Evaluation of CDM and RBM Methods to Estimate Small Q-matrices

RAPHAEL JEONG-HIN CHIN

Introduction

Cognitive diagnosis models (CDMs) are psychometric models that assess one's mastery of latent skills being tested. CDMs provide detailed feedback, including the probability of mastering a certain topic. Owing to CDMs' effectiveness in determining strengths and weaknesses in the topics to be tested, researchers in the field are becoming more aware of CDMs and "assessment for learning rather than assessment of learning" (Ravand & Robitzsch, 2015).

Multiple formulations of CDMs have been proposed in psychometric literature such as deterministic inputs, noisy "and" gate (DINA) (de la Torre, 2009), generalized DINA (GDINA) (de la Torre, 2011), and log-linear cognitive diagnosis models (LCDM) (Henson et al., 2009). There are multiple packages to fit different CDMs, such as the cdmTools and CDM packages (Nájera et al., 2022; Robitzsch et al., 2022). These packages help researchers use CDMs to learn more about the examinees' latent attributes based on their responses.

An important component of CDMs is the Q-matrix that informs the dependency structure between the J test items and K latent attributes (Li et al., 2022; Xu & Shang, 2018) because the Q-matrix can be effectively used to design intervention strategies. An example Q-matrix is shown in Table 1. '1' in the matrix means that Skill K is required for mastery of Item J. Thus, Q-restricted latent class models have gained popularity in fields such as educational proficiency assessments, psychiatric diagnosis, and many more disciplines (Xu & Shang, 2018). A well-known usage of CDMs is to study the dependency between

---

Contact: Raphael Jeong-Hin Chin <jeonghin@umich.edu>

Items	Attribute 1	Attribute 2	Attribute 3	Attribute 4
A	1	0	0	0
B	0	0	1	0
C	0	1	0	0
D	0	0	0	1

Table 1: Q-matrix Corresponds to Four Items, Four Latent Attributes, and $2^4 = 16$ Latent Classes

mathematical questions (items) and their latent skills for the topic of fractions as shown in Table 2. Let the six attributes tested in this topic be:

1. Find the lowest common denominator.
2. Add fractions.
3. Subtract fractions.
4. Multiply fractions.
5. Divide fractions.
6. Convert mixed numbers to improper fraction.

The first item (mathematical question) in the test is $2\frac{3}{4}+1\frac{1}{2}$, where "find the lowest common denominator," "add fractions," and "convert mixed numbers to improper fractions" (skills 1, 2, and 6) are required for this question to be answered correctly. Thus, the rows of the Q-matrix corresponding to this item will contain the vector (1,1,0,0,0,1) as shown in Table 2.

The Q-matrix plays an important role in CDMs because it can be used to categorize test items and design future assessments (Li et al., 2022). However, not all assessments can be explicitly specified with a Q-matrix. Even if there is an explicitly specified Q-matrix, the Q-matrix may not be accurate due to the following reasons: (i) design error by the assessment provider; and (ii) one test item may be linked to multiple attributes, but not all attributes are found and identified. For example, error (i) is committed in the second row of Table 2 because skill 6 is

Questions	Skill 1	Skill 2	Skill 3	Skill 4	Skill 5	Skill 6
$2\frac{3}{4}+1\frac{1}{2}$	1	1	0	0	0	1
$2\frac{3}{4}-1\frac{1}{2}$	1	0	1	0	0	1
$2\frac{3}{4}-1\frac{1}{4}$	0	0	1	0	0	0

Table 2: Q-matrix Corresponds to Three Math Questions and Six Latent Attributes

not required to correctly answer $2\frac{3}{4}-1\frac{1}{2}$. Thus, it is important to be able to learn more about the Q-matrix from the responses in order to have a better understanding of the relationship between the test items and latent variables' attributes.

Models

In this paper, the models of interest are the deterministic inputs, noisy "and" gate (DINA) model, the generalized-DINA (GDINA) model, and the restricted Boltzmann machines (RBMs). These three models are used in this paper to perform the following:

I. Test the accuracy of RBMs used by Li et al. (2022) on the data generated with a small number of latent attributes $K \in \{3,4,5\}$.
II. Compare the outputs from (i) with the results from Xu & Shang (2018).
III. Compare the results generated from the "CDM" package with the results from (i) and (ii) (Robitzsch et al., 2022).

Deterministic Inputs, Noisy "and" Gate (DINA) Model

The DINA model assumes a conjunctive relationship among attributes, where it is necessary to possess all the attributes indicated by the Q-matrix for a positive response (Xu & Shang, 2018). For each cell of the Q-matrix, q_{jk} is 1 if the kth attribute is required to correctly answer the jth item. In this model, an examinee's skills vector and the Q-matrix produce a latent response vector $\eta_i = \{\eta_{ij}\}$, where

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$$

has a value of 1 if examinee i possesses all the skills required for item j or has a value of 0 if the examinee lacks at least one of the required skills (de la Torre, 2009). K here represents the number of latent skills. Let $R_{i,j} = \{0,1\}$ represent the examinee i answering item j correctly. The uncertainties in this model are the slipping parameter, s_j, and guessing parameter, g_j, where

$$s_j = P(R_{i,j} = 0 \mid \eta_{ij} = 1)$$

$$g_j = P(R_{i,j} = 1 \mid \eta_{ij} = 0).$$

Therefore, the probability of examinee i with skills vector α_i answering item j correctly is given by

$$P_j\left(\alpha_i\right) = P(X_{ij} = 1 \mid \alpha_i) = g_j^{1-\eta_{ij}}\left(1 - s_j\right)^{\eta_{ij}}.$$

Generalized-DINA (GDINA) Model

Similar to the DINA model, the GDINA model requires a $J \times K$ Q-matrix as well. For each cell of the Q-matrix, q_{jk} is 1 if the kth attribute is required to correctly answer the jth item. In addition, GDINA separates the latent classes into $2^{K_j^*}$ latent groups where $K_j^* = \sum_{k=1}^{K} q_{jk}$ represents the number of required attributes for item j (de la Torre, 2011). Let α_{lj}^* be the reduced attribute vector whose elements are the required attributes for item j, and then the probability that examinees with attribute pattern α_{lj}^* will answer item j correctly is denoted by

$$P(X_j = 1 \mid \alpha_{lj}^*) = P\left(\alpha_{lj}^*\right).$$

In the GDINA model, there are three types of link functions available. This paper focuses only on the identity link function given by

$$P\left(\alpha_{lj}^*\right) = \beta_{j0} + \sum_{k=1}^{K_j^*}\beta_{jk}\alpha_{lk}$$
$$+ \sum_{k'=k+1}^{K_j^*}\sum_{k=1}^{K_j^*-1}\beta_{jkk'}\alpha_{lk}\alpha_{lk'} \dots + \beta_{j12\dots K_j^*}\prod_{k=1}^{K_j^*}\alpha_{lk} \tag{1}$$

where

β_{j0} is the intercept for item j;
β_{jk} is the main effect due to α_k;
$\beta_{jkk'}$ is the interaction effect due to α_k and $\alpha_{k'}$; and
$\beta_{j12\dots K_j^*}$ is the interaction effect due to $\alpha_1, \dots, \alpha_{K_j^*}$.

2.3 *Restricted Boltzmann Machines (RBMs)*

RBMs are generative stochastic artificial neural network models that can learn probability distributions over a collection of inputs. RBMs were initially invented

by Paul Smolensky under the name Harmonium (Smolensky, 1986). RBMs used in this paper follow the model design in Li et al. Visible units are denoted by $R = \{R_1, \dots R_j\} \in \{0,1\}^J$, and hidden units are denoted by $\alpha = \{\alpha_1, \dots \alpha_j\} \in \{0,1\}^K$. RBMs are characterized by the energy functions with their joint probability distribution given by

$$P(R, \alpha; \theta) = \frac{1}{Z(\theta)} exp\{-E(R, \alpha; \theta)\} \tag{2}$$

where $Z(\theta)$ is the partition function given by

$$Z(\theta) = \sum_{R \in \{0,1\}^J} \sum_{\alpha \in \{0,1\}^K} exp\{-E(R, \alpha; \theta)\} \tag{3}$$

and $E(R, \alpha; \theta)$ is the energy function given by

$$E(R, \alpha; \theta) = -b^T R - c^T \alpha - R^T W \alpha = -\sum_{j=1}^{J} R_j b_j - \sum_{k=1}^{K} \alpha_k c_k$$
$$-\sum_{j=1}^{J} \sum_{k=1}^{K} R_j w_{j,k} \alpha_k \tag{4}$$

In Equations 2–4, $\theta = \{b, c, W\}$ are the model parameters, $b \in R^J$ are visible biases, $c \in R^K$ are hidden biases, and $W \in R^{J \times K}$ is the weight matrix describing the interactions between the visible and the hidden units. The hidden and visible units are conditionally independent as there are no "R-R" or "$\alpha - \alpha$" interactions (Li et al., 2022). $w_{j,k} \neq 0$ in the weight matrix, W, for RBMs indicates the presence of interaction between the visible and the hidden units. Although DINA and GDINA models violate the conditionally independent assumptions of RBM, it was shown in Li et al. that the Q-matrices for these models are estimable.

Data

The data are simulated with latent attributes dimension $K \in \{3, 4, 5\}$ and the number of test items, $J = 20$. The true Q-matrices chosen are identifiable and similar to those used in Xu & Shang (2018). The three true Q-matrices are

$$Q_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad Q_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad Q_5 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

In this study, the data is simulated from the DINA latent class model. The ground truth response probabilities for all items are between 0.2 and 0.8, and both the slipping and the guessing parameters are set to 0.2. The dependency of latent attributes, ρ, is set to $\rho \in \{0, 0.15, 0.25, 0.5\}$. The two-step simulation of true latent profiles follows those set in Xu & Shang (2018). First, x_i is generated with $x_i = (x_{i1}, ..., x_{iK}) \overset{\text{i.i.d.}}{\sim} N(0, \Sigma)$ for $i = 1, ..., N$ where $\Sigma = (1 - \rho)I_K + \rho 1_K 1_K^T$. The attribute profile α_{ik} is set to be 1 if $x_{ik} \geq 0$ and 0 otherwise. The response data is then generated using the '*sim.din*' function from the *CDM* package.

Estimating the Q-matrix

The Q-matrices are estimated using the *gdina* function from the *CDM* package. As the response data follows the DINA model, a GDINA model can be fitted as the GDINA model is a generalized version of the DINA model. The *gdina* function will be used to fit the response data using both LASSO and Truncated

LASSO Penalty (TLP). The delta matrix returned by the function will be converted into a $J \times (2^K - 1)$ binary matrix (intercept column removed). The idea behind this is that because $\delta = \beta \times q$, if δ is not 0, q is definitely not 0, where β and q are elements in Equation (1). Values that are close to 0 in the delta matrix (smaller than 0.1) will be forced to be 0 and everything else to be 1. The $J \times 2^K$ binary matrix will be collapsed into a $J \times K$ binary matrix by grouping up the latent attributes that are required to master the item J.

Let $\alpha \in \{0,1\}$, $1 \le k \le K$, and $\delta_{ji} = \alpha_{iK}...\alpha_{i1}$ be the binary representation index of ith element in the jth row of the delta matrix. δ_{ji} will be transformed to have a value of 1 if it is greater than the threshold and 0 otherwise.

$$t_{jk} = \sum_{k=1}^{K} \delta_{ji} \text{ where } \alpha_{ik} = 1 \tag{6}$$

$$\hat{Q}_{jk} = 1 \text{ iff } t_{jk} \ne 0 \tag{7}$$

For example, let $\delta = (1.4, 1.32, 0.08, 2.1, 0.0003, 0.0001, 0)$, $J = 1$, $K = 3$, and threshold = 0.1, then applying Equations (6), we get,

$$\delta = (1.5, 1.7, 0.01, 2.9, 0.008, 0.0021, 0) \Rightarrow (1,1,0,1,0,0,0)$$

$$t = (2, 2, 0)$$

In Equation (6), the columns of the $J \times (2^K - 1)$ binary matrix refer to (Attr1, Attr2, Attr3, Attr12, Attr13, Attr23, Attr123). The matrix is then collapsed into a $J \times K$ matrix by summing up all the 1s into their respective latent attributes, where the columns refer to (Attr1, Attr2, Attr3). The estimated Q-matrix in Equation (7) is expected to be identifiable only up to rearranging the orders of the columns. This is because when estimating the Q-matrix, the columns do not contain information about the latent attributes (e.g., the nth column of the Q-matrix might not refer to the nth latent attribute). Thus, the estimated Q-matrix will be reordered so that each column shows the lowest possible average congruent coefficient with the True Q-matrix's columns. This process is done using the *"orderQ"* function in cdmTools (Nájera et al., 2022).

Accuracy Measurement

To evaluate the estimation accuracy, the entry-wise overall error (OE), out-of-true positive percentage error (OTP), and out-of-true negative percentage error (OTN) are reported. Their formulae are as follows:

$$OE = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} 1\left\{\hat{q}_{j,k} \neq q_{j,k}\right\} \tag{8}$$

$$OTP = \frac{\sum_{j=1}^{J} \sum_{k=1}^{K} 1\left\{\hat{q}_{j,k} = 0, q_{j,k} = 1\right\}}{\sum_{j=1}^{J} \sum_{k=1}^{K} 1\left\{q_{j,k} = 1\right\}} \tag{9}$$

$$OTN = \frac{\sum_{j=1}^{J} \sum_{k=1}^{K} 1\left\{\hat{q}_{j,k} = 1, q_{j,k} = 0\right\}}{\sum_{j=1}^{J} \sum_{k=1}^{K} 1\left\{q_{j,k} = 0\right\}} \tag{10}$$

Results

In this study, the Q-matrices are estimated completely from the data. For $K = 3$, the following crossover design is applied for the DINA model, three sample sizes, and four attribute-dependent levels: $\{DINA\} \otimes \{N = 500, 1000, 2000\} \otimes \{\rho = 0, 0.15, 0.25, 0.5\}$. For $K = 4$ and $K = 5$, the designs are $\{DINA\} \otimes \{N = 1000, 2000\} \otimes \{\rho = 0, 0.15, 0.25, 0.5\}$.

Table 3 shows the simulation results for 50 replications. From Table 3, it can be observed that, on average, TLP and RBM outperform the LASSO method. In the case of small N and small ρ, RBM usually outperforms the TLP method. However, as N and ρ become larger, the TLP method is able to estimate a Q-matrix more similar to the true Q-matrix. As N increases, accuracy also increases. This is because if there is more response data, the model has more data to learn and train from, resulting in a more accurate model. Surprisingly, across the three

K	N	Model	Accuracy (1-Error)			
			$\rho = 0$	$\rho = 0.15$	$\rho = 0.25$	$\rho = 0.5$
3	500	Lasso	0.8253	0.8587	0.8610	0.8613
		TLP	0.8420	0.8650	0.8823	0.9023
		RBM	0.8420	0.8727	0.8957	0.9017
	1000	Lasso	0.9043	0.9117	0.9217	0.9220
		TLP	0.9037	0.9453	0.9450	0.9593
		RBM	0.8667	0.9123	0.9323	0.9400
	2000	Lasso	0.9300	0.9703	0.9667	0.9550
		TLP	0.9623	0.9813	0.9857	0.9930
		RBM	0.8893	0.9390	0.9440	0.9513

Table 3: Mean Accuracy (50 Repetitions) for $K = 3, 4, 5$, and $J = 20$

K	N	Model	Accuracy (1-Error)			
			$\rho = 0$	$\rho = 0.15$	$\rho = 0.25$	$\rho = 0.5$
4	1000	Lasso	0.7375	0.7930	0.8030	0.8220
		TLP	0.7528	0.8140	0.8515	0.8740
		RBM	0.8285	0.8395	0.8588	0.8970
	2000	Lasso	0.8323	0.8615	0.8738	0.8673
		TLP	0.8453	0.8918	0.9008	0.9185
		RBM	0.8553	0.8708	0.8928	0.9093
5	1000	Lasso	0.6500	0.6648	0.7006	0.7452
		TLP	0.6500	0.6784	0.7188	0.7736
		RBM	0.8282	0.8534	0.8404	0.8432
	2000	Lasso	0.7096	0.7768	0.8122	0.8348
		TLP	0.7228	0.8004	0.8330	0.8990
		RBM	0.8668	0.8730	0.8850	0.8714

Table 3: (Continued)

different methods, the accuracy increases as the correlation among attributes increases. This may be because the higher the dependency among the attributes, the lesser the number of possible attribute patterns, making estimation relatively easier (Li et al., 2022).

Conclusion and Future Direction

In conclusion, it is shown in Table 3 that the CDMs with TLP method outperformed the ones with LASSO method. Moreover, it is interesting to see that the RBM models have stable performance for $K \leq 5$. The RBM models always have an accuracy of 82% or more for these data while CDMs perform badly when N is small.

The future work of interest would be to explore different ways to include interactions between latent attributes so that the assumptions set in RBM will not be violated. In practice, it is hard to find latent attributes that do not correlate with one another. Thus, by addressing this latent attribute interaction problem, the RBM method that has higher accuracy can be created. One potential way to address this problem may be integrating deep learning into the RBM method.

Owing to the success and stability of the RBM method in learning dichotomous item responses, it will also be interesting to implement the RBM method in

research that uses polytomous item responses. This is because a lot of questionnaires contain responses in the form of a 5-point or 7-point Likert scale. It will be interesting to study how different levels of responses correlate with mastering a certain skill or how the slipping and guessing parameters are affected by the way the questions were phrased. For example, an examinee may have the skills to answer a mathematical question correctly, but because the questions contain ambiguity and poor word choices, the examinee may be unable to answer the question.

Acknowledgment

References

de la Torre, J. (2009). DINA Model and Parameter Estimation: A Didactic. *Journal of Educational and Behavioral Statistics*, *34*(1), 115–130.

de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika*, *76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a Family of Cognitive Diagnosis Models Using Log-Linear Models with Latent Variables. *Psychometrika*, *74*(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5

Li, C., Ma, C., & Xu, G. (2022). Learning Large Q-Matrix by Restricted Boltzmann Machines. *Psychometrika*. https://doi.org/10.1007/s11336-021-09828-4

Nájera, P., Sorrel, M. A., & Abad, F. J. (2022). *cdmTools: Useful Tools for Cognitive Diagnosis Modeling* (1.0.1) [Computer software]. https://CRAN.R-project.org/package=cdmTools

Ravand, H., & Robitzsch, A. (2015). *Cognitive Diagnostic Modeling Using R*. https://doi.org/10.7275/5G6F-AK15

Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2022). *CDM: Cognitive Diagnosis Modeling* (7.6–11) [Computer software]. https://CRAN.R-project.org/package=CDM

Smolensky, P. (1986). *Information Processing in Dynamical Systems: Foundations of Harmony Theory*. Colorado University at Boulder Dept of Computer Science. https://apps.dtic.mil/sti/citations/ADA620727

Xu, G., & Shang, Z. (2018). Identifying Latent Structures in Restricted Latent Class Models. *Journal of the American Statistical Association*, *113*(523), 1284–1295. https://doi.org/10.1080/01621459.2017.1340889