Hierarchical Diagnostic Classification Models

Jonathan Templin – University of Georgia
and
Laine Bradshaw – University of Georgia
James Madison University
• Diagnostic Classification Models (DCMs)
  – The Log-linear Cognitive Diagnosis Model (LCDM)
  – Analysis of English Language Rules Assessment
    • LCDM indicated possible attribute hierarchy

• Hierarchical Diagnostic Classification Model (HDCM)
  – Statistical test of hypothesis of attribute hierarchy
  – Simulation: parameter efficacy study; type-I error study; alternate DCM study

• Analysis of English assessment with HDCM
  – Hierarchy or No Hierarchy...
An Analysis of the Examination for the Certification of Proficiency in English (ECPE)

DIAGNOSTIC CLASSIFICATION MODELS

Templin and Bradshaw (2011; NCME)
Diagnostic Classification Models

• DCMs are psychometric models with categorical latent variables
  – The goal is classification
  – Term diagnostic classification model is used to distinguish class of models from common fields where used (as opposed to cognitive diagnosis models)

• Numerous DCMs have been developed
  – Our focus will be the Log-linear Cognitive Diagnosis Model (LCDM)

• The ANOVA model and the LCDM take the same approach
  – Predict a response using dummy coded variables
    • In LCDM dummy coded variables are latent attributes
  – Using a set of main effects and interactions
    • Links attributes to item response

Templin and Bradshaw (2011; NCME)
• The ECPE is a test developed and scored by the English Language Institute of the University of Michigan
  • Measures advanced English ability in respondents for which English is not their first language

• LCDM analysis of grammar section of the ECPE
  • 28 multiple choice items
  • 3 purported attributes: morphosyntactic, cohesive, and lexical rules
    – 19 items measure one attribute
    – 9 items measure two attributes
    – 0 items measure three attributes

Templin and Bradshaw (2011; NCME)
95% of base-rates of profiles indicated a linear attribute hierarchy.

Is remaining 5% meaningful or simply statistical noise?

Templin and Bradshaw (2011; NCME)
Notation Used Throughout Talk

- **Attributes**: \( a = 1, \ldots, A \)

- **Examinees**: \( e = 1, \ldots, E \)

- **Attribute Profiles**: \( \alpha_e = [\alpha_{e1}, \alpha_{e2}, \ldots, \alpha_{eA}] \)
  - \( \alpha_{ea} \) is 0 or 1

- **Latent Classes**: \( c = 1, \ldots, C \)
  - We have \( C = 2^A \) latent classes – one for each possible attribute profile

- **Items**: \( i = 1, \ldots, I \)
  - Today restricted to dichotomous item responses (\( X_{ei} \) is 0 or 1)

- **Q-matrix**: Elements \( q_{ia} \) for an item \( i \) and attribute \( a \)
  - \( q_{ia} \) is 0 or 1

Templin and Bradshaw (2011; NCME)
Item 7 Estimated LCDM Response Function

\[ \text{logit}(X_{e7} = 1 | \alpha_e) = \lambda_{7,0} + \lambda_{7,1,(1)} \alpha_{e1} + \lambda_{7,1,(3)} \alpha_{e3} + \lambda_{7,2,(1,3)} \alpha_{e1} \alpha_{e3} \]

- Morphosyntactic rules (Attribute 1)
- Lexical rules (Attribute 3)

• Parameter estimates:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_{7,0} )</td>
<td>-0.106</td>
<td>0.095</td>
</tr>
<tr>
<td>(\lambda_{7,1,(1)} )</td>
<td>2.855</td>
<td>0.208</td>
</tr>
<tr>
<td>(\lambda_{7,1,(3)} )</td>
<td>0.952</td>
<td>0.144</td>
</tr>
<tr>
<td>(\lambda_{7,2,(1,3)} )</td>
<td>-0.952</td>
<td>0.144</td>
</tr>
</tbody>
</table>

Possible Attribute Patterns

Templin and Bradshaw (2011; NCME)
Do the ECPE Results Indicate an Attribute Hierarchy?

- From the LCDM structural and item parameter estimates, there is evidence of an attribute hierarchy
  - Structural model: 95% of sample with 4 profiles
  - Measurement model: all items measuring 2 attributes had high standard errors or were at boundary values

- Needed: a statistical test for the presence of an attribute hierarchy
  - Hierarchical diagnostic classification model

Templin and Bradshaw (2011; NCME)
ATTRIBUTE HIERARCHIES
AND THE HDCM

Templin and Bradshaw (2011; NCME)
• Often attribute hierarchies exist in education
  – Part of cognitive theory

• Diagnostic modeling methods exist for hierarchies
  – Attribute Hierarchy Method
  – Rule Space Method

• Neither approach allows for:
  – Statistical hypothesis test for attribute hierarchies
  – Analysis of attribute hierarchies with latent class-based DCMs

Templin and Bradshaw (2011; NCME)
The Hierarchical Diagnostic Classification Model

- Whereas the LCDM represented a crossed-factors ANOVA model, the HDCM uses nested factors
  - Profiles not possible are not estimated (no longer $2^A$)
    - Reduces number of structural model parameters (probabilities)
    - Changes nature of item parameters (nested interactions)
  - HDCM is nested within LCDM
    - Allows for hypothesis test for attribute hierarchy

- Under LCDM Item 7:
  \[
  \text{logit}(X_{e7} = 1 | \alpha_e) = \lambda_{7,0} + \lambda_{7,1,(1)} \alpha_{e1} + \lambda_{7,1,(3)} \alpha_{e3} + \lambda_{7,2,(1,3)} \alpha_{e1} \alpha_{e3}
  \]

- Under HDCM Item 7 (Att. 1 nested within Att. 3):
  \[
  \text{logit}(X_{e7} = 1 | \alpha_e) = \lambda_{7,0} + \lambda_{7,1,(3)} \alpha_{e3} + \lambda_{7,2,(1(3))} \alpha_{e1(3)} \alpha_{e3}
  \]
HDCM/LCDM Hypothesis Test for Attribute Hierarchy

• The HDCM with an attribute hierarchy can be phrased as a model nested within the LCDM

• A nested-model comparison test can be constructed
  – Deviance test: -2 difference in model log-likelihood values

• Deviance test does not follow typical Chi-Square distribution
  – HDCM fixes LCDM model parameters at boundaries

• Test is mixture of Chi-Squares
  – Cannot easily be derived analytically
  – Simulation can approximate p-value
  – If naïve test used, likely result is conservative Type-I error rates

Templin and Bradshaw (2011; NCME)
A simulation study was conducted to evaluate the performance of the HDCM:

- Parameter recovery/efficacy
- Type-I Error rates for hypothesis test
- Power for hypothesis test
- Ability of other DCMs to detect hierarchies

Specifics:

- 30 item test with balanced Q-matrix (3 Attributes)
- Parameters fixed at same values across 500 replications
- Three models fit: no hierarchy/partial hierarchy/full hierarchy

HDCM estimated using marginal maximum likelihood in Mplus
HDCM Item Parameter Estimate Efficiency

Bias

Fewer Parameters = Better Estimation Accuracy (exception main effects for partial hierarchy)
HDCM Structural Parameter Estimate Efficiency

Bias

Root Mean Squared Error

Fewer Parameters = Better Estimation Accuracy

Templin and Bradshaw (2011; NCME)
HDCM Correct Classification Rates

Correct Classification Rate

True Model:
- Full
- Partial
- No

Estimated Model:
- Full Hierarchy (4 Profiles)
- Partial Hierarchy (6 Profiles)
- No Hierarchy (8 Profiles)

CCR is unaffected by too many profiles
CCR is reduced by too few profiles
H$_0$: Attribute Hierarchy Exists

Type-I Error Rates for Naïve Test Were Small – Conservative Test

Accurate Type-I Rate

Notes:
- Information criteria (AIC, BIC, SSABIC) correctly picked true model nearly 100%
- Power for detecting alternative hypothesis was 100% in all replications

Templin and Bradshaw (2011; NCME)
DINA and DINO Type-I Error Rates

DINA: Never works
DINO: Works in some situations

Templin and Bradshaw (2011; NCME)
EVALUATING THE ECPE FOR AN ATTRIBUTE HIERARCHY

Templin and Bradshaw (2011; NCME)
Analysis of the ECPE with the HDCM

• The suspected attribute hierarchy in the ECPE was evaluated with the HDCM
  – Morphosyntactic nested within Cohesive nested within Lexical

• Classification of examinees for HDCM and LCDM had high overlap: 93.6% agreement (.909 kappa)

• Results presented:
  – Structural model
  – Example item
  – Hypothesis test results
Most of 5% shifted into Classes 1 and 2

<table>
<thead>
<tr>
<th>Morphosyntactic</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohesive</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Lexical</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Attribute Profile $c$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>$\hat{\eta}_c$ - LCDM (solid)</td>
<td>.301</td>
<td>.129</td>
<td>.012</td>
<td>.175</td>
<td>.009</td>
<td>.018</td>
<td>.011</td>
<td>.346</td>
<td></td>
</tr>
<tr>
<td>$\hat{\eta}_c$ - HDCM (dashed)</td>
<td>.320</td>
<td>.144</td>
<td>-</td>
<td>.184</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.351</td>
<td></td>
</tr>
</tbody>
</table>
Item 7 HCDM and LCDM Estimates

- **Item 7:** \( \logit(X_{e7} = 1 | \alpha_e) = \lambda_{7,0} + \lambda_{7,1,(3)} \alpha_{e3} + \lambda_{7,2,(1(3))} \alpha_{e1(3)} \alpha_{e3} \)
  - Morphosyntactic rules (Attribute 1)
  - Lexical rules (Attribute 3)

- **Parameter estimates:**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>HDCM Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_{7,0} )</td>
<td>-0.040</td>
<td>0.088</td>
</tr>
<tr>
<td>( \lambda_{7,1,(3)} )</td>
<td>0.924</td>
<td>0.140</td>
</tr>
<tr>
<td>( \lambda_{7,2,(1(3),3)} )</td>
<td>1.949</td>
<td>0.229</td>
</tr>
</tbody>
</table>

HDCM does not model [1,,0] due to hierarchy
Hypothesis Test for Attribute Hierarchies

• Null hypothesis: HDCM model (full hierarchy)
  – \(-2\times\text{LogLikelihood} = 85,502.63\); 68 Parameters Estimated

• Alternative hypothesis: LDCM model (no hierarchy)
  – \(-2\times\text{LogLikelihood} = 85,479.02\); 81 Parameters Estimated

• Naïve deviance test: \(\chi^2 = 23.06, df = 13, p = 0.039\)

• Simulation based p-value: \(p = 0.013\)

• AIC picks HDCM; BIC picks HDCM

• Conclusion: HDCM has an attribute hierarchy present
Concluding Remarks

• Attribute hierarchies are present in many data sets
  – Can be indicative less dimensionality
  – More reasons not to use DINA (or DINO)

• Until the HDCM, latent class-based DCMs have not been able to be adapted for hierarchies
  – No model was able to test for existence of hierarchy

• The HDCM fills these gaps and can help researchers test theories about nature of hypotheses

Templin and Bradshaw (2011; NCME)
Additional Remarks

• DCMs are statistical models that can be used to test for the presence of attribute hierarchies
  – Categorical attributes can show hierarchical structures

• IRT models cannot find attribute hierarchies
  – Hierarchy must exist on a single dimension
  – Determination of where cut-points occur is post-hoc
  – No ability to falsify hypothesis of hierarchy being present

• HOWEVER: If a hierarchy is present
  – DCM may not be most appropriate model
    • LCDM has too many dimensions – may just need single continuous
  – Other models may be used instead of HDCM (Rule Space or Attribute Hierarchy Method)
    • More research needed to make comparisons
    • Use HDCM to confirm/disconfirm hierarchy first

Templin and Bradshaw (2011; NCME)
Thank you!

• Questions? Comments? Paper? References?

jtemplin@uga.edu

• Complaints? Criticisms?

laineb@uga.edu

Templin and Bradshaw (2011; NCME)