Improving Attribute Classification Accuracy in High Dimensional Data: A Four-Step Latent Regression Approach

Jimmy de la Torre

Division of Learning, Development & Diversity
Faculty of Education
The University of Hong Kong

November 1, 2018

Acknowledgment

This work is a collaboration with Yan Sun of Rutgers, The State University of New Jersey.

Outline

Introduction

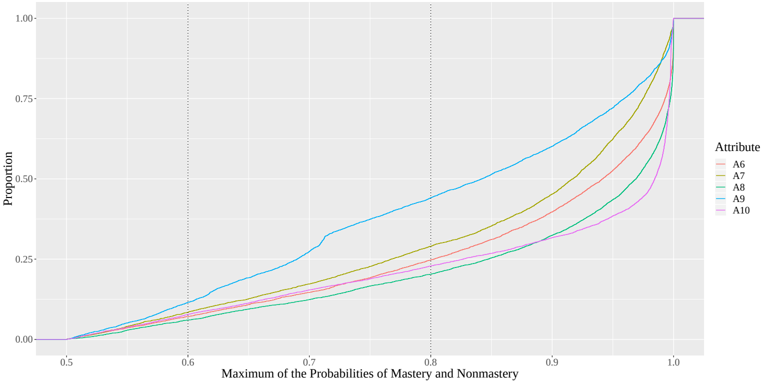
- Cognitive diagnosis models (CDMs) aim to provide detailed and actionable feedback on a set of discrete attributes
- For attributes to be diagnostically informative, they need to be sufficiently fine grained
- However, there is always a trade-off between grain size and the number of attributes - the finer grained the attributes are, the more attributes need to be measured
- Although, in theory the number of attributes that can be estimated by a CDM is unlimited, in practice this number typically cannot exceed 15 due to computational constraints

Introduction

- The accordion procedure (AP) has been proposed to address the issue of high dimensionality in situations where attributes can be partitioned into non-overlapping subsets
- AP focuses on diagnosing one subset of attributes at a time while collapsing attributes from other subsets into coarser attributes
- However, when the number of attributes is large, the posterior probabilities may not be sufficiently high/low to classify all examinees reliably

Introduction

- This graph shows uneven marginal posterior probabilities of 5 out of 24 attributes from a 60-item test



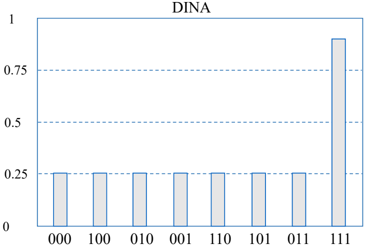
- In this study aims to improve the classification accuracy by using ancillary information

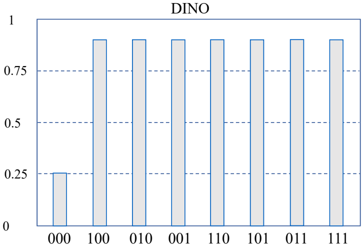
Outline

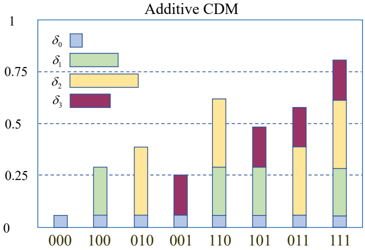
Cognitive Diagnosis Models (CDMs)

- A variety of reduced CDMs have been developed in the psychometric literature
  - The Deterministic Input, Noisy "And" Gate (DINA) model
  - The Deterministic Input, Noisy "Or" Gate (DINO) model
  - The Additive-CDM (A-CDM)
  - ...

Cognitive Diagnosis Models (CDMs)





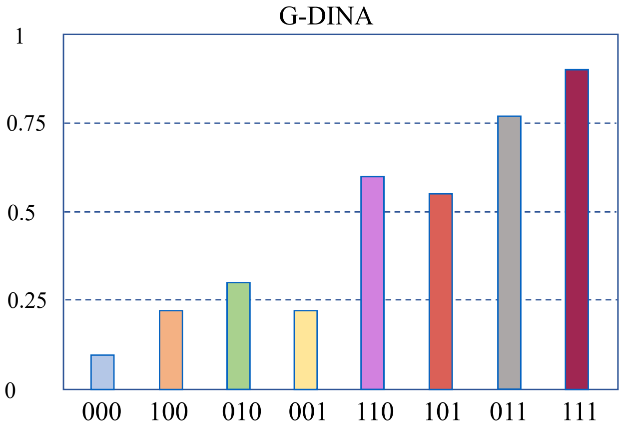


- In addition to specific models, general CDMs have also been developed to accommodate the complexity of real data
- The Generalized DINA (G-DINA) model is an example of general CDMs, and its item response function can be expressed as

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_{lj}) = \delta_{j0} + \sum_{k=1}^{K_j} \delta_{jk}\alpha_{ljk} + \sum_{k'=k+1}^{K_j} \sum_{k=1}^{K_j-1} \delta_{jkk'}\alpha_{ljk}\alpha_{ljk'} + \ldots + \delta_{j12\ldots K_j} \prod_{k=1}^{K_j} \alpha_{ljk},$$

where δ_{j0} is the intercept for item j, δ_{jk} is the kth attribute's main effect, $\delta_{jkk'}$ is the two-way interaction effect, and $\delta_{j12\ldots K_j}$ is the highest-order interaction effect

Cognitive Diagnosis Models (CDMs)



Cognitive Diagnosis Models (CDMs)

- Using other link functions, the G-DINA model is equivalent to the log-linear CDM (LCDM) and the general diagnostic model (GDM)
- Reduced CDMs can be derived from the G-DINA model using appropriate constraints
  - DINA: set all but δ_0 and $\delta_{j12...K_j}$ to 0
  - DINO: set $\delta_{jk} = -\delta_{jkk'} = ... = (-1)^{K_j}\delta_{j12...K_j}$
  - A-CDM, reduced reparameterized unified model (r-RUM) or linear logistic model (LLM): set all interaction effects to 0, using the identity, log, or logit link, respectively

Outline

The Higher-Order Structure

- Mastery or nonmastery of attributes are typically correlated
- The higher-order (HO) structure can be used to model correlated attributes, in which a continuous latent trait θ is posited as the general domain ability
- The HO formulation assumes that the attributes are independent conditional on θ
- This relationship can be expressed as:

$$P(\alpha_k = 1|\theta) = \frac{exp[a_k(\theta - b_k)]}{1 + exp[a_k(\theta - b_k)]}$$

where $P(\alpha_k = 1|\theta)$ represents the probability of mastering attribute k given θ, and a_k and b_k are the slope and the difficulty parameter of attribute k, respectively

The Accordion Procedure

- The AP attempts to address the issue of high dimensionality by partitioning attributes into non-overlapping subsets
- Two examples when AP can be used
  - Different abilities can be assumed to underlie different domains
  - Multiple finer-grained attributes at one time point can be collapsed into a single coarser attribute at another time point
- The AP focuses only on the attributes of one subset at a time, while the attributes of each of the remaining subsets are collapsed to create composite nuisance attributes

The Higher-Order Structure of Accordion Procedure

- Because attributes can be partitioned into multiple domains in AP, the attribute structure can be simplified by positing $\boldsymbol{\theta} = \{\theta_d\}$, where $d = 1, ..., D$
- The probability of mastering attribute k in domain d is

$$P(\alpha_{k(d)} = 1|\theta_d) = \frac{exp[a_{k(d)}(\theta_d - b_{k(d)})]}{1 + exp[a_{k(d)}(\theta_d - b_{k(d)})]}$$

where $a_{k(d)}$ and $b_{k(d)}$ are the slope and the difficulty parameter of attribute k in domain d and $k(d) = 1, ..., K(d)$

Example 1: Cross-Sectional Data

- Suppose there are J items measuring the $\sum_{d=1}^{D} K(d)$ attributes of interests, which can be partitioned into D mutually exclusive domains
- The Q-matrix of the complete knowledge is shown below:

$\sum_{d=1}^{D} K(d)$ attributes

J items

	$Domain_1$				$Domain_d$				$Domain_D$		
Item	α_{11}	…	$\alpha_{1K(1)}$	…	α_{d1}	…	$\alpha_{dK(d)}$	…	α_{D1}	…	$\alpha_{DK(D)}$
1	1	…	0	…	0	…	1	…	1	…	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
J	0	…	1	…	1	…	1	…	1	…	0

Example 1: Cross-Sectional Data

- When focusing on the first domain, the rest of the attributes are collapsed into one composite nuisance attribute per domain
- A variety of rules can be employed in creating the composite attributes (e.g., 1 when at least one attribute is required for an item)

$K(1)+(D-1)$ attributes

J items

	$Domain_1$				$Domain_d$		$Domain_D$
Item	α_{11}	…	$\alpha_{1K(1)}$	…	α_d	…	α_D
1	1	…	0	…	1	…	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
J	0	…	1	…	1	…	1

Example 2: Longitudinal Data

- Suppose the skills can be partitioned into "easy", "medium" and "hard" subsets based on the difficulty of mastering the attributes
- To track student learning progress at three different time points, it is not efficient to diagnose all the skills every time

Time 1

	Easy			*Medium*			*Hard*		
Item	α_{111}	$\cdots$	$\alpha_{11K(1)}$	α_{121}	$\cdots$	$\alpha_{12K(2)}$	α_{131}	$\cdots$	$\alpha_{13K(3)}$
1	1	…	0	0	…	1	1	…	1
⋮									
J									

Time 2

	Easy			*Medium*			*Hard*		
Item	α_{211}	$\cdots$	$\alpha_{21K(1)}$	α_{221}	$\cdots$	$\alpha_{22K(2)}$	α_{231}	$\cdots$	$\alpha_{23K(3)}$
1	1	…	0	0	…	1	1	…	1
⋮									
J									

Time 3

	Easy			*Medium*			*Hard*		
Item	α_{311}	$\cdots$	$\alpha_{31K(1)}$	α_{321}	$\cdots$	$\alpha_{32K(2)}$	α_{331}	$\cdots$	$\alpha_{33K(3)}$
1	1	…	0	0	…	1	1	…	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
J	0	…	1	1	…	1	1	…	0

Example 2: Longitudinal Data

- Instead of diagnosing all skills at the same time, the “easy” skills are examined first, then the “medium” skills are examined a later time, and the “hard” skills are examined at a much later time

Time 1

	Easy			*Medium*	*Hard*
Item	α_{111}	…	$\alpha_{11K(1)}$	α_{12}	α_{13}
1	1	…	0	1	1
⋮					
J					

Time 2

	Easy	*Medium*			*Hard*
Item	α_{21}	α_{221}	…	$\alpha_{22K(2)}$	α_{23}
1	1	0	…	1	1
⋮					
J					

Time 3

	Easy	*Medium*	*Hard*		
Item	α_{31}	α_{32}	α_{331}	…	$\alpha_{33K(3)}$
1	1	1	1	…	1
⋮	⋮	⋮	⋮	⋮	⋮
J	1	1	1	…	0

The Performance of Accordion Procedure

- The performance of AP was examined in a previous study
- The results showed that
  - AP accurately classified a large number of attributes when test length is long or the item is at least of medium quality
  - Compared to complete-profile estimation, AP provided as good a classification, but at a much shorter computing time
- Although AP is promising, long test length or medium item quality may not always be available, and classification accuracy may be uneven
- Hence this study considers the use of covariates to improve the performance of AP

Outline

Incorporating Covariates

- Covariates may be related to the students' attribute mastery and nonmastery classifications
- Previous literature has offered two ways to incorporate covariates: one-step approach and three-step approach
- The one-step approach estimates the CDM and the latent regression model simultaneously, and provides unbiased estimates of the relationship between examinee classification and the covariates
- However, any modifications to either part require refitting the entire model, and the one-step approach is computationally challenging when data are high dimensional

Original Three-Step Approach

- Steps involved in the three-step approach:
  1) fitting a CDM
  2) assigning examinees to latent classes, and
  3) regressing attribute vector $\boldsymbol{\alpha}_l$ (or each attribute α_k) on the covariates
- Regressing estimated attributes or attribute vectors on the covariates leads to biased regression coefficients
- To correct for the bias, the classification uncertainty needs to be taken into consideration

Original Three-Step Approach

- The three-step approach offers great flexibility in customizing the CDM and latent regression model separately, thus:
  - AP can be used in step 1 when the data are high dimensional
  - Regular CDM analyses such as, Q-matrix validation, model selection, or item fit can be conducted, and CDM modifications can be made
  - Adding or dropping covariates can be easily done
  - Variable selection techniques, such as least absolute shrinkage and selection operator or principal component regression, can be applied

Classification Error Probabilities

- Classification error probabilities (CEP) quantifies the amount of error in the classification, and can be either at the vector or attribute level
  - The vector-level CEP (VCEP) is a $2^K \times 2^K$ matrix of $P(\boldsymbol{\alpha}_s|\boldsymbol{\alpha}_l)$, where $\boldsymbol{\alpha}_s$ is the attribute vector assignment, $\boldsymbol{\alpha}_l$ is the true attribute vector, and l, $s = 1, 2, ..., 2^K$
  - The attribute-level CEP (ACEP),is a 2×2 matrix of $P(\alpha_q|\alpha_k)$ where α_k is the true attribute proficiency, α_q is the attribute assignment, and k, $q = 1, 2, ..., K$
- Correction weights can be obtained from VCEP or ACEP depending on either regressing $\boldsymbol{\alpha}_l$ or α_k on the covariates

Attribute Level Classification Error Probabilities

- In a previous study, VCEP and ACEP showed similar performance
- However, when K is large, regressing $\boldsymbol{\alpha}_l$ on the covariates becomes intractable, whereas regressing α_k on the covariates remains manageable
- For example, when $K = 15$, regressing $\boldsymbol{\alpha}_l$ on three covariates requires estimating close to 100,000 parameters, whereas using α_k only requires estimating 60 parameters
- Hence, we focus on the ACEP

Attribute Level Classification Error Proabbilities

- To calculate ACEP, marginal posterior probabilities, $P(\alpha_k|\boldsymbol{Y}_i)$ are calculated by aggregating $P(\boldsymbol{\alpha}_l|\boldsymbol{Y}_i)$ by attribute α_k
- Define $\boldsymbol{P}_{ik}$ as

$$\boldsymbol{P}_{ik} = [1 - P(\alpha_k|\boldsymbol{Y}_i)\ P(\alpha_k|\boldsymbol{Y}_i)]$$

and $\boldsymbol{P}_{iq}$ as

$$\boldsymbol{P}_{iq} = [1 - P(\alpha_q|\boldsymbol{Y}_i)\ P(\alpha_q|\boldsymbol{Y}_i)]$$

where $\boldsymbol{P}_{ik}$ is the vector of marginal probabilities and $\boldsymbol{P}_{iq}$ is a vector of attribute assignment of 0 or 1

Attribute Level Classification Error Probabilities

- ACEP can be computed at sample-level (SL_k) or individual posterior-distribution level (PDL_{ik})
- SL_k utilizes the marginal posterior probabilities of the whole sample and is the same for each examinee, whereas the PDL_{ik} utilizes each examinee's individual marginal posterior probabilities and is unique for each examinee
- SL_k and PDL_{ik} are calculated as:

$$SL_k = \frac{\sum_{i=1}^{N} \boldsymbol{P}_{ik} \boldsymbol{P}'_{iq}}{\sum_{i=1}^{N} \boldsymbol{P}_{ik}}, \qquad PDL_{ik} = \frac{\boldsymbol{P}_{ik} \boldsymbol{P}'_{iq} \times N}{\sum_{i=1}^{N} \boldsymbol{P}_{ik}}$$

Attribute Level Classification Error Probabilities

An example of SL_k:

$$SL_k = \begin{bmatrix} 0.78 & 0.22 \\ 0.23 & 0.77 \end{bmatrix}$$

- Rows represent the true proficiency of α_k and columns represent the attribute assignment α_q
- If the true $\alpha_k = 0$, 78% of the true, the examinee will be classified as nonmaster, and 22% of the true as master
- If the true $\alpha_k = 1$, 23% of the true, the examinee will be classified as nonmaster and 77% of the true as master

If an examinee is classified as nonmaster of attribute k, the first column, $w_{ik} = (0.78\ 0.23)$, is used as correction weights to obtain unbiased estimates of parameters

The Four-Step Approach

1. A CDM is fitted to data
2. Examinee latent classes are assigned, in this study, expected a posterior (EAP) is used
3. Logistic regression is fitted for each attribute

$$P(\alpha_{k(d)}|\boldsymbol{Z}_i) = \frac{exp(\beta_{k(d)0} + \boldsymbol{Z}_i'\boldsymbol{\beta}_{k(d)})}{1 + exp(\beta_{k(d)0} + \boldsymbol{Z}_i'\boldsymbol{\beta}_{k(d)})}$$

where $\boldsymbol{Z}_i$ are covariates, $\beta_{k(d)0}$ is the intercept parameter for $\alpha_{k(d)}$, and $\boldsymbol{\beta}_{k(d)}$ are slop parameters for $\alpha_{k(d)}$

$\hat{\beta}_{k(d)0}$ and $\hat{\boldsymbol{\beta}}_{k(d)}$ can be obtained by optimizing the objective function with the correction weights, as in,

$$logL_k = \sum_{i=1}^{N} log \sum_{\alpha_{k(d)}=0}^{1} P(\alpha_{k(d)}|\boldsymbol{Z}_i) w_{ik}$$

4 To combine information obtained from CDM, $P(\alpha_{k(d)}|\boldsymbol{Y}_i)$, and information from logistic regression, $L(\boldsymbol{Z}_i|\alpha_{k(d)})$, Bayes' theorem is applied:

$$P(\alpha_{k(d)}|\boldsymbol{Y}_i, \boldsymbol{Z}_i) = \frac{L(\boldsymbol{Z}_i|\alpha_{k(d)})P(\alpha_{k(d)}|\boldsymbol{Y}_i)}{\sum_{\alpha_{k(d)}=0}^{1} L(\boldsymbol{Z}_i|\alpha_{k(d)})P(\alpha_{k(d)}|\boldsymbol{Y}_i)}$$

$P(\alpha_{k(d)}|\boldsymbol{Y}_i, \boldsymbol{Z}_i)$ is the the updated posterior for examinee i from which the updated $\hat{\alpha}_{k(d)}$ is obtained

Outline

Goals

A simulation study was conducted to examine

- The extent to which incorporating covariates can improve classification accuracy, and
- How the four-step approach compares to other procedures

Methods

Five approaches are compared:

1) True: True α_k was used in the regression in the third step
2) PDL: Posterior-distribution level correction weight was used in the third step
3) SL: Sample level correction weight was used in the third step
4) UC: No correction was used in the third step
5) AP: Covariates were not incorporated

Design

Manipulated factors

- Number of domains: $D = 2$ and 4
- Number of attributes per domain: $K(d) = 5$ and 8
- Test length: short: $2 \times D \times K(d)$ and long: $4 \times D \times K(d)$
- $N = 500$ and 2000
- Item quality:
  - Low: $P_{\mathbf{0}}{=}U(.25,.35)$, $P_{\mathbf{1}}{=}U(.65,.75)$
  - Medium: $P_{\mathbf{0}}{=}U(.15,.25)$, $P_{\mathbf{1}}{=}U(.75,.85)$
  - High: $P_{\mathbf{0}}{=}U(.05,.15)$, $P_{\mathbf{1}}{=}U(.85,.95)$
- Association between α_k and $\boldsymbol{Z}$: McFadden's pseudo R^2 from regressing α_k on $\boldsymbol{Z}$: Strong (.45) and weak (.05)

Data Generation

HO $\boldsymbol{\theta}$ and covariates $\boldsymbol{Z}$ were generated from

$$(\boldsymbol{\theta}, \boldsymbol{Z}) = (\theta_1, ..., \theta_D, Z_1, Z_2, Z_3) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma}$ was defined depending on the association of $\boldsymbol{\theta}$ and $\boldsymbol{Z}$

- An example when $D = 2$ and association is strong

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1.0 & .40 & .90 & .50 & .20 \\ .40 & 1.0 & .30 & .60 & .80 \\ .90 & .30 & 1.0 & .25 & .25 \\ .50 & .60 & .25 & 1.0 & .25 \\ .20 & .80 & .25 & .25 & 1.0 \end{bmatrix} \Rightarrow \text{Pseudo } R^2 = 0.45$$

  $\boldsymbol{\Sigma}$ can be partitioned into the variance-covariance matrix of $\boldsymbol{\theta}$ (in red), the covariance matrix between $\boldsymbol{Z}$ and $\boldsymbol{\theta}$ (in brown), and the variance-covariance of $\boldsymbol{Z}$ (in blue)

- Weak: the covariances between $\boldsymbol{Z}$ and $\boldsymbol{\theta}$ were set to 0.2

Data Generation

- Q-matrices were created satisfying following requirements:
  1) At least one identity matrix is included
  2) Each attribute is measured the same number of times
  3) Each domain is measured by the same number of items
  4) Q-matrices of long test are obtained by doubling the Q-matrices of short test
- Attributes were generated using the HO model, where $a_{k(d)} = 3.5$ and $b_{k(d)}$ was sampled from $N(0, 0.5)$ across all domains
- Responses were generated using the complete attribute profiles and Q-matrix
- 25 replications for each condition were generated

Q-matrix for $D = 2$, $K(d) = 5$, Short Test

Item	$\alpha_{1(1)}$	$\alpha_{2(1)}$	$\alpha_{3(1)}$	$\alpha_{4(1)}$	$\alpha_{5(1)}$	$\alpha_{1(2)}$	$\alpha_{2(2)}$	$\alpha_{3(2)}$	$\alpha_{4(2)}$	$\alpha_{5(2)}$
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1
11	1	1	0	0	0	1	0	0	0	0
12	0	1	1	0	0	0	1	0	0	0
13	0	0	1	1	0	0	0	1	0	0
14	0	0	0	1	1	0	0	0	1	0
15	1	0	0	0	1	0	0	0	0	1
16	1	0	0	0	0	0	0	0	1	1
17	0	1	0	0	0	0	0	1	1	0
18	0	0	1	0	0	0	1	1	0	0
19	0	0	0	1	0	1	1	0	0	0
20	0	0	0	0	1	1	0	0	0	1

Classification accuracy

- Domain-level correct vector-wise classification accuracy (CVC_d)

$$\text{CVC}_d = \sum_{i=1}^{N} \sum_{d=1}^{D} \frac{I[\boldsymbol{\alpha}_{i(d)} = \hat{\boldsymbol{\alpha}}_{i(d)}]}{ND}$$

- Correct attribute-wise classification accuracy (CAC)

$$\text{CAC} = \sum_{i=1}^{N} \sum_{d=1}^{D} \sum_{k(d)=1}^{K(d)} \frac{I[\alpha_{ik(d)} = \hat{\alpha}_{ik(d)}]}{NDK(d)}$$

Evaluation Indices

Classification Certainty

- The certainty of examinee's attribute classification is examined by checking examinee's posterior probabilities
- Define

$$P^*(\alpha_{k(d)}|\boldsymbol{Y_i}) = max(1 - P(\alpha_{k(d)}|\boldsymbol{Y_i}), P(\alpha_{k(d)}|\boldsymbol{Y_i}))$$

  which indicates the certainty of an examinee being classified as either nonmaster or master
- Different cutoffs, namely, 0.6, 0.7, and 0.8, were used to distinguish certain and uncertain classification as follows:
  - Certain, if $P^*(\alpha_{k(d)}|\boldsymbol{Y}_i) \geq$ Cutoff
  - Uncertain, otherwise
- The proportion of *certain* classifications was computed

Results: Classification Accuracy

- Results of $D = 2$, $N = 500$ and 2000, are presented below
- Similar patterns were found for CVC_d and CAC results, hence only the former rsults are presented

Results: Classification Accuracy

CVC_d, $K(d) = 5$, $N = 500$

J	Item Quality	Assoc.	True	PDL	SL	UC	AP
20	Low	Weak	.22	.20	.20	.20	.20
		Strong	.35	.25	.25	.23	.20
	Medium	Weak	.44	.42	.42	.42	.41
		Strong	.52	.47	.46	.45	.41
	High	Weak	.68	.67	.67	.68	.67
		Strong	.71	.69	.69	.70	.67
40	Low	Weak	.38	.36	.36	.36	.36
		Strong	.51	.45	.44	.41	.35
	Medium	Weak	.64	.64	.64	.64	.63
		Strong	.70	.68	.68	.68	.64
	High	Weak	.86	.86	.86	.86	.86
		Strong	.87	.86	.87	.87	.86

Results: Classification Accuracy

$\text{CVC}_d,\ K(d) = 5,\ N = 2000$

J	Item Quality	Assoc.	True	PDL	SL	UC	AP
20	Low	Weak	.30	.27	.27	.27	.26
		Strong	.49	.39	.38	.33	.26
	Medium	Weak	.52	.51	.51	.51	.50
		Strong	.60	.57	.57	.57	.51
	High	Weak	.72	.72	.72	.72	.72
		Strong	.73	.71	.72	.73	.72
40	Low	Weak	.47	.47	.47	.47	.48
		Strong	.57	.55	.55	.54	.47
	Medium	Weak	.68	.67	.67	.68	.69
		Strong	.72	.70	.71	.72	.69
	High	Weak	.87	.87	.87	.87	.87
		Strong	.88	.87	.87	.88	.88

Results: Classification Accuracy

CVC_d, $K(d) = 8$, $N = 500$

J	Item Quality	Assoc.	True	PDL	SL	UC	AP
32	Low	Weak	.06	.06	.06	.06	.06
		Strong	.09	.07	.07	.07	.06
	Medium	Weak	.24	.23	.23	.23	.23
		Strong	.29	.26	.26	.25	.23
	High	Weak	.53	.52	.52	.52	.51
		Strong	.56	.53	.53	.53	.51
48	Low	Weak	.25	.23	.23	.23	.22
		Strong	.35	.29	.28	.26	.21
	Medium	Weak	.53	.52	.52	.52	.52
		Strong	.58	.55	.55	.55	.50
	High	Weak	.81	.81	.81	.81	.80
		Strong	.81	.80	.80	.80	.79

Results: Classification Accuracy

$\text{CVC}_d,\ K(d) = 8,\ N = 2000$

J	Item Quality	Assoc.	True	PDL	SL	UC	AP
32	Low	Weak	.13	.12	.12	.12	.12
		Strong	.20	.15	.15	.14	.11
	Medium	Weak	.34	.33	.33	.33	.32
		Strong	.42	.38	.37	.37	.31
	High	Weak	.61	.61	.61	.61	.59
		Strong	.64	.62	.62	.63	.60
48	Low	Weak	.35	.35	.35	.35	.34
		Strong	.44	.42	.43	.42	.34
	Medium	Weak	.58	.58	.58	.58	.58
		Strong	.61	.60	.60	.60	.58
	High	Weak	.82	.82	.82	.82	.82
		Strong	.83	.83	.83	.83	.83

Results: Classification Accuracy

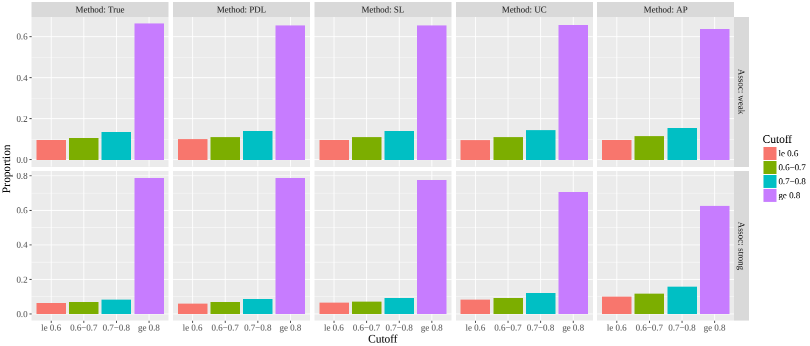
- When test is not sufficiently informative (i.e., short tests, poor item qualities), incorporating covariates can improve classification accuracy
- The improvement of CVC_d can be 1% - 23%, depending on different sample sizes, when the item quality is low, test length is short and association is strong
- The performance of the four-step approach with PDL or SL become similar to the true α as the sample size increases, and the difference between PDL and SL is negligible
- When the association is weak, incorporating covariates may perform slightly worse than AP

Results: Classification Certainty

- Results are presented using bar charts where the x-axis represents three cutoffs, y-axis difference between the proportion of certain classification of four methods and the AP, and color the different test lengths
- Results were compared in each panel across associations (vertically), methods (horizontally)
- Patterns across different D and $K(d)$ were very similar, so for illustration purposes, only results of $D = 2$ and $K(d) = 5$ are presented below

Results: Classification Certainty

$D = 2$, $K = 5$, Short Test, Low Item Quality

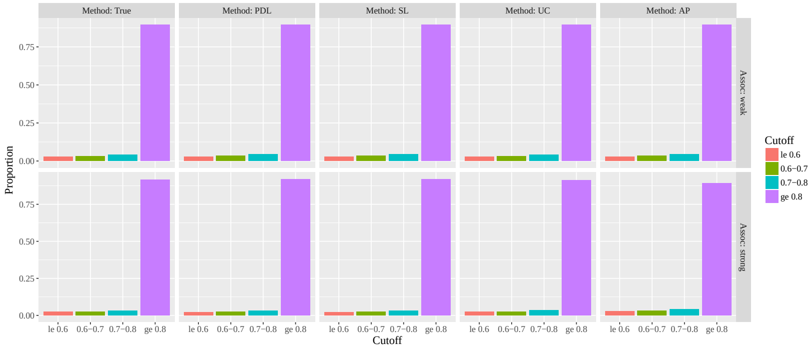


Results: Classification Certainty

$D = 2$, $K = 5$, Short Test, Medium Item Quality

Results: Classification Certainty

$D = 2$, $K = 5$, Short Test, High Item Quality



Results: Classification Certainty

- When association is strong, incorporating covariates increases the certainty of classification by moving the posterior probabilities towards the two extremes
- When the association is weak, the improvement is negligible
- The performance of PDL and SL are similar, and better than UC and AP alone

Outline

Conclusion and Discussion

- The availability of covariates is not unusual in many testing situations
- When covariates are related to the attributes of interest, they can be used to reinforce the information obtained from the assessment
- The four-step approach with correction weights can be used to improve classification when the test alone cannot provide sufficient information to classify examinees accurately

Conclusion and Discussion

- To make the most effective use of the covariates, the association between covariates and attributes needs to be tested:
  - when the association is strong, improvement can be achieved
  - however, when the association is weak, the performance can be equivalent to or only slightly worse than the AP
- Adding covariates can dramatically improve posterior probabilities and decrease the classification uncertainty when the association is strong and item quality is poor

Conclusion and Discussion

- AP, combined with the four-step approach, provides a feasible, as well as flexible way to analyze diagnostic assessment data with large number of attributes
- A future research direction should consider integrating various approaches to efficient testing (i.e., four-step approach + CD-CAT)
- Covariates can be used to tighten the examinees' prior distributions to optimize item selection at the early stages of CD-CAT

FIN.