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

Ken Koedinger
Human-Computer Interaction Institute
Carnegie Mellon University

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

Mean Error Rate - 158 Goals in Lesson

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 …

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

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)

Representing Knowledge Components as factors of items

- Problem: How to represent KC model?
- Solution: Q-Matrix (Tatsuoka, 1983)

<table>
<thead>
<tr>
<th>Item</th>
<th>Skills</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>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2*8 - 3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></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
Additive Factors Model Assumptions

- Logistic regression to fit learning curves (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)
  - i students, j problems/items, k skills (KCs)

Comparing Additive Factor Model to other psychometric techniques

- Instance of generalized linear regression, binomial family
  or “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

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*log-likelihood + numPar * log(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.

<table>
<thead>
<tr>
<th>Item</th>
<th>Skill</th>
<th>Add</th>
<th>Sub</th>
<th>Mul</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2 \times 8$</td>
<td>$0$</td>
<td>$0$</td>
<td>$1$</td>
<td></td>
</tr>
<tr>
<td>$2 \times 8 - 3$</td>
<td>$0$</td>
<td>$1$</td>
<td>$1$</td>
<td></td>
</tr>
<tr>
<td>$2 \times 8 - 30$</td>
<td>$0$</td>
<td>$1$</td>
<td>$1$</td>
<td></td>
</tr>
<tr>
<td>$3 + 2 \times 8$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td></td>
</tr>
</tbody>
</table>

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: $Q' = Q[, 2] \times P[,1]$

<table>
<thead>
<tr>
<th>Q- Matrix after add $P[, 1]$</th>
<th>Q- Matrix after splitting $P[, 1], Q[,2]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Skill</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>$2 \times 8$</td>
<td>$0$</td>
</tr>
<tr>
<td>$2 \times 8 - 3$</td>
<td>$0$</td>
</tr>
<tr>
<td>$2 \times 8 - 30$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

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

  **Original cognitive model in tutor:**
  - Parallelogram-area
  - Parallelogram-side
  - Pentagon-area
  - Pentagon-side
  - Trapezoid-area
  - Trapezoid-base
  - Trapezoid-height
  - Triangle-area
  - Triangle-side

  **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>2.14</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>0.45</td>
<td>4.3</td>
<td>0.2</td>
<td>0.63</td>
<td>0.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>student0</td>
<td>1.18</td>
</tr>
<tr>
<td>student1</td>
<td>0.82</td>
</tr>
<tr>
<td>student2</td>
<td>0.21</td>
</tr>
</tbody>
</table>

AFM Statistics

- AIC = 3,950
- BIC = 4,285
- MAD = 0.083

AFM Results for original KC model

- 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, Junker, 2007)

AFM Results for original KC model

- Initial error rate 12%
- Reduced to 8%
- After 18 times of practice

AFM Results for original KC model

- Initial error rate 76%
- Reduced to 40%
- After 6 times of practice

Log Data Input to LFA

- Items = steps in tutors with step-based feedback
- Q-matrix in single column: works for single KC items
- Opportunities Student has had to learn KC

<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>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-addition</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
“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

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

LFA –Model Search Process

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

Example LFA Results: 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>Model 1</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Model 2</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Model 3</td>
<td>2</td>
<td>17</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

Lawn Sprinkler: Section Four

Problem Statement:
If a lawn sprinkler sprays 360 degrees and has a 17.5 spray, find the area of lawn that will be watered if the sprinkler is at an angle of 50 degrees.

<table>
<thead>
<tr>
<th>Question 1</th>
<th>0</th>
<th>44.7</th>
<th>44.8</th>
<th>44.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of Spray (OD) - Rod</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watered Lawn - Circle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area of Total</td>
<td>4,248.86</td>
<td>4,248.86</td>
<td>4,251.07</td>
<td></td>
</tr>
</tbody>
</table>
Other Geometry problem examples

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

<table>
<thead>
<tr>
<th>Intersect</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>
</tr>
<tr>
<td>CMarea</td>
<td>.009</td>
<td>.17</td>
<td>9</td>
<td>.64</td>
<td>.86</td>
</tr>
<tr>
<td>CMsegment</td>
<td>-1.42</td>
<td>.48</td>
<td>1.9</td>
<td>.32</td>
<td>.54</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)
    - But 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 more problems to address CM-segment?

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

Some curves “curves”, when curves are flat => bad KC model

Scaffolded vs. unscaffolded “compose-by-addition” problems

- Scaffolded
  - Prompts are given for subgoals
- Unscaffolded
  - Prompts are not given for subgoals (initially)
Scaffolded vs. unscaffolded composition problems

- Scaffolded
  - Columns given for area subgoals

- Unscaffolded
  - Columns not given for area subgoals

Two Circles in a Square: Section Five, #6

Problem Statement
In the diagram, two identical circles, O and S, are enclosed in Square ABCD. Circle O has a radius (OR) of 1.0 inch. Circle S has a radius (OS) of identical length to Circle O. Use this information to find the area of the shaded region.

<table>
<thead>
<tr>
<th>Question</th>
<th>Subgoal 1</th>
<th>Subgoal 2</th>
<th>Area of Triangle</th>
<th>Area of Circle</th>
<th>Area of Shaded Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Q2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Before unpacking compose-by-addition

After -- unpacked into subtract, decompose, remaining compose-by-addition

If time:
DataShop Demo and/or Video

- See video on “about” page
- “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)

Knowledge Decomposability 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!

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
END