Learning from Learning Curves: Item Response Theory & Learning Factors Analysis

Ken Koedinger
Human-Computer Interaction Institute
Carnegie Mellon University

Knowledge Decomposibility Hypothesis

- Human acquisition of academic competencies can be decomposed into units, called knowledge components (KCs), that predict student task performance & transfer
- Performance predictions
  - If item I1 only requires KC1 & item I2 requires both KC1 and KC2, then item I2 will be harder than I1
  - If student can do I2, then they can do I1
- Transfer predictions
  - If item I1 requires KC1, & item I3 also requires KC1, then practice on I3 will improve I1
  - If item I1 requires KC1, & item I4 requires only KC3, then practice on I4 will not improve I1
- Fundamental EDM idea:
  - We can discover KCs (cog models) by working these predictions backwards!

Domain-Specific Cognitive Models

- Question: How do students represent knowledge in a given domain?
- Answering this question involves deep domain analysis
- The product is a cognitive model of students’ knowledge
  - Recall cognitive models drive ITS behaviors & instructional design decisions

Student Performance As They Practice with the LISP Tutor

Example of Items & KCs

<table>
<thead>
<tr>
<th></th>
<th>KC1</th>
<th>KC2</th>
<th>KC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1: 5+3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I2: 15+7</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>I3: 4+2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I4: 5-3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Mean Error Rate - 158 Goals in Lesson

Goal Number in Lesson (25 Exercises)
Production Rule Analysis

Evidence for Production Rule as an appropriate unit of knowledge acquisition

Using learning curves to evaluate a cognitive model

- Lisp Tutor Model
  - Learning curves used to validate cognitive model
  - Fit better when organized by knowledge components (productions) rather than surface forms (programming language terms)
- But, curves not smooth for some production rules
  - “Blips” in leaning curves indicate the knowledge representation may not be right
  - Let me illustrate …

Curve for “Declare Parameter” production rule

What’s happening on the 6th & 10th opportunities?

- How are steps with blips different from others?
- What’s the unique feature or factor explaining these blips?

Can modify cognitive model using unique factor present at “blips”

- Blips occur when to-be-written program has 2 parameters
- Split Declare-Parameter by parameter-number factor:
  - Declare-first-parameter
  - Declare-second-parameter

(defun second (lst) (append lst (list lst)))

Can learning curve analysis be automated?

- Learning curve analysis
  - Identify blips by hand & eye
  - Manually create a new model
  - Qualitative judgment

- Need to automatically:
  - Identify blips by system
  - Propose alternative cognitive models
  - Evaluate each model quantitatively

Learning Factors Analysis (LFA): A Tool for KC Analysis

- LFA is a method for discovering & evaluating alternative cognitive models
  - Finds knowledge component decomposition that best predicts student performance & learning transfer

- Inputs
  - Data: Student success on tasks in domain over time
  - Codes: Factors hypothesized to drive task difficulty & transfer
    - A mapping between these factors & domain tasks

- Outputs
  - A rank ordering of most predictive cognitive models
  - For each model, a measure of its generalizability & parameter estimates for knowledge component difficulty, learning rates, & student proficiency

Learning Factors Analysis (LFA) draws from multiple disciplines

- Machine Learning & AI
  - Combinatorial search (Russell & Norvig, 2003)
  - Exponential-family principal component analysis (Gordon, 2002)

- Psychometrics & Statistics
  - Q Matrix & Rule Space (Tatsuoka 1983, Barnes 2005)
  - Item response learning model (Draney, et al., 1995)
  - Item response assessment models (DiBello, et al., 1995; Embretson, 1997; von Davier, 2005)

- Cognitive Psychology
  - Learning curve analysis (Corbett, et al. 1995)
Steps in Learning Factors Analysis

1. Q-Matrix
   - Representing Knowledge Components as factors of items
     - Problem: How to represent KC model?
     - Solution: Q-Matrix (Tatsuoka, 1983)

<table>
<thead>
<tr>
<th>Items</th>
<th>Add</th>
<th>Sub</th>
<th>Mul</th>
<th>Div</th>
</tr>
</thead>
<tbody>
<tr>
<td>2^8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2^8-3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- Single KC item = when a row has one 1
  - 2^8 above
- Multi-KC item = when a row has many 1's
  - 2^8 – 3

What good is a Q matrix? Can predict student accuracy on items not previously seen, based on KCs involved.

The Statistical Model

- Generalized Power Law to fit learning curves
  - Logistic regression (Draney, Wilson, Pirolli, 1995)
- Assumptions
  - Some skills may easier from the start than others
    => use an intercept parameter for each skill
  - Some skills are easier to learn than others
    => use a slope parameter for each skill
  - Different students may initially know more or less
    => use an intercept parameter for each student
  - Students generally learn at the same rate
    => no slope parameters for each student
- These assumptions are reflected in a statistical model ...

Simple Statistical Model of Performance & Learning

- Problem: How to predict student responses from model?
- Solutions: Additive Factor Model (Draney, et al. 1995)

\[ p_{ij} = \Pr(Y_{ij} = 1 | \theta_i, \beta_j, \gamma) = \frac{\exp(\theta_i + \sum_{k=1}^{K} q_k \beta_k \gamma_k T_{ik})}{1 + \exp(\theta_i + \sum_{k=1}^{K} q_k \beta_k \gamma_k T_{ik})} \]

\[ q_k = \begin{cases} 1 & \text{item } j \text{ uses skill } k \\ 0 & \text{otherwise} \end{cases} \]

γ = the response of student i on item j
θ = coefficient for student i
β = coefficient for skill k
γ = coefficient for the learning rate of skill k
T = the number of practice opportunities student i has had on the skill k
Comparing Additive Factor Model to other psychometric techniques

- Instance of generalized linear regression, binomial family
  - "logistic regression"
  - R code: `glm(success ~ student+skill+skill:opportunity, family=binomial,...)`
- Extension of item response theory
  - IRT has simply a student term (theta-i) + item term (beta-j)
  - R code: `glm(success ~ student+item, family=binomial,...)`
- The additive factor model behind LFA is different because:
  - It breaks items down in terms of knowledge component factors
  - It adds term for practice opportunities per component

\[
p_{ij} = \Pr(Y_{ij} = 1 | \theta_i, \beta, \gamma) = \frac{\exp(\theta_i + \sum_{k=1}^{K} q_{i,k} \beta_k + \sum_{k=1}^{K} q_{i,k} \gamma_k T_k)}{1 + \exp(\theta_i + \sum_{k=1}^{K} q_{i,k} \beta_k + \sum_{k=1}^{K} q_{i,k} \gamma_k T_k)}
\]

\( q_{i,k} = \begin{cases} 
1 & \text{item } j \text{ uses skill } k \\
0 & \text{otherwise}
\end{cases} \)

Model Evaluation

- How to compare cognitive models?
  - A good model minimizes prediction risk by balancing fit with data & complexity (Wasserman 2005)
- Compare BIC for the cognitive models
  - BIC is “Bayesian Information Criteria”
  - BIC = \(-2\times\text{log-likelihood} + \text{numPar} \times \log(\text{numOb})\)
  - Better (lower) BIC == better predict data that haven’t seen
- Mimics cross validation, but is faster to compute

Item Labeling & the “P Matrix”:
Adding Alternative Factors

- Problem: How to improve existing cognitive model?
- Solution: Have experts look for difficulty factors that are candidates for new KCs. Put these in P matrix.

Using P matrix to update Q matrix

- Create a new Q’ by using elements of P as arguments to operators
  - Add operator: Q’ = Q + P[,1]
  - Split operator: Q’ = Q[,2] * P[,1]
LFA: KC Model Search

- Problem: How to find best model given Q and P matrices?
- Solution: Combinatorial search

- A best-first search algorithm (Russell & Norvig 2002)
  - Guided by a heuristic, such as BIC
- Goal: Do model selection within logistic regression model space
  - Steps:
    1. Start from an initial “node” in search graph using given Q
    2. Iteratively create new child nodes (Q’) by applying operators with arguments from P matrix
    3. Employ heuristic (BIC of Q’) to rank each node
    4. Select best node not yet expanded & go back to step 2

Learning Factors Analysis: Example in Geometry Area

Area Unit of Geometry Cognitive Tutor

- Original cognitive model in tutor:
  - 15 skills:
    - Circle-area
    - Circle-circumference
    - Circle-diameter
    - Circle-radius
    - Compose-by-addition
    - Compose-by-multiplication

Log Data Input to LFA

<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>p1x1</td>
<td>Circle-area</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>p2x1</td>
<td>Circle-area</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2x2</td>
<td>Rectangle-area</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p3x2</td>
<td>Compose-by-addition</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>p3x1</td>
<td>Circle-area</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
AFM Results for original KC model

Higher intercept of skill -> easier skill
Higher slope of skill -> faster students learn it

<table>
<thead>
<tr>
<th>Skill</th>
<th>Intercept</th>
<th>Slope</th>
<th>Avg Opportunities</th>
<th>Initial Probability</th>
<th>Avg Probability</th>
<th>Final Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallelogram-area</td>
<td>2.16</td>
<td>0.01</td>
<td>14.9</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>Pentagon-area</td>
<td>-2.16</td>
<td>0.85</td>
<td>4.3</td>
<td>0.2</td>
<td>0.63</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Student Intercept

Higher intercept of student -> student initially knew more

Model Statistics

AIC = 3,950
BIC = 4,285
MAD = 0.085

The AIC, BIC & MAD statistics provide alternative ways to evaluate models

MAD = Mean Absolute Deviation

Application: Use Statistical Model to improve tutor

- Some KCs over-practiced, others under (Cen, Koedinger, Juner, 2007)

Initial error rate 12% reduced to 8% after 18 times of practice

Initial error rate 76% reduced to 40% after 6 times of practice

“Close the loop” experiment

- In vivo experiment: New version of tutor with updated knowledge tracing parameters vs. prior version
- Reduced learning time by 20%, same robust learning gains
- Knowledge transfer: Carnegie Learning using approach for other tutor units

Example in Geometry of split based on factor in P matrix

Original Q matrix

Factor in P matrix

New Q matrix

Revised Opportunity

<table>
<thead>
<tr>
<th>Original Q matrix</th>
<th>Factor in P matrix</th>
<th>New Q matrix</th>
<th>Revised Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step</td>
<td>Skill</td>
<td>Opportunity</td>
<td>Embed</td>
</tr>
<tr>
<td>1</td>
<td>p1</td>
<td>Circle-area</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>p2</td>
<td>Circle-area</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>p3</td>
<td>Rectangle-area</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>p1</td>
<td>Compose-by-add</td>
<td>2</td>
</tr>
</tbody>
</table>

Example: Use Statistical Model to improve tutor

- Some KCs over-practiced, others under (Cen, Koedinger, Juner, 2007)

Initial error rate 12% reduced to 8% after 18 times of practice

Initial error rate 76% reduced to 40% after 6 times of practice

“Close the loop” experiment

- In vivo experiment: New version of tutor with updated knowledge tracing parameters vs. prior version
- Reduced learning time by 20%, same robust learning gains
- Knowledge transfer: Carnegie Learning using approach for other tutor units

Example in Geometry of split based on factor in P matrix

Original Q matrix

Factor in P matrix

New Q matrix

Revised Opportunity

<table>
<thead>
<tr>
<th>Original Q matrix</th>
<th>Factor in P matrix</th>
<th>New Q matrix</th>
<th>Revised Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step</td>
<td>Skill</td>
<td>Opportunity</td>
<td>Embed</td>
</tr>
<tr>
<td>1</td>
<td>p1</td>
<td>Circle-area</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>p2</td>
<td>Circle-area</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>p3</td>
<td>Rectangle-area</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>p1</td>
<td>Compose-by-add</td>
<td>2</td>
</tr>
</tbody>
</table>
LFA – Model Search Process

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

Automates the process of hypothesizing alternative KC models & testing them against data

LFA Results 1: Applying splits to original model

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Splits</th>
<th>Number of Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>18</td>
</tr>
</tbody>
</table>

Common results:
- Compose-by-multiplication split based on whether it was an area or a segment being multiplied
- Circle-radius is split based on whether it is being done for the first time in a problem or is being repeated
- Made sense, but less than expected ...

Example of Tutor Design Implications

- LFA search suggests distinctions to address in instruction & assessment
  - With these new distinctions, tutor can
    - Generate hints better directed to specific student difficulties
    - Improve knowledge tracing & problem selection for better cognitive mastery
  - Example: Consider Compose-by-multiplication before LFA

Other Geometry problem examples

Example:

<table>
<thead>
<tr>
<th>Problem</th>
<th>Intercept</th>
<th>Slope</th>
<th>Avg Practice Opportunities</th>
<th>Initial Probability</th>
<th>Avg Probability</th>
<th>Final Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>-.15</td>
<td>.1</td>
<td>10.2</td>
<td>.65</td>
<td>.84</td>
<td>.92</td>
</tr>
</tbody>
</table>

With final probability .92, many students are short of .95 mastery threshold
Making a distinction changes assessment decision

- However, after split:
  - CM-area and CM-segment look quite different
    - CM-area is now above .95 mastery threshold (at .96)
    - CM-segment is only at .60
  - Implications:
    - Original model penalizes students who have key idea about composite areas (CM-area) -- some students solve more problems than needed
    - CM-segment is not getting enough practice
      - Instructional design choice: Add instructional objective & more problems or not?

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>slope</th>
<th>Avg Practice Opportunities</th>
<th>Initial Probability</th>
<th>Avg Probability</th>
<th>Final Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>-.15</td>
<td>.1</td>
<td>10.2</td>
<td>.85</td>
<td>.84</td>
<td>.92</td>
</tr>
<tr>
<td>CMarea</td>
<td>-.009</td>
<td>.17</td>
<td>9</td>
<td>.84</td>
<td>.86</td>
<td>.96</td>
</tr>
<tr>
<td>CMsegment</td>
<td>-1.42</td>
<td>.48</td>
<td>1.9</td>
<td>.32</td>
<td>.34</td>
<td>.60</td>
</tr>
</tbody>
</table>

Perhaps original model is good enough -- Can LFA recover it?

- Merge some skills in original model, to produce 8 skills:
  - Circle-area, Circle-radius => Circle
  - Circle-circumference, Circle-diameter => Circle-CD
  - Parallelogram-area, Parallelogram-side => Parallelogram
  - Pentagon-area, Pentagon-side => Pentagon
  - Trapezoid-area, Trapezoid-base, Trapezoid-height => Trapezoid
  - Triangle-area, Triangle-side => Triangle
  - Compose-by-addition
  - Compose-by-multiplication

- Does splitting by "backward" (or otherwise) yield a better model? Closer to original?

LFA Results 2: Recovery

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Splits: 4</td>
<td>Number of Splits: 3</td>
<td>Number of Splits: 4</td>
</tr>
<tr>
<td>Circle*area</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle *forward 2. Compose-by-addition is not split</td>
<td></td>
</tr>
<tr>
<td>Circle<em>radius</em>initial</td>
<td>1. Circle is split into Circle backward initial, Circle backward repeat, Circle *forward 2. Compose-by-addition is not split</td>
<td></td>
</tr>
<tr>
<td>Circle<em>radius</em>repeat</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle <em>forward 2. Compose-by-addition is split into Compose-by-addition and Compose-by-multiplication</em>segment</td>
<td></td>
</tr>
<tr>
<td>Compose-by-addition</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle <em>forward 2. Compose-by-addition is split into Compose-by-addition and Compose-by-multiplication</em>segment</td>
<td></td>
</tr>
<tr>
<td>Compose-by-addition*area-difference</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle <em>forward 2. Compose-by-addition is split into Compose-by-addition and Compose-by-multiplication</em>segment</td>
<td></td>
</tr>
<tr>
<td>Compose-by-addition*area-combination</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle <em>forward 2. Compose-by-addition is split into Compose-by-addition and Compose-by-multiplication</em>segment</td>
<td></td>
</tr>
<tr>
<td>Compose-by-multiplication*area-combination</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle <em>forward 2. Compose-by-addition is split into Compose-by-addition and Compose-by-multiplication</em>segment</td>
<td></td>
</tr>
<tr>
<td>Compose-by-multiplication*area-difference</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle <em>forward 2. Compose-by-addition is split into Compose-by-addition and Compose-by-multiplication</em>segment</td>
<td></td>
</tr>
<tr>
<td>Compose-by-multiplication*segment</td>
<td>All skills are the same as those in model 1 except that 1. Circle is split into Circle backward initial, Circle backward repeat, Circle <em>forward 2. Compose-by-addition is split into Compose-by-addition and Compose-by-multiplication</em>segment</td>
<td></td>
</tr>
<tr>
<td>Number of skills: 12</td>
<td>Number of skills: 11</td>
<td>Number of skills: 12</td>
</tr>
<tr>
<td>BIC: 4,149.315</td>
<td>BIC: 4,171.523</td>
<td>BIC: 4,171.786</td>
</tr>
</tbody>
</table>

- Only 1 recovery: Circle-area vs. Circle-radius
- More merged model fits better
  - Why? More transfer going on than expected or not enough data to make distinctions?
  - Other relevant data sets …

Research Issues & Summary
Open Research Questions: Technical

- What factors to consider? P matrix is hard to create
  - Enhancing human role: Data visualization strategies
  - Other techniques: Principal Component Analysis +
  - Other data: Do clustering on problem text
- Interpreting LFA output can be difficult
  - LFA outputs many models with roughly equivalent BICs
  - How to select from large equivalence class of models?
  - How to interpret results?

=> Researcher can’t just “go by the numbers”
  1) Understand the domain, the tasks
  2) Get close to the data

DataShop Case Study video

- “Using DataShop to discover a better knowledge component model of student learning”

Summary of Learning Factors Analysis (LFA)

- LFA combines statistics, human expertise, & combinatorial search to discover cognitive models
- Evaluates a single model in seconds, searches 100s of models in hours
  - Model statistics are meaningful
  - Improved models suggest tutor improvements
- Other applications of LFA & model comparison
- Used by others:
  - Individual differences in learning rate (Rafferty et. al., 2007)
  - Alternative methods for error attribution (Nwaigwe, et al. 2007)
  - Model comparison for DFA data in math (Baker; Rittle-Johnson)
  - Learning transfer in reading (Leszczenski & Beck, 2007)

Open Research Questions: Psychology of Learning

- Test statistical model assumptions: Right terms?
  - Is student learning rate really constant?
    - Does a Student x Opportunity interaction term improve fit?
    - What instructional conditions or student factors change rate?
  - Is knowledge space “uni-dimensional”?
    - Does a Student x KC interaction term improve fit?
    - Need different KC models for different students/conditions?
- Right shape: Power law or an exponential?
  - Long-standing hot debate
    - Has focused on “reaction time” not on error rate!
- Other predictor & outcome variables (x & y of curve)
  - Outcome: Error rate => Reaction time, assistance score
  - Predictor: Opportunities => Time per instructional event
Open Research Questions: Instructional Improvement

- Do LFA results generalize across data sets?
  - Is BIC a good estimate for cross-validation results?
  - Does a model discovered with one year's tutor data generalize to a next year?
  - Does model discovery work in ill-structured domains?
- Use learning curves to compare instructional conditions in experiments
- *Need more “close the loop” experiments*
  - EDM => better model => better tutor => better student learning

Overview

- Learning Factors Analysis algorithm
- A Geometry Cognitive Model and Log Data
  - Experiments and Results
  - DataShop case study video
- Learning curves between conditions

Future Goals in Discovering Domain Models

- Improve knowledge-discovery methods
  - Partial Order Knowledge Structures (POKS)
  - Exponential-family Principle Component Analysis
- Improve human-machine interaction
  - Better process for task difficulty factor labeling
- Show models yield improved student learning
Lots of interesting questions to be addressed with Ed Data Mining!!

- Assessment questions
  - Can on-line embedded assessment replace standardized tests?
  - Can assessment be accurate if students are learning during test?
- Learning theory questions
  - What are the "elements of transfer" in human learning?
  - Is learning rate driven by student variability or content variability?
  - Can conceptual change be tracked & better understood?
- Instructional questions
  - What instructional moves yield the greatest increases in learning?
  - Can we replace ANOVA with learning curve comparison to better evaluate learning experiments?
- Metacognition & motivation questions
  - Can student affect & motivation be detected in on-line click stream data?
  - Can student metacognitive & self-regulated learning strategies be detected in on-line click stream data?

---

**Past Summer School Project**

- Rafferty (Stanford) & Yudelson (Pitt)
- Analyzed a data set from Geometry
- Applied Learning Factors Analysis (LFA)
- Driving questions:
  - Are students learning at the same rate as assumed in prior LFA models?
  - Do we need different cognitive models (KC models) to account for low-achieving vs. high-achieving students?

---

**EXTRA SLIDES**

Past SS Project

Physics curves

Condition contrast

---

**Past Project Example**

- Rafferty (Stanford) & Yudelson (Pitt)
- Analyzed a data set from Geometry
- Applied Learning Factors Analysis (LFA)
- Driving questions:
  - Are students learning at the same rate as assumed in prior LFA models?
  - Do we need different cognitive models (KC models) to account for low-achieving vs. high-achieving students?
A Statistical Model for Learning Curves

- Predicts whether student is correct depending on knowledge & practice

\[
p_{ij} = \Pr(T_{ij} = 1 | \theta_i, \beta, \gamma) = \frac{\exp(\theta_i + \sum_{k=1}^{K} \beta_k q_{ik} + \sum_{k=1}^{K} \gamma_k T_{ik})}{1 + \exp(\theta_i + \sum_{k=1}^{K} \beta_k q_{ik} + \sum_{k=1}^{K} \gamma_k T_{ik})}
\]

- \( \gamma_k \) = the response of student i on item j
- \( \beta_k \) = coefficient for student i
- \( q_{ik} \) = coefficient for the learning rate of skill k
- \( T_{ik} \) = the number of practice opportunities student i has had on the skill k

Learning rate is different for different skills, but not for different students

Rafferty & Yudelson Results 2

- Is it “faster” learning or “different” learning?
  - Fit with a more compact model is better for low start high learn

Learning curve contrast in Physics dataset ...

Low-Start High-Learn (LSHL) group has a faster learning rate than other groups of students

Resulted in best Young Researcher Track paper at AIED07
Not a smooth learning curve -> this knowledge component model is wrong. Does not capture genuine student difficulties.

More detailed cognitive model yields smoother learning curve. Better tracks nature of student difficulties & transfer
(Few observations after 10 opportunities yields noisy data)

Best BIC (parsimonious fit) for Default (original) KC model

Better than simpler Single-KC model

And better than more complex Unique-step (IRT) model

The Long-Standing Transfer Debate

- General: Faculty theory of mind
  - Mind has faculties that can be exercised with Latin, Geometry, ... video games, n-back task*
  - Transfer is broad & general, across domains
- Specific: Thorndike’s identical elements
  - Mind is made up of stimulus-response elements
  - Transfer occurs between tasks with common elements
  - 1922 study: “Multiply x&a & x&b” fails to transfer to “multiply 4&a & 4&b”
  - Transfer is narrow, within domains & tasks
- More recent view
  - Singley & Anderson: Learning & transfer occurs at the grain size of the production rule
  - But begs the question: What’s the grain size of a production?

Learning curve analysis in an experiment

- Experiments on Expert vs. Intelligent Novice feedback in an Excel Programming Tutor

Spreadsheet formulas: *Absolute referencing* stalls many learners

<table>
<thead>
<tr>
<th>Hours Worked</th>
<th>Hourly Wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$10</td>
</tr>
<tr>
<td>4</td>
<td>$20</td>
</tr>
<tr>
<td>5</td>
<td>$30</td>
</tr>
<tr>
<td>6</td>
<td>$40</td>
</tr>
</tbody>
</table>

Theoretical interest: Difficult domain reveals edges of human learning

Faculties of mind theory of transfer?

- Exercising “reasoning faculty” or “programming faculty” does *not* (clearly) reveal improvement

Identical elements theory of transfer?

- Maybe, but perhaps too fine
- Are there groups of items across which we see transfer?
Pattern emerges at more coarse grain

- Knowledge acquired in some problems transfer to others
- What is that knowledge?

Broader rules also plausible

- Which one is right?
- Can we empirically compare knowledge models?

Which knowledge model is correct?

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Model parameters</th>
<th>Log Likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 (identical elements)</td>
<td>108</td>
<td>-2070</td>
<td>4356</td>
<td>5062</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>-2097</td>
<td>4314</td>
<td>4706</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>-2103</td>
<td>4313</td>
<td>4666</td>
</tr>
<tr>
<td>1 (faculty theory)</td>
<td>50</td>
<td>-2447</td>
<td>4994</td>
<td>5321</td>
</tr>
</tbody>
</table>

- Intermediate models are best …
- Learning curve “smoothness” selects for better cognitive models of learning

Experimental Design: Intelligent Novice vs. Expert Model

- Recast delayed vs. immediate feedback debate as contrasting “model of desired performance”

- Control: Expert Model & Tutor
  - Goal: students should not make errors
  - Feedback is immediate on all errors

- Treatment: Intelligent Novice Model & Tutor
  - Goal: students can make some errors, but should recognize & self-correct
  - Feedback appears delayed, but is immediate
    -> when student fails to self-correct own errors
Expert Feedback

Student enters incorrect formula

Gets immediate feedback from tutor

Intelligent Novice Feedback

No feedback after incorrect formula

Student copies formula, sees consequences

Instruction helps students reason about what went wrong

Results: IN better on all measures

Transfer

60% ± 5

74% ± 7

All differences significant: F(1,35)>13, p<0.001
Can we detect a qualitative difference from learning curves?

Hypotheses:

■ Shallow rule model better accounts for EX student performance
■ Deep rule model better accounts for IN student performance
  □ Supporting error self-correction promotes deeper feature encoding
  => more general knowledge
  => greater transfer

Which knowledge component transfers to new situation?

Shallow vs. deep learning: Depth of encoding

■ Shallow feature encoding*:
  Rule S1: If you are copying down a column, then put a $ in front of the number.
■ Deep feature encoding:
  Rule D: If a cell index needs to stay fixed but would change when copied, then put a $ in front of that index.

*Note: These describe mental structures that students may not be able to verbalize.

Shallow rule model provides an OK fit for two groups

■ IN group generally less erratic
■ Better fit here for EX than in deep model …
Deep model: Fit gets better for IN, worse for EX!

Error self-correction support (IN) leads to deeper feature encoding

Intelligent Novice Tutor => More General Knowledge

- Beyond assessments of ability or achievement
  - Assess abilities during learning
  - Crease assessments of knowledge
- Beyond single test score
  - What do students know
  - What is range of their transfer