Bayesian Networks in Educational Assessment Tutorial

Session IV: ACED: ECD in Action

Russell Almond, Bob Mislevy, David Williamson and Duanli Yan

Unpublished work © 2002-2014 ETS
<table>
<thead>
<tr>
<th>SESSION</th>
<th>TOPIC</th>
<th>PRESENTERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session 1</td>
<td>Evidence Centered Design</td>
<td>David Williamson</td>
</tr>
<tr>
<td>Session 2</td>
<td>Bayesian Networks</td>
<td>Russell Almond</td>
</tr>
<tr>
<td>Session 3</td>
<td>Bayes Net Tools &amp; Applications</td>
<td>Duanli Yan</td>
</tr>
<tr>
<td>Session 4</td>
<td>ACED: ECD in Action</td>
<td>Russell Almond &amp; Duanli Yan</td>
</tr>
<tr>
<td>Session 5</td>
<td>Refining Bayes Nets with Data</td>
<td>Bob Mislevy</td>
</tr>
</tbody>
</table>
Section Outline

• ACED Background
• PM-EM Algorithm for Scoring
• Scoring Exercise
• Weight of Evidence
• Expected Weight of Evidence
• Task Selection Exercise
ACED Background

• ACED (Adaptive Content with Evidence-based Diagnosis)
• Val Shute (PD), Aurora Graf, Jody Underwood, Eric Hansen, Peggy Redman, Russell Almond, Larry Casey, Waverly Hester, Steve Landau, Diego Zapata
• Domain: Middle School Math, Sequences
• Project Goals:
  – Adaptive Task Selection
  – Diagnostic Feedback
  – Accessibility
ACED Features

Valid Assessment. Based on evidence-centered design (ECD).

Adaptive Sequencing. Tasks presented in line with an adaptive algorithm.

Diagnostic Feedback. Feedback is immediate and addresses common errors and misconceptions.

Aligned. Assessments aligned with (a) state and national standards and (b) curricula in current textbooks.
ACED Proficiency Model
Typical Task

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Blue</th>
<th>Red</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>

Katie is a biochemist. During her last trip to the Amazon rainforest, she brought back some leaves from an exotic plant. She extracted a substance from those leaves that had some amazing properties. One drop of the substance on a given cell produced a doubling of the cell, along with a smaller bonus cell (see Stages 1 and 2, below). The same pattern was found in consecutive trials (see Stages 3-4).

She made a table of her findings. Your task is to figure out how many blue, red, and total cells would be present in the 8th sequence. Complete the table by filling in the values for A, B, and C (where N = 8).

Enter the value for A: __________

Enter the value for B: __________

Enter the value for C: __________
ACED Design/Build Process

• Identify Proficiency variables
• Structure Proficiency Model
• Elicit Proficiency Model Parameters
• Construct Tasks to target proficiencies at Low/Medium/High difficulty
• Build Evidence Models based on difficulty/Q-Matrix
Parameterization of Network

• Proficiency Model:
  – Based on Regression model of child given parent
  – SME provided correlation and intercept
  – SME has low confidence in numeric values

• Evidence Model Fragment
  – Tasks Scored Right/Wrong
  – Based on IRT model
  – High/Medium/Low corresponds to $\theta = +1/0/-1$
  – Easy/Medium/Hard corresponds to difficulty $-1/0/+1$
  – Discrimination of 1
  – Used Q-Matrix to determine which node is parent
PM-EM Algorithm for Scoring

- Master Bayes net with just proficiency model (PM)
- Database of Bayes net fragments corresponding to evidence models (EMs), indexed by task ID
- To score a task:
  - Find EM fragment corresponding to task
  - Join EM fragment to PM
  - Enter Evidence
  - Absorb evidence from EM fragment into network
  - Detach EM fragment
An Example

- Five proficiency variables
- Three tasks, with observables \{X_{11}\}, \{X_{21}, X_{22}, X_{23}\}, \{X_{31}\}. 
Q: Which observables depend on which proficiency variables?
A: See the Q-matrix (Fischer, Tatsuoka).

<table>
<thead>
<tr>
<th></th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
<th>$\theta_5$</th>
<th>$X_{23}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{11}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>$X_{21}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$X_{22}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$X_{23}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>$X_{31}$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>--</td>
</tr>
</tbody>
</table>
Proficiency Model / Evidence Model Split

• Full Bayes net for proficiency model and observables for all tasks can be decomposed into fragments.
  ▪ Proficiency model fragment(s) (PMFs) contain proficiency variables.
  ▪ An evidence model fragment (EMF) for each task.
  ▪ EMF contains observables for that task and all proficiency variables that are parents of any of them.
• Presumes observables are conditionally independent between tasks, but can be dependent within tasks.
• Allows for adaptively selecting tasks, docking EMF to PMF, and updating PMF on the fly.
On the way to PMF and EMFs...

Proficiency variables

Observables and proficiency variable parents for the tasks
Marry parents, drop directions, and triangulate (in PMF, with respect to all tasks)
Footprints of tasks in proficiency model
(figure out from rows in Q-matrix)
Result:

- Each EMF implies a join tree for Bayes net propagation.
  - Initial distributions for proficiency variables are uniform.
- The footprint of the PM in the EMF is a clique intersection between that EMF and the PMF.
- Can “dock” EMFs with PMF one-at-a-time, to …
  - absorb evidence from values of observables to that task as updated probabilities for proficiency variables, and
  - predict responses in new tasks, to evaluate potential evidentiary value of administering it.
Docking evidence model fragments

PMF
## Scoring Exercise

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Task Name</th>
<th>Proficiency Variable</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong</td>
<td>tCommonRatio1a.xml</td>
<td>CommonRatio</td>
<td>Easy</td>
</tr>
<tr>
<td>Right</td>
<td>tCommonRatio2b.xml</td>
<td>CommonRatio</td>
<td>Medium</td>
</tr>
<tr>
<td>Wrong</td>
<td>tCommonRatio3b.xml</td>
<td>CommonRatio</td>
<td>Hard</td>
</tr>
<tr>
<td>Wrong</td>
<td>tExplicitGeometric1a.xml</td>
<td>ExplicitGoemetric</td>
<td>Easy</td>
</tr>
<tr>
<td>Right</td>
<td>tExplicitGeometric2a.xml</td>
<td>ExplicitGoemetric</td>
<td>Medium</td>
</tr>
<tr>
<td>Wrong</td>
<td>tExplicitGeometric3b.xml</td>
<td>ExplicitGoemetric</td>
<td>Hard</td>
</tr>
<tr>
<td>Wrong</td>
<td>tRecursiveRuleGeometric1a.xml</td>
<td>RecursiveRuleGeometric</td>
<td>Easy</td>
</tr>
<tr>
<td>Wrong</td>
<td>tRecursiveRuleGeometric2b.xml</td>
<td>RecursiveRuleGeometric</td>
<td>Medium</td>
</tr>
<tr>
<td>Wrong</td>
<td>tRecursiveRuleGeometric3a.xml</td>
<td>RecursiveRuleGeometric</td>
<td>Hard</td>
</tr>
<tr>
<td>Right</td>
<td>tTableExtendGeometric1a.xml</td>
<td>TableGeometric</td>
<td>Easy</td>
</tr>
<tr>
<td>Right</td>
<td>tTableExtendGeometric2b.xml</td>
<td>TableGeometric</td>
<td>Medium</td>
</tr>
<tr>
<td>Right</td>
<td>tTableExtendGeometric3a.xml</td>
<td>TableGeometric</td>
<td>Hard</td>
</tr>
<tr>
<td>Wrong</td>
<td>tVerbalRuleExtendModelGeometric1a.xml</td>
<td>VerbalRuleGeometric</td>
<td>Easy</td>
</tr>
<tr>
<td>Wrong</td>
<td>tVerbalRuleExtendModelGeometric1b.xml</td>
<td>VerbalRuleGeometric</td>
<td>Easy</td>
</tr>
<tr>
<td>Right</td>
<td>tVerbalRuleExtendModelGeometric2a.xml</td>
<td>VerbalRuleGeometric</td>
<td>Medium</td>
</tr>
<tr>
<td>Wrong</td>
<td>tVisualExtendGeometric1a.xml</td>
<td>VisualGeometric</td>
<td>Easy</td>
</tr>
<tr>
<td>Wrong</td>
<td>tVisualExtendGeometric2a.xml</td>
<td>VisualGeometric</td>
<td>Medium</td>
</tr>
<tr>
<td>Wrong</td>
<td>tVisualExtendGeometric3a.xml</td>
<td>VisualGeometric</td>
<td>Hard</td>
</tr>
</tbody>
</table>
Weight of Evidence

- Good (1985)
- $H$ is binary hypothesis, e.g., *Proficiency* > *Medium*
- $E$ is evidence for hypothesis
- Weight of Evidence (WOE) is

$$W(H : E) = \log \frac{\Pr(E \mid H)}{\Pr(E \mid \bar{H})} = \log \frac{\Pr(H \mid E)}{\Pr(H \mid \bar{E})} - \log \frac{\Pr(H)}{\Pr(\bar{H})}$$
Properties of WOE

- “Centibans” (log base 10, multiply by 100)
- Positive for evidence supporting hypothesis, negative for evidence refuting hypothesis
- Movement in tails of distribution as important as movement near center
- Bayes theorem using log odds
Conditional Weight of Evidence

• Can define Conditional Weight of Evidence

\[ W(H : E_2|E_1) = \log \frac{\Pr(E_2|H, E_1)}{\Pr(E_2|\bar{H}, E_1)} \]

• Nice Additive properties

\[ W(H : E_1, E_2) = W(H : E_1) + W(H : E_2|E_1) \]

• Order sensitive

• WOE Balance Sheet (Madigan, Mosurski & Almond, 1997)
## Evidence Balance Sheet

### P(Solve Geom Sequences)

<table>
<thead>
<tr>
<th>Task</th>
<th>Acc</th>
<th>H</th>
<th>M</th>
<th>L</th>
<th>WOE for H vs. M, L</th>
</tr>
</thead>
<tbody>
<tr>
<td>SolveGeometricProblems2a</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SolveGeometricProblems3a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SolveGeometricProblems3b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SolveGeometricProblems2b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VisualExtendTable2a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SolveGeometricProblems1a</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SolveGeometricProblems1b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VisualExtendVerbalRule2a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ModelExtendTableGeometric3a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ExamplesGeometric2a</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VisualExplicitVerbalRule3a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VerbalRuleModelGeometric3a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

63 tasks total

1. Easy
2. Medium
3. Hard
a. Item type
b. Isomorph

Unpublished work © 2006 by Educational Testing Service
Expected Weight of Evidence

When choosing next “test” (task/item) look at expected value of WOE where expectation is taken wrt $P(E|H)$.

$$EW(H : E) = \sum_{j=1}^{n} W(H : e_j) \Pr(e_j | H)$$

where $\{e_j, j = 1, ..., n\}$ represent the possible results.
Calculating EWOE

Madigan and Almond (1996)

- Enter any observed evidence into net

1. Instantiate Hypothesis = True (may need to use virtual evidence if hypothesis is compound)
2. Calculate $P(E_i|H)$ for each candidate item
3. Instantiate Hypothesis = False
4. Calculate $P(E_i|\overline{H})$ for each candidate item
Related Measures

- Value of Information

\[ \text{VoI}(T) = E_T \left[ \max_d E_S u(d, S) - \max_d E_{S|T} u(d, S) \right] \]

- S is proficiency state
- d is decision
- u is utility
Related Measures (2)

- Mutual Information
- Extends to non-binary hypothesis nodes

\[
\sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)}
\]

- Kullback-Liebler distance between joint distribution and independence

\[
\sum_x P(x) \sum_y P(y|x) \log \frac{P(y|x)}{P(y)}
\]
Task Selection Exercise 1

- Use ACEDMotif1.dne
  - Easy, Medium, and Hard tasks for Common Ratio and Visual Geometric
- Use Hypothesis SolveGeometricProblems > Medium
- Calculate EWOE for six observables
- Assume candidate gets first item right and repeat

- Next assume candidate gets first item wrong and repeat
- Repeat exercise using hypothesis SolveGeometricProblems > Low
Task Selection Exercise 2

- Use Network ACEDMotif2.dne
- Select the SolveGeometricProblems node
- Run the program Network>Sensitivity to Findings
- This will list the Mutual information for all nodes

- Select the observable with the highest mutual information as the first task
- Use this to process a person who gets every task right
- Use this to process a person who gets every task wrong
ACED Evaluation

• Middle School Students
• Did not normally study geometric series
• Four conditions:
  – Elaborated Feedback/Adaptive (E/A; n=71)
  – Simple Feedback/Adaptive (S/A; n=75)
  – Elaborated Feedback/Linear (E/L; n=67)
  – Control (no instruction; n=55)
• Students given all 61 geometric items
• Also given pretest/posttest (25 items each)
ACED Scores

- For Each Proficiency Variable
  - Marginal Distribution
  - Modal Classification
  - EAP Score (High=1, Low=-1)
ACED Reliability

<table>
<thead>
<tr>
<th>Proficiency (EAP)</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Solve Geometric Sequences (SGS)</em></td>
<td>0.88</td>
</tr>
<tr>
<td>Find Common Ratio</td>
<td>0.90</td>
</tr>
<tr>
<td>Generate Examples</td>
<td>0.92</td>
</tr>
<tr>
<td>Extend Sequence</td>
<td>0.86</td>
</tr>
<tr>
<td>Model Sequence</td>
<td>0.80</td>
</tr>
<tr>
<td>Use Table</td>
<td>0.82</td>
</tr>
<tr>
<td>Use Pictures</td>
<td>0.82</td>
</tr>
<tr>
<td>Induce Rules</td>
<td>0.78</td>
</tr>
<tr>
<td>Number Right</td>
<td>0.88</td>
</tr>
</tbody>
</table>

- Calculated with Split Halves (ECD design)
- Correlation of EAP score with posttest is 0.65 (close to reliability of posttest)
- Even with pretest forced into the equation, EAP score accounted for 17% unique variance
- Reliability of modal classifications was worse
Effect of Adaptivity

• For adaptive conditions, correlation with posttest seems to hit upper limit by 20 items

• Standard Error of Correlations is large

• Jump in linear case related to sequence of items
Effect of feedback

- E/A showed significant gains
- Others did not
- Learning and assessment reliability!!!!!
Acknowledgements

• Special thanks to Val Shute for letting us use ACED data and models in this tutorial.

• ACED development and data collection was sponsored by National Science Foundation Grant No. 0313202.

• Complete data available at: http://ecd.ralmond.net/ecdwiki/ACED/ACED