.13 per minute. PPS Cellular Phone Company has offered me $15.00 worth of free calls a month if I switch, but the charge is $.39 per minute.

4. After how many minutes of calls will the cost for both companies be the same?

Authentic problems  Feedback within complex solutions  Progress…

Challenging questions  Personalized instruction  … individualization
Cognitive model drives instruction

- **Cognitive Model**: A system that can solve problems in the various ways students can.

  \[ 3(2x - 5) = 9 \]

  - If goal is solve \( a(bx+c) = d \)
    - Then rewrite as \( abx + ac = d \)

  \[ 6x - 15 = 9 \]

  - If goal is solve \( a(bx+c) = d \)
    - Then rewrite as \( abx + c = d \)

  \[ 2x - 5 = 3 \]

  - If goal is solve \( a(bx+c) = d \)
    - Then rewrite as \( bx+c = d/a \)

  \[ 6x - 5 = 9 \]

- **Model Tracing**: Follows student through their individual approach to a problem -> context-sensitive instruction
Cognitive model drives instruction

- **Cognitive Model**: A system that can solve problems in the various ways students can

  
  \[
  3(2x - 5) = 9 \\
  6x - 15 = 9 \quad \text{Known? = 85% chance} \\
  2x - 5 = 3 \quad \text{Known? = 45%}
  \]

  - If goal is solve \( a(bx+c) = d \)
    - Then rewrite as \( abx + ac = d \)
    - Hint message: “Distribute \( a \) across the parentheses.”
  - If goal is solve \( a(bx+c) = d \)
    - Then rewrite as \( abx + c = d \)
    - Bug message: “You need to multiply \( c \) by \( a \) also.”

- **Model Tracing**: Follows student through their individual approach to a problem -> context-sensitive instruction

- **Knowledge Tracing**: Assesses student's knowledge growth -> individualized activity selection and pacing

  - Skillometer:
    - Calculate input value.
    - Writing expression, any form.
    - Set axis bounds.
You’ve had lots of experience with the English language.

You might say you know English.

But, do you know what you know?
Do educators know what they know?

Which is harder for algebra students?

*Story Problem*
As a waiter, Ted gets $6 per hour. One night he made $66 in tips and earned a total of $81.90. How many hours did Ted work?

*Word Problem*
Starting with some number, if I multiply it by 6 and then add 66, I get 81.90. What number did I start with?

*Equation*
x * 6 + 66 = 81.90

Math educators say: story or word is hardest

Students: equations are hardest

![Bar chart showing percent correct for different problem representations: Story 70%, Word 61%, Equation 42%]

Expert blind spot!
Algebra teachers, especially, incorrectly think equations are easy
Experts can describe <30% of what they know! (Clark et al)

What we know about our own learning

What we do not know we know

You can’t design well for what you don’t know!
Repeat the mantra with me:

“I don’t know what I know!”

“I don’t know what I know!”

“I don’t know what I know!”
Real World Impact of Data-Driven Cognitive Science

*Algebra Cognitive Tutor*

- Widespread intensive use
  ~600K students per year
  ~80 minutes per week

- Enhances student learning

- Still:
  Too many decisions driven by intuition


An Infrastructure for Course-Based Data Collection & Improvement

Ed tech + wide use = “Basic research at scale”

Since 2004
> 450 ed tech data sets in DataShop
> 320 in vivo experiments


Data-driven learner models of cognition

- LearnLab collaborators have also modeled
  - Metacognition (Aleven, Roll, ...)
  - Motivation (Baker, Nokes-Malach, ...)
  - Student discussions (Rose, Resnick, ...)

- Focus today: Discovering better *models* of what is hard for students to learn
Closing the data-improvement loop

• Performance on many steps in many problems is logged & labeled ...
Log data: Grade student steps & apply *label-based cognitive model*

<table>
<thead>
<tr>
<th>Student</th>
<th>Step (Item)</th>
<th>Textbook KCs</th>
<th>Opportunity</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>prob1step1</td>
<td>Circle-area</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>prob2step1</td>
<td>Circle-area</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>prob2step2</td>
<td>Rectangle-area</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>prob2step3</td>
<td>Compose-by-addition</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>prob3step1</td>
<td>Circle-area</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Steps within problems that are assessed

Map of steps to knowledge comp’s (Q-matrix)

Opportunities student has had to learn KC

Was student’s first attempt on this step a correct one?

<table>
<thead>
<tr>
<th>Single skill</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometry</td>
<td>1</td>
</tr>
<tr>
<td>Geometry</td>
<td>2</td>
</tr>
<tr>
<td>Geometry</td>
<td>3</td>
</tr>
<tr>
<td>Geometry</td>
<td>4</td>
</tr>
<tr>
<td>Geometry</td>
<td>5</td>
</tr>
</tbody>
</table>

Different KC hypothesis changes prediction of student success
Data mining using DataShop’s learning curve analytics

Without decomposition, using just a single “Geometry” KC, no smooth learning curve.

But with decomposition, 12 KCs for area concepts, a smoother learning curve.
Visualizing learning curves to find opportunities for improvement

Low, long curve => remove busy work
High rough curve => concept/skill is more complex

Cen et al (2007)
Statistical version of cognitive model

\[
\log \frac{p_{ij}}{1 - p_{ij}} = \theta_i + \sum_k \beta_k Q_{kj} + \sum_k Q_{kj} (\gamma_k T_{ik})
\]

GIVEN:
- \(p_{ij}\) = probability student i gets step j correct
- \(Q_{kj}\) = each knowledge component k needed for this step j
- \(T_{ik}\) = opportunities student i has had to practice k

ESTIMATED:
- \(\theta_i\) = proficiency of student i
- \(\beta_k\) = difficulty of KC k
- \(\gamma_k\) = gain for each practice opportunity on KC k

Cen, Koedinger, & Junker (2006)
Spada & McGaw (1985)
Machine learning to improve cognitive models, lower prediction error (RMSE)

Learning Factors Analysis algorithm uses **beam search**
- feature generation ops
- stat model as evaluation function

Like a genetic algorithm


<table>
<thead>
<tr>
<th>Dataset</th>
<th>Content area</th>
<th>RMSE Orig in-use</th>
<th>RMSE Best-hand</th>
<th>RMSE Best-LFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometry 9697</td>
<td>Geometry area</td>
<td>0.4129</td>
<td>0.4033</td>
<td>0.4011</td>
</tr>
<tr>
<td>Hampton 0506</td>
<td>Geometry area</td>
<td>NA</td>
<td>0.4022</td>
<td>0.4012</td>
</tr>
<tr>
<td>Cog Discovery</td>
<td>Geometry area</td>
<td>NA</td>
<td>0.3250</td>
<td>0.3244</td>
</tr>
<tr>
<td>DFA-318</td>
<td>Story problems</td>
<td>0.4461</td>
<td>0.4407</td>
<td>0.4405</td>
</tr>
<tr>
<td>DFA-318-main</td>
<td>Story problems</td>
<td>0.4376</td>
<td>0.4287</td>
<td>0.4266</td>
</tr>
<tr>
<td>Digital game</td>
<td>Fractions</td>
<td>0.4442</td>
<td>0.4396</td>
<td>0.4346</td>
</tr>
<tr>
<td>Self-explanation</td>
<td>Equation solving</td>
<td>NA</td>
<td>0.4014</td>
<td>0.3927</td>
</tr>
<tr>
<td>IWT 1</td>
<td>English articles</td>
<td>0.4262</td>
<td>0.4110</td>
<td>0.4068</td>
</tr>
<tr>
<td>IWT 2</td>
<td>English articles</td>
<td>0.3854</td>
<td>0.3854*</td>
<td>0.3806</td>
</tr>
<tr>
<td>IWT 3</td>
<td>English articles</td>
<td>0.3970</td>
<td>0.3965</td>
<td>0.3903</td>
</tr>
<tr>
<td>Statistics-Fall09</td>
<td>Statistics</td>
<td>0.3648</td>
<td>0.3527</td>
<td>0.3353</td>
</tr>
</tbody>
</table>

Variety of domains & technologies

11 of 11 improved models
Where do meaningful task features come from?

- Combine psychology & domain expertise

Stick with me here ... particularly if you’re in CS or stats
Say it with with me:
“I don’t know what I know”!
Same math but different difficulty!?

To make metal cans, the ends of the cans are stamped out of square pieces of metal. The part of the square that is left over is then recycled as scrap. The side length of each square piece of aluminum is 5.6 centimeters. The diameter of the can is equal to the side length of the square piece of aluminum.

Use 3.14 for π.

1. What is the area of the scrap metal?

<table>
<thead>
<tr>
<th>Diagram Label</th>
<th>Side of the metal square (cm)</th>
<th>Area of the metal square (cm²)</th>
<th>Radius of the bottom of the can (cm)</th>
<th>Diameter of the bottom of the can (cm)</th>
<th>Area of the bottom of the can (cm²)</th>
<th>Area of the metal circle (cm²)</th>
<th>Area of the Scrap Metal (cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>2.8</td>
<td>31.36</td>
<td>CA</td>
<td>CN</td>
<td>24.6176</td>
<td>6.7424</td>
<td></td>
</tr>
</tbody>
</table>
Answer: Different planning needs

Plan to combine areas

Subtract

Combine Areas

Scaffolding
Use data-driven model to redesign tutor

1. Change adaptive prob selection
   - New bars for discovered skills
   - Adjust optimization parameters

2. Sequence for gentle slope
   - Simple problems first

3. Create new problems to focus on planning KCs
   - Next slide ..
New problem type to *focus instruction* on greatest need

- Practice planning step *only*
In vivo experiment: Random assignment A/B test embedded in ed tech used in schools

- Compare new vs. old tutors part of normal classroom use
- 103 high school students
- Measures: Pre-test, post-test, time to mastery
Model-based instruction is a better student experience

- More efficient: 25% less time
- And better learning of planning skills

**Instructional time (minutes) by step type**

- Control: Original tutor
- Treatment: Model-based redesign

**Post-test % correct by item type**

- Control: Original tutor
- Treatment: Model-based redesign
Where do meaningful task features come from?

- Past successes combine psychology & domain expertise
- Matrix factorization, deep belief networks do not work
  - Better predictions, no actionable insight
- Crowdsource task labeling
  - What makes tasks A & B harder than C & D
- Model actual student learning!
SimStudent learns like students do
- Representations from co-occurrence frequency via grammar induction
- Skills from imitation & feedback via hybrid of logic induction mechanisms
SimStudent Discovers Hidden Skills

- Tutor SimStudent to create cognitive models
- Evaluate model
  - Productions => features
  - Predict learning curves
- Result: More accurate cognitive models
- Interpretable & actionable
SimStudent discovered this tacit student difficulty in learning the algebraic representation!

<table>
<thead>
<tr>
<th>Step</th>
<th>2x=12</th>
<th>6=3x</th>
<th>-x=5</th>
<th>-24=-4x</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data to model to better education

• You don’t know what you know!
  – Intuitive design is not reliable

• Need data-driven *explanatory models* of learners
  – SimStudent potential: create intelligent tutors by teaching, pedagogically more effective than human-built tutors

• Improve science & practice of learning
Thank you!

*And thanks to many funders & colleagues:*

- **Funders**
  - ONR, Heinz/Grable/Mellon/Buhl/Pittsburgh foundations, Dept of Ed IES, National Science Foundation

- **Cognitive Tutors**
  - John R. Anderson, Albert Corbett, Steve Ritter, ...

- **Algebra & Geometry Studies**
  - Mitchell Nathan, Mimi McLaughlin, John Stamper, Tristan Nixon ...

- **LearnLab**
  - Charles Perfetti, David Klahr, Lauren Resnick, Vincent Aleven, Kurt VanLehn, Carolyn Rose ...
  - All 200+ past & current members!
Different learned representation => different learned skills

Skill simSt-divide (e.g. $-3x = 6$)
- Retrieval path:
  - Left side ($-3, -3x$)
  - Right side (6)
- Precondition:
  - Left side ($-3x$) does not have constant term
- Function sequence:
  - Divide both sides with the coefficient (-3)

Skill simSt-divide-1 (e.g. $-x = 3$)
- Retrieval path:
  - Left side (-x)
  - Right side (3)
- Precondition:
  - Left side (-x) is of the form -v
- Function sequence:
  - Generate one (1)
  - Divide both sides with -1
Request: Help build learning data infrastructure

• Recommendation #4 of 5 in Science:

• Worth doing?

• Looking for “DataLab partners”
  – Contribute data, methods, or infrastructure
  – Share same
  – LearnLab DataShop is seed
Abstract

We have developed analytic methods to discover barriers to student learning from data for educational technology use (see learnlab.org). Such discoveries can guide the redesign of instruction and our online experiments demonstrate enhanced learning outcomes. Our analytic methods span issues of student skill acquisition, metacognition, and motivation. Focusing on the first, I will illustrate how alternative cognitive models of learning can be evaluated by translating them to statistical models and predicting learning curve data for model comparison. We have used machine learning in a couple of ways to generate alternative cognitive models, one more practical and the other more cutting edge. The second involves a computational model of student learning, SimStudent, that learns as students do by using an intelligent tutoring system. The cognitive models SimStudent acquires have been demonstrated to yield empirically-verified discoveries not present in the human-designed cognitive models behind the intelligent tutoring systems. In other words, with SimStudent is the potential to not only create intelligent tutoring systems without AI programming, but to also produce systems that are pedagogically more effective than human-built tutoring systems.
Data-driven learner models

• Accurate assessment during learning
• Models of cognition, metacognition, motivation, discussion
• Discover better models of what is hard to learn
Data => discovery => outcomes

- Discover barriers to student learning
  - Change instruction & produce better learning
  - Domain skills, learning strategies, dispositions

- Use data from graded online activities
  - Translate cognitive to statistical models
  - Predict student performance over time

- Machine learning generates cognitive models
  - Feature gen & search (“genetic algorithm”)
  - SimStudent learns as students do

- Proving SimStudent works
  - Creates intelligent tutors without programming
  - More effective than human-built tutoring systems
Expertise is represented as display-based production rules

• Skill divide (e.g. -3x = 6)

  - What:
    • Left side (-3x)
    • Right side (6)
  - When:
    • Left side (-3x) does not have constant term

=>

  - How:
    • Get-coefficient (-3) of left side (-3x)
    • Divide both sides with the coefficient

Learning mechanisms to generalize

• Perceptual information retrieval
• Preconditions on actions
• Action plans

• Original model required strong domain-specific functions, like Get-coefficient
  - More knowledge engineering required
  - Did not capture distinctions in student learning errors
    Like 3x=6 vs. -3x = 6