Student Variability and Automated Instructional Policies

Emma Brunskill

Carnegie Mellon
69 Million
1/6 to 1/4 of 1\textsuperscript{st} year college students take at least 1 remedial class
The instructors

Sebastian Thrun
Sebastian Thrun is a Research Professor of Computer Science at Stanford University, a Google Fellow, a member of the National Academy of Engineering and the German Academy of Sciences. Thrun is best known for his research in robotics and machine learning.

Peter Norvig
Peter Norvig is Director of Research at Google Inc. He is also a Fellow of the American

Check out our new classes at www.udacity.com!

Class has ended, but you can still log in here. Sign in Visitor

The subjects are clearly explained and taught with a lot of enthusiasm...I really appreciate what you are doing and I quite enjoy the lectures. Thanks for all the effort you put into this.

Elias

Enrollment is closed. We hope to offer more online classes in the future, and you can watch the lectures for this course on youtube.

A bold experiment in distributed education, "Introduction to Artificial Intelligence" will be offered free and online to students worldwide from October 10th to December 18th 2011. The course will include feedback on progress and a statement of accomplishment. Taught by Sebastian Thrun and Peter Norvig, the curriculum draws from that used in Stanford's introductory Artificial Intelligence course. The instructors will offer similar materials, assignments and exams.
Up to Two Standard Deviations Better Outcomes
Intelligent Tutoring System ≠ Sequential Decision making Under uncertainty ≠ Individual Variation in Learning Matters ≠ Leveraging Crowds
• Depend on student’s learning state
• Existing approaches often myopic: maximize immediate learning gain
State Exam Passed!

• Long term objectives
State Exam Passed!

- Adaptive, individualized, tutor decision policies
- Depend on student’s learning state & the objective
Sequential Decision Making Under Uncertainty

Find the sum:
\[
\frac{1}{6} + \frac{2}{6}
\]

\[
\frac{4}{6}
\]
Sequential Decision Making Under Uncertainty

Find the sum: $\frac{1}{6} + \frac{2}{6}$
Sequential Decision Making Under Uncertainty

Find the sum: \( \frac{1}{6} + \frac{2}{6} \)

4/6

Observation

Intelligent Tutoring System

Student mental state estimation

decision policy

Action
Intelligent Tutoring System ≠ Sequential Decision making Under uncertainty

Individual Variation in Learning Matters

Leveraging Crowds
To Create a Policy, Need Model of Student Learning
Existing Work

Same learning parameters
Proposed Model

- Student #1 parameters
- Student #2 parameters
- Student #3 parameters
Student Model: Knowledge Tracing

Not Mastered → Not Mastered → Mastered

Wrong → Correct → Wrong

Hidden state

Observation

Corbett & Anderson (1995)
2 State, 2 Observation HMM

Corbett & Anderson (1995)
Bayesian Filtering

\[ p(L_t) = P(\text{mastered skill} \mid \text{first } t \text{ observations}) \]

Corbett & Anderson (1995)
Parameter Estimation

Student Data

Population Model

\(< p(L_0), p(T), p(S), p(G) > \)
Parameter Estimation

Population Model

\[ < p(L_0), p(T), p(S), p(G) > \]

Individual Models

\[ < p_1(L_0), p_1(T), p_1(S), p_1(G) > \]
\[ < p_2(L_0), p_2(T), p_2(S), p_2(G) > \]
\[ < p_3(L_0), p_3(T), p_3(S), p_3(G) > \]
Parameter Estimation

Population Model

Individual Models

- Expectation maximization
- Brute force search
Model Evaluation

Population Model

Individual Models

• Standard approaches
  – Likelihood ratio test
  – Predictive accuracy
New Approach to Model Evaluation: Does this Change How to Teach?

Student #1 parameters

Student #2 parameters

Student #3 parameters
Objective: Mastery learning

• Popular objective
• Provide practice opportunities until probability student has mastered topic exceeds threshold
Expected Necessary Practice: Individual Model

• Given the estimated student parameters $p_i$
• Compute the expected number of practice opportunities a student needs for her to reach mastery
• Define “mastery” to be $p_i(L_t) \geq 95$
• Continuous-state Markov decision process policy evaluation
Recursively Compute Expected # Practice Opportunities Needed

\[
EO_i(p_i(L_t)) = \begin{cases} 
0 & \text{if } p_i(L_t) \geq d \\
\text{Probability of mastery has reached threshold} & 
\end{cases}
\]
Recursively Compute Expected # Practice Opportunities Needed

\[ EO_i(p_i(L_t)) = \begin{cases} 
0 & \text{if } p_i(L_t) \geq d \\
1 + p_i(c \mid p_i(L_t)) \cdot EO_i(p_i(L_{t+1} \mid c)) + p_i(w \mid p_i(L_t)) \cdot EO_i(p_i(L_{t+1} \mid w)) & \text{otherwise}
\end{cases} \]

Probability of mastery has reached threshold
Expected Necessary Practice: Individual Model

\[ p_i(L_1) \geq d \text{ ?  (d=threshold)} \]
Expected Necessary Practice: Individual Model

\[ p_i(L_1) \geq d \]?

- yes → Stop
- 0 Practice Opportunity needed
Expected Necessary Practice: Individual Model

$p_i(L_1) \geq d$ ?

- **yes**
  - Stop
  - 0 Practice Opportunity needed

- **no**
  - Give practice
    - correct
      - $p_i(c | p_i(L_1))$
      - $EO_i(p_i(L_2 | c))$
    - wrong
      - $p_i(w | p_i(L_1))$
      - $EO_i(p_i(L_2 | w))$
Expected Necessary Practice: Individual Model

\[ EO_i(p_i(L_t)) = \begin{cases} 
  0 & \text{if } p_i(L_t) \geq d \\
  1 + p_i(c|p_i(L_t)) \cdot EO_i(p_i(L_{t+1}|c)) + p_i(w|p_i(L_t)) \cdot EO_i(p_i(L_{t+1}|w)) & \text{otherwise} 
\end{cases} \]

Probability of mastery has reached threshold

0 Practice Opportunity needed

Stop

Give practice

Correct

Wrong

\[ p_i(L_1) \geq d ? \]

yes

no
Expected Necessary Practice: Individual Model

\[ p_i(L_1) \geq d ? \]

- Yes: Stop
  - 0 Practice Opportunity needed
- No: Give practice
  - Correct: \[ p_i(c|p_i(L_1)) \]
  - Wrong: \[ p_i(w|p_i(L_1)) \]

- \[ EO_i(p_i(L_2|c)) \]
  - Yes: Stop
  - No: \[ EO_i(p_i(L_2|c)) = 0 \]

- \[ p_i(L_2|c) \geq d ? \]
  - Yes: Stop
  - No: \[ EO_i(p_i(L_2|c)) = 0 \]

- \[ EO_i(p_i(L_2|w)) \]
  - Yes: Stop
  - No: \[ EO_i(p_i(L_2|w)) = 0 \]

Keep expanding the tree recursively.
Expected Necessary Practice: Individual Model

\[ < p_i(L_o), p_i(T), p_i(S), p_i(G) > \]
Expected Necessary Practice: Population Model

\[ \langle p_{\text{pop}}(L_0), p_{\text{pop}}(T), p_{\text{pop}}(S), p_{\text{pop}}(G) \rangle \]

\[ \langle p_i(L_0), p_i(T), p_i(S), p_i(G) \rangle \]
Expected Necessary Practice: Population Model

\[ <p_{\text{pop}}(L_0), p_{\text{pop}}(T), p_{\text{pop}}(S), p_{\text{pop}}(G)> \]

\[ p_{\text{pop}}(L_t) \]
Recursively Compute Expected # Practice Opportunities Needed

\[ EO = \begin{cases} 
0 & \text{if } p_{pop}(L_j) \geq d \\
\end{cases} \]
Expected Necessary Practice: Population Model

$p_{pop}(L_1) \geq d$ ?

- yes: Stop
  - 0 Practice Opportunity needed
- no: Give practice
  - correct
  - wrong
Expected Necessary Practice: Population Model

\[ p_{\text{pop}}(L_1) \geq d \] ?

- yes
  - Stop
  - 0 Practice Opportunity needed

- no
  - Give practice
    - correct
      - \( p_i(c \mid L_1) \)
    - wrong
      - \( p_i(w \mid L_1) \)
Expected Necessary Practice: Population Model

Given: \( p_{\text{pop}}(L_1) \geq d \)?

- yes: Stop
  - 0 Practice Opportunity needed
- no: Give practice
  - correct
    - \( p_i(c | L_1) \)
    - \( p_i(L_2 | c) \)
    - \( p_{\text{pop}}(L_2 | c) \)
  - wrong
    - \( p_i(w | L_1) \)
    - \( p_i(L_2 | w) \)
    - \( p_{\text{pop}}(L_2 | w) \)
Recursively Compute Expected # Practice Opportunities Needed

\[ EO = \begin{cases} 
0 & \text{if } p_{pop}(L_j) \geq d \\
1 + p(c|p_i(L_j)) \ast EO_c + p(w|p_i(L_j)) \ast EO_w & \text{otherwise}
\end{cases} \]
Does using the population model predict a significantly different number of expected practice opportunities compared to using the individual model?
Experiments

• Dataset on math problems (7\textsuperscript{th}-12\textsuperscript{th} grade) from ASSISTment system (Pardos & Heffernan, 2010)
• 42 problem sets, each corresponding to 1 skill
• Selected a subset of 265 students who did problems on 10 or more skills
• For student $i$, computed set of parameters for all skills she did problems on
  – Approximation, but multiple skills often modeled using the same parameters
Student Parameters: Init & Learn

![Histogram of Initial Probability](image1)

![Histogram of Probability of Learning](image2)
Student Parameters: Init & Learn

- Individual models provide a significantly better data fit than population model (LRT, p<=0.0001)
Student Parameters: Init & Learn

Does this variation matter?

- Individual models provide a significantly better data fit than population model (LRT, $p<=0.0001$)
Amount of Practice if Use Population Instead of Individual Params
Ratio of the expected number of practice opportunities
+20% expected to be made to do at least double the practice they need
Expected mastery
17% of students <=60% prob of mastery when mastery declared

![Histogram showing expected mastery and frequency of occurrence.](chart.png)
Individual Variability

• Significant variability in student parameters
• Experiments suggest this variability could impact instructional policy for +1/3 of students
• New method for evaluating student model variations by their effect on policy

best paper nominee, Lee & Brunskill, Conference on Education Data Mining 2012
Intelligent Tutoring System ≠ Sequential Decision making Under uncertainty $\neq$ Individual Variation in Learning Matters $\neq$ Leveraging Crowds
Looking Forward

<???>

< p_i(L_o), p_i(T), p_i(S), p_i(G) >
Crowds of Student Data
Intelligent Tutoring System

Sequential Decision making Under uncertainty ≠ Individual Variation in Learning Matters

Leveraging Crowds
To Achieve Scalable, Quality Education for All
Acknowledgements

• Graduate student Jung In Lee
• Pittsburgh Science of Learning Center
Intelligent Tutoring System ≠ Sequential Decision making Under uncertainty ≠ Individual Variation in Learning Matters ≠ Leveraging Crowds