Front. Comput. Sci., 2023, 17(1): 171604
https://doi.org/10.1007/s11704-022-1128-3

REVIEW ARTICLE

New development of cognitive diagnosis models

Yingjie LIU, Tiancheng ZHANG (✉), Xuecen WANG, Ge YU, Tao LI
School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China

Abstract Cognitive diagnosis is the judgment of the student's cognitive ability, is a wide-spread concern in educational science. The cognitive diagnosis model (CDM) is an essential method to realize cognitive diagnosis measurement. This paper presents new research on the cognitive diagnosis model and introduces four individual aspects of probability-based CDM and deep learning-based CDM. These four aspects are higher-order latent trait, polytomous responses, polytomous attributes, and multilevel latent traits. The paper also sorts on the contained ideas, model structures and respective characteristics, and provides direction for developing cognitive diagnosis in the future.

Keywords higher-order latent traits, polytomous responses, polytomous attributes, multilevel latent traits, cognitive diagnosis

1 Introduction to cognitive diagnosis and its models

In a broad sense, cognitive diagnosis refers to the relationship between observation scores and students' internal cognition. In surveys, people usually call the diagnosis of individual cognitive process, processing skills, or knowledge structure as cognitive diagnostic assessment or cognitive diagnosis [1]. Cognitive diagnosis is not only a main role of educational psychology but also a primary application of pedagogy. In the theory of educational psychology, the latent trait factors are found by exploring the relationships between behavior, psychology, and cognition. In pedagogy, the basic rules of cognitive process are applied to modeling learners' cognitive situations and improving teaching quality. Cognitive diagnosis theory is combined with the calculation method of educational statistics to reflect the matching degree between and item responses the learners' cognition.

2 Classical model of cognitive diagnosis model

In the past 40 years, the development of cognitive diagnosis models has been going on. About 70 models have been constructed, forming a relatively complete system. Tu et al. [2] summarized four classic cognitive diagnosis models, namely, the linear logistic trait model (LLTM,[3]), the rule space model (RSM, [4]), the unified model [5] and the fusion model [6]. The key points are shown in Table 1. Among them, i denotes the i-th student, j denotes the j-th item and k represents the k-th attribute.

It can be observed that with development of the cognitive diagnosis models, the parameters of the models have characteristics as follows:

1. The model refers to additional factors. Firstly, in LLTM model, the item's difficulty factors are obtained by linear fusion attribute difficulty. Secondly, the unified model considers the completeness of the Q-matrix, that is, whether the item inspects all the attributes in the Q-matrix. Finally, the fusion model calculates whether the attributes are mastered in the proportion, corresponding to the probability of the item's correct answer. In this process, more and more factors are taken into consideration.
2. The model follows the pattern of saturation prior for simplification. It has lower accuracy because of too many parameters. So the fusion model redesigned its parameters and abandoned some complicated and less interfering factors in the response strategy. The new design model reduces the large parameter system and has significant advantages of increasing the accuracy.

3 New development of cognitive diagnosis models

The classical model mainly tests students' cognition according to the general conditions. Recently, the cognitive diagnosis model has been improved due to some new scientific issues. According to different scientific issues, this paper introduces the method and sorts out the model relationship from four aspects: the higher-order latent traits, polytomous responses, polytomous attributes, and multilevel latent traits.

Each field is closely related to practical problems and has an impact factor in reality. Meanwhile, domains are not independent of each other, but rather influence each other, thereby a new research system is produced for cognitive diagnosis.

3.1 Higher-order latent traits

In many fields, latent traits have a hierarchical structure.In the

Received March 19, 2021; accepted December 9, 2021

E-mail: tczhang@mail.neu.edu.cn

Table 1 Four classical cognitive diagnostic models

Model	Mathematical expression	Feature
Linear Logistic Trait Model (LLTM,[3])	$P\left(Y_{ij}=1\|\theta_i\right)=\exp(\theta_i-b_j^*)/[1+\exp(\theta_i-b_j^*)]$ $b_j^*=\sum\eta_k q_{jk}+d$ represents the difficulty of the item.	The model depicts item difficulty through the linear combination of attribute difficulty, thereby realizing the combination of cognition and measurement.
Rule Space Model (RSM,[4])	Computes a set of order pairs (θ, ζ) where $\zeta=f(x)/[\mathrm{Var}\,f(x)]^{\frac{1}{2}}$ and $f(x)=[P(\theta)-T(\theta)]'[P(\theta)-X]$. θ: is a variable that represents students. ζ: indicates the degree to which the reaction mode of the actual test items of a students with ability deviates from his or her ability levels. $P(\theta)$: represents the probability of correct answers with θ ability. X: represents the binary response of the student's answers for the items. $T(\theta)$: represents the mean vector of the probability with ability θ of correct answer to the items.	The model evaluates a student's ability, creatively puts forward the Q-matrix theory, and can also distinguish and diagnose a student's attribute mastery mode.
Unified Model [5]	$P\left(Y_{ij}=1\|\alpha_i,\theta_i\right)=\left(1-s_j\right)\left\{d_j\prod_{k=1}^{K}\left[\pi_{jk}^{\alpha_{ik}q_{jk}} \quad r_{jk}^{(1-\alpha_{ik})q_{jk}}\right]P_{c_j}(\theta_i)+\left(1-d_j\right)\right.$ $\left.P_{b_j}(\theta_i)\right\}$. s_j : represents that probability of errors on item j. d_j : selects the strategy described by the Q-matrix to solve item j. $p_c(\theta)$: represents the probility of correctly using the external attributes of the Q-matrix when the ability is θ.	The attributes are divided into Q-matrix inner and outer attributes, and the mastery mode and problem-solving strategy of the Q-matrix inner attributes and the latent residual ability of items aiming at the outer attributes θ are considered. However, the model is too complex, and some parameters in the model can at times not be identified or estimated.
Fusion Model [6]	$P(x_{ij}=1\|\alpha_i,\theta_i)=\pi_j^*\prod_{k=1}^{K}r_{jk}^{*(1-\alpha_{ik})q_{jk}}P_{c_j}(\theta_i)$. π_j^*: represents the item difficulty parameter of item j based on the Q-matrix. r_{jk}^*: represents the probability ratio of a student lacking attribute k and mastering attribute k but answering item j correctly.	The model is a simplified and unified model, which conforms to three basic and important characteristics: 1) estimating the mastery mode of students, 2) describing the relationship between items and various attributes, and 3) identifying the model parameters.

previous cognitive diagnosis models, higher-order latent traits are usually divided into two categories: a lower-order trait and a higher-order trait [7]. Higher-order latent traits are important factors in many tests and has rich practical significance. For example, we pay more attention to the higher-order listening, speaking, reading, and writing traits in English tests rather than the lower-order latent trait like whether English grammar uses correctly.

Considering the latent level, the cognitive diagnosis model consists of two parts: **the latent structural model** and **the measurement model** [8]. The former one describes the structural relationship among the latent traits, while the latter one defines the probability to obtain the correct answers.

The hierarchical nature of **latent structural model** is reflected in that students have the specific ability for each attribute, at same time, students also have an overall ability, which affects their specific ability. Correspondingly, this specific ability may affect the overall ability, thus the two-order nature appears among the latent traits. There are generally two assumptions about the relationship of hierarchy [9] shown as follows:

Assumption 1 Assuming that each specific ability is a linear function of overall ability, the relationship is expressed as Eq. (1):

$$\theta_{ik}=\beta_k\theta_i^{(2)}+\varepsilon_{ik}^{(1)}. \tag{1}$$

In Eq. (1), i represents the i-th student, and k represents the first-order k attribute. θ_{ik} denotes that the i-th student has a specific ability at the k-th attribute of order one. The student i's overall ability $\theta_i^{(2)}$, satisfies $\sim N(0,1)$. β_k is the regression vector between the attribute k, which $0<\beta_k<1$. $\varepsilon_{ik}^{(1)}$ is the ability residual of the student i in first-order attribute k, which obeys $\varepsilon_{ik}^{(1)}\sim N(0,1)$.

Assumption 2 Assuming that overall ability is a linear combination of all the specific abilities, the relationship is expressed by the following Eq. (2):

$$\theta_i^{(2)}=\sum_k\lambda_v\theta_{ik}+\varepsilon_i^{(1)}, \tag{2}$$

where λ_v is the regression weight between θ_{ik} and $\theta_i^{(2)}$. The remaining parameters have the same meanings as the parameters used in Assumption 1.

Although the above two equations hold that there is a linear relationship between abilities, their correlation properties are different. In the simulation experiment, the two models are slightly different according to parameter recovery and are generally superior to the traditional model. The developed hierarchical latent traits model usually divides latent trait into two levels, which are not supported by a perfect pedagogical framework, and the interpretability between the orders is not strong. In reality, latent traits cannot be simply divided into two levels, which ignores the complexity and openness of latent cognitive structure.

The third-order latent trait model is also studied in [10–12]. The liner assumption is adopted for the relationship among the various orders, and the third-order expression [10] is obtained through the nesting of Eq. (1):

$$\theta_{ik}=\beta_k^{(2)}\beta_{d[k]}^{(3)}\theta_i^{(3)}+\beta_k^{(2)}\varepsilon_{id[k]}^{(2)}+\varepsilon_{ik}^{(1)}. \tag{3}$$

In this equation, $\beta_{d[k]}$, β_k represents the regression vector between abilities, while $\varepsilon_{ik}^{(h)}$ and $\varepsilon_{id[k]}^{(h)}$ represent the ability residual. Due to the addition of a new layer, the second-order attribute is no longer unique one; d represents the d-th ability of the second-order. Therefore, we construct a linear relationship among the three orders.

There are many application scenarios of the higher-order latent traits, including the Program for the International Assessment of Adult Competencies (PIAAC, [13]), the Trend Survey of International Mathematics and Scientific Research (TIMSS), and the International Student Ability Assessment Program (PISA).

Frank R. et al. [10] applied the third-order latent trait model to the TIMSS data set. The TIMSS is an international comparative education survey, which is dedicated to improving mathematics and science teaching for students around the world. The TIMSS believes that every mathematical problem belongs to a content and cognition domain. There are four categories of **content domains**: Algebra, Data and Probability, Geometry, and Number and three kinds of categories of **cognitive domains**: Knowing, applying, and Reasoning. Content domains can also be subdivided into **topic domains**. For example, Algebra can continue to be divided into Patterns, Algebraic expression, Formula and Function. Table 2 shows the classification of TIMSS mathematical problems in the domains of content and cognition. Each number represents the number of questions in the corresponding domain.

Rijmen et al. constructed two hierarchical models. The first model is a two-order latent trait model based on the domain of content, cognition or topic as the second order. The second model is a three-order latent trait model, which takes the topic as the second order, as well as the content as the third order.

Zhan P. et al. [11] proposed a Multi-Order Cognitive Diagnosis Model (MO-CDM) based on the PISA 2015 theory and realized a three-order latent trait model focused on a theoretical basis. The Program for International Student Assessment (PISA) is carried out by the Organization for Economic Cooperation and Development and evaluates student scientific literacy ability worldwide. The MO-CDM model mainly adopts PISA 2015 scientific literacy theory, which is divided into three levels. Three-layer latent relationship abstracts the general structure of individual scientific latent traits from bottom to top.

The selection and design of **the measurement model** is also the key issue of cognitive diagnosis models. Sheng and Wikle et al. [9] firstly proposed a higher-order two-parameter normal ogive hierarchical model (2PNOHM). The measurement model of the 2PNOHM model adopts the normal ogive model of 2PL, shorted 2PN. The model expression is shown in the following Eq. (4):

$$P\left(Y_{ijk}\right)=\phi\left(\alpha_{jk}\theta_{ik}-\gamma_{jk}\right) = \int_{-\infty}^{\alpha_{jk}\theta_{ik}-\gamma_{jk}} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} \mathrm{d}t. \tag{4}$$

In practical applications, when the IRT model defaults to D = 1.7, the difference between the ogive model and the IRT model is the smallest one. Therefore, in order to facilitate numerical processing, the popular 2PL IRT model is proposed. In Eq. (4), α_{jk} and γ_{jk} are the component of the attribute k of the discrimination and difficulty on item j, respectively, and θ_{ik} is the ability of the student i in the k-th attribute. In the end, we get the probability that the i-th student has the k-th attribute on item j. By combining Eqs. (1) and (4), the 2PNOHM probability expression is shown as follows:

$$P\left(Y_{ijk}\right)=\phi\left(\alpha_{jk}\beta_k\theta_i^{(2)}-\gamma_{jk}+\alpha_{jk}\varepsilon_{ik}^{(1)}\right). \tag{5}$$

The model assumes $\alpha_{jk}>0$, $p(\gamma_{jk})\propto 1$. The parameter calculation of the 2PNOHM model has the following deficiencies. First of all, for the item parameters of the model, such as α_{jk},γ_{jk}, the estimated effect is poor under real experiments. Secondly, under the condition of conjugate distribution, the model is insensitive to the prior distribution of the β through sensitivity analysis. If the prior information is relatively abundant, the posteriori estimation of the item and student parameters can be obtained by Gibbs sampling. 2PNOHM has the following two shortcomings. In addition to considering the small order range, first of all, in practical applications, the measurement model of 2PNOHM is not as wide as that of the three-parameter model. Secondly, 2PNOHM does not limit the correlation between hierarchial orders. Torre et al. proposed a higher-order three-parameter logistic model to overcome the deficiencies of 2PNOHM and make HO-IRM more widely applicable.

The **measurement model** of HO-IRM adopts the 3PL IRT model [12], and the model expression is shown in the following Eq. (6):

Table 2 Classification of TIMSS mathematics problems in terms of content and cognitive domains

Content domain	Cognitive domain			
	Knowing	Applying	Reasoning	Total
Algebra	32	15	17	64
Data and probability	14	18	8	40
Geometry	8	27	12	47
Number	27	28	8	63
Total	81	88	45	214

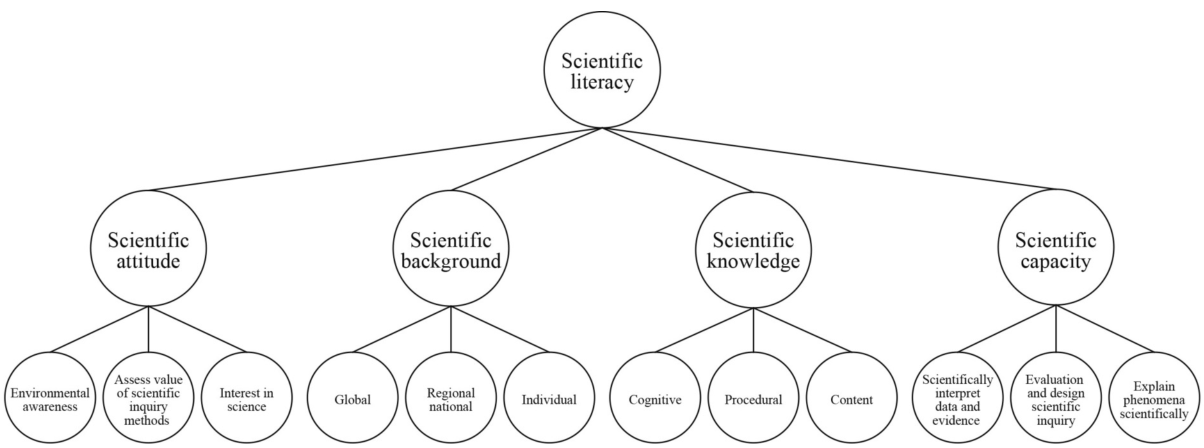

Fig. 1 PISA 2015 scientific literacy theory

$$P_{ijk} = c_{jk} + \left(1 - c_{jk}\right) \frac{\exp\left[\alpha_{jk}\left(\theta_{ik}^{(1)} - b_{jk}\right)\right]}{1 + \exp\left[\alpha_{jk}\left(\theta_{ik}^{(1)} - b_{jk}\right)\right]}. \tag{6}$$

The 3PL model is an important model in cognitive diagnosis. The difference between the 3PL and the 2PL is that the guessing parameter c. c is often considered in cognitive diagnosis to represent the guessing degree of the item, which refers to the attribute k that can be mastered with a certain probability without cognitive latent trait. The linear latent structural model is still used among the orders, thus forming a higher-order IRT model. Therefore, HO-IRM can be obtained by combining Eqs. (1) and (6). Experiments show that the number of students is small, when the number of tests is large and the test length is short. This method is superior to the continuous method. The continuous method does not use the linear calculation of higher-order latent traits, and only uses the single factor analysis method to describe the higher-order latent traits.

The measurement model also uses the DINA model with attribute mastering as the connection function [11]. The selection of different measurement models affects the fitting effect of the parameters to a certain extent. Parameter recovery experiments demonstrate the superiority of the higher-order model.

The higher-order latent traits model has been greatly developed since its inception. For example, there has been research on testlets based on the higher-order response models [14]; Huo et al. [15] proposed a hierarchical, multi-unidimensional two-parameter logistic item response theory (2PL-MUIRT) extended model for a large number of groups; There has been research conducted on the evaluation of longitudinal data obtained from repeated measurements of individuals [16]; And Zhang et al. [17] tested the performance of item parameter fitting based on chi-square fitting to test the HO-IRT model statistics. And he pointed out that the appropriate parameter distribution should be selected according to different attribute structures. In the study of Fu et al. [18], the parameter estimation of the higher-order latent traits model was tested, and the accuracy of the enhanced sampling method was improved using the data-enhanced Gibbs sampler.

3.2 Polytomous responses

Polytomous responses are core consideration in the cognitive diagnosis model. The previous models are most suitable for questions with scores of 0 or 1, such as multiple-choice questions. In fact, polytomous responses are common, and the phenomenon of mixed responses is also very common. In most cases, multiple-choice questions and subjective questions coexist in a common test, and the scoring rules are not limited to 0 and 1. Therefore, polytomous responses have great practical value.

There are two main methods for solving the polytomous responses, namely, the polytomous responses model based on the item response and the processing process [19]:

Method 1 The polytomous responses model **based on the item response** is more suitable for continuous models. The score difficulty does not correspond to the attribute but only takes the item category characteristics as consideration, such as item discrimination and item difficulty. Linking the item factors with the score to obtain a scientific calculation model.

The models for solving the polytomous responses item include: the rule space model (RSM,[4]), the graded response model (GRM,[20]), the partial credit model (PCM,[21]) and the generalized partial credit model (GPCM,[22]).

GRM The item response function of the GRM holds that the probability of score t is the difference between the cumulative probability of score t and score $t-1$ [20]:

$$P_{jt}^{*}(\theta) = \frac{\exp(Da_j(\theta_i - b_{jt}))}{1 + \exp(Da_j(\theta_i - b_{jt}))}. \tag{7}$$

$P_{jt}^{*}(\theta)$ indicates the probability that student i get the score t or higher on the item j. This function is called the cumulative category response function (CCRF). b_{jt} represents the difficulty parameter when the score of the item j is defined as t. This difficulty parameter is different from that used previously. It is generally considered that the difficulty has been previously reflected at the item level or attribute level. When the score is different according to the item, the corresponding difficulty is also different, which is more in line with the actual situation. Thus, the probability of the score t is shown as follows:

$$P_{jt}(\theta) = P_{jt}^{*}(\theta) - P_{jt+1}^{*}(\theta). \tag{8}$$

Combining the above two equations could construct the different relationship between scores. Thus, the expression of GRM is as follows:

$$\begin{aligned} &P_{jt}(\theta_i) \\ &= \frac{\exp(-Da_j(\theta_i - b_{jt+1})) - \exp(Da_j(\theta_i - b_{jt}))}{[1 + \exp(-Da_j(\theta_i - b_{jt}))][1 + \exp(Da_j(\theta_i - b_{jt+1}))]}. \end{aligned} \tag{9}$$

PCM PCM is a model that has been developed for ordered scoring [21]. It sets each score as a rating class. Next, we will introduce the derivation process:

$$P_j(\theta_i) = \frac{\exp\left[\left(\theta_i - b_j\right)\right]}{1 + \exp\left[\left(\theta_i - b_j\right)\right]}. \tag{10}$$

The probability of student i correctly answering item j is usually shown in the above equation. For each adjacent reaction score, under the condition of obtaining $t-1$ points, the conditional probability of t is as follows:

$$\begin{aligned} C_{jt} &= P_{jt|t-1,t}(\theta_i) \\ &= \frac{P_{jt}(\theta_i)}{P_{jt-1}(\theta_i) + P_{jt}(\theta_i)} \\ &= \frac{\exp\left[\left(\theta_i - b_{jt}\right)\right]}{1 + \exp\left[\left(\theta_i - b_{jt}\right)\right]}. \end{aligned} \tag{11}$$

When $k = 2, 3, \ldots, m_j$, it is transformed into the follows:

$$P_{jt}(\theta_i) = \frac{C_{jt}}{1 - C_{jt}} P_{jt-1}(\theta_i). \tag{12}$$

Assuming:

$$P_{j1}(\theta_i) = \frac{1}{G}, \tag{13}$$

where G is defined as a normalizing factor. Through the following equations:

$$P_{j2}(\theta_i) = \frac{\exp(\theta_i - b_{j2})}{G}, \tag{14}$$

$$P_{jm_j}(\theta_i) = \frac{\exp\left[\left(\sum_{v=2}^{m_j}(\theta_i - b_{jv})\right)\right]}{G}, \tag{15}$$

and after obtaining the full score m_j of item j. Considering $\sum_{k=1}^{m_j} P_{jt}(\theta) = 1$,

$$G = 1 + \sum_{c=2}^{m_j} \exp\left[\sum_{v=2}^{c} a_j\left(\theta_i - b_{jv}\right)\right], \tag{16}$$

then the reaction function of PCM is shown in the following equation:

$$P_{jt}(\theta_i) = \frac{\exp\left[\sum_{v=1}^{t} D\left(\theta_i - b_{jv}\right)\right]}{\sum_{c=1}^{m_j} \exp\left[\sum_{v=1}^{c} D\left(\theta_i - b_{jv}\right)\right]}. \tag{17}$$

In Eq. (17), $P_{jt}(\theta_i)$ represents the probability of student i obtaining a score t on item j. PCM is used to construct the proportion between score t and the sum of all the scores by recursiving the equation. This method rather than GRM which constructs the different relationships between the scores.

GPCM GPCM extends the item discrimination parameter [22] based on PCM, thereby forming a generalized polytomous responses model:

$$P_{jt}(\theta_i) = \frac{\exp\left[\sum_{v=1}^{t} Da_j\left(\theta_i - b_{jv}\right)\right]}{\sum_{c=1}^{m_j} \exp\left[\sum_{v=1}^{c} Da_j\left(\theta_i - b_{jv}\right)\right]}. \tag{18}$$

According to GPCM, it is easy to obtain the proportion between the score t and the score $t-1$, as shown in the following Eq. (19):

$$\log\left(\frac{P_{ijt}}{P_{ij(t-1)}}\right) = \alpha_j\theta_i - b_{jt}. \tag{19}$$

Huang et al. [12] added the function of polytomous responses to HO-IRM and successfully applied the model to an item test with polytomous responses. In the test, the appraoch extended the universality of the HIRT model and its robustness to deal with real data. The model links the two latent traits with a linear latent structure model; thus, by combining Eqs. (1) and (19), we obtain the relationship between the higher-order latent traits and item scores. The higher-order generalized partial scoring model is as follows:

$$\log\left(\frac{P_{ijtk}}{P_{ij(t-1)k}}\right) = \alpha_{jk}\left(\beta_k\theta_i^{(2)} + \varepsilon_{ik}^{(1)}\right) - \left(b_{jk} + \tau_{jtk}\right). \tag{20}$$

This paper points out several aspects of using the HO-IRM model flexibility:

1. Different items use different polytomous responses models. For example, items 1 to 10 could use a one-parameter IRT model as the measurement model, items 11 to 15 could use PCM, and items 16 to 20 could use GRM.
2. Different latent trait models adopted according to a different number of orders. For example, both the second-order and the third-order hierarchical relationships are used for model design.

In addition to replacing the measurement models and polytomous responses models, polytomous responses also have a rich development, combined with multilevel group structure to form a comprehensive latent trait model.

Although HO-IRM has the higher-order and polytomous responses functions reflected in item difficulty and student latent traint parameter, respectively. It is theoretically feasible since there is no conflict between the two models. If the latent traint has both higher-order and multilevel characteristics, it will produce too many parameters to learn, and result in insignificant parameter estimation, which has no practical value.

Other scholars have also conducted excellent research on polytomous responses models. Tu et al. [23] proposed the P-DINA model, which is based on the idea of PCM, combining with the DINA model, includes slip and guess parameters. In addition, the accuracy of the model parameter estimation is high in both non-structural and structural attribute hierarchy. Chen et al. [24] presented a combination of G-DINA and GRM when solving the polytomous responses item problem and proposed a reduction model to reduce the number of parameters. The effectiveness, adaptability, and portability of this model are demonstrated in the simulation experiment.

Method 2 The previous polytomous responses method took the details of attribute structure and attribute type into account and established a connection between the attribute and the scoring process, but it cannot dynamically build the difficulty of an item. Therefore, we propose a polytomous responses model related to attribute difficulty **based on the processing process** is proposed.

For example, if an item text is $4\frac{1}{8} - \frac{3}{8}$ =? and the item score is 3 points. We first point examines attribute A1 using the borrow method, i.e., $4\frac{1}{8}$ into $3\frac{9}{8}$. The second point examines attribute A2, fraction subtraction, and obtains the result of $3\frac{6}{8}$. The third part examines attribute A3, the approximate score, and obtains the approximate score result of $3\frac{3}{4}$. Therefore, the restricted Q-matrix [25] is constructed, as shown in Table 3.

The restricted Q-matrix is different from the previous Q-matrix. In general, the Q-matrix is a $J \times K$-sized matrix, where J is the size of the items, and K is the attributes size. The restricted Q-matrix creates a Q-matrix for each item. The Q-matrix size of each item is $H_j \times K$, where H_j is the highest

Table 3 Restricted Q-matrix

Step	Score	Attribute		
		A1	A2	A3
$3\frac{9}{8} - \frac{3}{8}$	1	1	0	0
$3\frac{6}{8}$	0	1	0	
$3\frac{3}{4}$	3	0	0	1

score of each item j.

The Seq-G-Dina model [25] adds a polytomous responses function based on the generalized DINA model. The model holds that under the premise of score $t-1$, the attribute mastery is defined as α_c to answer item j, and the probability of scoring t is the processing function of scoring $S_j(t|\alpha_c)$ for the next three important conditions:

1. The initial score and edge score: the probability of a score of 0 is 1. When the highest score of the item is H_j, the probability of a score of H_j+1 is 0.
2. The relationship between the two adjacent scores: when calculating the probability of correct answers and assuming that the student has scored $t-1$, it is certain the student has successfully completed the score attributes of 1,..., $t-1$. $t-1$ is not the highest score and that the t score is incorrect.
3. The sum of all the score's probability should be 1.

According to the above three conditions, the probability equation is listed as follows:

$$S_j(t|\alpha_c)=\begin{cases}1, & \text{if} \quad t=0,\\ 0, & \text{if} \quad t=H_j+1,\end{cases} \tag{21}$$

$$P(X_j=t|\alpha_c)=[1-S_j(t+1|\alpha_c)]\prod_{x=0}^{t}S_j(x|\alpha_c), \tag{22}$$

$$\sum_{t=0}^{H_j}P\left(X_j=t|\alpha_c\right)=1,\forall c. \tag{23}$$

The Seq-G-Dina model could be obtained by knowing the score probability expression of the restricted mastery level and combining it with the generalized DINA model.

The generalized DINA model (G-DINA) was proposed by Torre et al. [26]. Compared to the DINA model, the G-DINA model reduces the restrictions regarding the conditions, and loose the requirements between the theoretical assumptions and attributes. Its saturation model is equivalent to that of the general model based on other connection functions. Under appropriate constraints, some commonly used cognitive diagnostic models can be regarded as some special cases of the general models. That means the G-DINA model provides a general framework for the cognitive diagnosis model. The probability expression is as follows:

$$\begin{aligned}P\left(\alpha_{ij}\right)&=\delta_{j0}+\sum_{k=1}^{K_j^*}\delta_{jk}\alpha_{ik}\\&+\sum_{k'=k+1}^{K_j^*}\sum_{k=1}^{K_j^*-1}\delta_{jkk'}\alpha_{ik}\alpha_{ik'}\cdots+\delta_{j12\cdots k_j^*}\prod_{k=1}^{K_j^*}\alpha_{ik}.\end{aligned} \tag{24}$$

The expression of $K_j^*=\sum_{k=1}^{K}q_{jk}$ represents the number of attributes examined in items j. δ_{j0} is the intercept of item j, which is a baseline probability, that is, the probability that the item can be answered correctly without mastering the required attributes; however, it cannot be equal to the item's guess, which is only one of the baseline factors, including other undiscovered factors that affect the probability of answering the item correctly without mastery. δ_k indicates the probability of mastering an attribute α_k. $\delta_{kk'}$ suggests that mastering two adjacent attribute levels of α_k and $\alpha_{k'}$ simultaneously affects the probability of a correct answer. Similarly, $\delta_{j12\cdots k_j^*}$ suggests that mastering K^* attributes simultaneously is influential when all the attributes investigated by the item j are grasped. Therefore, the G-DINA model is a high-saturation classification model that considers the different probability contributions from mastering a single attribute to mastering multiple attributes.

Similar to the G-DINA model, the probability of the scoring t under the premise of scoring $t-1$ is as follows:

$$\begin{aligned}S\left(t|\alpha_{ijt}\right)&=\delta_{jt0}+\sum_{k=1}^{K_{jt}^*}\delta_{jtk}\alpha_{ik}\\&+\sum_{k'=k+1}^{K_{jt}^*}\sum_{k=1}^{K_{jt}-1^*}\delta_{jtkk'}\alpha_{ik}\alpha_{ik'}+\cdots+\delta_{jt12\ldots K_{jt}^*}\prod_{k=1}^{K_{jt}^*}\alpha_{ik}.\end{aligned} \tag{25}$$

In this expression, K_{jt}^* refers to the number of the attributes item j at the score t, and the number of intercept parameters is $2^{K_{jt}^*}$. All the other parameters have the same explanation as that given for the G-DINA model parameters, except that the restriction of the score t is introduced.

Based on the restricted Q-matrix, the model intuitively obtains the attribute corresponding to the scores, adds the restricted relationship between the attributes and the scores, conducts the attribute mastery by using the longitudinal dimension of the score. Combined with the above three conditions, the probability of the item j under the attribute mastery α_c is $P(X_j=t|\alpha_c)$. In the previous part, we have pointed out that the model can also be used with nominal response data [27] to obtain a wider range of application. Nominal response data is encoded data. For example, there are several options available for multiple-choice questions. Each multiple-choice question is coded for the correct choices, but the choices do not have a sequence relationship. Therefore, it is suitable to construct a nonordered but restrained Q-matrix to represent the relationship between the attributes and scores. In a subsequent development, Ma [28] considered various student strategies based on the restricted Q-matrix and proposed a multi strategy diagnostic tree model to track student problem-solving processes. Thus, the polytomous responses method based on processing has a wide application range and strong mobility, which is a basic analytical method used for classifying data.

3.3 Polytomous attributes

In practice, a fine-grained division of attributes is required. Polytomous attributes hold that the level of attribute mastery and the level of examining of items for attributes are not simply coded as either 0 or 1 (mastered or not mastered), rather, multiple values are used to represent different levels, with a value of $0\sim M$ (where M indicates the highest mastery level of the attribute). Compared to second-level attributes, polytomous attributes have much more detailed diagnostic information, as shown in Table 4.

Table 4 Polytomous attributes

Attributes	Level 1	Level 2
Comparision and order of fractions	Students should be able to compare two fractions and determine whether one of these fractions is equal to, less than, or greater than the other.	Students should be able to order three or more fractions.
The proportion constructed according to the situation	Given an item situation involving ratios, students should be able to construct a single ratio to describe the situation	Given a proportional situation, students should be able to construct an appropriate proportion.

In this table, there are two inspection attributes: the comparison and ranking of scores and the construction of proportions according to the situation. There are also two levels of mastery. Level 2 is a further extension of Level 1 and a reflection of a higher level. Notably, Level 0 does not consider mastery level of this attribute. Therefore, the primary mastery level starts with Level 1. Polytomous attributes add abundant measurement features, and describe the latent classess of student's attributes in more detail, which is more in line with the actual situation. In general, the polytomous status of attributes is set by experts.

The processing idea of polytomous attributes is to extract the key mastery without lossing any information, so as to convert the polytomous attributes into second-level attributes, which is convenient for calculation. Since polytomous attributes affect the student's attribute mastery and item attribute examinng level, the general processing method is used to define a new discrete attribute mastery level α_{ik}^{*}:

$$\alpha_{ik}^{*}=\begin{cases}0, & \text{if} \quad \alpha_{\mathrm{ik}}<\mathrm{q}_{\mathrm{jk}},\\ 1, & \text{otherwise.}\end{cases} \tag{26}$$

When the attribute master α_{ik} is less than the level of the item examined q_{jk}, the attribute k is not mastered, and α_{ik}^{*} thus is 0. While the attribute mastery α_{ik} and q_{jk} are equal or α_{ik} is greater than q_{jk}, then the attribute k is mastered, and it is coded as 1.

The pG-DINA model proposed by Chen and Torre et al. [29], based on the G-DINA model, the function of processing polytomous attributes is added, which is an important extension of polytomous attributes. The mathematical expression of the pG-DINA model is as follows:

$$\begin{aligned}P\left(\alpha_{ik}^{**}\right)=&\ \delta_{j0}+\sum_{k=1}^{K_j^*}\delta_{jk}\alpha_{ik}^{**}\\&+\sum_{k'>k}^{K_j^*}\sum_{k=1}^{K_j^*}\delta_{jkk'}\alpha_{ik}^{**}\alpha_{iik'}^{**}+\cdots+\delta_{j1,\ldots,K_j^*}\prod_{k=1}^{K_j^*}\alpha_{iK_j^*}^{**}.\end{aligned} \tag{27}$$

Compared with the G-DINA model, the structure and function of this model are roughly the same as the G-DINA model. The attribute mastery level of the input model can be divided into two stages: the secondary mastery level α_{ik}^{*} and the reduction mastery level α_{ik}^{**}. The reduced item attribute mastery α_{ik}^{**}, which is called the folding attribute vector, is used to reduce the number of mastering vectors. For example, if the total number of all attributes is K, and the master level M = 3, then the number of latent classes is K^3. While we only consider the attributes examined by item j, that is, K_j^* in this case; Thus, the number of parameters has been reduced to K_j^{3*}.

When restraining the saturation model parameters to zero, such as δ_{jk}, $\delta_{jkk'}$, the pG-DINA model can be converted to a DINA model with polytomous attributes. This is defined as follows:

$$P(\alpha_{ij}^{**})=\delta_{j0}+\delta_{j1,\ldots,K_j^*}\prod_{k=1}^{K_j^*}\alpha_{ik}^{**}. \tag{28}$$

It can be seen from the equation that the DINA model with polytomous attributes is a non-compensatory model. Only when the student has mastered all the attributes can they answer this item correctly with a high probability. As long as an attribute is not mastered, the probability of being correctly answered is equal to the probability of guessing, which is δ_{j0}. In this way, we can obtain $g_j=\delta_{j0}$ and $1-s_j=\delta_{j0}+\delta_{j12\ldots K^*}$.

Scholars have carried out a series of studies on polytomous attributes and have conducted in-depth studies in some fields. Examples include how experts set each level to reasonably reflect the mastery level and evaluate items after setting each level. Regarding mathematical proportional reasoning problems (PR), Tjoe and Torre et al. [30] put forward of a reasonable three-stage gradual transformation method to meet the polytomous attributes. The design of the Q-matrix in polytomous attributes is not limited to the polytomous test Q-matrix. Cai and Tu et al. [31] proposed the pa-DINA model and considered that the test Q-matrix contains an R-matrix that can successfully achieve the distinguishing degree for each student with a different knowledge state. Under the premise of considering polytomous attributes, the idea of calculating the attribute R-matrix is proposed based on the Q-matrix theory to combines attribute hierarchy. Zhao et al. [32] proposed the PRPa-DINA model, which expands the polytomous attributes DINA model (RPa-DINA model) to a polytomous responses, ensuring that specifically processing polytomous responses data with no information lost.

3.4 Multilevel latent traits

Multilevel can be applied to attributes, items, or latent traits. The horizontal structure of attributes and items usually setted by experts. And the horizontal structure of the latent traits is divided according to the latent class of the test group. The students with similar latent levels are divided into one class, and those with large differences in their latent levels are divided into different classes to obtain the latent structure of the test group. The common method is to use a mixed latent trait model to calculate the multilevel latent traits. The mixed latent trait model has the function of simultaneously processing latent classified variables and latent continuous variables. It overcomes the shortcomings of the two methods and obtains hidden latent trait classification. It is used to evaluate the measurement structure's multidimensionality and to detect the differences in the respondent styles.

mixIRT The mixture Rasch model (MRM, 1990) [33] is the most original mixed model. It was developed by Rost to improve the fitting of the model to data. It is also a mixed IRT model which is widely used in education and psychological testing.The model's expression is as follows:

$$P\left(Y_{ijg}=1|\theta_{ig},\beta_{jg},g\right)=\frac{\exp\left(\theta_{ig}-\beta_{jg}\right)}{1+\exp\left(\theta_{ig}-\beta_{jg}\right)}. \tag{29}$$

$P\left(Y_{ijg}=1\right)$ denotes the probability that student i who belongs to the latent class g will answer item j correctly. It can be observed that the students have different item strategies in different classes and that the same item has different item parameters in different classes. Assuming that all the classes are mutually exclusive, we get the correct answer probability for item j for student i under all latent classes:

$$\begin{aligned}P\left(Y_{ij}=1|\theta_{ig},\beta_{jg},g\right)&=\sum_{g=1}^{G}\pi_g\left[\frac{\exp\left(\theta_{ig}-\beta_{jg}\right)}{1+\exp\left(\theta_{ig}-\beta_{jg}\right)}\right]\\&=\sum_{g=1}^{G}\pi_g\left[\frac{1}{1+\exp\left[-\left(\theta_{ig}-\beta_{jg}\right)\right]}\right].\end{aligned} \tag{30}$$

Among them, θ_{ig} is the latent trait of the student i, and β_{jg} is the difficulty coefficient of item j. The model allows different classes have different mixing ratios. Finally, g is the latent class of the student ($g=1,2,...,G$), where π_g represents the mixing ratio of latent class g, $0<\pi_g<1$.

mmixIRT Cho and Cohen et al. [34] proposed a multilevel mixed item response theory model (mmixIRT). Compared to the mixirt model, this model extends the latent classification of students from the horizontal direction and abstracts a new classification-school. As a high-level latent classification, there is a latent classification of student class and school class. The model can deal with discontinuous and continuous variables to meet the multilevel distribution structure provided by samples. The mathematical expression of the model is as follows:

$$\mathrm{P}\left(Y_{ijh}=1|g,k,\theta_{ihgk}\right)=\frac{1}{1+\exp\left[-\left(\theta_{ihgk}-\beta_{jgk}\right)\right]}. \tag{31}$$

In this equation, g is the index of latent classes of the student class, k is the index of latent classes of the school class, and i refers to the ith student. h refers to the h school. $P\left(Y_{ijh}=1\right)$ is the probability that student i belonging to the school h answers correctly item j. Students from different schools have two latent class features: latent school class and latent student class. After mining the students' latent school classes, characteristics of students' classes are explored. The latent school class k is not necessarily the same as the actual school class. The latent school class K is mined according to the students' latent traits. Like g, it is not a statistic but obtained through statistical methods. θ_{ihgk} is student i from school h and belongs to latent school class g and latent student class k. β_{jgk} is the difficulty of item j of latent school class g and latent student class k. In the general MMixIRT, the item difficulty parameter has both student-level and school-level attributes, which represent the interaction between the latent student class g and the latent school class k, respectively. There are two mixing ratios in different classes, which are $\pi_{g|k}$ and π_k. The calculation structure of the mixing ratio is shown in Table 5.

Through simulation research, the model's performance under the condition of the actual item function difference (DIF) test has been discussed. For the considered conditions, the generated parameters are recovered. Meij et al. [35] applied the mixed IRT model to the extroversion and neuro-quality tables of the Amsterdam Biographical Questionnaire to model the heterogeneity of the population and complete the migration to personality test. Louis et al. [36] adopted characteristics such as years of work experience and gender as covariates, combined with a mixed model, to study the influence of constant but influential parameters in the model. Huang et al. [16] studied the multilevel latent factors in the higher-order latent model, considering multilevel factors from the horizontal and vertical aspects.

More in-depth research has also been carried out based on the MMixIRTM. For example, Kim et al. [37] studied the invariance of unobservable groups to reduce multilevel model classification error. The horizontal latent trait model is usually defined to link the latent trait classes with the item class. Wang et al. [38] believed that the basic analysis process of the mixed model is as follows: "Firstly, the best number of latent trait classes is determined by the model adaptability test. Then according to the students' response patterns, the students are divided into the corresponding latent trait class and their latent trait are estimated. Therefore, the mixed model includes two important aspects: model selection and parameters estimate." The commonly included parameter estimation methods are the EM and MCMC algorithms. Therefore, the model can deal with the continuous data applicable to the latent trait model and the category data applicable to the latent classification model, which applies to the homogeneity within groups and the heterogeneity between groups.

4 Further study

Deep Latent Trait Due to the strong feature representation ability of deep learning technology, the model is expert at extracting features from many sparse data points and could find and characterize structures to obtain more abundant prior information. Therefore, using deep learning to calculate latent traits is also a research trend.

DIRT model was put forward by Cheng et al. [39]. On the basis of IRT model, this model extends the parameter-solving method, adopting popular deep learning method to solve the item and latent trait parameters, combining the interpretability of statistical methods with the robust feature extraction ability of the deep learning method. The final experimental result is

Table 5 The mixing ratio of mmixIRT

	K=1	K=2	⋯	⋯	K=K
G=1	$\pi_{1\|1}$	$\pi_{1\|2}$	⋯	⋯	$\pi_{1\|K}$
G=2	$\pi_{2\|1}$	$\pi_{2\|2}$	⋯	⋯	$\pi_{2\|K}$
⋮	⋮	⋮	⋮	⋮	⋮
G=G	$\pi_{G\|1}$	$\pi_{G\|1}$	⋯	⋯	$\pi_{G\|K}$
Sum	$\sum_{g=1}^{G}\pi_{g\|1}=1$	$\sum_{g=1}^{G}\pi_{g\|2}=1$	⋯	⋯	$\sum_{g=1}^{G}\pi_{g\|K}=1$

far superior to other models. The deep neural network (DNN) was used to analyze latent trait extraction:

$$\theta = \mathrm{DNN}_\theta(\Theta), \Theta = \sum_{K_k \in \mathcal{K}_q} \alpha_k K_k. \tag{32}$$

K_k and α_k are the higher-order vectors of attribute k and the latent trait component of attribute k, respectively. The calculation expression of the item discrimination parameter a in this model is as follows:

$$a = 8 \times (sigmoid(\mathrm{DNN}_a(A)) - 0.5), A = \sum_{K_k \in \mathcal{K}_q} K_k. \tag{33}$$

Parameter A adopts the multidimensional item discrimination method proposed by Yao et al. [40] to sum the attributes of one item to obtain the total attribute and vector A. The model uses another DNN to extract the item discrimination. A is inputed into DNN for feature extraction. According to the traditional education model, the item's discrimination range is [−4,4]; Thus, using the sigmoid function, subtracting 0.5 and multiplying it by 8 can meet the range requirements.

The item difficulty parameter b is closely related to the item text. The model mainly uses the item difficulty calculation method proposed by Huang [41] to mine the text information to calculate the difficulty. Firstly, since LSTM can process long-term sequence text from a semantic perspective, LSTM should be selected as the perfect process model. Secondly, the depth and breadth of the attributes also affect the difficulty prediction of the item. Based on the attention, LSTM is used to calculate the difficulty of the item, considering the different influence of the weight parameter. The attention mechanism part of this model is defined as follows:

$$x_t = \sum_{K_k \in \mathcal{K}_q} softmax\left(\frac{\xi_j}{\sqrt{d_0}}\right) K_k + w_t, \xi_j = w_t^T K_k. \tag{34}$$

The weight of the attribute K_k is given by the text word w of the item. d_o is the scale factor of the attention. The softmax function obtains different attribute proportions and determines the size of attention. This model selects the most classical LSTM model composed of the following four parts: input gate, output gate, forget gate and state control. Input x_t into lstm. The prediction part uses the IRT model. The previously calculated parameters are input into the IRT model.

The optimizer optimizes the objective function of negative log-likelihood.The latent trait parameters of the students and the weight parameters of the two DNN and LSTM are also updated.

$$\mathcal{L} = r_{ij} \log r_{ij} + \left(1 - r_{ij}\right) \log\left(1 - r_{ij}\right). \tag{35}$$

The paper does not explain how to improve the IRT model; On the contrary, it places much emphasis on data embedding, feature extraction, and deep network processing, which is a relatively novel method for calculating the latent trait model parameters.

Recently, a combination of neural networks and cognitive diagnosis has gradually emerged. Initially, Shu et al. [42] found that when calculating the DINA model parameters on small samples, the neural network is greatly affected by the difficulty of the Q-matrix and the item. Therefore, when using only neural networks, it is necessary to estimate the distribution of the student's attribute mastery level as prior information. Lamb et al. [43] studied the cognitive diagnosis in serious educational games (SEG). Due to a large amount of data in the game log, IRT was firstly used to verify the item parameters and the ability parameters. Then, combined with an artificial neural network (ANN) and Q-matrix, a hierarchy of attributes was established through the propagation weights and the model fitting. Wang et al. [44] proposed a generalized neural cognitive framework (NeuralCD). Diagnosis framework incorporates neural networks to learn the complex interactions between the student factor vectors and exercise factor vectors. Neural networks can comprehensively learn the relationship among students, items, and interactions. To sum up, the applications of the neural network in cognitive diagnosis is mostly reflected in neural network-aided calculation, which mainly uses neural networks to extract features and completes end-to-end calculation of cognitive parameters combined with statistical models.

5 Model comparison

For the mentioned above, we select the recent influential models and compare them according to five aspects: the number of dimensions of the models, the compensation among attributes, the representation of attributes, the reduction models, and the saturation of the models, as shown in Table 6.

Model dimension refers to the number of attributes considered by the model. Compensation between attributes refers to the probability that when a certain attribute is not mastered, the students can still answer the questions correctly by mastering other attributes. However, non-compensatory means that when one attribute is not mastered, the mastery of other attributes cannot make up for its deficiency; Therefore, the probability of the correct answer given by the students is low. The representation of attributes includes the continuous and discrete categories. The continuous category refers to modeling the students' latent trait of as continuous parameters, while the discrete category represents the cognitive state of the

Table 6 Comparison of seven new advanced cognitive diagnosis models

Model	Model dimensions	Attribute compensation	Attribute representation	Model reduction	Model saturation
HO-IRM	Multidimensional	Compensatory	Continuous	No	No
HO-CDM	Multidimensional	Compensatory	Discrete	No	No
G-DINA	Multidimensional	Compensatory and Non-compensatory	Discrete	Yes	Yes
pG-DINA	Multidimensional	Compensatory and Non-compensatory	Discrete	Yes	Yes
Seq-G-DINA	Multidimensional	Compensatory and Non-compensatory	Discrete	Yes	Yes
mixIRT	Multidimensional	Non-compensatory	Continuous	No	No
mmixIRT	Multidimensional	Non-compensatory	Continuous	No	No

student with multiple discrete vectors. The reduction model considers whether the number of attributes examined by all items is due to different items and considers different attributes to reduce the parameters. Saturation refers to a general model that does not strictly limit the role among the attributes and fully meets various attribute conditions. Each of the above measurement indexes is the key feature of the recent cognitive diagnosis models.

We can see that:

1. Due to the complexity of testing, the new developments are based on the multidimensional attributes.
2. In the G-DINA model, $\sum_{k=1}^{K_j^*}\alpha_{ik}$ denotes that all the attributes have the influence of the baseline probability of being answered correctly and that the attributes have compensation. $\prod_{k=1}^{K_j^*}\alpha_{ik}$ holds that there is no compensation between attributes and that attributes are independent. Due to a baseline probability between two attributes and a baseline probability between three attributes, the restriction of conditions in the G-DINA model is greatly relaxed. Therefore, the G-DINA model is a comprehensive model that integrates the compensation and non-compensation as well as saturation and attribute reduction. The central bodies of pG-DINA and Seq-G-DINA are also the model of G-DINA; Thus, they have the same characteristics.
3. In HO-IRM, there is a linear relationship between the higher-order attributes and the lower-order attributes, and the attributes are not independent; thus, they also have saturation.

6 Research outlook

With the development of cognitive diagnosis research, educational data mining with cognitive characteristics has attracted wide attention and has shown great potential in many works. This paper proposes four research prospects for implementing the cognitive diagnosis model technology from the perspective of the research process.

Data

Cognitive diagnosis should not be limited to cognitive judgments based on test scores. Researchers should try to collect all records in regard to an attribute. The more detailed the records are, the more precise the student cognitive track will be, and the more accurate the cognitive diagnosis results will be.

Data source aspect Complete semester analysis consists of the tests and the exercises.

For the tests, it is also necessary to combine the student behavioral logs during the tests, such as the duration of the student's answers, the number of alterations, the order of the students' recorded answers, and other comprehensive behaviors.

In practice, tracking the practice behaviors in exercises is also a supplement to the cognitive situation. The exercises should be divided into fixed exercises and free exercises. Fixed exercises should be designed for all the students with the same amount of exercises; however, the exercise time should not be fixed, and there should be no time limit during the test, which is are parameters similar to the school homework task mode. In free exercises, when there is no restriction on the number of exercises, the choice of question type is free. At this time, it is necessary to introduce parameters such as the I-matrix (student-item matrix) to modeling.

Data type aspect According to different application scenarios, the types of data are also different. Data types should consider both inherent data types and real-time changed data types.

For inherent data types, the unchangeable factors should include gender, age, education level, marital status, and other factors that affect latent cognitive traits but cannot be changed in a short period. Previous studies have defined these factors as covariates to control the calculation of cognitive parameters.

Real-time changed data types should include variables such as the online learning platform's click rate. When considering these factors, the combination of long-term data and short-term data could be followed, and the static and dynamic cognitive characteristics could be described at the same time.

Parameter calculation

Machine learning methods, such as neural networks are used for calculation. The neural network has become a hot research topic. Although the neural network model's generalization and interpretation are not comparable to those of the statistical model, the neural network reflects an accuracy and efficiency that traditional statistics cannot achieve. These characteristics make the case for big data mining, regression, and classification tasks.

More and more cognitive diagnosis models choose neural networks as the core model for parameter calculation. However, deep learning also has its special problems. The neural network model needs a large amount of training data to prove its superior performance. In addition, the interpretability of deep learning technology is poor. In most cases, it can only explain what the model output is, while the intermediate process is a black box, which cannot explain why such an output result is obtained.

Model

The model is introduced from six aspects, namely, the sample, the attributes' structures, the compensatory aspect of the attributes, the Q-matrix, the reduction model, and the saturation model.

The sample aspect Firstly, the parameters can be estimated by controlling the experimental conditions. Sample selection is the critical point that needs to be paid attention. The samples should be randomly selected from different test groups to ensure the sample's normal prior distribution, which is a vital prerequisite for statistical and machine learning.

The attributes structure aspect The structure between the attributes is an important consideration. Different attribute structures have different calculation results. The basic attribute structure is linear, convergent, divergent, and unstructured types. The more complex relationship among attributes can form an attribute structure network. Attention should be paid to determining the attribute structure reasonably from the perspective of trees or graphs.

The compensatory of attributes aspect From the perspective of different attributes that contribute differently to mastery, we should focus on compensatory and non-compensatory attributes. [45] considered that the item satisfies the compensation and the non-compensatory attributes, that the vector of the attribute mastery is different, and that it is represented by a fuzzy set. Other models are representedly use the multiplication or addition of the mastery level of attributes to indicate the compensatory and non-compensatory attributes.

The Q-matrix aspect The Q-matrix should be designed reasonably and normatively; For example, in the Seq-G-DINA model, the restricted Q-matrix is used to express the relationship between the attributes and the scores. The restricted Q-matrix can also fully consider the relation items, attributes, and scores. The original Q-matrix only considered the two factors items and attributes. When designing a Q-matrix, it can be similar to a restricted Q-matrix. According to different application scenarios, multiple factors are considered to create a multidimensional Q-matrix +.

The reduction model aspect The reduced model can reduce the number of parameters and improve the accuracy of parameter estimation. It is also possible to further reduce the parameters of different sizes according to the different attributes examined in different items.

The saturation model aspect Attention should be paid to the degree of saturation. Excessive saturation may not calculate the parameters, and considering the conditions of multi-attribute relations is counterproductive. The neural network can calculate the highly concerned attribute relations, and then design the model for calculation.

Application scenario

There is a wide, complicated range of application scenarios. Classic application scenarios include serious educational games [46], difference item function (DIF [47]), computer adaptive testing (CD-CAT [48]), cognitive diagnosis evaluation (cognitive diagnosis assessment) etc. With the development of online education and online education platforms, student performance prediction based on online judgement platforms and knowledge tracking [49] are gradually becoming the research focus. Cognitive diagnosis is gradually being applied in a wide range of fields. Under the application scenarios related to personalized education, various online learning systems have emerged in an endless stream, thereby creating an open and free learning environment for students, providing rich learning resources, and accumulating data from many students learning at the same time. Open learning environments and more research application scenarios have brought about more opportunities for and challenges to the research of cognitive diagnosis solutions, which are of great significance.

7 Conclusion

This paper presents new research on the cognitive diagnosis model and introduces four aspects of probability-based CDM and deep learning-based CDM. These four aspects are higher-order latent traits, polytomous responses, polytomous attributes, and multilevel latent traits. We also sort out the contained ideas, model structures and respective characteristics, which provides the direction for developing cognitive diagnosis in the future.

Acknowledgements This work was partially supported by the National Natural Science Foundation (Grant Nos. U1811261, 62137001, 61902055) and the Fundamental Research Funds for the Central Universities (N180716010, N2117001).

References

1. Leighton J P, Gierl M J. Cognitive Diagnostic Assessment for Education: Theory and Applications. Cambridge: Cambridge University Press, 2007
2. Tu D B, Cai Y, Dai H Q, Qi S Q. A review on cognitive diagnostic models under modern test theory. Psychological Exploration, 2008, 28(2): 64–68
3. Fischer G H. The linear logistic test model as an instrument in educational research. Acta Psychologica, 1973, 37(6): 359–374
4. Tatsuoka K K. Rule space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 1983, 20(4): 345–354
5. DiBello L V, Stout W F, Roussos L A. Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In: Nichols P D, Chipman S F, Brennan R L, eds. Cognitively Diagnostic Assessment. Hillsdale: Erlbaum, 1995, 361–389
6. Hartz S M. A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality. University of Illinois at Urbana-Champaign, Dissertation, 2002
7. Schmid J, Leiman J M. The development of hierarchical factor solutions. Psychometrika, 1957, 22(1): 53–61
8. Rupp A A, Templin J, Henson R A. Diagnostic Measurement: Theory, Methods, and Applications. New York: Guilford Press, 2010
9. Sheng Y Y, Wikle C K. Bayesian multidimensional IRT models with a hierarchical structure. Educational and Psychological Measurement, 2008, 68(3): 413–430
10. Rijmen F, Jeon M, Von Davier M, Rabe-Hesketh S. A third-order item response theory model for modeling the effects of domains and subdomains in large-scale educational assessment surveys. Journal of Educational and Behavioral Statistics, 2014, 39(4): 235–256
11. Zhan P D, Yu Z H, Li F M, Wang L J. Using a multi-order cognitive diagnosis model to assess scientific literacy. Acta Psychologica Sinica, 2019, 51(6): 734–746
12. Huang H Y, Wang W C, Chen P H, Su C M. Higher-order item response models for hierarchical latent traits. Applied Psychological Measurement, 2013, 37(8): 619–637
13. Organisation for Economic Co-operation and Development. Technical report of the survey of adult skills (PIAAC). Paris: OECD, 2013
14. Huang H Y, Wang W C. Higher order testlet response models for hierarchical latent traits and testlet-based items. Educational and Psychological Measurement, 2013, 73(3): 491–511
15. Huo Y, De La Torre J, Mun E Y, Kim S Y, Ray A E, Jiao Y, White H R. A hierarchical multi-unidimensional IRT approach for analyzing sparse, multi-group data for integrative data analysis. Psychometrika, 2015, 80(3): 834–855
16. Huang H Y. A multilevel higher order item response theory model for measuring latent growth in longitudinal data. Applied Psychological Measurement, 2015, 39(5): 362–372
17. Zhang X, Wang C, Tao J. Assessing item-level fit for higher order item response theory models. Applied Psychological Measurement, 2018, 42(8): 644–659
18. Fu Z H, Zhang X, Tao J. Gibbs sampling using the data augmentation scheme for higher-order item response models. Physica A: Statistical Mechanics and its Applications, 2020, 541: 123696

19. Tu D B. Advanced Cognitive Diagnosis. Beijing: Beijing Normal University Publishing House, 2019
20. Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika, 1969, 34(1): 1–97
21. Masters G N. A rasch model for partial credit scoring. Psychometrika, 1982, 47(2): 149–174
22. Muraki E. A generalized partial credit model: application of an EM algorithm. Applied Psychological Measurement, 1992, 16(2): 159–176
23. Tu D B, Cai Y, Dai H Q, Ding S L. A polytomous cognitive diagnosis model: P-DINA model. Acta Psychologica Sinica, 2010, 42(10): 1011–1020
24. Chen J S, De La Torre J. Introducing the general polytomous diagnosis modeling framework. Frontiers in Psychology, 2018, 9: 1474
25. Ma W C, De La Torre J. A sequential cognitive diagnosis model for polytomous responses. British Journal of Mathematical and Statistical Psychology, 2016, 69(3): 253–275
26. De La Torre J. The generalized DINA model framework. Psychometrika, 2011, 76(2): 179–199
27. Templin J, Henson R, Rupp A, Jang E, Ahmed M. Cognitive diagnosis models for nominal response data. See researchgate.net/profile/Robert-Henson/publication/228894528_Cognitive_diagnosis_models_for_nominal_response_data/links/0a85e5332fadc2ef60000000/Cognitive-diagnosis-models-for-nominal-response-data.pdf website, 2008
28. Ma W C. A diagnostic tree model for polytomous responses with multiple strategies. British Journal of Mathematical and Statistical Psychology, 2019, 72(1): 61–82
29. Chen J S, De La Torre J. A general cognitive diagnosis model for expert-defined polytomous attributes. Applied Psychological Measurement, 2013, 37(6): 419–437
30. Tjoe H, De La Torre J. Designing cognitively-based proportional reasoning problems as an application of modern psychological measurement models. Journal of Mathematics Education, 2013, 6(2): 17–26
31. Cai Y, Tu D B. Extension of cognitive diagnosis models based on the polytomous attributes framework and their Q-matrices designs. Acta Psychologica Sinica, 2015, 47(10): 1300–1308
32. Zhao S Y, Chang W, Wang L J, Zhan P D. A polytomous extension of reparametrized polytomous attributes DINA. CNKI (in Chinese), 2019
33. Rost J. Rasch models in latent classes: an integration of two approaches to item analysis. Applied Psychological Measurement, 1990, 14(3): 271–282
34. Cho S J, Cohen A S. A multilevel mixture IRT model with an application to DIF. Journal of Educational and Behavioral Statistics, 2010, 35(3): 336–370
35. Meij A M M D, Kelderman H, Van Der Flier H. Fitting a mixture item response theory model to personality questionnaire data: characterizing latent classes and investigating possibilities for improving prediction. Applied Psychological Measurement, 2008, 32(8): 611–631
36. Tay L, Newman D A, Vermunt J K. Using mixed-measurement item response theory with covariates (MM-IRT-C) to ascertain observed and unobserved measurement equivalence. Organizational Research Methods, 2011, 14(1): 147–176
37. Kim E S, Joo S H, Lee P, Wang Y, Stark S. Measurement invariance testing across between-level latent classes using multilevel factor mixture modeling. Structural Equation Modeling: A Multidisciplinary Journal, 2016, 23(6): 870–887
38. Wang X, Tan G H, Wang X, Zhang M Q, Luo C. The mixture item response theory models and its application traces. Advances in Psychological Science, 2014, 22(3): 540–548
39. Cheng S, Liu Q, Chen E H, Huang Z, Huang Z Y, Chen Y Y, Ma H P, Hu G P. DIRT: deep learning enhanced item response theory for cognitive diagnosis. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019, 2397-2400
40. Yao L H, Schwarz R D. A multidimensional partial credit model with associated item and test statistics: an application to mixed-format tests. Applied Psychological Measurement, 2006, 30(6): 469–492
41. Huang Z Y, Liu Q, Chen E H, Zhao H K, Gao M Y, Wei S, Su Y, Hu G P. Question difficulty prediction for READING problems in standard tests. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017, 1352−1359
42. Shu Z, Henson R, Willse J. Using neural network analysis to define methods of DINA model estimation for small sample sizes. Journal of Classification, 2013, 30(2): 173–194
43. Lamb R L, Annetta L, Vallett D B, Sadler T D. Cognitive diagnostic like approaches using neural-network analysis of serious educational videogames. Computers & Education, 2014, 70: 92–104
44. Wang F, Liu Q, Chen E H, Huang Z Y, Chen Y Y, Yin Y, Huang Z, Wang S J. Neural cognitive diagnosis for intelligent education systems. In: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. 2019, 6153−6161
45. Liu Q, Wu R Z, Chen E H, Xu G D, Su Y, Chen Z G, Hu G P. Fuzzy cognitive diagnosis for modelling examinee performance. ACM Transactions on Intelligent Systems and Technology, 2018, 9(4): 48
46. Boyle E A, Hainey T, Connolly T M, Gray G, Earp J, Ott M, Lim T, Ninaus M, Ribeiro C, Pereira J. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education, 2016, 94: 178–192
47. Andrich D, Hagquist C. Real and artificial differential item functioning. Journal of Educational and Behavioral Statistics, 2012, 37(3): 387–416
48. Yu X F, Cheng Y, Chang H H. Recent developments in cognitive diagnostic computerized adaptive testing (CD-CAT): a comprehensive review. In: Von Davier M, Lee Y S, eds. Handbook of Diagnostic Classification Models. Cham: Springer, 2019, 307–331
49. Liu H Y, Zhang T C, Wu P W, Yu G. A review of knowledge tracking. Journal of East China Normal University: Natural Science, 2019(5): 1–15

Yingjie Liu received a BE degree from Shanxi Normal University, China in 2018. She is now majoring in computer technology at the College of Computer Science and Engineering, Northeastern University, China. Her main research area is intelligent education, cognitive diagnosis, and recommended systems.

Tiancheng Zhang received a MS degree in 2001 and a PhD degree in computer science from Northeastern Unversity, China in 2008. He is currently an associate professor at Northeastern University, China. His research interests include spatiotemporal data mining, artificial intelligence, intelligent education.

Xuecen Wang received a BE degree from Shandong University, China in 2018. She is now majoring in computer technology at the College of Computer Science and Engineering, Northeastern University, China. Her main research interest is in intelligent education and information systems.

Ge Yu received an MS degree in computer science from Northeastern University, China in 1986 and a PhD degree in computer science from Kyushu University, Japan in 1996. He is currently a professor and PhD supervisor at Northeastern University, China, a member of ACM and IEEE, and a CCF senior member. His research interests include database theory and technology, and distributed system, as well as parallel computing and cloud computing.

Tao Li received the MA degree in sociolinguistics from Northeastern University, China. The main research interests are computer-aided teaching, etc.