This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2023.3302848

Multivariate Cognitive Response Framework for Student Performance Prediction on MOOC

Lianhong Wang, Xiaoyao Li, Zhihui Luo, Zinan Hu, Qing Yan

Abstract—Based on student's cognitive structure, the cognitive diagnostic models (CDMs) can reveal the potential relationships among the student's knowledge level, test item features and the corresponding item scores, and then predict each student's future performance. However, due to the simplistic prior information and deficient cognitive mechanism, most of the existing CDMs have limited prediction performance. To address the issues, we propose the multivariate cognitive response framework (MvCRF). We firstly collect student's learning activity logs to calculate the corresponding effort trait. Considering both student's ability trait and effort trait, MvCRF then introduces the compensation mechanism to calculate student's knowledge level. In addition, we introduce not only the slip and guessing parameters in prediction but also the skill weakness parameter related with the student's knowledge level and the importance of each skill on solving specific item. Experimental results on both simulation study and real-data application on MOOC demonstrate that MvCRF achieves better prediction performance, robustness and interpretability than the baseline CDMs.

Index Terms—Educational data mining, cognitive diagnostic model, effort trait, compensation mechanism, skill weakness.

1 INTRODUCTION

WITH the rapid development of education informatization and online learning platforms, students from all walks of life can break through the limits of space and time to acquire knowledge [1]. Educators then apply big data to track the whole learning process of students and evaluate their own instructional practices. Educational data mining and learning analytics are two areas specific to the use of big data in education [2], [3], [4], [5], [6]. Generally, educational data mining develops new models to analyze educational data by employing techniques like data mining, machine learning, statistics, psychology and so on, while learning analytics enables educators to adopt existing methods to understand learning performance and optimize education approaches according to the need and ability of each student. Recently, educational data mining draws increasing attention in numerous educational platforms, such as massive open online courses (MOOC) [7] and intelligent tutoring systems [8].

As a fundamental issue in educational data mining, cognitive diagnosis seeks to measure student's skill proficiency by quantifying the learning status and assessing the cognitive structure so as to identify student's strengths and weaknesses. Therefore, cognitive diagnostic models (CDMs) receive massive attention in intelligent education. The typical CDMs include item response theory (IRT) [9], deterministic inputs, noisy and gate model (DINA) [10], fusion model [11], attribute hierarchy model [12], etc. The widely-used DINA and IRT first apply student's item scores, test item features and other prior information to quantitative estimate student's knowledge level and then predict student's future performance. Their difference is that IRT denotes the knowledge level as a continuous scalar variable while DINA denotes it as a binary vector variable (or a binary multidimensional variable). Based on DINA, Torre and Douglas [13] introduced a higher-order latent trait to evaluate student's knowledge level and proposed the higher-order DINA model (HO-DINA). Tu et al. [14], [15], [16], [17] extended the DINA model to polytomous scoring with a binary variable to denote student's knowledge level. Employing fuzzy set theory, Liu et al. [18] improved the work of Tu et al. by fuzzifying the student's knowledge level determined by latent trait to a continuous variable between 0 to 1, and proposed the fuzzy cognitive diagnosis framework (Fuzzy-CDF). Besides, Li et al. [19] proposed a revised fuzzy cognitive diagnosis framework by introducing the skill importance level according to the frequency of each skill in a given question bank and assumed that the skill importance level would affect the student's cognitive diagnosis. In addition, many efforts were made to apply neural networks for cognitive diagnosis in intelligent education. Wang et al. [20] proposed an interpretable neural cognitive diagnosis (NeuralCD) framework to explore the complex interactions between students and test items. Cheng et al. [21] introduced the skill importance level into NeuralCD to enhance the fitting degree of student-item interactions. Gao et al. [22] integrated the traditional CDMs with deep learning to capture the internal relationship between skills and items. In addition to the aforementioned static cognitive diagnosis, another research focus in the field of cognitive diagnosis is dynamic cognitive diagnosis (i.e. knowledge tracing) [23], [24], [25]. For example, the popular Bayesian knowledge tracing (BKT) [26] and deep knowledge tracing (DKT) [27] can continually update student's knowledge level in real time as the students solve the questions one by one.

Lianhong Wang is corresponding author.

- *L. Wang, Z. Luo and Z. Hu are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China. E-mail: 292386791@qq.com, luo1998@hnu.edu.cn, 664291984@qq.com.*
- *X. Li is with the School of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha 410082, China. E-mail: lxyimayday@gmail.com.*
- *Q. Yan is with the Urtrust Insurance Co., Ltd., Guangzhou 510000, China. E-mail: 545281420@qq.com*

In the review and analysis of previous works on cognitive diagnosis for educational assessment, there still exist some challenges in the following aspects:

- *How to capture more prior information.* For the present cognitive models, only student's test scores and skill-item correlation matrix are taken as the prior information for parameter estimation. The lack of prior information always affects the quality and robustness of parameters.
- *How to characterize student's knowledge level.* By introducing the higher-order latent trait, the higher-order models (e.g. HO-DINA and Fuzzy-CDF) significantly improves the interpretability and prediction accuracy of the lower-order CDMs (e.g. IRT and DINA). However, the existing higher-order cognitive models consider only student's innate ability without other potential factors. These often lead to limited performance in educational assessment. For example, student's acquired effort can compensate their innate ability and also plays an important role in calculating knowledge level.
- *How to represent student's item mastery.* Most of the CDMs take only student's skill proficiency into consideration in estimating test item mastery and equally treat each skill in solving a question. These always result in low prediction accuracy because the importance level of each skill varies in different test items. For example, the least-mastered skill may seem fairly unimportant for a test item while the most necessary for another one.

To solve these problems, we propose a new higher-order cognitive diagnostic model named the multivariate cognitive response framework (MvCRF) mainly focusing on the static cognitive diagnosis. The main contributions are summarized as below:

- Besides student's item scores and skill-item correlation matrix, we collect student's learning activity logs (i.e. homework, classroom tests, class discussions and learning hours) for MvCRF to determine the effort trait.
- Unlike the existing CDMs using only student's innate ability, MvCRF first introduces the effort trait combined with ability trait and utilizes a compensation mechanism between the two latent traits to calculate knowledge level. Furthermore, we analyze the compensation mechanism based on two assumptions reflecting the interactions between ability and effort traits.
- Integrating the skill importance level of each test item with knowledge level, MvCRF applies skill weakness to estimate item mastery level and then predicts corresponding item score by considering slip and guess.
- We propose a Markov Chain Monte Carlo (MCMC) sampling algorithm to estimate the parameters of MvCRF and verify its effectiveness and interpretability by extensive simulated and real experiments.

The remainder of the paper is organized as follows. Section 2 reviews several typical CDMs. Section 3 specifies the whole framework of MvCRF. Section 4 gives the experimental results of MvCRF in both simulation study and real-data application. Finally, Section 5 concludes this paper.

2 Related Work

2.1 Static Cognitive Diagnosis

2.1.1 IRT

IRT model [9] takes advantages of the examinee's knowledge level and the multidimensional feature of each test item to predict the examinee's response. IRT model is formulated as:

$$P\left(r_{ij}=1|\alpha_i,\lambda_j,d_j,c_j\right)=c_j+\frac{1-c_j}{1+e^{-D\lambda_j(\alpha_i-d_j)}} \tag{1}$$

where r_{ij} denotes the score of examinee i on item j. $r_{ij}=0$ (or 1) means that the answer is wrong (or correct). $P\left(r_{ij}=1|\cdot\right)\in[0,1]$ denotes the probability of examinee i answering item j correctly. The scalar α_i denotes the knowledge level of examinee i. The item discrimination parameter λ_j denotes the ability of item j to distinguish between higher ability examinees and lower ability examinees. The difficulty parameter d_j denotes the difficulty of item j. The guessing parameter c_j denotes the probability of an examinee correctly answering item j by guessing. The scale parameter D is usually set to 1.7.

As mentioned above, IRT model considers examinee's knowledge level as one-dimensional data or a scalar. However, the knowledge level of an examinee is actually multidimensional because it also varies in skills. Therefore, the traditional IRT model usually results in poor fitting result in practice. To address this issue, an IRT-like higher-order cognitive diagnostic model [13], [18], [19] was proposed to formulate the relationship between examinee's latent trait and knowledge level

$$P\left(\alpha_{ik}=1\right)=\frac{1}{1+e^{-1.7\lambda_{1k}(\theta_i-\lambda_{0k})}} \tag{2}$$

where $P\left(\alpha_{ik}=1\right)\in[0,1]$ denotes the probability that examinee i has fully mastered skill k. θ_i denotes the latent trait parameter of examinee i. λ_{1k} and λ_{0k} denote the discrimination and difficulty parameters of skill k, respectively.

2.1.2 DINA

DINA model [10] considers both examinee's multidimensional knowledge level and multidimensional feature of test item to predict examinee's test result. The DINA model is espressed as

$$P\left(r_{ij}=1|\delta_{ij},s_j,g_j\right)=\left(1-s_j\right)^{\delta_{ij}}\cdot g_j^{1-\delta_{ij}} \tag{3}$$

$$\delta_{ij}=\prod_{k=1}^{K}\alpha_{ik}^{q_{kj}} \tag{4}$$

where δ_{ij}, α_{ik} and q_{kj} are binary variables. The item mastery level (or the latent response) δ_{ij} denotes whether examinee i has mastered item j or not. The knowledge level α_{ik} denotes whether examinee i has mastered skill k or not. The correlation parameter q_{kj} denotes whether skill k is required to correctly answer item j. s_j and g_j are the slip parameter and guessing parameter in terms of item j, respectively. $s_j=P\left(r_{ij}=0|\delta_{ij}=1\right)$ denotes the probability of examinee i incorrectly answering item j because

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2023.3302848

of carelessness or slipping when examinee i has mastered all the required skills for item j. $g_j = P(r_{ij} = 1|\delta_{ij} = 0)$ denotes the probability of examinee i correctly answering item j by guessing when examinee i lacks at least one of the required skills for item j.

2.1.3 HO-DINA

As a variant version of DINA model, HO-DINA model [13] is formulated as

$$P(r_{ij} = 1|\delta_{ij}, s_j, g_j) = (1 - s_j)\,\delta_{ij} + g_j\,(1 - \delta_{ij}) \tag{5}$$

where δ_{ij} is obtained by Eqs. (2) and (4).

Based on HO-DINA model, Fuzzy-CDF model [18] improves Eq. (4) by treating examinee's knowledge level $\alpha_{ik} \in [0, 1]$ as a continuous variable instead of a binary variable. Fuzzy-CDF uses the following equation to formulate the conjunctive interaction [28] of all required skills on item j

$$\delta_{ij} = \min(\alpha_{ik}),\, k \in \{k|q_{kj} = 1, k \in [1, K], k \in \mathbb{Z}\} \tag{6}$$

where $\mathbb{Z}$ is the integer domain. Eq. (6) indicates that δ_{ij}, the mastery level of examinee i on item j, depends on the least-mastered skill.

2.2 Knowledge Tracing

BKT model [26] associates examinee's knowledge level with a binary latent variable (i.e. mastered or non-mastered) and examinee's response with a binary observed variable (i.e. correct or wrong), and utilizes the hidden Markov model to update the knowledge level along with the process of solving questions. In addition, BKT model also considers the cases of guessing the right answer and missing the right answer by carelessness.

Compared with hidden Markov model, the recurrent neural networks (RNN) are more suitable for building complex model due to its high dimensional and continuous representation of latent trait. Accordingly, DKT model [27] takes advantages of RNN to track the time-series information in examinee's responses, then evaluates the changes of examinee's knowledge level over time and finally predict his future performance.

Inspired by DKT, extensive efforts were made in knowledge tracing. The dynamic key-value memory networks (DKVMN) [29] extended DKT by using a static matrix to store all the skills and a dynamic matrix to maintain and update the student's knowledge level. To balance the interpretability and accurate performance prediction for personalized learning, the attentive knowledge tracing method [30] utilizes a series of attention networks to extract connections between the current question and the previous questions the student has answered. Taking advantages of graph neural network (GNN), [31] reformulated knowledge tracing (GKT) as a GNN task and proposed graph-based knowledge tracing to address complex and multiple relationships between skills. Since DKT and DKVMN fail to capture the high-level semantic information and GKT ignores the influence between test items, [32] proposed a joint graph convolutional network by exploring the implicit relationships between test items and skills.

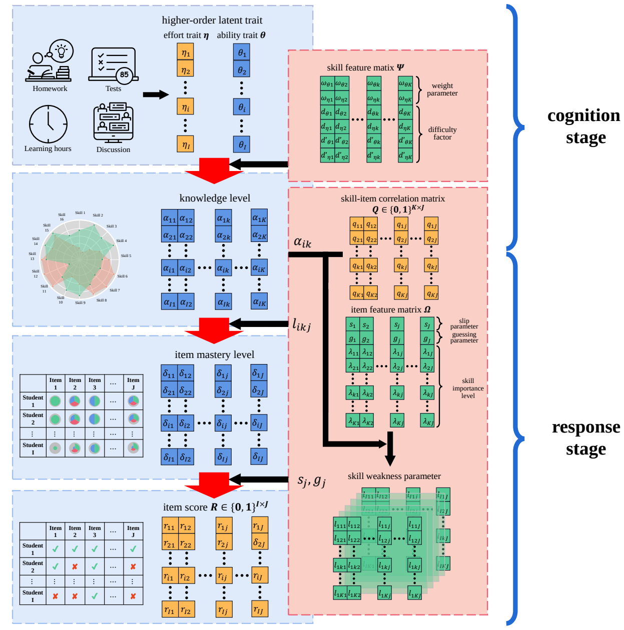


Fig. 1. The structure of MvCRF.

TABLE 1
Notation Summary

Notation	Description	Variable type
θ_i	the ability trait of examinee i	continuous
η_i	the effort trait of examinee i	continuous
α_{ik}	the proficiency of examinee i on skill k	continuous
δ_{ij}	the mastery level of examinee i on item j	continuous
r_{ij}	the predicted score of examinee i on item j	binary
q_{kj}	the requisite of skill k for item j	binary
λ_{kj}	the importance level of skill k for item j	continuous
l_{ikj}	the skill weakness parameter	continuous
s_j	the slip parameter of item j	continuous
g_j	the guessing parameter of item j	continuous
$\boldsymbol{\Psi}$	the skill feature matrix	continuous
$\boldsymbol{\Omega}$	the test item feature matrix	continuous
$\boldsymbol{Q}$	the skill-item correlation matrix	binary
$\boldsymbol{R}$	the score matrix	binary
$\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\alpha}_i, \boldsymbol{r}_i$	the corresponding vector of $\theta_i, \eta_i, \alpha_{ik}$ and r_{ij}	-
$\boldsymbol{\Psi}_k, \boldsymbol{\Omega}_j, \boldsymbol{q}_j$	the kth or jth column vector of $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$ or $\boldsymbol{Q}$	-
I, J, K	the numbers of examinees, items and skills	-

3 Multivariate Cognitive Response Framework

After reviewing the typical CDMs, we propose a novel higher-order CDM, called the **M**ulti**v**ariate **C**ognitive **R**esponse **F**ramework (MvCRF). The model applies a compensation mechanism to integrate ability and effort traits to estimate examinee's knowledge level. Furthermore, MvCRF introduces the skill weakness parameter to take both examinee's knowledge level and the importance of each skill to predict the mastery level of test item.

3.1 The Structure of MvCRF

As shown in Fig. 1, MvCRF is a four-tier higher-order model to catch the relationship between examinee's knowledge level and their test performance. In MvCRF, the whole learning process is divided into two stages, i.e. the cognition stage and the response stage. The cognition stage is the process of acquiring knowledge and understanding through

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2023.3302848

individual learning activities, while the response stage is the process of taking a test based on the acquired knowledge.

For easy reading, Table 1 summarizes the main notations used in the following sections.

3.2 Cognition Stage

The cognition stage includes determining higher-order latent trait and calculating knowledge level. Since the examinee's knowledge level is correlated with the ability trait, effort trait and skill feature, we formulate the cognitive model as

$$\boldsymbol{\alpha}_i = f_c\left(\theta_i, \eta_i, \boldsymbol{\Psi}\right), \tag{7}$$

where θ_i and η_i denote the ability trait and effort trait of examinee $i \in \{1, 2, \cdots, I\}$, respectively. $\boldsymbol{\Psi}$ denotes the skill feature matrix of K skills. The knowledge level α_{ik} denotes the proficiency of examinee i on skill $k \in \{1, 2, \cdots, K\}$. $\boldsymbol{\alpha}_i = [\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{iK}]^T$ denotes the knowledge level vector of examinee i. I and K are the total numbers of examinees and skills, respectively.

3.2.1 Determining Higher-order Latent Trait

Unlike the typical higher-order CDMs in Section 2, the higher-order latent trait in MvCRF consists of two independent components, i.e. the ability trait θ_i and effort trait η_i. If the whole learning process of an individual can be divided into several phases, people has the ability trait before starting a learning phase and develop the effort trait through daily experience over the given period of time. In other words, the effort trait is regarded as the acquired, explicit and dynamic quality of individual, and ability trait as the innate, implicit and stable quality during a given learning phase. Obviously, the examinee's knowledge level increases monotonically with θ_i or η_i increasing. In addition, the ability trait θ_i is not related with the difficulty of skill or test item. Different from the ability trait, the effort trait η_i can be reflected by the learning activity logs, such as homework, classroom tests, class discussions and learning hours in Fig. 1.

Ability trait. Let P_{ik} denote the ratio of the examinees with knowledge level no more than α_{ik} to the total, then we have

$$P_{ik} = \frac{\#\{i_0 | \alpha_{i_0 k} \leq \alpha_{ik}, i_0 \in \mathbb{Z}, i_0 \in [1, I]\}}{I} \tag{8}$$

where # is the count operator. Since the rank order of $\{\alpha_{1k}, \alpha_{2k}, \cdots, \alpha_{Ik}\}$ is fixed for different skill k, the equation $P_{i1} = P_{i2} = \cdots = P_{iK} = P_i$ holds. Suppose examinee's ability trait θ_i obeys standard normal distribution, the relationship between θ_i and P_i can be formulated as

$$P_i = \int_{-\infty}^{\theta_i} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt. \tag{9}$$

According to [33], we have $\int_{-\infty}^{\theta_i} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \approx \frac{1}{1+e^{-1.7\theta_i}}$ with approximate error less than 0.01 for any θ_i in real domain. Therefore, Eq. (9) can be expressed as

$$P_i = \frac{1}{1 + e^{-1.7\theta_i}}. \tag{10}$$

Effort trait. Like the ability trait θ_i, the effort trait η_i also obeys the standard normal distribution as

$$1 - \frac{rank(\eta_i)}{I+1} = \int_{-\infty}^{\eta_i} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \approx \frac{1}{1 + e^{-1.7\eta_i}} \tag{11}$$

where $rank(\eta_i)$ denotes the rank of examinee i based on the effort level. After using Eq. (11), the ranking order of η_i stays unchanged.

3.2.2 Calculating Knowledge Level

Extensive empirical works [34], [35], [36] have given solid support for three common mathematical laws in the field of human psychology. The three laws are adding, multiplying and averaging shown in Table 2. As can be seen, all the three laws can reflect the compensation relationship between ability and effort. In the adding law, the effort is equally effective in improving performance no matter examinees are low or high in ability. Besides, the examinee with high ability performs better than that with low ability even both of them make no effort. However, in the multiplying law, the effort can greatly improve performance only for examinees with high ability. Examinees, whether with high or low ability, gain almost nothing by no effort. In the averaging law, there are two kinds of parameters, i.e. the weight w and scale value s. The weight w_θ and w_η correspond to the relative importance of ability and effort, respectively. The scale value s_θ and s_η denote the numerical values of ability and effort on a rating scale, respectively. w_0 and s_0 denote the initial state. Different from the adding law, the averaging law introduces the weight to adjust the effect of ability and effort on performance. The sum of all weights equals to 1.

TABLE 2
Three laws of information integration

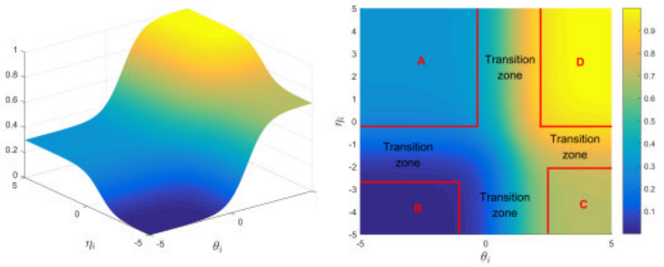
Law	Calculation
Adding	Performance = Ability + Effort
Multiplying	Performance = Ability $\times$ Effort
Averaging	Performance = $\frac{w_\theta s_\theta + w_\eta s_\eta + w_0 s_0}{w_\theta + w_\eta + w_0}$

Considering the pros and cons of the three laws, we first introduce the following assumptions:

Assumption 1 *The effort trait can compensate the ability trait in improving examinee's knowledge level and the compensation effect is limited.*

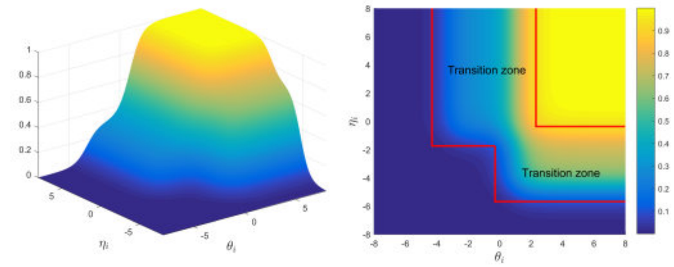
Assumption 2 *The examinee's knowledge level approaches to zero when one of the ability trait and effort trait goes to the negative infinity.*

Assumption 1 indicates that the examinee with low ability can catch up the examinee with the high ability by effort. In addition, none of the examinees can improve their knowledge level limitlessly. Assumption 2 indicates that the examinees can hardly improve their knowledge levels when they make no effort or have almost no ability to learn. Based on the two assumptions, we then propose the compensation mechanism to calculate examinee's knowledge level. The mechanism includes combined compensation and decaying compensation, and the details are presented as below.



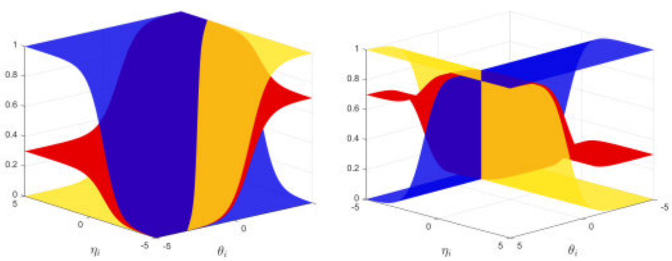
(a) (b)

Fig. 2. The 3D graphs of α_{ik} vs. (θ_i, η_i) by Eq. (12): (a) front view; (b) top view.



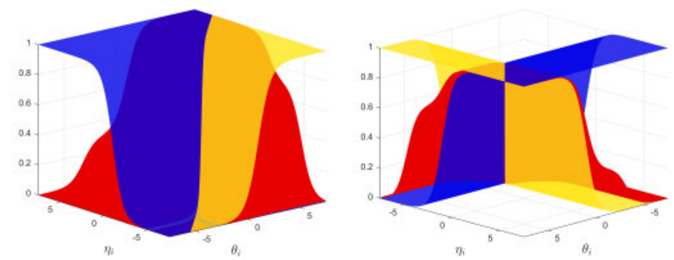
(a) (b)

Fig. 4. The 3D graphs of α_{ik} vs. (θ_i, η_i) by Eq. (16): (a) front view; (b) top view.



(a) (b)

Fig. 3. Three surfaces of α_{ik} vs. (θ_i, η_i) (red) by Eq. (12), $\hat{\alpha}_{ik}$ vs. θ_i (blue) and $\check{\alpha}_{ik}$ vs. η_i (yellow): (a) front view; (b) back view.



(a) (b)

Fig. 5. Three surfaces of α_{ik} vs. (θ_i, η_i) (red) by Eq. (16), $\hat{\alpha}_{ik}$ vs. θ_i (blue) and $\check{\alpha}_{ik}$ vs. η_i (yellow): (a) front view; (b) back view.

Combined compensation. Suppose the ability trait and effort trait are independent, we have the following equation to calculate the knowledge level α_{ik}

$$\alpha_{ik} = \frac{w_{\theta k}}{1 + e^{-1.7(\theta_i - d_{\theta k})}} + \frac{w_{\eta k}}{1 + e^{-1.7(\eta_i - d_{\eta k})}}, \tag{12}$$

where $w_{\theta k} + w_{\eta k} = 1$. $d_{\theta k}$ and $d_{\eta k}$ are the difficulty factors. $w_{\theta k}$ and $w_{\eta k} \in [0, 1]$ are the weight parameters of θ_i and η_i on skill k, respectively.

For example, given $w_{\theta k} = 0.7, w_{\eta k} = 0.3, d_{\theta k} = -d_{\eta k} = 1$ in Eq. (12), we can plot the 3D graphs of the knowledge level α_{ik} versus two latent traits (θ_i, η_i) in Fig. 2. As can be seen, there are more than one value pair of (θ_i, η_i) corresponding to each α_{ik}. Furthermore, the whole region is divided into four regions, i.e. Region A with lower θ_i and higher η_i, Region B with lower θ_i and η_i, Region C with higher θ_i and lower η_i, and Region D with higher θ_i and η_i. The intra-regional parts are flat while the inter-regional parts are the transition zones from one flat region to the other. Let $\hat{\alpha}_{ik} = \frac{1}{1 + e^{-1.7(\eta_i - d_{\eta k})}}$ and $\check{\alpha}_{ik} = \frac{1}{1 + e^{-1.7(\theta_i - d_{\theta k})}}$, we obtain three surfaces of α_{ik}, $\hat{\alpha}_{ik}$ and $\check{\alpha}_{ik}$ as shown in Fig. 3. Since $0 \leq w_{\theta k}, w_{\eta k} \leq 1$, we have

$$\min(\hat{\alpha}_{ik}, \check{\alpha}_{ik}) \leq \alpha_{ik} \leq \max(\hat{\alpha}_{ik}, \check{\alpha}_{ik}). \tag{13}$$

As can be observed, the yellow surface is located between the red and blue surfaces. Therefore, Assumption 1 holds.

Decaying compensation. In terms of Eq. (12), we have

$$\lim_{\theta_i \to -\infty} \alpha_{ik} = \frac{w_{\eta k}}{1 + e^{-1.7(\eta_i - d_{\eta k})}} \neq 0, \tag{14}$$

$$\lim_{\eta_i \to -\infty} \alpha_{ik} = \frac{w_{\theta k}}{1 + e^{-1.7(\theta_i - d_{\theta k})}} \neq 0. \tag{15}$$

where $\equiv$ is the equivalence sign. However, it does not satisfy Assumption 2. To address this issue, we further introduce two decay factors $\beta_{\theta k}$ and $\beta_{\eta k}$ as below

$$\alpha_{ik} = \frac{\beta_{\theta k} w_{\theta k}}{1 + e^{-1.7(\theta_i - d_{\theta k})}} + \frac{\beta_{\eta k} w_{\eta k}}{1 + e^{-1.7(\eta_i - d_{\eta k})}}, \tag{16}$$

$$\beta_{\theta k} = \frac{1}{1 + w_{\eta k} e^{-1.7(\eta_i - d'_{\eta k})}}, \beta_{\eta k} = \frac{1}{1 + w_{\theta k} e^{-1.7(\theta_i - d'_{\theta k})}} \tag{17}$$

where $w_{\theta k} + w_{\eta k} = 1$. $d_{\theta k}$, $d_{\eta k}$, $d'_{\theta k}$ and $d'_{\eta k}$ are the difficulty factors. As shown in Eq. (7), $\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \cdots, \boldsymbol{\Psi}_K]$ and $\boldsymbol{\Psi}_k = [w_{\theta k}, w_{\eta k}, d_{\theta k}, d_{\eta k}, d'_{\theta k}, d'_{\eta k}]^T$. According to Eq. (17), $\beta_{\theta k}$ is positively correlated with η_i and inversely correlated with $w_{\eta k}$, because the effect of ability trait on knowledge level decreases as the effort trait becomes more and more important. The same goes for $\beta_{\eta k}$ in Eq. (17).

For example, given $w_{\theta k} = 0.7, w_{\eta k} = 0.3, d_{\theta k} = -d_{\eta k} = 1$, $d'_{\theta k} = -3$ and $d'_{\eta k} = -4$ in Eq. (16), we can plot the 3D graphs of α_{ik} vs. (θ_i, η_i) in Fig. 4 and three surfaces of α_{ik}, $\hat{\alpha}_{ik}$ and $\check{\alpha}_{ik}$ as shown in Fig. 5. As can be seen, α_{ik} tends to 0 as θ_i or η_i decreases and satisfies Eq. (13). Thus, Assumptions 1 and 2 hold.

3.3 Response Stage

The response stage consists of estimating item mastery level and predicting item score. Applying the test feature and the examinee's knowledge level obtained in the cognitive stage, the response model can be expressed as

$$\boldsymbol{r}_i = f_r(\boldsymbol{\alpha}_i, \boldsymbol{\Omega}, \boldsymbol{Q}), \tag{18}$$

where $\boldsymbol{\Omega}$ and $\boldsymbol{Q}$ denote the test item feature matrix and the correlation matrix between skill and test item, respectively. r_{ij} is a binary variable and indicates the predicted score of examinee i on item $j \in \{1, 2, \cdots, J\}$.

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2023.3302848

$\boldsymbol{r}_i = [r_{i1}, r_{i2}, \cdots, r_{iJ}]^T$ denotes the predicted score vector of examinee i. J is the total number of test items.

3.3.1 Estimating Item Mastery Level

Similar to Liebig's law of the minimum, Fuzzy-CDF [18] determines the mastery level of a test item by the least-mastered skill as Eq. (6) under the assumption that the skills' interaction on objective questions is conjunctive. However, the examinee still probably has a good master of the test item when the least-mastered skill is of less importance for the item than the other required skills. Moreover, when the least-mastered skill gets more important, the examinee becomes more likely to miss the item. Therefore, to characterize item mastery more reasonably, we introduce the skill weakness parameter by considering not only the knowledge level but also the importance of skill in estimating item mastery level.

Let λ_{kj} denotes the importance level of skill k in correctly answering item j, we have the skill weakness parameter l_{ikj} as

$$l_{ikj} = \lambda_{kj}(1 - \alpha_{ik}), \tag{19}$$

where l_{ikj} denotes the weakness degree of skill k for examinee i to correctly answer item j. Eq. (19) indicates that the skill weakness parameter l_{ikj} goes higher when skill k becomes more important to item j while the proficiency of examinee i on skill k becomes lower. Then, we obtain the mastery level δ_{ij} of examinee i on item j as

$$\delta_{ij} = 1 - \max\left(l_{ikj}\right), k \in \{k | q_{kj} = 1, k \in [1, K], k \in \mathbb{Z}\}, \tag{20}$$

where the skill-item correlation parameter q_{kj} denotes whether skill k is required to correctly answer item j. Besides, $\boldsymbol{q}_j = [q_{1j}, q_{2j}, \cdots, q_{Kj}]^T$ and $\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_J]$. Eq. (20) means that the mastery level δ_{ij} of examinee i on skill j is determined by the skill with the maximum l_{ikj}. Accordingly, the item mastery in MvCRF involves both the skills' conjunction interaction and the importance of each skill required by a specific item.

3.3.2 Predicting Item Score

Considering both the item mastery level and some occasional cases, such as carelessness and guessing, we can obtain the score r_{ij} of examinee i on item j as

$$P\left(r_{ij} = 1 | \delta_{ij}, s_j, g_j\right) = (1 - s_j)\delta_{ij} + g_j(1 - \delta_{ij}) \tag{21}$$

where s_j and g_j are the slip parameter and guessing parameter in terms of item j, respectively. In addition, we have $s_j = P\left(r_{ij} = 0 | \delta_{ij} = 1\right)$ and $g_j = P\left(r_{ij} = 1 | \delta_{ij} = 0\right)$. In Eq. (18), $\boldsymbol{\Omega}_j = [s_j, g_j, \lambda_{1j}, \lambda_{2j}, \cdots, \lambda_{Kj}]^T$ and $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \cdots, \boldsymbol{\Omega}_J]$.

3.4 Parameter Estimation for MvCRF

According to Bayes rule [10], the joint posterior distribution of $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_I]^T$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ given score matrix $\boldsymbol{R} = [\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_I]$ is

$$P\left(\boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega} | \boldsymbol{R}\right) \propto P(\boldsymbol{\Psi})P(\boldsymbol{\theta})P(\boldsymbol{\Omega})L\left(\boldsymbol{R} | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right), \tag{22}$$

where the joint likelihood function L is expressed as

Algorithm 1 Parameter estimation for MvCRF

Input: $\boldsymbol{R}$, $\boldsymbol{Q}$ and $\boldsymbol{\eta} = [\eta_1, \eta_2, \cdots, \eta_I]^T$
Output: samples of $\boldsymbol{\theta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$

Generate L parallel Markov chains
for $l = 1 : L$ **do**
  Initialize $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\Psi}^{(0)}$ and $\boldsymbol{\Omega}^{(0)}$ with random values;
  for $t = 1 : T$ **do**
    for $k = 1 : K$ **do**
      Draw the candidates $w_{\theta k}^{(t)} \sim U(w_{\theta k}^{(t-1)} - \delta_w, w_{\theta k}^{(t-1)} + \delta_w)$, $d_{\theta k}^{(t)} \sim N(d_{\theta k}^{(t-1)}, \sigma_{d_\theta}^2)$ and $d_{\eta k}^{(t)} \sim N(d_{\eta k}^{(t-1)}, \sigma_{d_\eta}^2)$;
      Calculate $w_{\eta k}^{(t)} = 1 - w_{\theta k}^{(t)}$ and ${d'}_{\theta k}^{(t)}, {d'}_{\eta k}^{(t)}$ by Eqs. (27);
      Obtain $\boldsymbol{\Psi}_k^{(t)} = [w_{\theta k}^{(t)}, w_{\eta k}^{(t)}, d_{\theta k}^{(t)}, d_{\eta k}^{(t)}, {d'}_{\theta k}^{(t)}, {d'}_{\eta k}^{(t)}]^T$;
    end for
    Accept $\boldsymbol{\Psi}^{(t)} = [\boldsymbol{\Psi}_1^{(t)}, \boldsymbol{\Psi}_2^{(t)}, \cdots, \boldsymbol{\Psi}_K^{(t)}]$ with probability

$$P\left(\boldsymbol{\Psi}^{(t-1)}, \boldsymbol{\Psi}^{(t)}\right) = \min\left\{\frac{P\left(\boldsymbol{\Psi}^{(t)} | \boldsymbol{\theta}^{(t-1)}, \boldsymbol{\Omega}^{(t-1)}, \boldsymbol{R}\right)}{P\left(\boldsymbol{\Psi}^{(t-1)} | \boldsymbol{\theta}^{(t-1)}, \boldsymbol{\Omega}^{(t-1)}, \boldsymbol{R}\right)}, 1\right\}$$

    for $i = 1 : I$ **do**
      Draw $\theta_i^{(t)} \sim N(\theta_i^{(t-1)}, \sigma_\theta^2)$;
    end for
    Accept $\boldsymbol{\theta}^{(t)} = [\theta_1^{(t)}, \theta_2^{(t)}, \cdots, \theta_I^{(t)}]^T$ with probability

$$P\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}^{(t)}\right) = \min\left\{\frac{P\left(\boldsymbol{\theta}^{(t)} | \boldsymbol{\Psi}^{(t)}, \boldsymbol{\Omega}^{(t-1)}, \boldsymbol{R}\right)}{P\left(\boldsymbol{\theta}^{(t-1)} | \boldsymbol{\Psi}^{(t)}, \boldsymbol{\Omega}^{(t-1)}, \boldsymbol{R}\right)}, 1\right\}$$

    for $j = 1 : J$ **do**
      Draw the candidates $s_j^{(t)} \sim U(s_j^{(t-1)} - \delta_s, s_j^{(t-1)} + \delta_s)$ and $g_j^{(t)} \sim U(g_j^{(t-1)} - \delta_g, g_j^{(t-1)} + \delta_g)$;
      for $k = 1 : K$ **do**
        Draw $\lambda_{kj}^{(t)} \sim U(\lambda_{kj}^{(t-1)} - \delta_\lambda, \lambda_{kj}^{(t-1)} + \delta_\lambda)$;
      end for
      Obtain $\boldsymbol{\Omega}_j^{(t)} = [s_j^{(t)}, g_j^{(t)}, \lambda_{j1}^{(t)}, \lambda_{j2}^{(t)}, \cdots, \lambda_{kj}^{(t)}]^T$;
    end for
    Accept $\boldsymbol{\Omega}^{(t)} = [\boldsymbol{\Omega}_1^{(t)}, \boldsymbol{\Omega}_2^{(t)}, \cdots, \boldsymbol{\Omega}_J^{(t)}]$ with probability

$$P\left(\boldsymbol{\Omega}^{(t-1)}, \boldsymbol{\Omega}^{(t)}\right) = \min\left\{\frac{P\left(\boldsymbol{\Omega}^{(t)} | \boldsymbol{\Psi}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{R}\right)}{P\left(\boldsymbol{\Omega}^{(t-1)} | \boldsymbol{\Psi}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{R}\right)}, 1\right\}$$

    if convergence criterion meets **then**
      return
    end if
  end for
end for
return

$$L\left(\boldsymbol{R} | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right) = \prod_{i=1}^{I} \prod_{j=1}^{J} \left[P\left(r_{ij} = 1 | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right)\right]^{r_{ij}} \cdot \left[1 - P\left(r_{ij} = 1 | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right)\right]^{1 - r_{ij}}. \tag{23}$$

Then, the full conditional distributions of each parameter given $\boldsymbol{R}$ and the rest of the parameters are as follows

$$P\left(\boldsymbol{\Psi} | \boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{R}\right) \propto P(\boldsymbol{\Psi})L\left(\boldsymbol{R} | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right), \tag{24}$$

$$P\left(\boldsymbol{\theta} | \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{R}\right) \propto P(\boldsymbol{\theta})L\left(\boldsymbol{R} | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right), \tag{25}$$

$$P\left(\boldsymbol{\Omega} | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{R}\right) \propto P(\boldsymbol{\Omega})L\left(\boldsymbol{R} | \boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega}\right). \tag{26}$$

At last, we apply the Markov Chain Monte Carlo (MCMC) method [13], [18], [37] to estimate the parameters for MvCRF as shown in Algorithm 1. Given the score matrix $\boldsymbol{R}$, skill-item correlation matrix $\boldsymbol{Q}$ and effort trait vector $\boldsymbol{\eta}$, we first initialize $\boldsymbol{\theta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ with random values. For each iteration t, we draw a random sample of each parameter from a uniform distribution $U(min, max)$ or normal distribution $N(\mu, \sigma^2)$ within a preset interval. μ and σ^2 denote mathematic expectation and variance, respectively. Next, we calculate the full conditional probability of $\boldsymbol{\theta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ by Eqs. (24)-(26) and then the corresponding acceptance probability. After T iterations of sampling, we obtain the final parameter estimation of $\boldsymbol{\theta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$.

3.5 Discussion

3.5.1 Skill Feature vs. Skill Weakness

As shown in Fig. 1, the skill feature matrix $\boldsymbol{\Psi}$ and the skill weakness parameter l_{ikj} both can characterize the skills. To better understand the similarities and differences between $\boldsymbol{\Psi}$ and l_{ikj}, we compare the two parameters from the following three aspects.

- **Stage.** $\boldsymbol{\Psi}$ is obtained in cognition stage while l_{ikj} is calculated in response stage.
- **Definition.** $\boldsymbol{\Psi}$ includes the weight parameters $\omega_{\theta k}, \omega_{\eta k}$ and difficult factors $d_{\theta k}, d_{\eta k}, d'_{\theta k}, d'_{\eta k}$ of effort and ability traits for skill $k \in \{1, 2, \cdots, K\}$, while l_{ikj} denotes the weakness degree of skill k for examinee i to correctly answer test item j related with skill k.
- **Purpose.** $\boldsymbol{\Psi}$ is used to compute examinee's knowledge level, i.e. the proficiency on a specific knowledge, while l_{ikj} is used to estimate examinee's mastery level on each item.

Therefore, the skill feature matrix $\boldsymbol{\Psi}$ depends only on skills while the skill weakness parameter l_{ikj} depends on not only skills but also examinees and items.

3.5.2 Comparison with Several Exisiting CDMs

First, we compare MvCRF with some static CDMs in module design. The parameters or modules considered by each method are summarized in Table 3. As can be seen, DINA and NeuralCD consider only two of the eight aspects. IRT, Fuzzy-CDF and HO-DINA adopt the ability trait (abi.), the difficulty factor (dif.) and the guessing parameter (guess). Besides, the former two both apply the discrimination factor (dis.) while the latter two utilize the slip parameter (slip). However, MvCRF includes the ability trait, the effort trait (eff.), the difficulty factor, the skill weakness parameter (weak.), the slip parameter, the guessing parameter and the compensation mechanism (comp.) simultaneously. Furthermore, MvCRF also considers the importance level of each skill to a specific item to exploit the implicit relation between them.

TABLE 3
The summary of methods

Method\Parameter	abi.	eff.	dif.	dis.	weak.	slip	guess	comp.
IRT	√		√	√			√	
DINA						√	√	
HO-DINA	√		√			√	√	
Fuzzy-CDF	√		√	√		√	√	
NeuralCD			√	√				
MvCRF-I	√	√	√			√	√	c
MvCRF-II	√	√	√			√	√	d
MvCRF-III	√		√		√	√	√	
MvCRF	√	√	√		√	√	√	d

Second, we list the pros and cons of MvCRF over two dynamic CDMs, i.e. BKT and DKT models, as below.

- BKT and DKT use the skill-item correlation information and examinee's score as the input and lack information in other learning activities. MvCRF further collects examinee's learning activity logs of homework, tests and discussions to determine another input feature, i.e. the effort trait.
- BKT evaluates a single skill at one time and cannot analyzes multiple skills simultaneously. DKT can model the connections among multiple skills by examinee's historical performance and score. MvCRF introduces the concept of skill weakness to associate the importance level of each skill with each test item and each examinee.
- The binary hypothesis of BKT ignores fuzzy and continuous representation of examinee's knowledge level. DKT evaluates only examinee's general knowledge level and lacks the mastery level on each skill. MvCRF can estimate the fuzzified knowledge level on each skill.
- BKT and DKT are suitable for examinee's dynamic learning process while MvCRF focuses on the comprehensive analysis of examinee's current knowledge level during a given period of time.

4 EXPERIMENTS

In this section, we first investigate the parameter recovery performance of MvCRF with some simulated datasets, and then compare MvCRF with some baseline methods on the examinee performance prediction task with real datasets to verify the effectiveness and interpretability of the proposed cognitive model.

4.1 Simulation Study

4.1.1 Experimental Setup

In this experiment, we first input $\boldsymbol{\eta}, \boldsymbol{Q}$ and $\boldsymbol{R}$ and run L parallel Markov chains for T iterations. To guarantee the convergence of Markov chains generated in Algorithm 1, we then estimate $\boldsymbol{\theta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ using the last half of the iterations. Finally, we compare the estimated value and the true value of each parameter and analyze the error. The detailed simulation setup is given as follows.

Dataset. Accoding to [10], [13], we randomly generate the binary matrix $\boldsymbol{Q}$ and other simulation parameters using the prior distributions as

$$
\begin{aligned}
&\theta_i \sim N(0,1), \eta_i \sim N(0,1), w_{\theta k} \sim B(2,2), d_{\theta k} \sim N(0,1),\\
&d_{\eta k} \sim N(0,1), 1-\lambda_{kj} \sim B(0.5, 5.25),\\
&g_j \sim 4 - B(0, 0.6), 1 - s_j \sim 4 - B(0.4, 1),\\
&d'_{\theta k} = d_{\theta k} + \frac{\log(0.001) - \log(w_{\theta k}) - \log(0.999)}{1.7},
\end{aligned}
$$

$$d'_{\eta k} = d_{\eta k} + \frac{\log(0.001) - \log(w_{\eta k}) - \log(0.999)}{1.7}, \tag{27}$$

and then apply Eqs. (16)-(17) and (20)-(21) to calculate the corresponding probability $P\left(r_{ij}=1|\cdot\right)$ and the score matrix $\boldsymbol{R}$.

Hyperparameter settings. The hyperparameters in Algorithm 1 are set as

$$\begin{aligned} &L = 5, T = 10000, \\ &\delta_w = \delta_\lambda = \delta_g = \delta_s = 0.05, \sigma^2_{d_\theta} = \sigma^2_{d_\eta} = 0.3, \sigma^2_\theta = 1.1, \end{aligned} \tag{28}$$

where $B(a,b)$ generates a random value in the range $(0,1)$ from Beta distribution with parameters specified by a and b.

Assessment metrics. As shown below, we calculate the mean absolute deviation (ABSE) and root mean squared error (RMSE) to evaluate parameter recovery.

$$ABSE = \frac{\sum_{t=t_0}^{T_0} \left|\hat{x}^{(t)} - x\right|}{T_0 - t_0 + 1}, RMSE = \sqrt{\frac{\sum_{t=t_0}^{T_0} \left(\hat{x}^{(t)} - x\right)^2}{T_0 - t_0 + 1}}, \tag{29}$$

where $\hat{x}^{(t)}$ and x denote the estimated value of the parameter in the tth iteration and the corresponding true value, respectively. $[t_0, T_0]$ denotes the selected range of the iteration number. For example, $[t_0, T_0] = [5001, 10000]$ in the experiment.

4.1.2 Results

Table 4 shows the ABSE and RMSE results of parameter estimation by MvCRF in different cases, such as different numbers of examinees, items or skills.

Case 1 presents the estimation errors of examinee's knowledge level and item feature parameters by MvCRF with 500 examinees, 100 items and 3 skills (i.e. $I = 500$, $J = 100$ and $K = 3$). In terms of ABSE and RMSE, most of the results below 0.1 demonstrate good parameter recovery performance of MvCRF.

In Case 2, we analyze the estimation errors of the item feature parameters by MvCRF with $J = 100$ and $K = 3$ when $I = 300$, 500 and 800, respectively. As can be seen, both ABSE and RMSE decrease as I increases, because more examinees involved in the experiment indicate that more feature information per item can be obtained to bring the estimated value closer to the corresponding real value when the numbers of items and skills are fixed.

TABLE 4
ABSE and RMSE of parameter estimation by MvCRF in different cases

Parameter		ABSE			RMSE		
Case 1	α_{ik}	0.0760			0.0973		
	s_j	0.0633			0.0815		
	g_j	0.0712			0.0883		
	λ_{kj}	0.1086			0.1740		
Case 2	I	300	500	800	300	500	800
	s_j	0.0769	0.0633	0.0470	0.0961	0.0815	0.0605
	g_j	0.0740	0.0731	0.0712	0.0991	0.0930	0.0883
	λ_{kj}	0.1166	0.1086	0.0990	0.1804	0.1740	0.1645
Case 3	J	50	80	100	50	80	100
	α_{ik}	0.0774	0.0638	0.0542	0.0970	0.0835	0.0739
Case 4	K_j	1	2	3	1	2	3
	α_{ik}	0.0688	0.0677	0.0712	0.0823	0.0810	0.0933
	s_j	0.0396	0.0603	0.0517	0.0505	0.0843	0.0665
	g_j	0.0713	0.0539	0.0532	0.0907	0.0668	0.0680
	λ_{kj}	0.0127	0.0236	0.0455	0.0428	0.0560	0.0830

In Case 3, we give the estimation error results of the examinee's knowledge level by MvCRF with $I = 500$ and $K = 3$ when $J = 50$, 80 and 100, respectively. As can be observed, the more items we use, the lower ABSE and RMSE become. This is because, for the same examinees, the examinee's performance on more test items reflects the knowledge level in more detail.

Case 4 shows the ABSE and RMSE results of the examinee's knowledge level and item feature parameters by MvCRF with $I = 500$, $J = 100$ and $K = 6$ when $K_j =$ 1, 2 and 3, respectively. K_j denotes the average number of skills contained in each item. On the one side, the increasing number of skills leads to more complicated structure of each item, therefore the estimation error of skill importance level λ_{kj} increases. On the other side, since the examinee's performance on each item indicates more information about the knowledge level as K_j increases, the estimation error of guessing parameter g_j decreases.

4.2 Real-data Application

4.2.1 Experimental Setup

Before the performance comparison, we first give a brief introduction to the experimental setup including the following four parts.

Dataset. There are numerous public datasets for educational data mining, such as educational processing mining dataset (EPM) [38], ASSISTments [39], EdNet [40], FrcSub [41] and Math1&2 [42]. The detailed analysis is listed below.

- EPM contains the grades of the students' assignments for calculating effort trait and the results of the final exam for training and testing model, but lacks the skill-item correlation matrix to calculate knowledge level and skill weakness parameter.
- ASSISTments and EdNet contain the skill-item correlation matrix and question solving logs for each student. The question solving logs of each student are time-series data. The student can learn the tutoring material before answering each question and get the corresponding explanations after solving the given question. It would lead to unbearable computational burden if the effort trait is calculated using the question solving logs and has to be updated for each tutoring request.
- FrcSub and Math1&2 contain the skill-item correlation matrix and the test scores for each student, but lack learning logs to obtain effort trait.

Since MvCRF is a static cognitive diagnostic model and requires both examinee learning activity logs and skill-item correlation matrix to derive effort trait, we select the examinee learning logs, score matrix $\boldsymbol{R}$ and skill-item correlation matrix $\boldsymbol{Q}$ with $I = 37318$, $J = 16383$ and $K = 654$ test items from ASSISTments dataset[1] to verify the effectiveness of MvCRF according to the above analysis. We choose the number of attempts each examinee

1. ASSISTments dataset (May to June in 2020-2021 school year) is available in https://osf.io/qevpu.

made to answer the item (i.e. attempt_count in plogs.csv), the total number of items completed by each examinee (i.e. answered_problems_count in sdets.csv) and the average time on task of an examinee on every item (i.e. mean_problem_time_on_task in sdets.csv) in ASSISTments dataset to compute effort trait. After normalizing the three groups of data separately, we calculate the weighted average of the three values and then take the order of the average as $rank(\eta_i)$. Finally, we can obtain the effort trait η_i for each examinee by Eq. (11).

In addition to the public ASSISments dataset, we collect the HNU_SYS1 and HNU_SYS2 datasets[2] from the Signals and Systems course through MOOC[3]. HNU_SYS1 dataset contains the learning logs, $\boldsymbol{R}$ and $\boldsymbol{Q}$ with $I = 466$, $J = 90$ and $K = 6$. Similarly, HNU_SYS2 dataset includes the learning logs, $\boldsymbol{R}$ and $\boldsymbol{Q}$ with $I = 770$, $J = 221$ and $K = 50$. For effective prediction of MvCRF, the learning activity logs and score matrix $\boldsymbol{R}$ in HNU_SYS1&2 have the following characteristics.

- The learning activity logs cover diverse learning activities, such as homework, classroom tests, class discussions and effective learning hours.
- The learning logs are easy enough to calculate examinee's effort trait by Eq. (11) rather than ability trait, for example, the questions associated with the basic concepts or similar to examples in the tutoring materials.
- The score matrix $\boldsymbol{R}$ is obtained by a phase test covering all the required skills learned by examinees currently, such as the mid-term exam and final exam.
- All the items in the phase test are objective questions.

Since the examinees may just leave the web page opening or tutoring video playing without learning, we use the evaluation results of homework, tests and discussions instead of learning hours to derive effort trait.

Fig. 6 shows a preview of matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ in HNU_SYS1&2 datasets, and effort trait η_i obtained from learning activity logs in HNU_SYS1&2 and ASSISTments datasets. We randomly select more than half items for each examinee. The 1's and 0's in $\boldsymbol{Q}$ denote each skill is and is not required for a specific item, respectively. The 0's, 1's and 2's in $\boldsymbol{R}$ denote the non-selected, incorrectly-answered and correctly-answered items, respectively. According to the performance of examinee learning activities, the effort trait η_i of examinee i can be obtained by Eq. (11). As shown in Figs. 6(c)-(e), the probability density distribution of η_i generally follows the standard normal distribution (i.e. the red curve). In Figs. 6(d)-(e), the minimum values of η_i concentrate between -1.5 to -1.25 because a number of examinees scored zero on the daily homework, tests and discussions, and shared the last place.

Baseline methods. Our MvCRF are compared with five baseline CDMs, including DINA [10], HO-DINA [13], Fuzzy-CDF [18], NeuralCD [20], probabilistic matrix factorization (PMF) [43] and DKT model [27]. All the methods

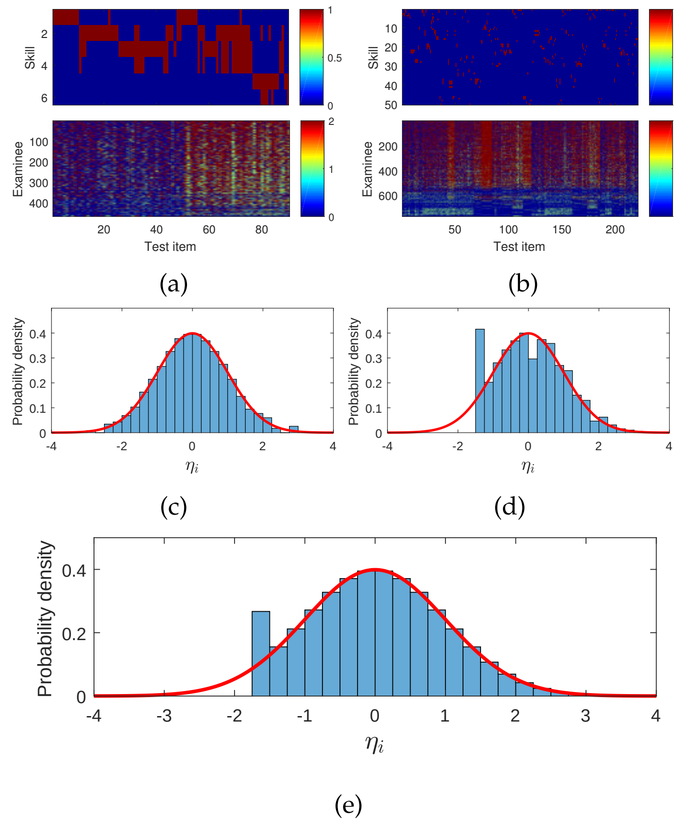


Fig. 6. The preview of matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ from (a) HNU_SYS1 dataset and (b) HNU_SYS2 dataset, and the corresponding probability density distributions of effort trait η_i from and (c)-(d) HNU_SYS1&2 and (e) ASSISTments datasets.

are implemented via Matlab on a Core i7-7700U 3.60GHz machine.

Hyperparameter settings. The hyperparameters in the real-data application are the same as Eq. (28).

Assessment metrics. We adopt two metrics to assess the fitting capability of all the models, i.e. mean absolute error (MAE) and RMSE, and other five metrics to evaluate their prediction performance, i.e. the accuracy (ACC), precision (PRE), recall, F1 score and the area under the ROC curve (AUC) [44]. The corresponding calculation formulas are shown as below

$$MAE = \frac{\sum_{i,j=1}^{I,J} |P(r_{ij}=1)-r_{ij}|}{IJ}, RMSE = \sqrt{\frac{\sum_{i,j=1}^{I,J} |P(r_{ij}=1)-r_{ij}|^2}{IJ}}, \tag{30}$$

$$PRE = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}, \tag{31}$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}, F1 = \frac{2 \cdot PRE \cdot Recall}{PRE + Recall}, \tag{32}$$

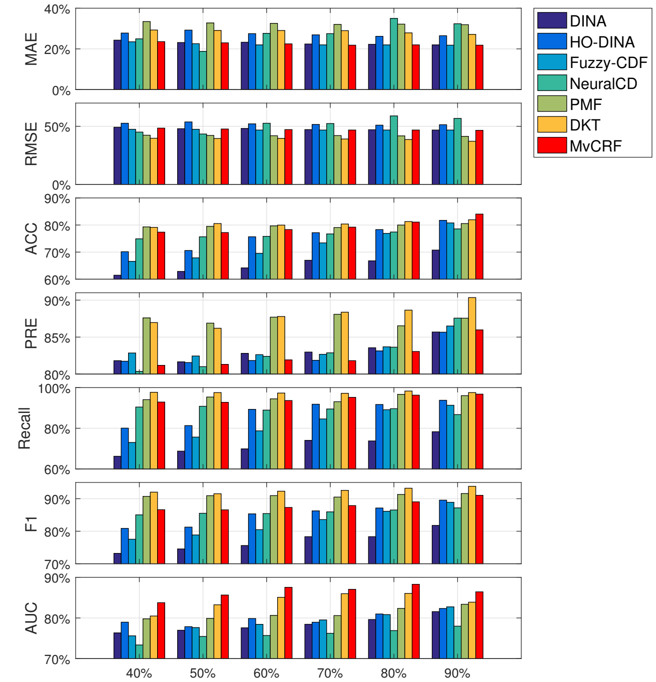
where TP, TN, FP and FN denote the numbers of true positives, true negatives, false positives and false negatives, respectively.

4.2.2 Results

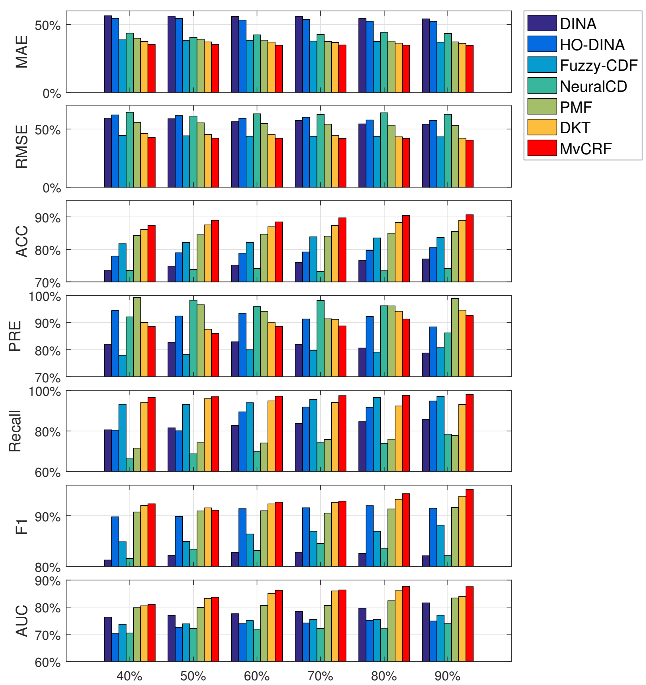
Prediction performance. To evaluate the fitting and prediction performance of the model at different training ratios, we randomly select 40% $\sim$ 90% of each row from score matrix $\boldsymbol{R}$ as the training datasets and the rest as the test datasets, respectively.

2. HNU_SYS1&2 datasets are available in https://drive.google.com/drive/folders/1Opdyqit0SIq8pogmPnwsbibVJbwLET7K?usp=share_link.

3. https://www.icourse163.org/course/HNU-1003363020



(a)



(b)

Fig. 7. MAE, RMSE, ACC, PRE, recall, F1 score and AUC results on (a)HNU_SYS1 and (b) HNU_SYS2 datasets with different training ratios.

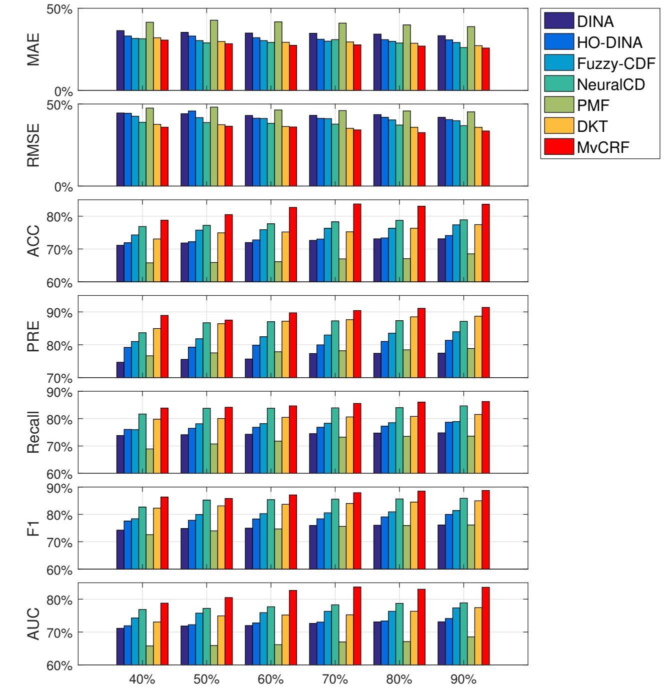


Fig. 8. MAE, RMSE, ACC, PRE, recall, F1 score and AUC results on ASSISTments dataset with different training ratios.

The MAE and RMSE results in Figs. 7 and 8 illustrate how well each model fits to HNU_SYS1&2 and ASSISTments datasets, respectively. Lower MAE and RMSE indicate better fitting capacity. In Fig. 7(a), Fuzzy-CDF and MvCRF outperform other static CDMs in MAE and RMSE. When the ratio of the training set increases from 40% to 90%, MvCRF changes within a range of 1.74% according to MAE while DINA, HO-DINA, Fuzzy-CDF and NeuralCD varies within the ranges of 2.33%, 3.05%, 1.67% and 16.22%, respectively. In terms of RMSE, MvCRF fluctuates within a range of 1.90% compared with DINA, HO-DINA, Fuzzy-CDF and Neural changing within the ranges of 2.41%, 2.79%, 0.63% and 15.60%, respectively. As can be seen, DKT and PMF outperform MvCRF in terms of RMSE, because DKT and PMF use loss functions to minimize the cross entropy or the sum-of-squared-errors between the prediction result and the corresponding true value while MvCRF aims at drawing samples of each target parameter from a probability distribution of interest. In Figs. 7(b) and 8, MvCRF performs the best in both metrics. Compared with other competing methods, MvCRF decreases the MAE and RMSE by 1.86%∼20.54% and 2.53%∼17.85% in HNU_SYS1 dataset, and by 1.58%∼13.10% and 1.55%∼11.70% in ASSISTments dataset, respectively. As can be seen, MvCRF achieves the best fitting performance in the two datasets.

In addition, Figs. 7 and 8 also present the ACC, PRE, recall, F1 score and AUC results of MvCRF and the competing models on HNU_SYS1 and HNU_SYS2 datasets with different training ratios. As can be seen in Fig. 7(a), MvCRF achieves the best results in all cases in terms of AUC and outperforms the other static CDMs in ACC, recall and F1 score. When the ratio of training set declines from 90% to 40%, the decreases in prediction accuracy of DINA, HO-DINA, Fuzzy-CDF, NeuralCD, PMF, DKT and MvCRF are 9.27%, 11.58%, 14.19%, 3.66%, 1.23%, 2.81% and 6.67%, respectively. The corresponding decreases in recall are 12.08%, 13.71%, 18.30%, -3.71%, 1.97%, -0.13% and 3.90%, respectively, those in F1 score are 8.60%, 8.68%, 11.36%, 2.15%, 0.88%, 1.80% and 4.43%, and those in AUC are 5.23%, 3.38%, 7.13%, 4.62%, 3.58%, 3.40% and 2.67%, respectively. For the prediction precision, the results of five static CDMs are close to each other and the differences among them are no more than 2.5%. In Fig. 7(b), MvCRF performs the best in

TABLE 5
The quantitative results (%) of ablation study

Method	MvCRF-I	MvCRF-II	MvCRF-III	MvCRF
MAE	26.39±0.31	25.48±0.46	24.65±0.40	22.44±0.72
RMSE	51.42±0.42	49.21±0.58	48.44±0.62	47.24±0.71
ACC	76.65±4.16	77.95±4.01	77.21±2.93	79.54±2.62
PRE	82.28±1.84	82.68±1.66	82.39±1.73	82.54±1.81
Recall	89.00±4.38	91.35±3.25	89.87±3.57	94.60±1.74
F1	85.49±2.95	86.79±2.23	85.95±2.43	88.08±1.72
AUC	78.68±1.51	83.38±1.75	83.19±2.13	86.45±1.60

*Blue font denotes the best performance.

TABLE 6
Runtime comparison of MvCRF and the competing methods

Method	Training time(s)	Test time(ms)
DINA	0.3864	0.1376
HO-DINA	0.4234	0.1462
Fuzzy-CDF	0.4298	0.1876
NeuralCD	0.1245	0.0988
PMF	0.0189	0.0683
DKT	0.0835	0.0737
MvCRF	0.3574	0.1462

*Blue font denotes the best performance.

almost all cases in terms of ACC, recall, F1 score and AUC. The average ACC, recall, F1 score and AUC results of HO-DINA, Fuzzy-CDF, NeuralCD, PMF, DKT and MvCRF in six cases are (75.55%, 83.12%, 82.28%, 78.42%), (79.19%, 90.09%, 90.99%, 73.45%), (82.85%, 94.84%, 86.37%, 75.07%), (73.73%, 74.27%, 83.06%, 72.09%), (84.69%, 86.59%, 91.01%, 81.10%), (87.54%, 94.02%, 92.58%, 84.12%) and (89.27%, 97.23%, 93.07%, 85.38%), respectively. The improvements of MvCRF over other methods are 1.74% ∼15.54%, 2.39%∼14.11%, 0.49%∼10.81% and 1.27%∼13.29% in terms of ACC, recall, F1 score and AUC, respectively. In terms of ASSISTments dataset in Fig. 8, MvCRF increases the ACC, PRE, recall, F1 and AUC by 4.11%∼15.32%, 2.62%∼13.51%, 1.44%∼13.12%, 2.34%∼12.58% and 4.11%∼15.32%, respectively, compared with other methods.

Therefore, the quantitative results not only indicate the excellent fitting capability and low sensitivity of MvCRF to training ratios, but also demonstrate the competitivity of MvCRF over competing methods in ACC, recall, F1 score and AUC while keeping a good performance in PRE.

Ablation study. As can be seen in Table 5, MvCRF-I and MvCRF-II introduce only the effort trait η_i without the skill weakness parameter l_{ikj} by applying the combined (c) and decaying (d) compensation mechanism, respectively. They calculate the item mastery level δ_{ij} as Eq. (6). MvCRF-III considers only l_{ikj} without η_i. Table 5 shows the quantitative results of ablation experiments of MvCRF in HNU_SYS1 dataset with training ratio ranging from 40% to 90%. The figures are displayed in the form of 'average ± standard deviation'. Compared with MvCRF-I, MvCRF-II increases by about -0.91%, -2.21%, 1.30%, 0.40%, 2.35%, 1.29% and 4.70% in MAE, RMSE, ACC, PRE, recall, F1 score and AUC, respectively. It proves the advantages of decaying compensation over combined compensation. Furhtermore, the improvement of MvCRF over MvCRF-II and MvCRF-III indicates the effectiveness of skill weakness parameter and effort trait, respectively.

Runtime comparison. Our experiments on the runtime comparison of different methods are conducted on a PC with an Intel Core i9-10900K CPU and 32GB RAM. Table 6 lists the average training time and test time per sample of MvCRF and the competing methods on HNU_SYS1 dataset. As can be seen, PMF is the fastest while Fuzzy-CDF is the slowest. MvCRF ranks fourth in training time and sixth in test time.

Case analysis. Fig. 9 presents an example of four examinees' diagnostic results via MvCRF in HNU_SYS1 dataset. The upper five subfigures visualize the skill weakness matrices $(l_{ikj})^{6\times 90}$ with i =15, 165, 278 and 325, and the

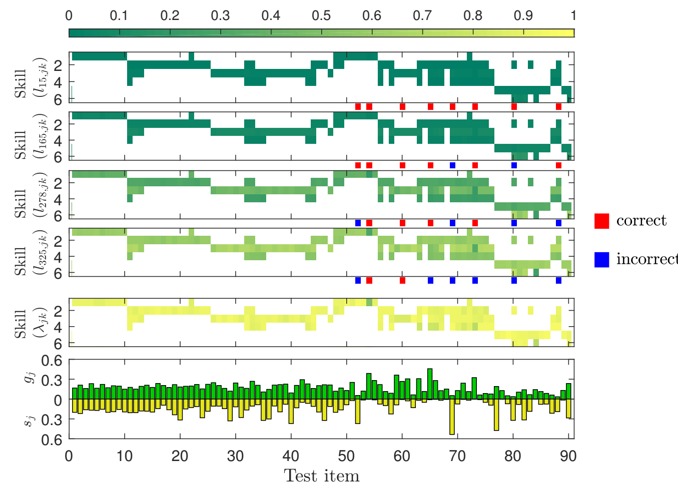


Fig. 9. Diagnosis example of 4 examinees in HNU_SYS1.

skill importance level matrix $(\lambda_{kj})^{6\times 90}$, respectively. The darker color means smaller value. The last subfigure shows the slip and guessing parameters of each item. Eight item scores of each examinee are shown below the corresponding skill weakness matrices with red and blue denoting correct and incorrect answers, respectively. On the one side, the examinees are more likely to guess the correct answers of an item but not to miss it carelessly when the skills covered by the item (e.g. j = 54, 60, 65 and 73) are less important or necessary. Although such items are easy to score, the examinees (e.g. i = 325) tend to give the wrong answers as his skill proficiency is far from satisfying the requirement of the items (e.g. j = 65 and 73). On the other side, it is harder for the examinees to guess the correct answer and easier to lose score by carelessness due to the higher importance of the skills involved in the item (e.g. j = 52, 69, 80 and 88). However, the examinees with higher skill proficiency (e.g. i = 15) are more likely to work out such difficult items.

4.3 Discussion

According to the above simulation study and real-data application, we find that MvCRF can produce interpretable and accurate cognitive diagnosis results. When applied on MOOC, MvCRF can diagnose and evaluate each student's knowledge level in the course, and thus help students to identify their weaknesses. In addition, MvCRF can recommend the test items with appropriate difficulty to each student by predicting the student's score on each item.

However, there are still some aspects to be improved. First, we may collect both online and offline learning activity logs and effective learning time information (e.g. the current total time and frequency of watching tutoring video) to further optimize the effort trait. Second, diverse expressions of test items (e.g. objective and subjective questions) can be considered for cognitive diagnosis. Third, the historical information based models usually suffer from the ”cold start” problem [45], [46]. Fourth, we will manage to improve the efficiency of MvCRF due to its high computational complexity.

5 Conclusion

In this paper, we proposed MvCRF for cognitive diagnosis. Firstly, we captured the prior information to calculate examinee's effort trait and integrated the ability and effort traits by compensation mechanism to estimate each examinee's knowledge level. Combining the examinee's knowledge level with the importance of each skill on a specific test item, we then introduced the skill weakness parameter to predict the item score. Compared with the existing CDMs, MvCRF takes more consideration of the skill learning and application processes of the examinees. Furthermore, extensive experimental results verified the superiority of MvCRF in prediction performance, interpretability and robustness over the baseline CDMs. We hope this work could lead to further research in the future, such as how to efficiently apply effort trait into knowledge tracing.

Acknowledgment

This work was funded in part by the National Key Research and Development Program of China (2019YFE0105300), by China Association of Higher Education (21SZYB15) and by Key Scientific Research Project of the Education Department of Hunan Province (22A0021).

References

[1] Zhiting Zhu and Fei Wei, “Educational informatization 2.0: Starting on a journey of intelligence education guided by smart education,” *Journal of Education*, vol. 39, no. 09, pp. 5–16, 2018.

[2] Marie Bienkowski, Mingyu Feng, and Barbara Means, “Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.” *Office of Educational Technology*, 2012.

[3] Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su, and Guoping Hu, “Ekt: Exercise-aware knowledge tracing for student performance prediction,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 1, pp. 100–115, 2021.

[4] José A Ruipérez-Valiente, Daniel Jaramillo-Morillo, Srećko Joksimović, Vitomir Kovanović, Pedro J Muñoz-Merino, and Dragan Gašević, “Data-driven detection and characterization of communities of accounts collaborating in moocs,” *Future Generation Computer Systems*, vol. 125, pp. 590–603, 2021.

[5] Yupei Zhang, Rui An, Shuhui Liu, Jiaqi Cui, and Xuequn Shang, “Predicting and understanding student learning performance using multi-source sparse attention convolutional neural networks,” *IEEE Transactions on Big Data*, pp. 1–1, 2021.

[6] Guiyun Feng, Muwei Fan, and Yu Chen, “Analysis and prediction of students' academic performance based on educational data mining,” *IEEE Access*, vol. 10, pp. 19 558–19 571, 2022.

[7] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec, “Engaging with massive online courses,” in *Proceedings of the 23rd International Conference on World Wide Web*. New York, NY, USA: Association for Computing Machinery, 2014, p. 687–698.

[8] Hugh Burns, Carol A. Luckhardt, James W. Parlett, and Carol L. Redfield, *Intelligent tutoring systems: Evolutions in design*. Psychology Press, 2014.

[9] Xitao Fan, “Item response theory and classical test theory: An empirical comparison of their item/person statistics,” *Educational and psychological measurement*, vol. 58, no. 3, pp. 357–381, 1998.

[10] Jimmy De La Torre, “Dina model and parameter estimation: A didactic,” *Journal of educational and behavioral statistics*, vol. 34, no. 1, pp. 115–130, 2009.

[11] Sarah McConnell Hartz, *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*. University of Illinois at Urbana-Champaign, 2002.

[12] Jacqueline P Leighton, Mark J Gierl, and Stephen M Hunka, “The attribute hierarchy method for cognitive assessment: A variation on tatsuoka's rule-space approach,” *Journal of educational measurement*, vol. 41, no. 3, pp. 205–237, 2004.

[13] Jimmy De La Torre and Jeffrey A Douglas, “Higher-order latent trait models for cognitive diagnosis,” *Psychometrika*, vol. 69, no. 3, pp. 333–353, 2004.

[14] Dongbo Tu, Yan Cai, Haiqi Dai, and Shuliang Ding, “A polytomous cognitive diagnosis model: P-dina model,” *Acta Psychologica Sinica*, vol. 42, no. 10, pp. 1011–1020, 2010.

[15] Xuliang Gao, Daxun Wang, Fang Wang, Yan Cai, and Dongbo Tu, “Development of a generalized cognitive diagnosis model for polytomous responses based on partial credit model,” *Acta Psychologica Sinica*, vol. 51, no. 12, pp. 1386–1397, 2019.

[16] Yan Cai, Yang Zhao, Shuchang Liu, Shufang Zhang, and Dongbo Tu, “An extended polytomous cognitive diagnosis model,” *Journal of Psychological Science*, vol. 40, no. 06, pp. 1491–1497, 2017.

[17] Dongbo Tu, Yan Cai, and Haiqi Dai, “A polytomous extension of higher-order dina momdel,” *Journal of Psychological Science*, vol. 36, no. 04, pp. 984–988, 2013.

[18] Qi Liu, Runze Wu, Enhong Chen, Guandong Xu, Yu Su, Zhigang Chen, and Guoping Hu, “Fuzzy cognitive diagnosis for modelling examinee performance,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 9, no. 4, pp. 1–26, 2018.

[19] Youxi Li, Yimin Wen, Xinhe Yi, and Zhi Xu, “Revised model of fuzzy cognitive diagnosis framework,” *Journal of Data Acquisition and Processing*, vol. 32, no. 05, pp. 958–968, 2017.

[20] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang, “Neural cognitive diagnosis for intelligent education systems,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 6153–6161.

[21] Yan Cheng, Meng Li, Haomai Chen, Yingying Cai, Huan Sun, Gang Wu, Zhuang Cai, and Guanghe Zhang, “Neural cognitive modeling based on the importance of knowledge point for student performance prediction,” in *2021 16th International Conference on Computer Science Education (ICCSE)*, 2021, pp. 495–499.

[22] Lina Gao, Zhongying Zhao, Chao Li, Jianli Zhao, and Qingtian Zeng, “Deep cognitive diagnosis model for predicting students' performance,” *Future Generation Computer Systems*, vol. 126, pp. 252–262, 2022.

[23] Ghodai Abdelrahman, Qing Wang, and Bernardo Nunes, “Knowledge tracing: A survey,” *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–37, 2023.

[24] Xiangyu Song, Jianxin Li, Taotao Cai, Shuiqiao Yang, Tingting Yang, and Chengfei Liu, “A survey on deep learning based knowledge tracing,” *Knowledge-Based Systems*, vol. 258, p. 110036, 2022.

[25] Qi Liu, Shuanghong Shen, Zhenya Huang, Enhong Chen, and Yonghe Zheng, “A survey of knowledge tracing,” *arXiv preprint arXiv:2105.15106*, 2021.

[26] Albert T Corbett and John R Anderson, “Knowledge tracing: Modeling the acquisition of procedural knowledge,” *User modeling and user-adapted interaction*, vol. 4, pp. 253–278, 1994.

[27] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein, “Deep knowledge tracing,” *Advances in neural information processing systems*, vol. 28, 2015.

[28] Zachary Pardos, Neil Heffernan, Carolina Ruiz, and Joseph Beck, “The composition effect: Conjuntive or compensatory? an analysis of multi-skill math questions in its,” in *Educational Data Mining 2008*, 2008.

[29] Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung, “Dynamic key-value memory networks for knowledge tracing,” in *Proceedings of the 26th international conference on World Wide Web*, 2017, pp. 765–774.

[30] Aritra Ghosh, Neil Heffernan, and Andrew S Lan, “Context-aware attentive knowledge tracing,” in *Proceedings of the 26th ACM*

SIGKDD international conference on knowledge discovery and data mining, 2020, pp. 2330–2339.

[31] Hiromi Nakagawa, Yusuke Iwasawa, and Yutaka Matsuo, "Graph-based knowledge tracing: modeling student proficiency using graph neural network," in *IEEE/WIC/ACM International Conference on Web Intelligence*, 2019, pp. 156–163.

[32] Xiangyu Song, Jianxin Li, Yifu Tang, Taige Zhao, Yunliang Chen, and Ziyu Guan, "Jkt: A joint graph convolutional network based deep knowledge tracing," *Information Sciences*, vol. 580, pp. 510–523, 2021.

[33] Yanwen Song, "Combination forecasing of short-term traffic flow by robust statistics," Master's thesis, Tianjing University, 2012.

[34] Norman H Anderson, *Foundations of information integration theory*. New York: Academic Press, 1981.

[35] ——, *Methods of information integration theory*. New York: Academic Press, 1982.

[36] Colleen F Surber, "Measuring the importance of information in judgment: Individual differences in weighting ability and effort," *Organizational Behavior and Human Decision Processes*, vol. 35, no. 2, pp. 156–178, 1985.

[37] Walter R. Gilks, *Markov Chain Monte Carlo*. John Wiley & Sons, Ltd, 2005.

[38] Mehrnoosh Vahdat, Luca Oneto, Davide Anguita, Mathias Funk, and Matthias Rauterberg, "A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator," in *Design for Teaching and Learning in a Networked World: 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Toledo, Spain, September 15-18, 2015, Proceedings 10*. Springer, 2015, pp. 352–366.

[39] Neil T Heffernan and Cristina Lindquist Heffernan, "The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching," *International Journal of Artificial Intelligence in Education*, vol. 24, pp. 470–497, 2014.

[40] Youngduck Choi, Youngnam Lee, Dongmin Shin, Junghyun Cho, Seoyon Park, Seewoo Lee, Jineon Baek, Chan Bae, Byungsoo Kim, and Jaewe Heo, "Ednet: A large-scale hierarchical dataset in education," in *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21*. Springer, 2020, pp. 69–73.

[41] Lawrence T DeCarlo, "On the analysis of fraction subtraction data: The dina model, classification, latent class sizes, and the q-matrix," *Applied Psychological Measurement*, vol. 35, no. 1, pp. 8–26, 2011.

[42] Runze Wu, Qi Liu, Yuping Liu, Enhong Chen, Yu Su, Zhigang Chen, and Guoping Hu, "Cognitive modelling for predicting examinee performance," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[43] Andriy Mnih and Russ R Salakhutdinov, "Probabilistic matrix factorization," *Advances in neural information processing systems*, vol. 20, 2007.

[44] Andrew P Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[45] Qi Liu, Enhong Chen, Hui Xiong, Yong Ge, Zhongmou Li, and Xiang Wu, "A cocktail approach for travel package recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 2, pp. 278–293, 2012.

[46] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He, "A survey on knowledge graph-based recommender systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3549–3568, 2022.

Lianhong Wang received the B.S., M.S. and Ph.D. degrees from Hunan University, Changsha, China, in 1993, 2002, and 2009, respectively.

She is currently an Associate Professor with the College of Electrical and Information Engineering, Hunan University, Changsha, China. She was a visiting scholar at Brandeis University from 2011 to 2012. She has published more than 30 papers in journals and conferences. Her research interests include signal/image processing, data mining, and artificial intelligence.

Xiaoyao Li received the B.S. degree in mathematics and applied mathematics from China University of Petroleum, Beijing, China, in 2012, and the Ph.D degree in control science and engineering from Hunan University, Changsha, China, in 2022.

She is currently an Assistant Professor with the School of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha, China. She was a Research Assistant with the Department of Computer and Information Science at University of Macau in 2017. Her research interests include signal/image processing and data mining.

Zhihui Luo received the B.S. degree in Nanjing Agricultural University, Nanjing, China, in 2020.

He is currently pursuing the M.S. degree with Electronic Science and Technology, College of Electrical and Information Engineering, Hunan University, Changsha, China. His research interests include educational data mining and machine learning.

Zinan Hu is currently pursuing the B.S. degree with Electronic Information Engineering, College of Electrical and Information Engineering, Hunan University, Changsha, China. His research interests include educational data mining and machine learning.

Qing Yan received the B.S. degree in financial engineering from Huazhong University of Science and Technology, Wuhan, China, in 2019, and the M.S. degree in business analytics from University of Maryland College Park, Maryland, USA, in 2020.

She currently works in the Urtrust Insurance Co., Ltd., in Guangzhou, China. She was a Teaching Assistant with Robert H. Smith School of Business at University of Maryland College Park in 2020. Her research interests include data mining and machine learning.