Crowdsourcing Cognitive Models for Assessment, Tutoring, and In-Game Support

Ken Koedinger
Human-Computer Interaction & Psychology
Carnegie Mellon University

Talk for
Crowdsourcing Personalized Online Education
July 2012
Argument Overview

• Student models drive effective instruction
• Data can be used to create better models
  – LearnLab’s DataShop: Datasets, learning curve visualizations, model comparison leaderboards, data mining algorithms
• Result: Combined human & machine intelligence discovers new models across domains
• Practical consequence: Focused redesign to improve instruction

Devil’s in the details

In the special case of linear relationship, use the least squares regression line as a summary of the overall pattern, and use it to make predictions.

Let's go back now to our motivating example, in which we wanted to predict the maximum distance at which a sign is legible for a 60-year-old. Now that we have found the least squares regression line, this prediction becomes quite easy:

\[
\text{Distance} = 557 - 0.09\ \text{km}
\]

People all over the world know fables of Aesop, but there is very little information about his famous Greek storyteller. Scholars agree that Aesop was born in 620 B.C. In his early years, he was a slave, and he lived on an island in the Aegean Sea. Even as a slave, Aesop had wisdom and knowledge. His master respected him so much that he freed him. When Aesop became a free man, he traveled to many countries in order to learn and to teach. In Lydia, a king invited him to stay in that country and gave Aesop some difficult jobs in government. In his work, Aesop often used fables to convince people of his ideas. One time, the king sent Aesop to Delphi with gold for people of that city. Aesop became disgusted with people's greed, so he sent gold back to the king. People of Delphi were very angry at Aesop for this, and they killed him. After his death, Aesop became a famous sculptor, and you see in photo above.
Intelligent tutoring goes to school

Algebra Cognitive Tutor

- Based on computational models of student thinking & learning
- Enhances student learning
- Widespread intensive use
  ~600K students use
  ~80 minutes per week

- 1998 Spin-off:

Cognitive Tutor Technology
Use cognitive model to individualize 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 $bx + c = d/a$

  $2x - 5 = 3$

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

  $6x - 5 = 9$

• **Model Tracing**: Follows student through their individual approach to a problem -> context-sensitive instruction
Cognitive Tutor Technology
Use cognitive model to individualize 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 \(abx + ac = d\)
  - Hint message: “Distribute \(a\) across the parentheses.”
  - Known? = 85% chance

  6x - 15 = 9

  2x - 5 = 3

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

  If goal is solve \(a(bx+c) = d\):
  - Then rewrite as \(abx + c = d\)
  - Bug message: “You need to multiply \(c\) by \(a\) also.”
  - Known? = 45%

  6x - 5 = 9

- **Knowledge Tracing**: Assesses student's knowledge growth -> individualized activity selection and pacing
Why data is important to improving student learning

• If we knew everything about students’ learning challenges, we would not need data

• But, there is a lot we do not know about student learning

• In fact, there’s a lot we don’t know about our own learning
  – You’ve had lots of experience with the English language
  – You might say you know English
  – But, do you know what you know?
Which kind of problem is most difficult 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
\[ x \times 6 + 66 = 81.90 \]
Data contradicts common beliefs of researchers and teachers


Why have cognitive/student models?

• Drive personalization decisions
  • Better student or cognitive model
    => better feedback & context-sensitive instruction
    => better adaptive problem selection

• Relevant to improving instruction more generally

So, how can we develop & improve student models?
Cognitive Task Analysis

**Methods:** Expert interviews, think alouds

**Benefits:** Better cognitive models & better instruction, *1.7 sd!* (Clark et al, 2008)

**But:** Time & expertise intensive

**Instead:** Use ed tech datasets: [learnlab.org/datashop](http://learnlab.org/datashop)

**Insight:** Build statistical models of student performance: *latent variables* represent *components of symbolic cognitive model*
Problem Circle N in the area unit of the Geometry Cognitive Tutor

Point B is the center of circle B. Point A and point C lie on circle B.

Answer each question using the given information. Use 3.14 for \( \pi \).

1. The area of the circle is 7234.56 square millimeters. What are the radius and diameter of the circle? What is the circumference of the circle?

2. The circumference of the circle is 295.16 millimeters. What are the radius and diameter of the circle? What is the area of the circle?

<table>
<thead>
<tr>
<th>Unit</th>
<th>Diagram Label</th>
<th>Radius of Circle B</th>
<th>Diameter of Circle B</th>
<th>Circumference of Circle B</th>
<th>Area of Circle B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AB</td>
<td>millimeter</td>
<td>millimeter</td>
<td>millimeter</td>
<td>7234.56</td>
</tr>
</tbody>
</table>
Learning Curves

Dataset: Geometry Area (1996-97)
Sample(s): All Data

All Selected Knowledge Components

Error Rate (%)

<table>
<thead>
<tr>
<th>Opportunity Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>59</td>
<td>59</td>
<td>59</td>
<td>58</td>
<td>56</td>
<td>55</td>
<td>48</td>
<td>48</td>
<td>47</td>
<td>45</td>
<td>42</td>
<td>42</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>40</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

Opportunity Number

<table>
<thead>
<tr>
<th>All Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
</tr>
</tbody>
</table>
DataShop’s learning curve tools aid discovery of better cognitive models.

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

But with decomposition, 12 KCs for area concepts, a smoother learning curve.

**Upshot:** Cognitive models of student learning can be discovered from data.
DataShop’s “leaderboard” ranks discovered cognitive models 100s of datasets coming from ed tech in math, science, & language

Some models are *machine* generated (based on human-generated learning factors)

Some models are *human* generated
Statistical Model of Learning

Additive Factors 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)
Metrics for model prediction

- AIC & BIC penalize for more parameters, fast & consistent
- Cross Validation, 10 fold
  - Target root mean squared error (RMSE)
  - *Stratified* by student, item, or not
  - Student models predict future performance
    => item stratified is best choice
Learning Factors Analysis (LFA)  
(Cen, Koedinger, & Junker, 2006)

• Method for discovering & evaluating cognitive models  
• Finds Q matrix that best predicts learning  
• Inputs  
  ➢ Data: Student success on tasks over time  
  ➢ P matrix: Factors hypothesized to explain learning  
  ➢ Initial Q matrix  
• Outputs  
  ➢ Rank order of most predictive Q matrix  
  ➢ Parameter estimates for each
Simple search process example: modifying Q matrix by factor in P matrix to get new Q’ matrix

- Q matrix factor Sub *split* by P matrix factor Neg-result
- Produces new Q matrix
- Two new KCs (Sub-Pos & Sub-Neg) replace old KC (Sub)
  - Redo opportunity counts
LFA: Best First Search Process

- Search algorithm guided by a heuristic: BIC
- Start from an existing cog model (Q matrix)

But where does P matrix come from?

Frequent questions about LFA:
• Isn’t all the feature labeling too much work?
• Wouldn’t it be better to have a data mining algorithm that doesn’t require it?

Answers: No and no!
• Better prediction without interpretation does not help to redesign instruction!

=> Extract labels from “normal” course of work ...
Scientist “crowd”sourcing: Feature input (P) comes “for free”

Some models are human generated

Union of all hypothesized KCs in human generated models
Does it work?

## Applying LFA across domains

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Content area</th>
<th>RMSE</th>
<th>Median Learning slope (logit)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Orig in-use</td>
<td>Best-hand</td>
</tr>
<tr>
<td>Geometry 9697</td>
<td>Geometry area</td>
<td>0.4129</td>
<td>0.4033</td>
</tr>
<tr>
<td>Hampton 0506</td>
<td>Geometry area</td>
<td>NA</td>
<td>0.4022</td>
</tr>
<tr>
<td>Cog Discovery</td>
<td>Geometry area</td>
<td>NA</td>
<td>0.3250</td>
</tr>
<tr>
<td>DFA-318</td>
<td>Story problems</td>
<td>0.4461</td>
<td>0.4407</td>
</tr>
<tr>
<td>DFA-318-main</td>
<td>Story problems</td>
<td>0.4376</td>
<td>0.4287</td>
</tr>
<tr>
<td>Digital game</td>
<td>Fractions</td>
<td>0.4442</td>
<td>0.4396</td>
</tr>
<tr>
<td>Self-explanation</td>
<td>Equation solving</td>
<td>NA</td>
<td>0.4014</td>
</tr>
<tr>
<td>IWT 1</td>
<td>English articles</td>
<td>0.4262</td>
<td>0.4110</td>
</tr>
<tr>
<td>IWT 2</td>
<td>English articles</td>
<td>0.3854</td>
<td>0.3854*</td>
</tr>
<tr>
<td>IWT 3</td>
<td>English articles</td>
<td>0.3970</td>
<td>0.3965</td>
</tr>
<tr>
<td>Statistics-Fall09</td>
<td>Statistics</td>
<td>0.3648</td>
<td>0.3527</td>
</tr>
</tbody>
</table>

### Key findings:
- **Variety of domains & technologies**
- **11 of 11 improved models**
- **9 of 11 equal or greater learning**
Interpreting student model improvements

Method
• Isolate improvement in KCs from base model to new models by computing reduction in RMSE

Example analysis from Geometry 9697 dataset

<table>
<thead>
<tr>
<th>Original model KCs</th>
<th>% reduction in RMSE</th>
<th>orig-&gt;hand</th>
<th>hand-&gt;LFA</th>
<th>orig-LFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRCLE-RADIUS</td>
<td>5.8%</td>
<td>4.0%</td>
<td>9.5%</td>
<td></td>
</tr>
<tr>
<td>COMPOSE-BY-ADDITION</td>
<td>5.2%</td>
<td>0.3%</td>
<td>5.5%</td>
<td></td>
</tr>
<tr>
<td>TRIANGLE-SIDE</td>
<td>10.0%</td>
<td>1.2%</td>
<td>11.1%</td>
<td></td>
</tr>
<tr>
<td>Range of the 12 other KCs</td>
<td>-.5 to 3.4%</td>
<td>-.3 to 1.0%</td>
<td>-.2 to 3.1%</td>
<td></td>
</tr>
</tbody>
</table>
New KC revealed from LFA search

- **Uses forward strategy**
- **Uses backward strategy**
- **Greater slope indicates better model of learning and transfer**

### KC Values for DecompArithDiam Model

<table>
<thead>
<tr>
<th>KC Name</th>
<th>Intercept (logit)</th>
<th>Intercept (probability)</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>circle-area</td>
<td>0.31</td>
<td>0.58</td>
<td>0.068</td>
</tr>
</tbody>
</table>

### KC Values for LFASearchModel1-renamed Model

<table>
<thead>
<tr>
<th>KC Name</th>
<th>Intercept (logit)</th>
<th>Intercept (probability)</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>circle-area</td>
<td>0.44</td>
<td>0.61</td>
<td>0.106</td>
</tr>
<tr>
<td>radius-from-area</td>
<td>-0.38</td>
<td>0.41</td>
<td>0.165</td>
</tr>
</tbody>
</table>
Working backgrounds is not a general skill (in this domain)

Circle-area backward (54%) harder than forward (80%)

But no backward vs. forward diff in general
  pentagon area: 66% vs. 62%
  trapezoid area: 54% vs. 55%

Interpretation:
• Unique feature of circle-area backwards is need for a square root operation
• More practice & instruction needed
Implications for tutor redesign

Recommendations:
• More focus on problems requiring square root
  — More area-to-radius & square-area-to-side problems
• Merge all other forward/backward distinctions
• Modify skill bars

Expected Outcomes:
• Reduce time to mastery for other area formulas
• More time for practice on square root
Conclusions

• Combine human & machine intelligence
  – Student crowdsourcing: data from widely fielded systems
  – Scientist crowdsourcing (in the small): Use leaderboard to gain human participation in feature labeling
  – Results in better prediction AND interpretable and useful models

• Better predictions across 11 datasets & a variety of domains
  – Models overlap => small prediction error reduction, but ...

• Isolated improvements are large and useful
  – Yield tutor redesign recommendations
    => more efficient & effective learning & transfer
Comment on the Zeitgeist...

- Lots of great excitement, new ideas & smart people coming to the educational technology table
- Keep it coming!
- Find partners & do not get discouraged as it gets tough
  - Education is a **hugely** complex system
  - Complex systems have many failure joints
  - Devil’s in the details

- Complexity/opportunity in instruction options too:
Thank you
Crowdsourcing Cognitive Models for Assessment, Tutoring, and In-Game Support

Cognitive Tutors personalize online education by adapting problem-based instruction both within problems, by providing feedback and instruction specific to students' solution progress, and between problems, by selecting and pacing activities based on students' performance history. The approach relies on a cognitive model of the unobservable mental skills and concepts needed for success across problem-solving tasks. Widespread use of tutors (and educational games and simulations too) is producing vast data that can be used to create and improve cognitive models to drive better adaptive instruction. LearnLab's DataShop (http://learnlab.org/datashop) is the world's largest open repository of such data. I will discuss educational data mining algorithms to create, test, and improve cognitive models for better personalized assessment and tutoring within educational technologies. Large crowds of students and, more recently, small crowds of learning researchers are being leveraged in these approaches. In a recent study, we employed our Learning Factors Analysis algorithm on eleven different DataShop datasets to automatically discover better cognitive models in all cases.
Extra slides
Sample of log data used to generate learning curves

<table>
<thead>
<tr>
<th>Student</th>
<th>Step (Item)</th>
<th>Skill (KC)</th>
<th>Opportunity</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-addition</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Example in Geometry of “split” based on factor in P matrix

Original Q matrix

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
<th>Opportunity</th>
<th>Embed</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area</td>
<td>1</td>
<td>alone</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area</td>
<td>2</td>
<td>embed</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-add</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area</td>
<td>3</td>
<td>alone</td>
</tr>
</tbody>
</table>

Factor in P matrix

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
<th>Opportunity</th>
<th>Embed</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-add</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

After Splitting Circle-area by Embed

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area-alone</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area-embed</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-add</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area-alone</td>
<td>2</td>
</tr>
</tbody>
</table>

New Q matrix

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area-alone</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area-embed</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-add</td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area-alone</td>
</tr>
</tbody>
</table>

Revised Opportunity

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area-alone</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area-embed</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-add</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area-alone</td>
<td>2</td>
</tr>
</tbody>
</table>

LAWN_SPRINKLER: SECTION FOUR, #2

Problem Statement

If a lawn sprinkler spins 360 degrees and has a 17 ft spray, find the area of lawn that will be watered if the sprinkler is not moved.

<table>
<thead>
<tr>
<th>Length of Spray (OD) - Radius</th>
<th>Watered Lawn - Circle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units</td>
<td>feet</td>
</tr>
<tr>
<td>Question 1</td>
<td>17</td>
</tr>
</tbody>
</table>

LAWN_SPRINKLER_2: SECTION FOUR, #5

Problem Statement

A lawn sprinkler spins 360 degrees and has a 1 ft spray. If the rectangular lawn is 5 feet by 9 feet, find the area of lawn that will not be watered if the sprinkler is not moved.

<table>
<thead>
<tr>
<th>Length of Spray (OD) - (Radius)</th>
<th>Watered Lawn (Circle)</th>
<th>Area of Total Lawn (Rectangle)</th>
<th>Area of Unwatered Lawn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units</td>
<td>cm</td>
<td>cm</td>
<td>sq. cm</td>
</tr>
<tr>
<td>Question 1</td>
<td>7</td>
<td>3.1416</td>
<td>45</td>
</tr>
</tbody>
</table>
DataShop visualizations help identify potentially bad KCs

Many curves show a reasonable decline

Some do not => Opportunity to improve model!
Example model modification with implications for tutor redesign

KC Values For Original Model

<table>
<thead>
<tr>
<th>KC Name</th>
<th>Intercept (logit)</th>
<th>Intercept (probability)</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALT:COMPOSE-BY-ADDITION</td>
<td>1.04</td>
<td>0.74</td>
<td>0</td>
</tr>
</tbody>
</table>

non-smooth, not low, not declining

Original Model KC split into 3 KCs

KC Values For DecompArithDiam Model

<table>
<thead>
<tr>
<th>KC Name</th>
<th>Intercept (logit)</th>
<th>Intercept (probability)</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtract</td>
<td>2.05</td>
<td>0.89</td>
<td>0</td>
</tr>
<tr>
<td>compose-by-addition</td>
<td>0.84</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>decompose</td>
<td>-0.56</td>
<td>0.36</td>
<td>0.148</td>
</tr>
</tbody>
</table>
Splitting and combining of circle-radius and other related hypothesized KCs

- Circle-radius splits to circle-area, circle-diam-from-subgoal, and circle-diam-from-given
- Target skill for all 3 is computing radius
- Circle-area further splits to circle-area and radius-from-area
Future Work

• Variations on Learning Factors Analysis
  – Use add operator in search (models partial transfer)
  – Try other statistical models: BKT, PFA ...
  – Other search methods: beam search, new heuristics ...
• In vivo experiments to test model discoveries
  – compare learning with revised vs. original ed tech
• Apply human-machine discovery approach in other domains
  – Humans do feature engineering; machine uses
  – Humans & machine compete & cooperate through leaderboard