Graduate
Management
Admission
Council®

Practical Issues Concerning the Application of the DINA Model to CAT Data

Alan Huebner, Bo Wang, & Sung Lee

ACT, Inc.

Presented at the Diagnostic Testing Paper Session, June 3, 2009

Abstract

There has been much interest in cognitive diagnostic models (CDMs) in recent years. These models aim to provide examinees with information about multiple fine-grained discrete skills. Understandably, there is interest in applying CDMs to existing assessments developed under an item response theory framework, a practice known as retrofitting. This paper demonstrates that this is not suitable if the data come from items that have been administered adaptively. We also demonstrate the feasibility of calibrating item parameters for CDMs under a pretesting scheme.

Acknowledgment

Presentation of this paper at the 2009 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.

Citation

Huebner, A., Wang, B., & Lee, S. (2009). Practical issues concerning the application of the DINA model to CAT data. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/

Author Contact

Alan Huebner, 500 ACT Dr, Iowa City, IA, 52243. Email: Alan.Huebner@act.org.

Practical Issues Concerning the Application of the DINA Model to CAT Data

Recent years have seen much interest in cognitive diagnostic modeling (or, diagnostic classification modeling, which has been deemed a more appropriate label by some researchers) for the purpose of formative assessment, i.e., assessment done while the teaching/learning is taking place. Cognitive diagnostic models (CDMs) give information about an examinee's mastery status of several discretely defined skills (sometimes called attributes), as opposed to assigning a single score representing a broadly defined latent trait, as is the case with item response theory (IRT). The information CDMs provide may be a vector of 1/0 mastery/non-mastery statuses or a vector of probabilities that the examinee possesses each skill. For example, for an exam diagnosing five skills, an examinee exhibiting mastery of the first three skills and non-mastery of the last two skills might be assigned the vector (1,1,1,0,0) or perhaps (0.85,0.91,0.79,0.26,0.13) as a score.

In order to enhance examinee score reporting, there has been much interest in applying CDM methodology to existing large scale IRT-based assessments, including computerized adaptive testing (CAT) assessments. For example, von Davier (2005) fit his general diagnostic model (GDM) to the Test of English as a Foreign Language (TOEFL) assessment, and McGlohen and Chang (2008) considered item selection rules in a simulated CAT setting using item parameters derived from fitting both a three-parameter logistic (3PL) model and a particular CDM, the Fusion Model, to a large scale state assessment.

This paper describes how a CDM may be fit to the GMAT Focus, an online diagnostic tool that prepares examinees for the quantitative portion of the GMAT exam. We also explore the need to calibrate the CDM parameters using randomly, rather than adaptively, administered test data. Specifically, section 2 introduces the DINA model, a CDM that has received much attention in the literature. Section 3 describes a simulation study examining the calibration of the DINA model parameters using responses simulated with real GMAT Focus item parameters, and section 4 demonstrates the method on real assessment data. Section 5 concludes with a review and discussion.

Methodology

The DINA model

We introduce some mathematical notation and concepts common to many CDMs. For a dataset with i = 1, … ,N examinees responding to an exam with $j = 1, \ldots, J$ items diagnosing K skills, α_{ik} represents the mastery level for the k^{th} skill of the i^{th} examinee. The complete skill profile of the i^{th} examinee is given by the vector $\boldsymbol{\alpha}_i = (\alpha_{i1},\ldots,\alpha_{iK})$. For a set of K skills measured on a dichotomous scale, there are $L = 2^K$ possible skill patterns. These skill patterns are referred to as latent classes, and the primary objective of a CDM is to classify examinees into these latent classes. This study aimed to investigate the effectiveness of using CDM technology to glean diagnostic information from the GMAT Focus.

Almost all CDMs utilize a Q matrix, a $J \times K$ matrix which indicates the skills that are required by each item (Tatsuoka, 1985). In many situations, the elements of the matrix, q_{jk}, are valued 1 if the j^{th} item requires the k^{th} skill and 0 if not. For example, the following Q matrix describes an exam diagnosing $K = 5$ skills in which Item 1 requires skill 1, Item 2 requires skills 1 and 3, Item 3 requires all the skills, and so on.

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

The CDM implemented for this study is the Deterministic Input, Noisy-And (DINA) model, first proposed by Haertel (1984, 1990) (acronym introduced by Junker & Sijtsma, 2001). The DINA model is easily interpreted, and there are a relatively large number of recent papers that focus on the DINA model compared to other CDMs. The DINA models the probability of a correct response given by examinee i on item j, denoted X_{ij}, as

$$P(X_{ij} = 1 \mid \boldsymbol{\alpha}) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}, \tag{1}$$

where $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$, s_j is the probability of a slip (an incorrect response given, despite the examinee having mastered all required skills for that item), and g_j is the probability of a guess (a correct response given despite the examinee not having mastered all the required skills for that item). Note that the η_{ij}s are binary indicators signifying whether the i^{th} examinee possesses all the required skills for item j. For example, consider a 30-item exam diagnosing five skills. Suppose Item 1 requires skills 1, 2, 3, and 4, and examinee 1 possesses all five skills. Then,

$$\eta_{11} = \prod_{k=1}^{5} \alpha_{1k}^{q_{1k}} = 1^1 * 1^1 * 1^1 * 1^1 * 1^0 = 1,$$

indicating that the examinee possesses all the required skills. In contrast, suppose examinee 2 possesses skills 1, 2, and 3. Then, for Item 1,

$$\eta_{21} = \prod_{k=1}^{5} \alpha_{1k}^{q_{1k}} = 1^1 * 1^1 * 1^1 * 0^1 * 0^0 = 0,$$

indicating that the examinee is lacking at least one required skill.

The GMAT Focus

From www.gmac.com, the GMAT Focus is an "…online diagnostic tool for prospective students (which) identifies quantitative strengths and abilities using real GMAT questions and a computer adaptive process that mimics the GMAT exam." This online assessment consists of twenty-four retired GMAT quantitative items. Each item has one feature of each of three dichotomous categories: item type (problem solving or data sufficiency), content (arithmetic or algebra), and application (real or pure) (Talento-Miller & Rudner, 2009). Thus, six different diagnostic areas are identified. These diagnostic areas can be expressed as follows, where the various item types are translated into Q matrix rows :

$$\begin{array}{c} data\ sufficiency / algebra / pure \\ problem\ solving / algebra / pure \\ data\ sufficiency / arithmetic / pure \\ problem\ solving / arithmetic / pure \\ data\ sufficiency / algebra / real \\ problem\ solving / algebra / real \\ data\ sufficiency / arithmetic / real \\ problem\ solving / arithmetic / real \end{array} \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

The DINA, being a conjunctive model, requires that all necessary skills be mastered in order for the examinee to have a high probability of responding correctly. This seems to be a plausible constraint for a quantitative assessment and, moreover, the DINA has been successfully fit to other quantitative educational data (de la Torre, 2008). The DINO (Templin and Henson, 2006), a disjunctive version of the DINA that requires only one skill to be mastered in order for there to be a high probability of a correct response, seems to be less appropriate here, and in fact, has only been fitted successfully to diagnosis of psychological disorders, rather than educational assessments.

The GMAT Focus data available for this study consisted of responses from N = 7,235 examinees, each of whom responded to 24 out of a possible 136 items. Item parameters from the 3PL model had been calibrated previously. The data, however, had been collected from items that were administered adaptively. Thus, the DINA item parameters could not be calibrated directly from this dataset. Next, we consider whether it is feasible to calibrate DINA item parameters when the responses are collected as pretest data, i.e., when subsets of items are administered to examinees randomly, rather than adaptively.

Simulation Study

A simulation study was performed using the R statistical software environment. Three different sets of responses were generated using the known GMAT Focus 3PL estimates and the examinee θs. Each of these datasets was then calibrated to the DINA model using an R program based on the EM algorithm approach proposed by de la Torre (2009). For the first dataset, responses were simulated such that each examinee responded to all 136 items. The DINA item parameters estimated from this dataset were regarded as the “true” parameters. For the second dataset, each examinee responded to 24 items selected randomly, and for the third dataset each examinee responded to 24 items selected adaptively (a simple adaptive testing scheme selecting items by maximum information was used).

We denote as S^1 and G^1 the set of DINA guess and slip parameters calibrated from the first dataset, and $P^1 = (S^1,G^1)$, the set of all DINA parameters from the first simulation. Similarly, $P^2 = (S^2,G^2)$ and $P^3 = (S^3,G^3)$. Table 1 shows the correlation between P^1, the “true” parameters, and P^2 and P^3.

Table 1. Correlation Between DINA Parameters

Parameters	Correlation
(P^1,P^2)	0.993
(P^1,P^3)	0.585

Clearly, P^1 and P^2 are highly correlated, which means that estimating item parameters by administering a small subset of items to each examinee will recover the true item parameters to a high degree of accuracy. On the other hand, the low correlation between P^1 and P^3 indicates that the DINA item parameters cannot be calibrated using responses to items that have been administered adaptively.

Real Assessment Data Example

A similar experiment exploring the feasibility of using a pretest scheme to calibrate DINA item parameters was conducted using actual assessment data. The data consisted of a full response matrix for N = 3,776 examinees and 97 items diagnosing four skills. A *Q* matrix was constructed using the domain classification of each item; thus, each item required only one skill. A pretest scheme was mimicked by discarding a random subset of 47 of the 97 items for each examinee, so that each examinee was left with responses to 50 items. The DINA model was fit to both the full matrix data and the artificially created sparse matrix data. The correlation between the two sets of parameters was 0.997. This verifies again that using randomly administered subsets of items, i.e., pretesting, is an acceptable means of calibrating DINA item parameters.

Discussion

Cognitive diagnostic modeling has become an exciting and very active area of psychometric research over the past several years. Much of the research has been geared toward proposing different models and theoretical issues, such as estimation by either the EM algorithm or Markov chain monte marlo methods. Only very recently have attempts begun to be made to utilize CDM methodology in a CAT setting. Thus, there is relatively little information for practitioners who may wish to begin exploring the use of CDMs in large scale assessments. This paper has sought to illustrate the importance and feasibility of calibrating DINA model parameters from a randomly administered, or pretest, sample.

There are many other issues concerning the use of CDMs in a practical CAT setting that have been addressed very little or not at all. Minimum sample size for item calibrations for various models has not been mentioned, and, while various item selection rules have been proposed for CAT CDM, there has been nothing published on stopping rules. Also, there have been almost no large scale external validation studies for the use of CDMs, which the authors intend to explore in the near future.

References

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34*, 115-130.

Haertel, E. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement, 8*, 333-346.

Junker, B. & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25, 3*, 258-272.

McGlohen, M. & Chang, H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods, 40*, 808-21.

Talento-Miller, E. & Rudner, L. (2009). Focus on improvement: Development and analysis of a diagnostic tool for the quantitative section of the Graduate Management Admission Test. Graduate Management Admission Council.

Tatsuoka, K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics, 12*, 55-73.

Templin, J. & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*, 287-305

von Davier, M. (2005). *A general diagnostic model applied to language testing data*. ETS Research Report. Princeton, New Jersey: ETS.