Effort-based Tutoring: an Empirical Approach to Intelligent Tutoring

Ivon Arroyo¹, Hasmik Mehranian², Beverly P. Woolf¹

¹ Department of Computer Science
² Dept. of Mechanical and Industrial Engineering
University of Massachusetts Amherst
Making Wayang Math Tutor Smarter
From past student logs

Multimedia Adaptive Tutoring System for Standardized-Tests Math Problems

When a certain rectangle is divided in half, two squares are formed. If each of the squares has perimeter 48, what is the perimeter of the original rectangle?

A 96
B 72
C 36
D 24
E 12
What this paper is about

- A description of how we made the Wayang Tutor Smarter
  - From past student data
  - Given Content=Cases that students needed to know to solve
  - A wide range of students in public schools

- A concrete procedure to add smartness to any ILE
  - From past student data
  - Even for ill-defined domains
  - Regardless of amount of variety of content
  - High level of detail for replication

- Unveiling parameters that regulate ITS functioning
  - For optimization
What do students do in Wayang?

Three different ways to express their effort

Frequency of Behavior on one “easy” math problem in Wayang Outpost

484 cases

Incorrect Attempts

Hints

Time (each bar=5 seconds)

>=2.8min

What is x + y?

30°
What do students do in Wayang?
Three different ways to express their effort

Frequency of Behavior on one “harder” math problem in Wayang Outpost

Incorrect Attempts
(or violated constraints, etc.)

Hints

Time (each bar=5 seconds)

529 cases

In rectangle $ABCD$, sides $AB$ and $CD$ pass through the centers of the two circles. If $AB=12$ and $AD=16$ what is the area of the shaded region?
What is expected behavior?

In any problem $p_i \ i=1, \ldots, N \ N=$Total problems in system

Incorrect Attempts

Hints

Time (each bar=5seconds)

$E(I_i) = \text{mean (or median) Incorrect Attempts for problem } p_i$

$\delta_{IH} = \text{SD}(I_i) \times \theta_{\text{HIGH}} \quad \text{e.g. } \theta_{\text{HIGH}} = 0.5$

$\delta_{IL} = \text{SD}(I_i) \times \theta_{\text{LOW}} \quad \text{e.g. } \theta_{\text{LOW}} = 0.25$

Parameters
What is expected behavior?

In any problem $p_i$, $i=1, \ldots, N$  
$N=$Total problems in system

Incorrect Attempts

Hints

Time (each bar=5 seconds)

Within expected behavior

A new student encounters this problem…
Is their behavior within expectation, or atypical?
What is odd behavior?

In any problem $p_i$, $i=1, \ldots, N$  
$N=$Total problems in system

Odd behavior

Attempts < $E(I_i)$ — $\delta_{IL}$
Few Inc. Attempts

Hints > $E(H_i) + \delta_{HH}$
Lots of Hints

Time < $E(T_i)$ — $\delta_{TL}$
Little Time
## Pedagogical Moves in Wayang Outpost

### Student Model
Estimate most likely scenario for student on p_i

<table>
<thead>
<tr>
<th>Mistakes</th>
<th>Hints</th>
<th>Time</th>
<th>Most Likely</th>
<th>Pedagogical Model Moves Cognitive or Affective or Metacognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Decision</td>
<td>Other Actions</td>
</tr>
<tr>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>Mastery without effort</td>
<td>Increase Problem Difficulty Show learning progress</td>
</tr>
<tr>
<td>2</td>
<td>&lt;</td>
<td>&gt;</td>
<td>Toward</td>
<td>Maintain Problem Difficulty Affective feedback: Praise Effort</td>
</tr>
<tr>
<td>3</td>
<td>&lt;</td>
<td>&gt;</td>
<td>Mastery, low effort</td>
<td>Reduce Problem Difficulty Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>4</td>
<td>&lt;</td>
<td>&gt;</td>
<td>Towards mastery, effort</td>
<td>Maintain Problem Difficulty Praise effort</td>
</tr>
<tr>
<td>5</td>
<td>&gt;</td>
<td>&lt;</td>
<td>Quick guessing, low effort</td>
<td>Reduce Problem Difficulty Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>6</td>
<td>&gt;</td>
<td>&gt;</td>
<td>Hint avoidance and high effort</td>
<td>Reduce Problem Difficulty Offer hints upon incorrect answer in the next problem</td>
</tr>
<tr>
<td>7</td>
<td>&gt;</td>
<td>&lt;</td>
<td>Quick guess and hint abuse</td>
<td>Reduce Problem Difficulty Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>8</td>
<td>&gt;</td>
<td>&gt;</td>
<td>Low mastery and High Effort</td>
<td>Reduce Problem Difficulty Emphasize importance of effort and perseverance</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>Otherwise Expected Behavior</td>
<td>Maintain Problem Difficulty</td>
</tr>
</tbody>
</table>

- **Mistakes:** Number of mistakes made by the student.
- **Hints:** Number of hints given by the instructor.
- **Time:** Time taken by the student to complete the task.
- **Most Likely:** Description of the student's behavior and likely outcome.
- **Decision:** Strategy to adopt based on the student's behavior.
- **Other Actions:** Additional steps to take, if any.
## Pedagogical Moves in Wayang Outpost

<table>
<thead>
<tr>
<th>Student Model</th>
<th>Pedagogical Model Moves</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cognitive or Affective or Metacognitive</td>
</tr>
<tr>
<td>Mistakes</td>
<td>Hints</td>
</tr>
<tr>
<td>1</td>
<td>&lt;</td>
</tr>
<tr>
<td>2</td>
<td>&lt;</td>
</tr>
<tr>
<td>3</td>
<td>&lt;</td>
</tr>
<tr>
<td>4</td>
<td>&lt;</td>
</tr>
<tr>
<td>5</td>
<td>&gt;</td>
</tr>
<tr>
<td>6</td>
<td>&gt;</td>
</tr>
<tr>
<td>7</td>
<td>&gt;</td>
</tr>
<tr>
<td>8</td>
<td>&gt;</td>
</tr>
<tr>
<td>9</td>
<td>Otherwise</td>
</tr>
</tbody>
</table>
### Pedagogical Moves in Wayang Outpost

<table>
<thead>
<tr>
<th>Student Model</th>
<th>Pedagogical Model Moves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate most likely scenario for student on p_i</td>
<td>Cognitive or Affective or Metacognitive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mistakes</th>
<th>Hints</th>
<th>Time</th>
<th>Most Likely</th>
<th>Decision</th>
<th>Other Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>Mastery without effort</td>
<td>Increase Problem Difficulty</td>
<td>Show learning progress</td>
</tr>
<tr>
<td>2</td>
<td>&lt;</td>
<td>&lt;</td>
<td>Toward Mastery effort</td>
<td>Maintain Problem Difficulty</td>
<td>Affective feedback: Praise Effort</td>
</tr>
<tr>
<td>3</td>
<td>&lt;</td>
<td>&gt;</td>
<td>Hint avoidance, low effort</td>
<td>Reduce Problem Difficulty</td>
<td>Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>4</td>
<td>&lt;</td>
<td>&gt;</td>
<td>Toward mastery, effort</td>
<td>Maintain Problem Difficulty</td>
<td>Praise effort</td>
</tr>
<tr>
<td>5</td>
<td>&gt;</td>
<td>&lt;</td>
<td>Quick guessing, low effort</td>
<td>Reduce Problem Difficulty</td>
<td>Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>6</td>
<td>&gt;</td>
<td>&lt;</td>
<td>Hint avoidance, high effort</td>
<td>Reduce Problem Difficulty</td>
<td>Offer hints upon incorrect answer in the next problem</td>
</tr>
<tr>
<td>7</td>
<td>&gt;</td>
<td>&gt;</td>
<td>Quick guess and hint abuse</td>
<td>Reduce Problem Difficulty</td>
<td>Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>8</td>
<td>&gt;</td>
<td>&gt;</td>
<td>Low mastery and High Effort</td>
<td>Reduce Problem Difficulty</td>
<td>Emphasize importance of effort and perseverance</td>
</tr>
<tr>
<td>9</td>
<td>Otherwise</td>
<td>Expected Behavior</td>
<td>Maintain Problem Difficulty</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

11
### Pedagogical Moves in Wayang Outpost

<table>
<thead>
<tr>
<th>Student Model</th>
<th>Pedagogical Model Moves</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cognitive or Affective or Metacognitive</td>
</tr>
<tr>
<td>Mistakes</td>
<td>Hints</td>
</tr>
<tr>
<td>1</td>
<td>&lt;</td>
</tr>
<tr>
<td>2</td>
<td>&lt;</td>
</tr>
<tr>
<td>3</td>
<td>&lt;</td>
</tr>
<tr>
<td>4</td>
<td>&lt;</td>
</tr>
<tr>
<td>5</td>
<td>&gt;</td>
</tr>
<tr>
<td>6</td>
<td>&gt;</td>
</tr>
<tr>
<td>7</td>
<td>&gt;</td>
</tr>
<tr>
<td>8</td>
<td>&gt;</td>
</tr>
<tr>
<td>9</td>
<td>Otherwise</td>
</tr>
</tbody>
</table>
# Pedagogical Moves in Wayang Outpost

## Pedagogical Model Moves

<table>
<thead>
<tr>
<th>Student Model</th>
<th>Pedagogical Model Moves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate most likely scenario for student on $p_i$</td>
<td>Cognitive or Affective or Metacognitive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mistakes</th>
<th>Hints</th>
<th>Time</th>
<th>Most Likely</th>
<th>Decision</th>
<th>Other Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt;$</td>
<td>$&lt;$</td>
<td>$&lt;$</td>
<td>Mastery without effort</td>
<td>Increase Problem Difficulty</td>
<td>Show learning progress</td>
</tr>
<tr>
<td>$&lt;$</td>
<td>$&lt;$</td>
<td>$&gt;$</td>
<td>Toward mastery effort</td>
<td>Maintain Problem Difficulty</td>
<td>Affective feedback: Praise Effort</td>
</tr>
<tr>
<td>$&lt;$</td>
<td>$&gt;$</td>
<td>$&lt;$</td>
<td>Hint avoidance, low effort</td>
<td>Reduce Problem Difficulty</td>
<td>Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>$&lt;$</td>
<td>$&gt;$</td>
<td>$&gt;$</td>
<td>Towards mastery, effort</td>
<td>Maintain Problem Difficulty</td>
<td>Praise effort</td>
</tr>
<tr>
<td>$&gt;$</td>
<td>$&lt;$</td>
<td>$&lt;$</td>
<td>Quick guessing, low effort</td>
<td>Reduce Problem Difficulty</td>
<td>Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>$&gt;$</td>
<td>$&lt;$</td>
<td>$&gt;$</td>
<td>Hint avoidance and high effort</td>
<td>Reduce Problem Difficulty</td>
<td>Offer hints upon incorrect answer in the next problem</td>
</tr>
<tr>
<td>$&gt;$</td>
<td>$&gt;$</td>
<td>$&lt;$</td>
<td>Quick guess and hint abuse</td>
<td>Reduce Problem Difficulty</td>
<td>Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>$&gt;$</td>
<td>$&gt;$</td>
<td>$&gt;$</td>
<td>Low mastery and High Effort</td>
<td>Reduce Problem Difficulty</td>
<td>Emphasize importance of effort and perseverance</td>
</tr>
<tr>
<td>Otherwise</td>
<td>Expected Behavior</td>
<td>Maintain Problem Difficulty</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Increasing Problem Difficulty
At the next time step. Assume we know problem difficulty of items.

\[ H = \text{Sorted list of harder math problems} \]

\[ H[1..m], \gamma = \left\lceil \frac{m}{\gamma} \right\rceil \]

Parameter \( \gamma = 3 \) --> Challenge rate
Decreasing Problem Difficulty
At the next time step. Assume we know problem difficulty of items.

\[ E = \text{Sorted list of easier math problems} \]

Parameter \( \gamma = 3 \)
What is Problem Difficulty?
Which problems are easier or harder?

Correctness effort factor
\[ d_{ci} = \frac{E(I_i)}{\text{Max}_{j=1}^{N}(E(I_j))} \quad d_{ci} \in [0,1] \]

Hint effort factor
\[ d_{hi} = \frac{E(H_i)}{\text{Max}_{j=1}^{N}(E(H_j))} \quad d_{hi} \in [0,1] \]

Time effort factor
\[ d_{ti} = \frac{E(T_i)}{\text{Max}_{j=1}^{N}(E(T_j))} \quad d_{ti} \in [0,1] \]

\[ d_i = \text{mean}(d_{ci}, d_{ti}, d_{hi}) \quad d_i \in [0,1] \]

Reasonable… But is it really sound?
Accuracy of Problem Difficulty?

Just too important to get them right
  • They are too relevant to the Pedagogical Model moves

Just too likely they might be wrong
  • They will be biased to the problem selector in place!
    • However, if you have a variety of past problem selectors…

Need criteria that can guarantee they are good-enough
  • What kind of tests can we run?
Are difficulty estimations accurate?
Criteria for evaluating the soundness of estimations

Axiom:

“PAIRS OF SIMILAR PROBLEMS SHOULD HAVE SIMILAR PROBLEM DIFFICULTY ESTIMATES”

\[ p_{003} \]

\[ p_{033} \]

What is \( x + y \)?

What is \( x + y \)?

30 pairs of problems like this. It should happen that: \[ d_{p_i} \approx d_{p_{30+i}} \]
Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

Criteria 1: Similar to each other?

Pearson Correlation: $p<0.0001$, $R=.823$
Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

Criteria 1: Similar to each other?

Criteria 2: Different from rest?

Pearson Correlation: $p<0.0001$, $R=0.823$

Paired samples $t(29)=7.35$, $p<0.000$

\[
\left( d_{pi} - d_{p30+i} \right)^2 < \frac{\sum_{j=1}^{30} \left( d_{pi} - d_{pj} \right)^2}{3N}
\]

<table>
<thead>
<tr>
<th>$d_{pi}$</th>
<th>$d_{p30+i}$</th>
<th>$SE$</th>
<th>$MSE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.28</td>
<td>0.0001</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>30</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Are difficulty estimations accurate?
Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

Criteria 1: Similar to each other? Pearson Correlation: $p<0.0001$, $R=.823$
Criteria 2: Different from rest? Paired samples $t(29)=7.35$, $p<.000$
Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$

Criteria 1: Similar to each other?
Criteria 2: Different from rest?
Criteria 3: Different from each other?

Pearson Correlation: $p<0.0001$, $R=0.823$

Paired samples $t(29)=7.35$, $p<.000$

21 restrictions, $\chi^2(20)=5.25$, $p<0.05$
Are difficulty estimations accurate?

Criteria for evaluating the soundness of $d_i$ in relation to $d_{30+i}$
Progressing Through Knowledge Units

- Intelligent/Adaptive Tutor =
  - Discarding some of the content
    - Some of it is inappropriate
    - Student is ready to move on
  - When has the student seen enough?

- Content Organization
  - Where should they move on to?
  - What are Knowledge Units?
Switching through Knowledge Units

Parameters to Switch Topics or Knowledge Units

Knowledge Units are Chunks of Problems that use similar skills

<table>
<thead>
<tr>
<th>Topic Switch Criterion</th>
<th>Reason</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Topic Mastery was reached (e.g. enough &quot;hard&quot; problems answered correctly)</td>
<td>Cognitive</td>
<td>$M_{KU}$</td>
</tr>
<tr>
<td>2.2 Persistent failure to find a problem of desired difficulty</td>
<td>Content limitation</td>
<td>$F_{KU}$</td>
</tr>
<tr>
<td>2.3 Maximum time in Topic condition, or Maximum Number of Problems allowed</td>
<td>Classroom Implementation</td>
<td>$T_{KU}$, $N_{KU}$</td>
</tr>
</tbody>
</table>
Does it improve learning?
Randomized Controlled Experiment (N=56) Spring 2004

56 high school students from different math classes in a public school
Pretest and Posttest, 9 (medium and hard) SAT-M problems

Experimental

\[
\begin{align*}
\text{Within KU} \\
\gamma &= 2 \\
\theta_{\text{LOW}} &= 0 \\
\theta_{\text{HIGH}} &= 0
\end{align*}
\]

Control

Random

Between KU

\[N_{KU} = \text{fixed value}\]

Only problem difficulty moves, no “other actions”
Does it improve learning?
Randomized Controlled Experiment (N=56) Spring 2004

56 high school students from different math classes in a public school
Pretest and Posttest, 9 (medium and hard) SAT-M problems

ANCOVA for Posttest Score $F(55,1)=8.4$, $p=.006$
Conclusions

- We can discern expected from odd behavior
- We can unveil parameters that regulate ITS functioning
- A procedure to make Learning Environments Smarter
- Tests to evaluate accuracy of problem difficulty
- Evidence that adaptive problem sequencing works
  - Students learned more with adaptive problem selection
  - No learning companions in experiment!
Future Work

- A variety of new behaviors
  - Drawing tools
  - Example viewing
  - Emotion Self-reports

- Combinations of possible behaviors are too many
  - Partial combinations
  - Analyzing clusters of behaviors

- Sequential behavior
  - Motifs and Time-based Patterns
    - Identifying High-Level Student Behavior Using time-based Motif Discovery (Shanabrook et al., EDM 2010)
Effort-based Tutoring: an Empirical Approach to Intelligent Tutoring

Ivon Arroyo¹, Hasmik Mehranian², Beverly P. Woolf¹

¹ Department of Computer Science
² Dept. of Mechanical and Industrial Engineering
University of Massachusetts Amherst
What ITS research needs

- Sophisticated Integrated Student Models (SM)
  - *Discern* between Knowledge, Engagement, Affect, Meta-cognition

- Smarter Pedagogical Models (PM)
  - Juggle a variety of student models
  - Effective pedagogical components: in what situations do they work?

- Evidence of Success at Optimizing Learning
  - From the combination of SM and PM
    - Evaluating PM assuming a perfect SM (Beck, 2000)

- What parameters regulate ITS functioning?
  - For optimization
  - For replication: to add smartness to an ILE?
Why an Empirical Approach to ITS?  
Some reasons

- How students use ILEs is not known at design time
  - We don’t know how students misuse
  - We don’t know how much students misuse

- Because it can help us to see an integral view
  - Misuse can throw off our estimates of student knowledge
    - E.g. incorrect attempts may appear as unknowing
  - Specific behaviors may be attributed to different reasons

- To understand Domain better
  - *Objective* Problem Difficulty $<>$ *Subjective* Problem Difficulty

- We could learn to make the ILE smarter
An Integral Wholesome View
Factors that regulate/mediate the functioning of an ITS

<table>
<thead>
<tr>
<th>Item</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Behavior (Engagement, Use/Misuse)</td>
<td></td>
</tr>
<tr>
<td>Pedagogical Moves</td>
<td></td>
</tr>
<tr>
<td>Knowledge Tracing</td>
<td></td>
</tr>
<tr>
<td>Affect Models Emotions/Motivation</td>
<td></td>
</tr>
<tr>
<td>Metacognition Models</td>
<td></td>
</tr>
<tr>
<td>Interface (HCI) Answer format?</td>
<td></td>
</tr>
<tr>
<td>Content Availability</td>
<td></td>
</tr>
<tr>
<td>Content Quality Help quality</td>
<td></td>
</tr>
<tr>
<td>Item Difficulty</td>
<td></td>
</tr>
</tbody>
</table>

Student Models
The Problem of Intelligent Tutoring

What are we trying to solve?

- Optimizing Learning
  - And Motivation, and Meta-cognition…

What might help us to get us there?

- Modeling Student Knowledge and Others (SM)
- Good Pedagogical Material (Content)
- Good Pedagogical Decisions (PM)
Problems to Implementation of ITS

- Gaming throws off the Knowledge Estimates
- Parameters to Knowledge Tracing
  - Didn’t know exactly what to set
- Ill-defined:
  - Content does not break down nicely into skills
- Step back: Take an Wholesome, Integrated View
  - Can we write a “recipe” to follow to add smartness to an ILE?