15-780 - GRADUATE ARTIFICIAL INTELLIGENCE
AI AND EDUCATION I

Shayan Doroudi
April 24, 2017
Series on applications of AI to education.

**Lecture**  **Application**

4/24/17  Learning
4/26/17  Assessment
5/01/17  Instruction
Series on applications of AI to education.

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Application</th>
<th>AI Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/24/17</td>
<td>Learning</td>
<td>Machine Learning + Search</td>
</tr>
<tr>
<td>4/26/17</td>
<td>Assessment</td>
<td>Machine Learning + Mechanism Design</td>
</tr>
<tr>
<td>5/01/17</td>
<td>Instruction</td>
<td>Multi-Armed Bandits</td>
</tr>
</tbody>
</table>
Series on applications of AI to education.

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Application</th>
<th>AI Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Time</td>
<td>Learning</td>
<td>Machine Learning + Search</td>
</tr>
<tr>
<td>4/26/17</td>
<td>Assessment</td>
<td>Machine Learning + Mechanism Design</td>
</tr>
<tr>
<td>5/01/17</td>
<td>Instruction</td>
<td>Multi-Armed Bandits</td>
</tr>
</tbody>
</table>
• 1956: Dartmouth Workshop on AI.

*The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.*
• 1956: Dartmouth Workshop on AI.
• Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.
HISTORY OF AI AND EDUCATION AT CMU

• 1956: Dartmouth Workshop on AI.
• Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.
HISTORY OF AI AND EDUCATION AT CMU

• 1956: Dartmouth Workshop on AI.
• Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.
• John Anderson and Albert Corbett develop LISPITS in 1983.
HISTORY OF AI AND EDUCATION AT CMU

• 1956: Dartmouth Workshop on AI.
• Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.
• John Anderson and Albert Corbett develop LISPITS in 1983.
• Carnegie Learning founded in 1998 (including co-founders John Anderson and Ken Koedinger), which has taught math to over half a million students.
What We Know About Learning*

ABSTRACT

by copying their professors’ lectures. In spite of the invention of
printing not too long thereafter, students still continued to behave
in their classes as copyists — audiously taking notes, recording the
difficult words of professors as if they didn’t know printing had
been invented and was available. I have heard that there are some
universities where this happens even today.

1987
Learning mathematics from examples and by doing

Ximing Zhu
Carnegie Mellon University

Herbert Alexander Simon
Artificial Intelligence and Psychology Project.

Situated Learning and Education

JOHN R. ANDERSON  LYNNE M. REDER  HERBERT A. SIMON

Applications and Misapplications of Cognitive Psychology to Mathematics Education

John R. Anderson
Lynne M. Reder
Herbert A. Simon

Radical Constructivism and Cognitive Psychology

JOHN R. ANDERSON, LYNNE M. REDER, and HERBERT A. SIMON

Those who believe that education needs a foundation in the
modern science of cognitive psychology sometimes feel
that they are jousting with windmills. Virtually every educational
movement, whatever its merits, claims to have a scientific basis. However,
this is often not the case.
APPLICATIONS OF AI TO LEARNING
• Power Law: \( P = aT^b \)
  • \( P \) = performance (error rate, reaction time)
  • \( T \) = number of trials/opportunities
  • \( a, b \) constants
• Log-log form: \( \log P = b \log(T) + \log(a) \)

(Content of these slides taken and modified from Ken Koedinger's slides
www.learnlab.org/opportunities/summer/presentations/2012/2.Learning-curves2.ppt)
• Newell and Rosenbloom (1981) tested fits of various models to learning curves and gave explanation for power law of practice.
• Newell and Rosenbloom (1981) tested fits of various models to learning curves and gave explanation for power law of practice.

• Heathcote, Brown, and Mewhort (2000) give alternative explanation:
  • Each student's practice is better fit by an exponential curve
  • Aggregation of them fit a power law curve
How can we apply learning curves to model a student's learning in an intelligent tutoring system?
How can we apply learning curves to model a student's learning in an intelligent tutoring system?

• There may be individual differences in students.

(additive factors model (AFM))
How can we apply learning curves to model a student's learning in an intelligent tutoring system?

• There may be individual differences in students.

• Students learn different skills at different rates.
ADDITIVE FACTORS MODEL (AFM)

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

• There may be individual differences in students.

• Students learn different skills at different rates.

• Different problems may share some of the same skills.
How can we apply learning curves to model a student's learning in an intelligent tutoring system?

• There may be individual differences in students. ($\theta_i$: Ability of student $i$)

• Students learn different skills at different rates.

• Different problems may share some of the same skills.
How can we apply learning curves to model a student's learning in an intelligent tutoring system?

- There may be individual differences in students. ($\theta_i$: Ability of student $i$)
- Students learn different skills at different rates. ($\beta_k$: learning rate of skill $k$)
- Different problems may share some of the same skills.
How can we apply learning curves to model a student's learning in an intelligent tutoring system?

- There may be individual differences in students. ($\theta_i$: Ability of student $i$)
- Students learn different skills at different rates. ($\beta_k$: learning rate of skill $k$)
- Different problems may share some of the same skills. ($Q$ matrix: maps problems to skills)
### Q MATRIX

<table>
<thead>
<tr>
<th>Items</th>
<th>Add</th>
<th>Sub</th>
<th>Mul</th>
<th>Div</th>
</tr>
</thead>
<tbody>
<tr>
<td>a*b</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>a*b + c</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>a*b - c</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>c + a*b</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
• $p_{ij,T}$: Probability that student $i$ answers question $j$ correctly at opportunity $T$. 
ADDITIVE FACTORS MODEL (AFM)

- $p_{ij,T}$: Probability that student $i$ answers question $j$ correctly at opportunity $T$.
- AFM: $\log \left( \frac{p_{ij,T+1}}{1-p_{ij,T+1}} \right) = \theta_i + \sum_k Q_{jk}(\beta_k + \gamma_k T)$
ADDITIVE FACTORS MODEL (AFM)

- $p_{ij,T}$: Probability that student $i$ answers question $j$ correctly at opportunity $T$.
- AFM: $\log \left( \frac{p_{ij,T+1}}{1-p_{ij,T+1}} \right) = \theta_i + \sum_k Q_{jk}(\beta_k + \gamma_k T)$
- Poll: Which of the following is true about this model?
  - It is a linear regression model.
  - It is a logistic regression model.
  - It follows a power law of practice for $P = \log \left( \frac{p_{ij,T+1}}{1-p_{ij,T+1}} \right)$.
  - It follows an exponential law of practice for $P = \log \left( \frac{p_{ij,T+1}}{1-p_{ij,T+1}} \right)$. 
Many curves show a reasonable decline

Some do not => Opportunity to improve model!
LEARNING FACTORS ANALYSIS (LFA)

- Method for automatically improving a cognitive model.
• Method for automatically improving a cognitive model.
• Inputs: a cognitive model (Q matrix), a model with hypothesized new skills (P matrix), and student log data.
LEARNING FACTORS ANALYSIS (LFA)

- Method for automatically improving a cognitive model.
- Inputs: a cognitive model (Q matrix), a model with hypothesized new skills (P matrix), and student log data.
- Outputs: Cognitive models that fit the data best along with parameter estimates and model fits for those models.
### Q Matrix

<table>
<thead>
<tr>
<th>Items</th>
<th>Add</th>
<th>Sub</th>
<th>Mul</th>
<th>Div</th>
</tr>
</thead>
<tbody>
<tr>
<td>a*b</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>a*b + c</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>a*b - c</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>c + a*b</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

### P Matrix

<table>
<thead>
<tr>
<th>Items</th>
<th>Multi-Step</th>
<th>Order of Ops</th>
</tr>
</thead>
<tbody>
<tr>
<td>a*b</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a*b + c</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>a*b - c</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>c + a*b</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
We refine our $Q$ matrix by *adding* and/or *splitting* skills.

<table>
<thead>
<tr>
<th>New $Q$ Matrix</th>
<th>Skills</th>
<th>Items</th>
<th>Add</th>
<th>Sub</th>
<th>Mul</th>
<th>Div</th>
<th>Multi-Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a \times b$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a \times b + c$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a \times b - c$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c + a \times b$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We refine our $Q$ matrix by *adding* and/or *splitting* skills.

<table>
<thead>
<tr>
<th>Items</th>
<th>Add</th>
<th>Sub</th>
<th>Mul</th>
<th>Div</th>
<th>Multi-Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a\times b$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$a\times b + c$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$a\times b - c$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$c + a\times b$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
We refine our $Q$ matrix by *adding* and/or *splitting* skills.

### New $Q$ Matrix

<table>
<thead>
<tr>
<th>Items</th>
<th>Add</th>
<th>Sub</th>
<th>Mul-First</th>
<th>Skills</th>
<th>Mul-Second</th>
<th>Div</th>
<th>Multi-Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>a*b</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a*b + c</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>a*b - c</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c + a*b</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
1. Start with original $Q$ matrix.

2. Apply all possible add and split operations using $P$ matrix, evaluate model fit for each model, and add models to frontier.

3. Remove model from frontier with best fit, make that the new $Q$ matrix.

4. Go back to step 2.
1. Start with original $Q$ matrix.

2. Apply all possible add and split operations using $P$ matrix, evaluate model fit for each model, and add models to frontier.

3. Remove model from frontier with best fit, make that the new $Q$ matrix.

4. Go back to step 2.

What is the goal node?
• Log likelihood $l(\theta)$?
• Akaike Information Criterion (AIC): $2k - 2l(\theta)$, where $k$ is number of parameters.
• Akaike Information Criterion (AIC): $2k - 2l(\theta)$, where $k$ is number of parameters.
• Bayesian Information Criterion (BIC): $Nk - 2l(\theta)$, where $N$ is number of observations.
MODEL FIT

- Akaike Information Criterion (AIC): $2k - 2l(\theta)$, where $k$ is number of parameters.
- Bayesian Information Criterion (BIC): $Nk - 2l(\theta)$, where $N$ is number of observations.
- Cross-Validated Root Mean Squared Error
  - Ideal, but takes a lot longer to compute.
LEARNING FACTORS ANALYSIS (LFA)

LFA implements which of the following search algorithms?

- Uniform Cost Search
- Greedy (Best-First) Search
- A* Search
- None of the above
- Beats me
• Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.
SUMMARY

• Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.

• A combination of cognitive science/domain knowledge and machine learning can be used to model student learning.
Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.

A combination of cognitive science/domain knowledge and machine learning can be used to model student learning.

A combination of cognitive science/domain knowledge and AI can be used to automatically refine cognitive models.
SUMMARY

• Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.

• A combination of cognitive science/domain knowledge and machine learning can be used to model student learning.

• A combination of cognitive science/domain knowledge and AI can be used to automatically refine cognitive models.

• Next time: how statistics/machine learning and AI has been used to model and improve assessment of student knowledge.