



Predicting examinee performance based on a fuzzy cloud cognitive diagnosis framework in e-learning environment

Hua Ma^{1,2} · Zhuoxuan Huang¹ · Haibin Zhu³ · WenSheng Tang^{1,2} · Hongyu Zhang⁴ · Keqin Li⁵

Accepted: 19 March 2023 / Published online: 17 April 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The score profiles could be used to measure learners' skills proficiency via cognitive diagnosis models (CDMs) for predicting their performance in the future examination. The prediction results could provide important decision-making supports for personalized e-learning instruction. However, facing the possible complexity of skills, the uncertainty of learners' skill proficiency and the large-scale volume of score profiles, the existing CDMs have limitations in the measurement mechanisms and diagnostic efficiency. In this paper, we proposed an approach based on a fuzzy cloud cognitive diagnosis framework (FC-CDF) to predicting examinees' performance in e-learning environment. In this approach, the normal cloud models (NCMs) are utilized innovatively to measure the expectation, degree of variation and variation frequency of learners' skill proficiency, and each NCM is transformed into an interval fuzzy number to characterize the uncertainty of the skill proficiency for every learner. Combining the educational psychology hypothesis with the parameter estimation method, we could obtain the learners' skill proficiency level and the slip and guess factors relevant to each test item, based on which the learners' scores could be predicted in a future test. Finally, the experiments demonstrate that the proposed approach provides good accuracy and significantly reduces execution time for predicting examinee performance, compared with the existing methods.

Keywords Cloud model · Cognitive diagnosis · e-learning environment · Predicting examinee performance

1 Introduction

1.1 Motivation

Recently, e-learning has become popular worldwide due to the COVID-19 pandemic (Maatuk et al. 2022), and a massive amount of educational data has been collected in

e-learning platforms. Assessing the learners' cognitive abilities based on such data has potentials to help instructors and learners to obtain a comprehensive view of a learner's learning progress (Mai et al. 2021). The score profiles may be analyzed by cognitive diagnosis models (CDMs) to evaluate the learners' strengths and weaknesses by identifying the presence or absence of multiple fine-grained skills required for solving problems on a test (Torre

✉ Zhuoxuan Huang
lilhzx@hunnu.edu.cn

Hua Ma
huama@hunnu.edu.cn

Haibin Zhu
haibinz@nipissingu.ca

WenSheng Tang
tangws@hunnu.edu.cn

Hongyu Zhang
hyzhang@csu.edu.cn

Keqin Li
lik@newpaltz.edu

¹ Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, China

² College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China

³ Department of Computer Science and Mathematics, Nipissing University, North Bay, ON P1B 8L7, Canada

⁴ School of Business, Central South University, Changsha 410083, China

⁵ Department of Computer Science, State University of New York, New Paltz, NY 12561, USA

2008). The measurements of learners' skill proficiencies are helpful in predicting examinee performance in a future test (Ma et al. 2020; Song et al. 2022; Yao et al. 2019). This insight enables the possibility of personalized recommendation (Wan and Niu 2020; Yang et al. 2023; Yun et al. 2022), school precaution (Bernacki et al. 2020; Niyogisubizo et al. 2022), examination paper generation (Wu et al. 2020), collaborative learning team formation (Ma et al. 2023, 2022) and learning path recommendation (Jiang et al. 2022; Liu and Li 2020; Liu et al. 2019b).

Learners' performance is usually predicted by cognitive diagnosis in the existing research on educational psychology (DiBello et al. 2006; Sessoms and Henson 2018). Researchers and practitioners from both educational psychology and data mining fields have devoted many efforts to CDMs due to their outstanding advantages for analyzing the learners' cognitive status. Some classical CDMs, e.g., the item response theory (IRT) (Janssen et al. 2000) and the Deterministic Input, Noisy And-gate (DINA) model (Torre 2008), have been applied widely. Among them, IRT models a learner as a single proficiency variable, i.e., latent trait, and the performance of a learner in a test item could be predicted by combing it with some items attributes, e.g., degrees of difficulty and degree of distinction, and the DINA model introduces a Q -matrix (Tatsuoka and Kikumi 1983) to measure learners' proficiencies from the skill level. The Q -matrix is a matrix to describe the relationship between test items (i.e., questions) and the skills that are usually relevant to the specified knowledge point. Aiming at the shortcomings of classical CDMs, some improved models have been developed, such as HO-DINA (Torre and Douglas 2004), P-DINA (Tu et al. 2010), FuzzyCDF (Liu et al. 2018), R-FuzzyCDF (Li et al. 2017).

However, the above CDMs still have much room to improve the measurement mechanisms and diagnosis efficiency: (1) The DINA model only uses a binary value to measure one learner's skill proficiency, while FuzzyCDF and R-FuzzyCDF can improve the accuracy of predicting examinee performance by introducing a fuzzy number-based measurement method. However, an integer or a fuzzy real number fails to address the uncertainty of learner's skill proficiency. In reality, the real performance of learners might be unstable due to the complexities of one skill itself and the diversity of application scenarios. (2) Although recent FuzzyCDF and R-FuzzyCDF significantly improved the accuracy of performance prediction compared with IRT and DINA model and worked well in a small-scale data environment, their diagnosis efficiency is dramatically reduced due to the slow convergence speed and numerous estimated parameters. Particularly, more and more education programs have been moved to online channels due to the pandemic, and the massive amount of educational data could be automatically collected and used

for the investigation and prediction of the learners' learning performance on e-learning platforms. Thus, two key issues are still open for studying CDMs in an e-learning environment: (1) How could we measure a learner's skills proficiency comprehensively and accurately according to the relationship among learners, test items, and skills? (2) How could we improve the execution efficiency of cognitive diagnosis without a loss of the diagnosis accuracy for supporting large-scale data processing scenarios involving a great deal of learners, items and skills?

To address the above issues, we innovatively introduced the normal cloud models (NCMs) (Peng et al. 2018; Zhang et al. 2007) to model the learners' skills proficiency. The NCM (Wang et al. 2014; Yang et al. 2018) is an effective tool to realize the conversion between qualitative concepts and quantitative expressions and can be used to measure the fuzziness, the randomness and the relevance of uncertain concepts. At present, it has been applied to many fields, such as intelligent control and information processing (Wang et al. 2014; Yang et al. 2018), system evaluation (Liu et al. 2019a; Ma et al. 2021), recommendation system (Peng et al. 2018; Zhang et al. 2007) and decision-making (Gong et al. 2021; Wang et al. 2021). One NCM consisting of three numerical characteristics (i.e., expectation, entropy and hyper-entropy) is obtained via a reverse cloud generator (RCG) (Li et al. 1995).

Based on the NCMs, an approach via fuzzy cloud-based cognitive diagnosis framework (FC-CDF) is put forward to predict learners' performance for both subjective and objective test items in an e-learning environment. In this approach, a cognitive cloud is defined for measuring the expectation, degree of variation and variation frequency of learners' skill proficiency, and each NCM is transformed into an interval fuzzy number to characterize the fuzziness and uncertainty of the skill proficiency for every learner. Similar to FuzzyCDF (Liu et al. 2018), the learners' cognitive status on items is calculated by combining the educational hypotheses and fuzzy sets theory. The learners' performance could be predicted in a future examination by considering the slip and guess factors of each test item. Finally, the experiments based on three real-world datasets, a simulation dataset and three extended datasets are analyzed by comparing the other approaches, and the results demonstrate that the proposed approach provides good accuracy for predicting examinee performance and significantly reduces the execution time of cognitive diagnosis.

1.2 Our contributions

The main contributions of this paper are as follows:

- (1) To measure the learners' real skills proficiency more objectively and comprehensively, the cloud model

theory is introduced innovatively into FC-CDF. Different from the existing approaches that measure the learners' skill proficiency with an integer or a fuzzy real number, FC-CDF establishes an interpretable fuzzy measurement mechanism to characterize the uncertainty of learners' skills proficiency via two-parameter interval numbers transformed by cognitive clouds. Based on this mechanism, this proposed approach could predict the learners' performance accurately in future examinations.

- (2) To provide an efficient cognitive diagnosis for supporting the learners' autonomic learning online in their fragmented time, FC-CDF employs a reverse cloud generation algorithm to calculate the learners' skill proficiency. Based on it, the process of parameters estimation is simplified and the computation speed of cognitive diagnosis is accelerated significantly, in comparison to the existing approaches. The proposed approach could be applied to a large-scale data environment involving many learners, test items and skills.

The remainder of the paper is organized as follows: Sect. 2 reviews the related work. Section 3 gives the problem statement. Section 4 introduces the preliminaries. Section 5 proposes the framework of fuzzy cloud cognitive diagnosis. Section 6 puts forward the examinee performance prediction via FC-CDF. Section 7 analyzes the experimental results. Finally, Sect. 8 concludes the paper.

2 Related work

2.1 Cognitive diagnosis model

In psychological and educational measurement, the cognitive diagnosis is a diagnostic assessment of an individual's cognitive process, skill or knowledge structure (Nichols and Joldersma 2008). Researchers proposed the psychometric models with cognitive diagnosis function (i.e., CDMs) based on the cognitive psychology and psychometrics theory. Generally, CDMs could be classified into two categories, i.e., the continuous CDMs and the discrete CDMs. The representative continuous CDM is IRT (Janssen et al. 2000), and the DINA model (Torre 2008) is a classical discrete CDM.

2.1.1 Item response theory

IRT assumes that each learner has a "unique latent trait." Thus, it measures one learner via a one-dimensional continuous variable to predict her/his performance in test items with different discrimination and difficulty features. The 2-

parameter logistic model (2PL) (Janssen et al. 2000) is defined by

$$P(X_{ij} = 1|\theta) = \frac{1}{1 + \exp(D \times a(\theta - b))} \tag{1}$$

where $P(X_{ij} = 1|\theta)$ denotes the probability of learner #i with the latent trait #θ answering item #j correctly, a and b are the discrimination and difficulty of test item, respectively, θ represents the latent trait of learner, and D is an empirical scaling constant in logistic cognitive models, which is set as -1.7 (Gregory and Camilli 1994). IRT is built on a rigorous and complex mathematical model, and it has been applied to examinations and traditional intelligence tests. However, the unidimensionality assumption used by IRT is too coarse-grained to accurately measure learners' proficiency, since learners' proficiency is related to different skills. To address this issue, the DINA model is proposed to depict learner's proficiency from the skill level (Torre 2008).

2.1.2 DINA model

Different from IRT, the DINA model introduces a Q -matrix (Tatsuoka and Kikumi 1983), which indicates whether the skills are associated with items, and diagnoses the learner's cognitive status with a proficiency vector about multi-dimensional skills. Each dimension of this proficiency vector denotes whether a learner masters one specific skill. In this case, the mastery of learner #i on item #j is calculated as follows:

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \tag{2}$$

where α_{ik} is a binary value indicating whether the learner #i master the skill #k, 1 denotes that #i masters #k; otherwise, 0 is suggested; q_{jk} denotes whether the skill #k is required to correctly answer the item #j (1 means required and 0 otherwise); and $\eta_{ij} = 0$ means that the learner #i cannot master the item #j; otherwise, $\eta_{ij} = 1$.

Though the DINA model strives to overcome some shortcomings of IRT, there are still some limitations in this model, such as dichotomous description mechanism (Tu et al. 2010); (Cai et al. 2017), underexplored cognitive diagnosis on subjective items (Liu et al. 2018) and high computational cost (Wang et al. 2018). Recently, many efforts have been devoted to improving DINA models for understanding the learner's learning status comprehensively and providing effectively personalized learning guidance for learners.

2.1.3 Improved DINA model

The improved DINA models could be divided into several categories as follows:

- (1) Improve the dichotomous 0–1 scoring mechanism
The DINA model only supports the dichotomous 0–1 scoring mechanism, in which a learner gets either a full score or zero in a test item. Tu et al. (2010) developed a P-DINA model to provide a polychotomous scoring mechanism on a scale of zero to a full score. However, the score values easily focus on two extremes (i.e., 0 or a full score) when the P-DINA model is used in data analysis. Thus, Cai et al. (2017) presented an rP-DINA (revised P-DINA) model by reconstructing learner's ideal score, and this model support all kinds of ideal scores ranging from 0 to full score.
- (2) Adopt the new diagnosis mechanism for subjective test items

Considering that the traditional DINA model is unable to support the cognitive diagnosis for the subjective items, (Liu et al. 2018) proposed a fuzzy cognitive diagnosis framework (FuzzyCDF). FuzzyCDF measures the skill proficiency of a learner with the degree of membership (i.e., a real number in the range [0, 1]) and adopts the cognitive answer mode via the fuzzy intersection and the fuzzy union operations. In FuzzyCDF, the mastery of a learner on objective items is of conjunctive type and the mastery of a learner on subjective items is of compensatory type. For an objective item, a learner may answer it correctly if he/she possesses all the skills required for it. As for a subjective item, the more skills a learner possesses, the higher the score that the learner gets from it. Based on FuzzyCDF, Li et al. (2017) put forward a revised framework called R-FuzzyCDF by taking into consideration the influence of the importance of every skill. Although R-FuzzyCDF might provide a better accuracy of performance prediction than FuzzyCDF, more parameters being introduced makes the process of cognitive diagnosis more time-consuming than FuzzyCDF.

- (3) Extend to High-order CDMs

The high-order CDMs are developed to describe the possible structural relationship among multiple latent traits in CDM. Considering that the proficiency of a learner may be affected by one (or more) high-order latent characteristics, De la Torre and Douglas (2004) presented a high-order latent structure model

and combined it with the DINA model to design a high-order DINA model (HO-DINA) for reducing the number of estimated parameters. Considering that the HO-DINA model only processes data with a 2-order latent structure, Zhan et al. (2019) put forward a multi-order DINA model by combining a multi-order latent structural model and the DINA model for processing 3-order or higher order latent structure data.

- (4) Improve computational efficiency

The traditional CDMs usually work well when the small-scale sample data are involved in. However, in the large-scale e-learning scenario, the calculation of traditional CDMs will be significantly time-consuming due to their slow convergence speed. Aiming at this issue, Wang et al. (2018) proposed an approach by integrating Maximum Entropy DINA and Incremental Maximum Entropy DINA to accelerate the calculation speed of the DINA model.

2.1.4 Neural networks-based CDMs

The great success of neural networks in both computer version and natural language processing domains attracts researchers to apply it to the cognitive diagnosis field (Liu 2021).

Aiming at the problem of poor diagnosis performance of IRT on sparse data, (Cheng et al. 2019) presented a general deep item response theory (DIRT) to enhance the robustness of traditional IRT. In DIRT, the neural networks are used to mine the text information of each test item for estimating the parameters in the traditional IRT formula. Wang et al. (2020, 2022) proposed a general neural cognitive diagnosis model (NeuralCDM) by applying the neural networks into cognitive diagnosis. Based on the past performance of learners, NeuralCDM used the neural networks to enhance the model's learning ability for more accurately modeling learners' proficiency, items' factors (e.g., discrimination of the item and the difficulty of the skill required by the item) and their interactions. Gao et al. (2021) combined the graph neural networks with an attention mechanism to model student-exercise-skill hierarchical relations and improve the accuracy of cognitive diagnosis.

However, there remain some limitations for the neural networks-based CDMs due to the poor interpretability and the overfitting problem when regarding to small sample data.

2.2 Predicting examinee performance

2.2.1 Matrix factorization-based methods

The matrix factorization (MF) is widely used in recommendation systems and can also be used to predict examinee performance. For example, the MF techniques could measure the degree of learners’ performance in a low-dimensional space by mapping each learner and each item into a low-dimensional vector. Tscher et al. (2010) modeled learner by singular value decomposition. Thia-Nghe et al. (2010) leveraged MF model to predict learners’ performance to improve the prediction accuracy. Thia-Nghe et al. (2015) put forward a multi-relational factor factorization approach for modeling learners in the intelligent tutoring system and predicting the learners’ performance. However, the traditional MF-based examinee performance prediction methods usually lack a good interpretability. For example, each dimension of a learners’ latent vector could not correspond to a specific skill. Thus, it is hard to explain the relationship between skills and such a latent vector.

2.2.2 Cognitive diagnosis model-based methods

The cognitive diagnosis models (CDMs) usually use learners’ score profiles as input data to model their mastery of skills. Based on the diagnosed skill proficiencies, learners’ performance on the unfinished items could be predicted. Additionally, the accuracy of predicted results could be used to verify the effectiveness of CDMs. In the DINA model (Torre 2008), the mastery of learners on the item and the exception of slip and guess are incorporated to model learners’ real performance in the item. Specifically, the probability of learner #*i* with skill proficiency α_i answering item #*j* correctly could be modeled as:

$$\begin{aligned}
 P_j(\alpha_i) &= P(X_{ij} = 1 | \alpha_i) = g_j^{1-\eta_{ij}} (1 - s_j)^{\eta_{ij}}, \\
 P_j(\alpha_i) &= P(X_{ij} = 1 | \alpha_i) = g_j^{1-\eta_{ij}} (1 - s_j)^{\eta_{ij}}
 \end{aligned}
 \tag{3}$$

where s_j and g_j denote the slip and guess factors of item #*j*, respectively; η_{ij} is the mastery of learner #*i* on item #*j*; and X_{ij} is the normalized score of learner #*i* on item #*j*.

To predict the learners’ performance on both objective and subjective items, FuzzyCDF (Liu et al. 2018) employs the following equation:

$$P(X_{ij} = 1 | \eta_{ij}, s_j, g_j) = (1 - s_j)\eta_{ij} + g_j(1 - \eta_{ij})
 \tag{4}$$

where $(1-s_j)\eta_{ij}$ denotes the probability that learner #*i* with all necessary skills answers item #*j* correctly; $g_j(1-\eta_{ij})$ is the probability that learner #*i* who does not master all the skills required by item #*j* answers it correctly.

Different from the MF-based methods, the CDM-based methods enhance the interpretability of predicted results by measuring the learners’ cognitive status according to the relationships among skills, items and score profiles. However, the diagnostic efficiency of the existing CDMs is usually time-consuming since they need to spend considerable time in the process of parameter estimation. Moreover, the existing CDMs ignore the uncertainty of learners’ skill proficiency. It might bring some errors to the cognitive diagnosis, and consequently, these errors result in the biases of predicted results.

3 Problem statement

Suppose there is a mobile learning platform for “Java programming” course. After learners have completed a learning task about one knowledge point (i.e., one skill), this platform would provide them with some related exercises (i.e., test items) to check whether they have mastered the skill relevant to this knowledge point. Moreover, this platform also allows learners to spend their fragmented time on practicing or testing online freely and independently. Once a learner completes a test, the platform should recommend the new exercises for this learner’s next round of practice according to his/her latest score profiles. In this case, the platform is required to measure the skill proficiency of learners via cognitive diagnosis quickly and accurately.

Suppose two learners u_1 and u_2 learned three skills $k_1 \sim k_3$, e.g., “Array Programming,” “Switch Programming” and “Loop Programming,” and finished 12 related exercises in three days. These exercises are designed with different difficulties. For example, one exercise requires learners to add codes to complete an algorithm or figure out the execution result of a given code segment. Obviously, to get high scores on these exercises, one learner should grasp the skills’ essence. One learner should not only understand the theoretical basis related to the skills, but also use them flexibly to solve practical problems by programming. An example of cognitive diagnosis is shown in Fig. 1.

In Fig. 1, the score profiles of u_1 and u_2 and the corresponding *Q*-matrix are the input data. The *Q*-matrix given by the experts describes the relationship between exercises and skills. q_{jk} representing the element on the *j*th row and *k*th column of the matrix indicates whether skill #*k* is required to correctly answer exercise #*j*. $q_{jk} = 1$ means skill #*k* is required; otherwise, $q_{jk} = 0$. After the input data are processed by one CDM (e.g., DINA model or FuzzyCDF), the output, i.e., the diagnosis results describing learners’ mastery on every skill, is obtained. In this example, all CDMs are implemented by Python 3.6 on a Linux server with a 2.3 GHz Intel Xeon CPU and 16 GB memory.

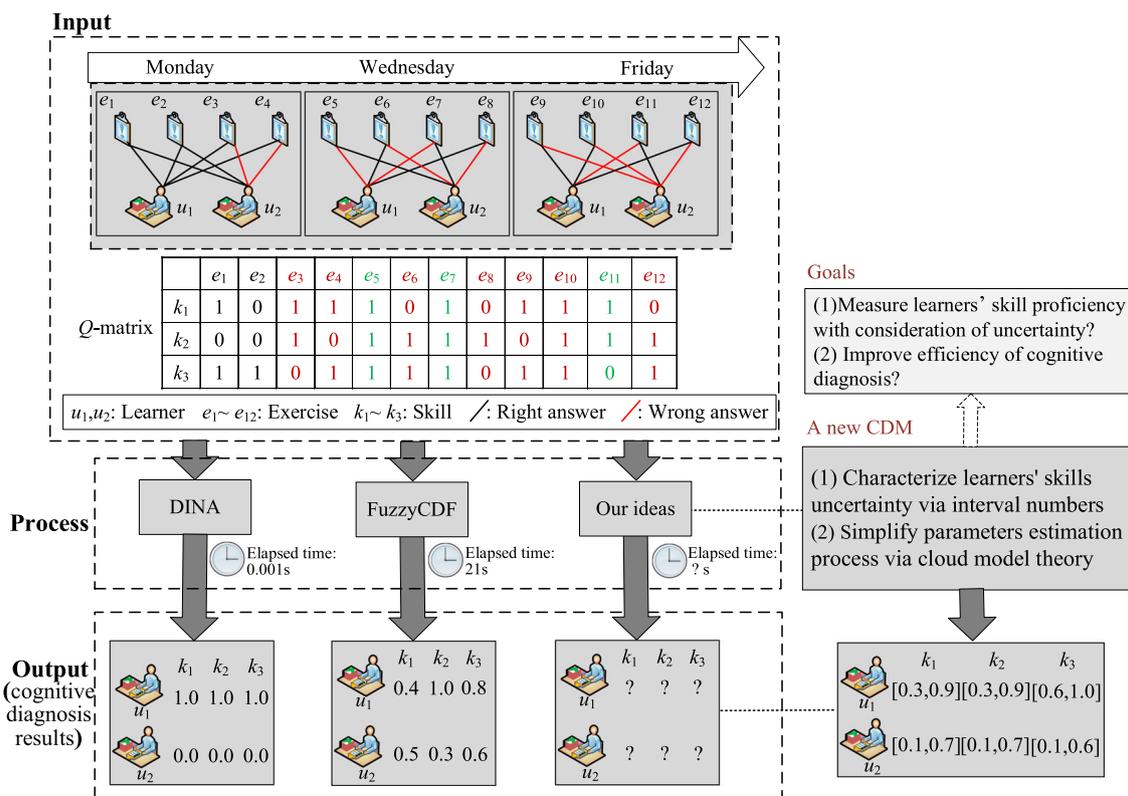


Fig. 1 An example of cognitive diagnosis

From Fig. 1, different CDMs generate different results with different processing efficiency, and the following analysis could be provided:

- 1) For u_2 , DINA measures u_2 's proficiency on k_3 with a fixed integer 0; that is, u_2 did not master k_3 totally. However, the score profiles show that u_2 gave the right responses to some exercises, i.e., e_1, e_2, e_5 and e_7 , related to k_3 .
- 2) For u_1 , DINA and FuzzyCDF measure u_1 's proficiency on k_2 with the number 1, which means that u_1 mastered k_2 totally. Nevertheless, actually u_1 gave the wrong responses to some exercises, i.e., e_5, e_7 and e_{11} , related to k_2 .
- 3) DINA diagnoses two users' skills proficiency on 3 skills with the expectation-maximization algorithm and takes 0.001 s. However, when more skills are involved in, the execution time of DINA would increase dramatically since its time complexity is exponential with the growing number of skills. Our experiments show that DINA takes 72.83 s when a real-world dataset Math2, including 16 skills, is used as the input data. It is foreseen that the DINA's performance is bound to be unbearable when 30 or 40 skills from the "Java programming" course are used.

- 4) Although FuzzyCDF could model the learners' cognitive status more precisely than DINA, it needs more time to estimate the parameters than DINA does. Based on the u_1 and u_2 's score profiles, FuzzyCDF takes 21 s to diagnose their skills proficiency. However, our experiments show that the FuzzyCDF's execution time increases to 9,729 s when the Math2 dataset, including 16 skills, is used as the input data. Obviously, it is unacceptable.

According to the above analysis, we could find two key issues are still underexplored: (1) The existing CDMs, e.g., DINA and FuzzyCDF, are not able to measure the uncertainty of skills proficiency for learners objectively. It should be helpful to predict learners' performance accurately if a reasonable mechanism could be used to measure the characteristic of fluctuation. (2) The execution efficiency of the existing CDMs needs to be significantly improved, especially when the input data cover many skills. It is necessary to explore an efficient and concise CDM to diagnose the learners' skill proficiency with high diagnosis accuracy.

In summary, a problem of cognitive diagnosis in an e-learning environment could be defined as follows: Given the learners' score profiles and the Q -matrix, how could we comprehensively measure the learners' skill proficiency

with the consideration of uncertainty, and how could we conduct an efficient cognitive diagnosis to support the large-scale data processing scenarios involving a great deal of learners, exercises and skills?

In dealing with this problem, the following ideas are proposed: (1) To describe the uncertain feature of learners' skill proficiency, we represent learners' skill proficiency in the form of interval numbers, which is shown in Fig. 1. (2) To improve the efficiency of cognitive diagnosis, we introduce the cloud model theory to calculate learners' skill proficiency and its uncertainty, which could simplify the parameter estimation process.

For easy reading, the key definitions used in the following sections are given in Table 1.

4 Preliminaries

4.1 Definition of the cognitive cloud

Definition 1. (Cognitive cloud). Define domain $X =$ "score profiles of learner U_i ($i = 1, 2, \dots, I$) on the item set $P = \{P_1, P_2, \dots, P_J\}$ which requires skill V_k ($k = 1, 2, \dots, K$)," domain $T =$ "mastery of learner U_i ($i = 1, 2, \dots, I$) on skill V_k ($k = 1, 2, \dots, K$)" and $\mu_T(x) \in [0, 1]$ is the certainty degree of $x \in X$ belonging to T (i.e., the value of $\mu_T(x)$ represents the degree to which the score x is close to the learner's real skill proficiency). The cognitive cloud is defined as the distribution of $\mu_T(x)$ on qualitative concept X , and it can be characterized by the following three numerical characteristics:

- (1) Cognitive Expectation E_x^C : it represents the expectation of a learner's skill proficiency, which can be

regarded as the most representative and typical sample of the qualitative concept.

- (2) Cognitive Entropy E_n^C : it represents the fluctuation range of a learner's skill proficiency, which reflects the uncertainty of cognitive expectation.
- (3) Cognitive Hyper-entropy H_e^C : it denotes the change frequency of a learner's skill proficiency, which is the uncertain degree of cognitive entropy.

Thus, a cognitive cloud is expressed as: $C^C = (E_x^C, E_n^C, H_e^C)$. The cognitive cloud of learner # i on skill # k is defined as C^{Cik} . To generate C^{Cik} , the related score profiles need to be firstly extracted from X . The score profiles of learner # i on the item set requiring skill # k are described as $X'_{ik} = \{x'_{ik1}, x'_{ik2}, \dots, x'_{ikN}\}$, in which N is the number of the test items requiring skill # k . The three numerical characteristics of C^{Cik} are obtained via RCG (Li et al. 1995) as follows.

$$\begin{cases} E_x^{Cik} = \frac{1}{N} \times \sum_{m=1}^N x'_{ikm} \\ E_n^{Cik} = \sqrt{\frac{\pi}{2}} \times \frac{1}{N} \sum_{m=1}^N |x'_{ikm} - E_x^{Cik}| \\ H_e^{Cik} = \sqrt{\left| \frac{1}{N-1} \sum_{m=1}^N (x'_{ikm} - E_x^{Cik})^2 - (E_n^{Cik})^2 \right|} \end{cases} \quad (5)$$

4.2 Case analysis of cognitive cloud

Suppose a learner's performance is normalized to $[0, 1]$, where 0 means that the learner gives a totally wrong response to the item, and 1 otherwise. Now, there are two learners # a and # b . The performance of the former on skill # k is $[0.90, 0.80, 0.85, 0.90, 0.82]$, and the performance of

Table 1 Key definitions

Definitions	Description
X_{ij}	The score of learner # i on item # j
α_{ik}	The proficiency of learner # i on skill # k
I, J, K	The number of learners, the number of test items, the number of skills
α_{ik}^L	The lower limit of proficiency of learner # i on skill # k
α_{ik}^H	The higher limit of proficiency of learner # i on skill # k
η_{ij}	The mastery of learner # i on item # j
η_{ij}^L	The lower limit of mastery of learner # i on item # j
η_{ij}^H	The higher limit of mastery of learner # i on item # j
C^{Cik}	The cognitive cloud of learner # i on skill # k
E_x^{Cik}	The cognitive expectation of cognitive cloud of learner # i on skill # k
E_n^{Cik}	The cognitive entropy of cognitive cloud of learner # i on skill # k
H_e^{Cik}	The cognitive hyper-entropy of cognitive cloud of learner # i on skill # k
s_j, g_j	The slip and guess factor of item # j

the latter on skill # k is [1.00, 0.00, 0.70, 0.60, 0.80]. According to Eq. (5), the cognitive clouds of learners # a and # b on skill # k can be computed as $C^{Cak} = (0.8540, 0.0461, 0.0069)$ and $C^{Cbk} = (0.6200, 0.3208, 0.1976)$, respectively. Taking the three numerical characteristics of each cognitive cloud as the input, the scatter diagram with 1,500 cloud drops generated by the forward normal cloud generator (Li et al. 2009) is displayed in Fig. 2, where the abscissa denotes learner’s performance on skill # k , and the ordinate represents the membership degree. Additionally, considering the characteristics of cognitive diagnosis analysis, the abscissa of cloud drops will be intercepted within [0, 1] in practice since a learner’s performance is normalized to [0, 1].

Figure 2a shows that the cloud drops are densely distributed. Most cloud drops are in a small interval abscissa ranging around E_x^{Cak} and have higher membership degree due to his good and stable performance on skill # k . For learner # b , Fig. 2b shows that the cloud drop distribution is scattered in the range of abscissa [0, 1] and with a higher membership degree. The reason is that learner # b ever obtained not only a full score but also a zero on skill # k . Figure 2b clearly indicates the uncertainty and fluctuation for the proficiency of learner # b on skill # k . Overall, the cognitive cloud can comprehensively and objectively characterize the learners’ real skill proficiency.

4.3 Fuzzy cognitive interval numbers

Definition 2. (Fuzzy cognitive interval numbers). Suppose b^L and b^H are two real numbers, if $B = [b^L, b^H]$, $0 \leq b^L \leq b^H \leq 1$, B can be called fuzzy cognitive interval numbers, where b^L is the lower limit of the fuzzy interval of learner’s cognitive status, and b^H denotes the higher limit of the fuzzy interval of learner’s cognitive status (i.e., how much a learner masters a skill or an

item). The fuzzy cognitive interval number B is a real number when $b^L = b^H$.

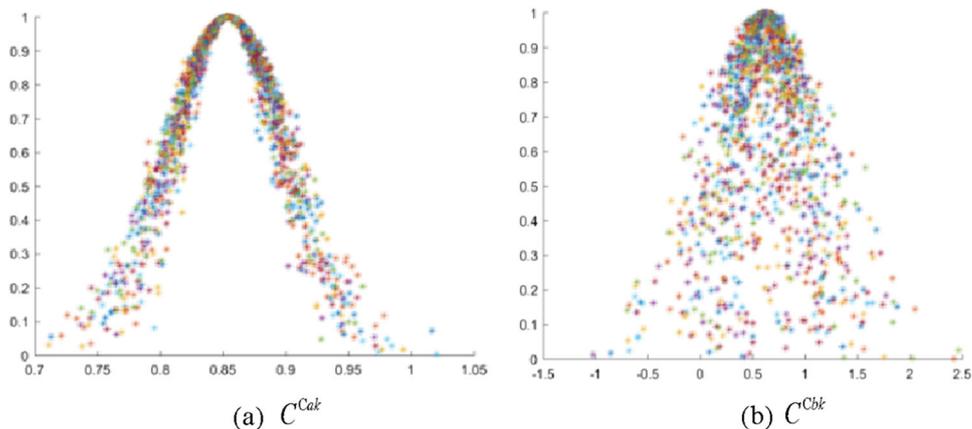
5 Fuzzy cloud cognitive diagnosis framework

By incorporating the cognitive cloud and fuzzy cognitive interval numbers, the fuzzy cloud cognitive diagnosis framework (FC-CDF) measures the learners’ skill proficiency objectively and comprehensively for predicting learners’ performance, shown in Fig. 3.

The input data of FC-CDF include the learners’ score matrix, i.e., X and a matrix that describes the relationship between skills and test items, i.e., Q . The diagnosis process of FC-CDF is mainly divided into 4 steps as follows:

- (1) Employ cognitive clouds to model the proficiency of each learner on each skill. For example, $C^{Ci1} = (E_x^{Ci1}, E_n^{Ci1}, H_e^{Ci1})$ represents the cognitive cloud of learner # i on skill #1.
- (2) Transform each cognitive cloud into the fuzzy interval numbers of learners’ proficiency on the skill level by incorporating the three numerical characteristics of cognitive clouds. The transformation results are affected by a parameter w , which is introduced in Subsect. 6.1 and discussed in Subsect. 7.2 in detail.
- (3) Combine the ideas of “fuzzy intersection” and “fuzzy union” to measure learners’ mastery of the items in the form of fuzzy interval numbers.
- (4) Calculate the interval numbers of learners’ performance in each item, and convert these interval numbers into learners’ final performance prediction results. The calculation results are affected by two

Fig. 2 Cognitive cloud scatter diagram of learner # a and learner # b on skill # k



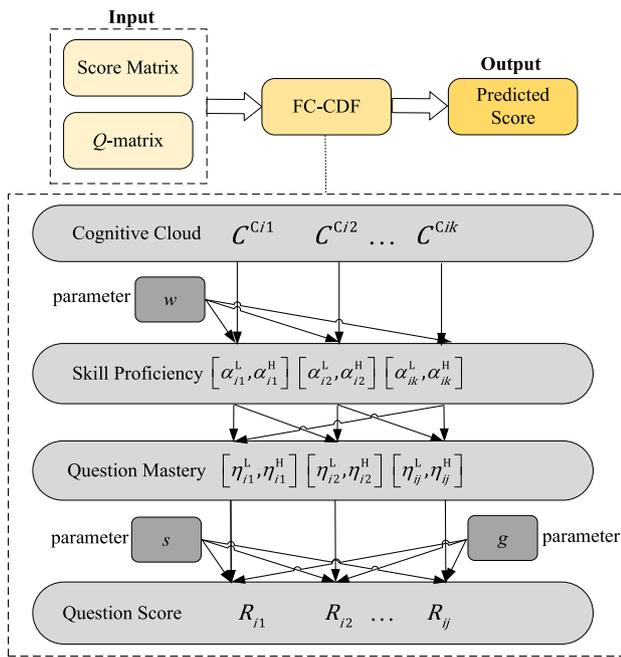


Fig. 3 4-tiers fuzzy cognitive diagnosis framework

parameters, i.e., the slip factor s and the guess factor g , which are discussed in Subsect. 6.3 in detail.

6 Examinee performance prediction via FC-CDF

In this section, an approach via FC-CDF is proposed to predict examinees’ performance in an e-learning environment by introducing the cloud model theory.

6.1 Diagnose learner’s skill proficiency

By introducing cognitive clouds, the learner’s skill proficiency is depicted with three numerical characteristics. Among these characteristics, the cognitive expectation denotes the normal level of one learner’s skill proficiency, and the cognitive entropy and the cognitive hyper-entropy depict the uncertainty of one learner’s skill proficiency. Therefore, the learner’s skill proficiency could be depicted comprehensively if these numerical characteristics are integrated into a measurement mechanism. In FC-CDF, the fuzzy interval numbers are introduced. By transforming the cognitive cloud’s three numerical characteristics into fuzzy

interval numbers, the measurement mechanism of FC-CDF could not only consider the normal level of learners’ skill proficiency, but also investigate the uncertainty of it. To be specific, the learners’ skill proficiency level matrix is defined as $A = [[\alpha_{ik}^L, \alpha_{ik}^H]]_{I \times K}$, where α_{ik}^L and α_{ik}^H denote the lower and upper limits of the skill proficiency level of learner # i on skill # k , respectively. $\alpha_{ik}^L, \alpha_{ik}^H \in [0, 1]$. Referring to (Ma et al. 2016), α_{ik}^L and α_{ik}^H could be calculated as follows: 1) If there are only two items in the item set requiring skill # k , i.e., a learner’s score profiles on skill # k obtain two values r_1 and r_2 , then $\alpha_{ik}^L = r_1, \alpha_{ik}^H = r_2, r_1 \leq r_2$. 2) When more than two items require skill # k , α_{ik}^L and α_{ik}^H are computed as follows:

$$\alpha_{ik}^L = \max\{0, E_x^{Cik} - E_n^{Cik} - w \times H_e^{Cik}\} \tag{6}$$

$$\alpha_{ik}^H = \min\{1, E_x^{Cik} + E_n^{Cik} + w \times H_e^{Cik}\} \tag{7}$$

In Eqs. (6) and (7), E_x^{Cik}, E_n^{Cik} and H_e^{Cik} denote the cognitive expectation, the cognitive entropy and the cognitive hyper-entropy of learner # i on skill # k , respectively, and w is the influence coefficient of the hyper-entropy of the cognitive cloud. The value of w is discussed in Subsect. 7.2 in detail.

6.2 Diagnose learner’s mastery of item

As mentioned before, the learner’s proficiency on the skills is measured in the form of fuzzy interval numbers since the skill proficiency is fuzzy and uncertain. Similarly, the learners’ mastery of the items is also fluctuating in practice. Inspired by FuzzyCDF (Liu et al. 2018), FC-CDF also assumes as follows: (1) The learners’ mastery of the objective and subjective items is conjunctive; namely, one learner could answer one objective item correctly only if he/she masters all the skills required by the objective item; (2) the learners’ mastery on the subjective items is compensatory; namely, the more skills required by the subjective items one learner has mastered, the higher score that he/she would get for this subjective item. Then, the fuzzy intersection and fuzzy union of fuzzy theory (Klir and Yuan 1995) are introduced to transform the conjunctive and compensatory assumptions into mathematical formulas. To be specific, one learner’s mastery of one subjective/objective item is the fuzzy intersection/union of his/her skills proficiency required by this item. Thus, the mastery of learners on the items could be defined as follows:

$$\eta_{ij}^L = \bigcap_{1 \leq k \leq K, q_{jk}=1} \alpha_k^L(i), \eta_{ij}^H = \bigcap_{1 \leq k \leq K, q_{jk}=1} \alpha_k^H(i) \tag{8}$$

$$\eta_{ij}^L = \bigcup_{1 \leq k \leq K, q_{jk}=1} \alpha_k^L(i), \eta_{ij}^H = \bigcup_{1 \leq k \leq K, q_{jk}=1} \alpha_k^H(i), \tag{9}$$

In the above equations, \cap and \cup represent the fuzzy intersection and union operators of the fuzzy theory, respectively. They are defined as follows:

$$(A \cap B)(x) = \min(A(x), B(x)) \tag{10}$$

$$(A \cup B)(x) = \max(A(x), B(x)) \tag{11}$$

Based on the fuzzy intersection/union operation, one learner’s mastery of one objective/subjective item could be the minimum/maximum of his/her skills proficiency required by the item.

6.3 Estimate item parameters

Intuitively, the learners with good item mastery give a correct response to the item and the ones with bad item mastery give a wrong response with a high probability. Nevertheless, it is possible that the learners with good item mastery might give a wrong response to the item by carelessness, and those with bad item mastery could correctly answer the item by a lucky guess. Thus, the existing research, such as the DINA model and FuzzyCDF, introduces the slip factor (s) and the guess factor (g) shown in Eqs. (3) and (4), respectively. The expectation–maximization (EM) algorithm and the Markov chain Monte Carlo (MCMC) algorithm (Chakraborty 2002) are often used in CDMs for parameter estimation, and MCMC performs better for multi-parameter estimation than EM does. In this section, an effective training algorithm using MCMC is proposed to estimate the slip and guess factors of each test item and the variance of the normalized score of learners on subjective test items for FC-CDF. By referring to the parameter settings of HO-DINA (Torre and Douglas 2004), the prior distribution of the parameters in FC-CDF is given as follows:

$$\begin{aligned} s_j &\sim \text{Beta}\left(m_s, n_s, \min_s, \max_s\right), \\ g_j &\sim \text{Beta}\left(m_g, n_g, \min_g, \max_g\right), \\ \frac{1}{\sigma^2} &\sim \Gamma(x_\sigma, y_\sigma), \end{aligned} \tag{12}$$

where $\text{Beta}(m, n, \min, \max)$ is a four-parameter Beta distribution, which has two shape parameters m and n and is supported on the range $[\min, \max]$; $\Gamma(x_\sigma, y_\sigma)$ is a Gamma

distribution which has shape parameter x_σ and scaling parameter y_σ . Then, the joint posterior distribution of s, g and σ^2 given the score matrix (i.e., X matrix) is as follows:

$$P(s, g, \sigma^2 | X) \propto L(s, g, \sigma^2) P(s) P(g) P(\sigma^2) \tag{13}$$

where L is the joint likelihood function of FC-CDF as follows:

$$L(s, g, \sigma^2) = L_o(s, g, \sigma^2) L_s(s, g, \sigma^2) \tag{14}$$

where L_o and L_s denote the joint likelihood functions of objective and subjective items, respectively. Based on Eqs. (12)–(14), L_o and L_s are defined by

$$L_o(s, g, \sigma^2) = \prod_i^I \prod_j^{P_o} (F_{ij})^{X_{ij}} (1 - F_{ij})^{1-X_{ij}} \tag{15}$$

$$L_s(s, g, \sigma^2) = \prod_i^I \prod_j^{P_s} N(X_{ij} | F_{ij}, \sigma^2) \tag{16}$$

In Eqs. (18) and (19), σ^2 is the variance of the normalized score of learners on subjective test items, and $N(\cdot | \mu, \sigma^2)$ denotes the probability density function of a Gaussian distribution with mean μ and variance σ^2 . The full conditional distributions of each parameter are as follows:

$$P(s, g | X, \sigma^2) \propto L(s, g, \sigma^2) P(s) P(g) \tag{17}$$

$$P(\sigma^2 | X, s, g) \propto L(s, g, \sigma^2) P(\sigma^2) \tag{18}$$

Finally, a Metropolis–Hastings (M-H)-based MCMC method for parameter estimation is proposed by Algorithm 1. First, each parameter is initialized according to the initial distribution settings. For each iteration, a uniformly random sample of each parameter is drawn within a predefined interval specified by δ_s, δ_g and δ_{σ^2} , and then, the full conditional probability of item slip factor s , item guess factor g and the variance of normalized scores of subjective items σ^2 are computed. Next, the acceptance probability of samples can also be calculated based on the M-H algorithm. Each sample will accept the transfer if the acceptance probability is greater than the random number generated from the uniform distribution $U(0, 1)$; otherwise, it will not accept the transfer. In this way, the parameters could be estimated after W iterations of sampling. In addition, to simplify the calculation, $P(s), P(g)$ and $P(\sigma^2)$ are equal to the probability density functions of their corresponding distributions.

Algorithm 1. Sampling Algorithm for FC-CDF

Require: Score profile matrix X and the Q -matrix.

Output: s, g and σ^2 .

1: Initialize s, g and σ^2 with random values

2: **for** $k=1, 2, \dots, W$ **do**

3: Draw $s_w \sim U(s_{w-1} - \delta_s, s_{w-1} + \delta_s), g_w \sim U(g_{w-1} - \delta_g, g_{w-1} + \delta_g)$ and accept s_w, g_w with the probability by

$$\min\left\{1, \frac{L(s_w, g_w, \sigma_{w-1}^2)P(s_w)P(g_w)}{L(s_{w-1}, g_{w-1}, \sigma_{w-1}^2)P(s_{w-1})P(g_{w-1})}\right\}.$$

4: Draw $\sigma_w^2 \sim U(\sigma_{w-1}^2 - \delta_{\sigma^2}, \sigma_{w-1}^2 + \delta_{\sigma^2})$ and accept σ_w^2 with the probability by

$$\min\left\{1, \frac{L(s_w, g_w, \sigma_w^2)P(\sigma_w^2)}{L(s_w, g_w, \sigma_{w-1}^2)P(\sigma_{w-1}^2)}\right\}.$$

5: end

6: end

6.4 Predict learner’s performance

Considering that learners’ mastery of the skills is fluctuant and uncertain, the learners’ performance in the items should be also fluctuant and uncertain. Inspired by FuzzyCDF (Liu et al. 2018), the learners’ performance in both objective and subjective test items can be predicted based on fuzzy interval numbers by

$$F_{ij}^L = (1 - s_j)\eta_{ij}^L + g_j(1 - \eta_{ij}^L) \tag{19}$$

$$F_{ij}^H = (1 - s_j)\eta_{ij}^H + g_j(1 - \eta_{ij}^H) \tag{20}$$

$$F_{ij} = \frac{F_{ij}^L + F_{ij}^H}{2} \tag{21}$$

where F_{ij}^L and F_{ij}^H denote the lower limit and the upper limits of learner $\#i$ ’s performance on item $\#j$, respectively; F_{ij} is the predicted score of learner $\#i$ on item $\#j$.

6.5 Deal with data deficiency problem

In practice, when the learners just start to learn a course, they have only completed a handful of test items. In this case, only a small amount of data about learners’ score profiles are available, which will lead to the limited accuracy of cognitive diagnosis and obviously influence the performance of prediction. Some experiments have confirmed it. For example, in the real-world dataset Math2,

which is introduced in Sect. 7.1, skill #2 is only required by item #2, i.e., there is only one score profile to depict learner’s cognitive status on skill #2. Then, if one learner’s response to item #2 is correct/wrong, his/her cognitive cloud on skill #2 would be $[1, 0, 0]/[0, 0, 0]$ in FC-CDF or his skill proficiency on this skill would be 1/0 in other CDMs. Obviously, the limited sample data increase difficulty to cognitive diagnosis. Aiming at the possible data deficiency problem, two strategies are proposed as follows:

- 1) If the number of learner $\#i$ ’s score profiles on skill $\#k$ is less than 4, i.e., $0 < |X'_{ik}| < 4$, the whole score profiles of learner $\#i$ will be regarded as input data to compute the cognitive cloud C^{Cik} .
- 2) If learner $\#i$ has not answered the item, which requires skill $\#k$, i.e., $|X'_{ik}| = 0$, aiming at the possible data deficiency, FC-CDF would combine PMF to guarantee the performance of examinee performance prediction. PMF was proposed by Ruslan et al. (Mnih and Salakhutdinov, 2007) to deal with the prediction tasks in the large, sparse and imbalanced datasets. By integrating the PMF method, the prediction result is computed as follows:

$$F_{ij}^* = (1 - \lambda) \frac{F_{ij}^L + F_{ij}^H}{2} + \lambda F_{ij}^{PMF}$$

where F_{ij}^{PMF} denotes the performance of learner $\#i$ on item $\#j$ predicted via PMF; λ is the adjustment coefficient. $\lambda = 1$ when $|X'_{ik}| = 0$, otherwise $= 0$.

7 Experiments

To verify the effectiveness of our FC-CDF, we conduct experiments to answer the following research questions (RQs):

RQ1: What is an appropriate value of the coefficient w when we diagnose learner's skill proficiency via cognitive clouds?

RQ2: How about the performance of FC-CDF on score prediction compared with the existing research?

RQ3: How about the performance of FC-CDF on execution efficiency compared with the existing research?

All procedures were implemented by Python 3.6 on a Linux server with a 2.3 GHz Intel Xeon CPU and 16 GB memory.

7.1 Experimental setup

7.1.1 Experimental datasets

The experiments use three real-world datasets, a simulation dataset and three extended datasets to verify the prediction performance and execution efficiency of FC-CDF. These datasets are introduced as follows:

- (1) The real-world datasets: They are FrcSub (DeCarlo 2011), Math1 and Math2. FrcSub records the score profiles committed by junior high school students about fraction subtraction objective test items. Math1 and Math2 are actually two independent subsets of Math2015,¹ and widely used in existing research, e.g., FuzzyCDF (Liu et al. 2018) and R-FuzzyCDF (Li et al. 2017). Math1 and Math2 are collected from two final mathematical examinations, including both objective and subjective test items, for high school students. Each dataset contains a normalized score matrix X and a Q -matrix.
- (2) The Simulation dataset: To verify the performance of FC-CDF in large-scale data environment, Simulation dataset² is generated. In the dataset, there are 1,000 learners, 100 items including 80 objective items and 20 subjective items, and 10 skills. All the learners answered all the items, and their score profiles obey the standard norm distribution. Each objective item

requires 2 to 3 skills randomly, and each subjective item requires 3 to 5 skills randomly. Every skill is required by more than 8 items.

- (3) Three extended datasets (i.e., Vir_FrcSub, Vir_Math1, Vir_Math2): The three extended datasets³ are generated based on FrcSub, Math1 and Math2, respectively, to analyze the FC-CDF's performance in a large-scale data environment. The data of every extended dataset contain two parts, i.e., the original real-world dataset and the simulated data. The extended data are constructed prudently. Taking Vir_Math1 as an example, the Q -matrix of Vir_Math1 is the same as Math1's, and the size of the simulated data part of Vir_Math1's score matrix is the same as that of Math1. In the simulated data part of Vir_Math1's score matrix, for the objective items, the score profiles of each objective item are the same as Math1's. As for subjective items, suppose the subjective item score matrix of Math1 and Vir_Math1 is defined as $R_s = [r_{ij}^s]_{I \times J_s}$ and $VR_s = [vr_{ij}^s]_{I \times J_s}$, respectively, where J_s denotes the number of subjective test items, and r_{ij}^s and vr_{ij}^s are the performance of learner $\#i$ on subjective test item $\#j$ in Math1 and Vir_Math1, respectively, $vr_{ij}^s \sim U(r_{ij}^s - 0.1, r_{ij}^s + 0.1)$.

The preview of the Q -matrix and the brief summary of all datasets are shown in Fig. 4 and Table 2, respectively. Figure 4 demonstrates those test items related to every skill. According to Fig. 4, these datasets fall into 2 categories, i.e., the symmetric datasets and asymmetric datasets. In the symmetric datasets (e.g., FrcSub, the simulation dataset), every skill is required by more than 1 test item. In the asymmetric datasets (e.g., Math1, Math2, Vir_Math1, Vir_Math2), some skills might be required by only 1 test items. For example, in Math1 and Vir_Math1, skill #4 is required by only item #11.

7.1.2 Baseline approach and parameter settings

The baseline approaches and their corresponding parameter settings are shown as follows:

- (1) PMF (Mnih and Salakhutdinov 2007): It maps the learners and items into low-dimensional latent vectors for performance prediction. The dimension of the latent vector is set as 10.
- (2) DINA (Torre 2008; Torre and Douglas 2004): It is one of the most representative CDMs which uses

¹ <http://staff.ustc.edu.cn/%7Eqliuq/data/math2015.rar>

² The simulation dataset is available on <https://pan.baidu.com/s/1sPm6eQcKbKvZ3WdJ5uW4oQ> Password: 4tb5.

³ The three extended datasets are available on <https://pan.baidu.com/s/1LcwTfQrzayPwDrYUI2uvBA> Password: mdbp.

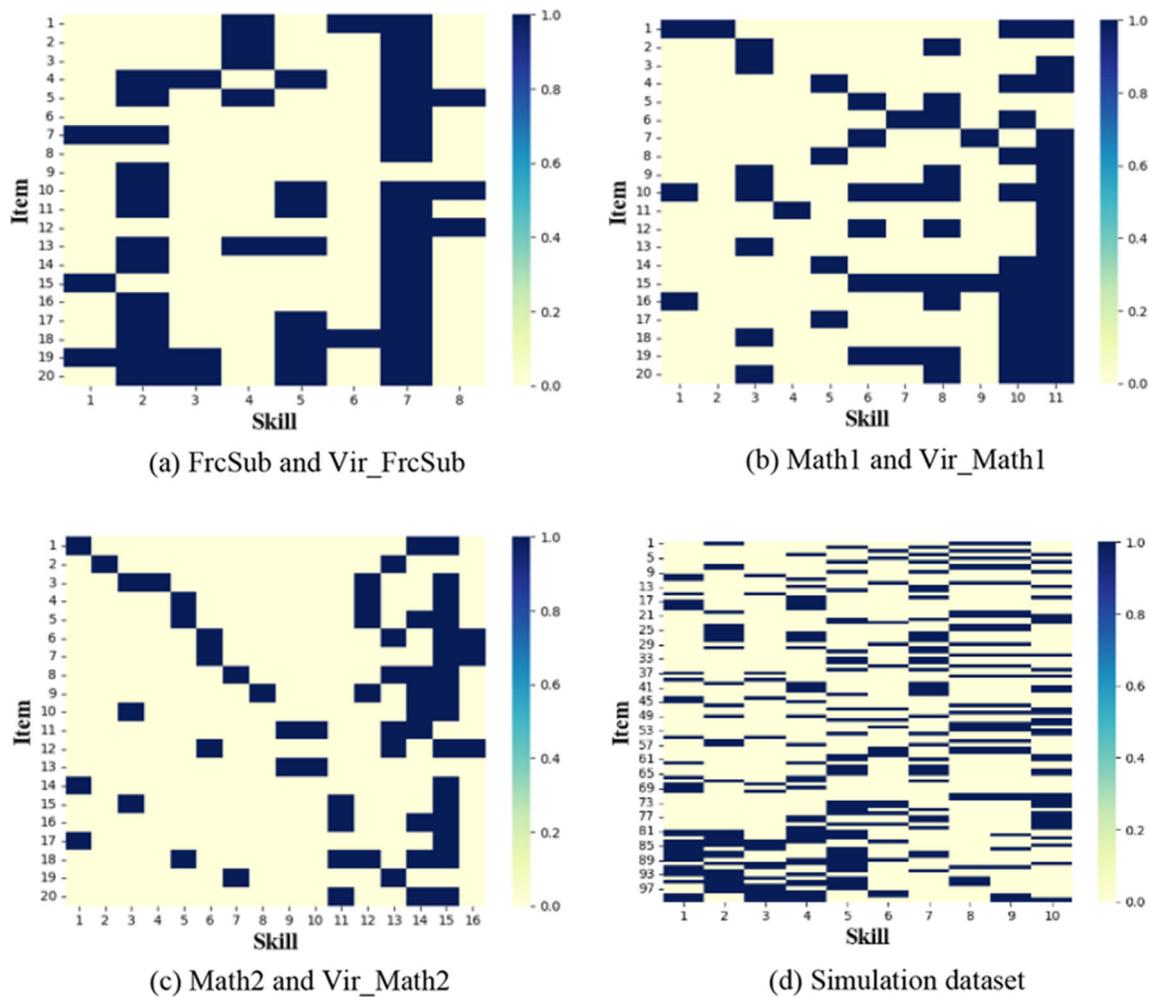


Fig. 4 Preview of Q -matrix of all datasets

Table 2 Summary of datasets information

Datasets	Learners	Skills	Items	
			Objective	Subjective
FrcSub	536	8	20	0
Math1	4209	11	15	5
Math2	3911	16	16	4
Vir_FrcSub	1072	8	20	0
Vir_Math1	8418	11	15	5
Vir_Math2	7822	16	16	4
Simulation	1000	10	80	20

binary vectors with multi-dimension to represent learners' skill proficiency. It could predict learners' performance by combing the slip and guess factors of items. In this paper, we use two methods to estimate the parameters of the DINA model. The DINA model

using the MCMC method for parameter estimation is denoted as MCMC-DINA, and the DINA model with the EM method for parameter estimation is denoted as EM-DINA.

- (3) FuzzyCDF (Liu et al. 2018): It is the first approach that applies fuzzy theory to cognitive diagnosis and address the problem of conducting cognitive diagnosis on subject items. Its parameter setting follows the original paper.
- (4) R-FuzzyCDF (Li et al. 2017): It improves the performance of FuzzyCDF by introducing the importance of skills. The parameter setting of R-FuzzyCDF follows its original paper.
- (5) NeuralCDM (Wang et al. 2020) It is one of the most representative CDMs that apply neural networks to cognitive diagnosis. Its parameter setting follows the original paper.

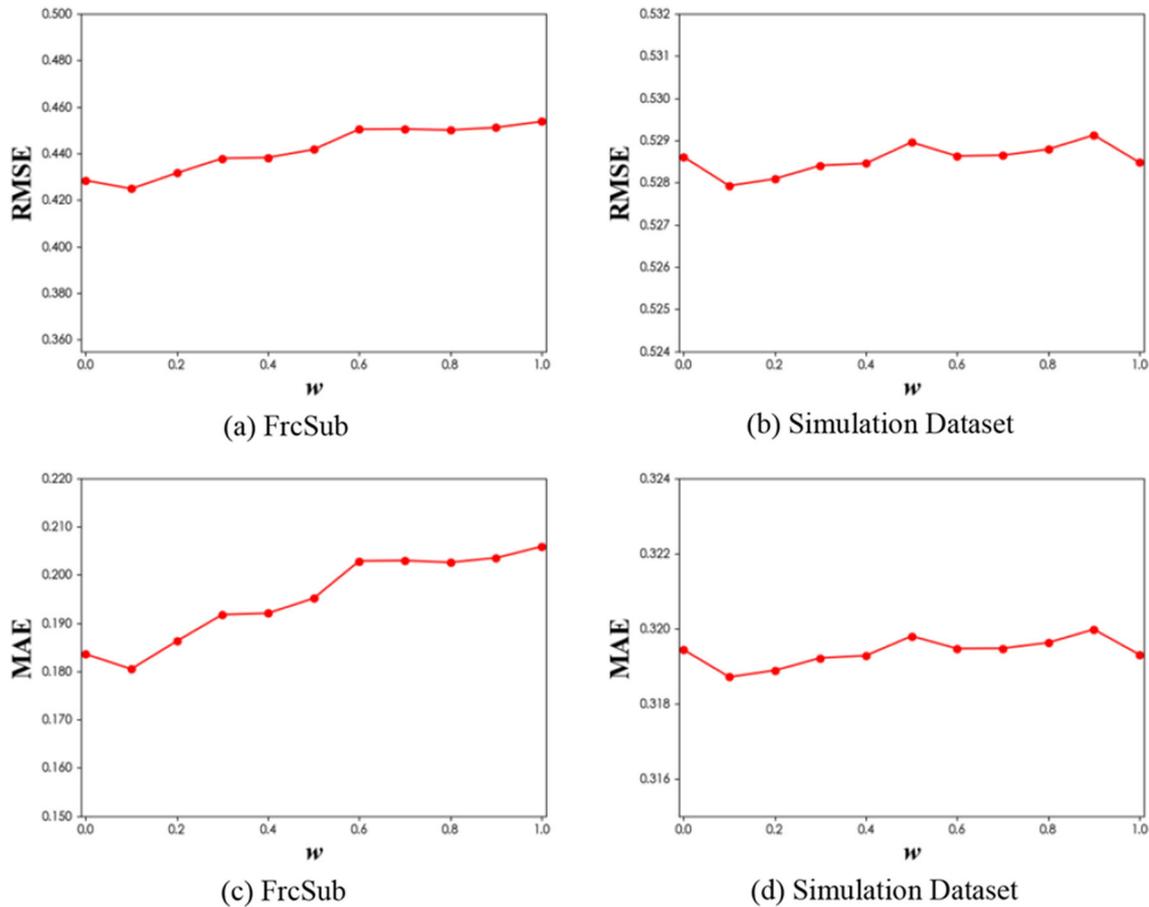


Fig. 5 Prediction performance with different values of w

For our FC-CDF, the hyperparameters of Algorithm 1 are set by referring to HO-DINA (Torre and Douglas 2004) as follows:

$$m_s = 1, n_s = 2, \min_s = 0, \max_s = 0.6, m_g = 1, n_g = 2, \min_g = 0, \max_g = 0.6, x_\sigma = 4, y_\sigma = 6, w = 0.1$$

Referring to FuzzyCDF (Liu et al. 2018), the number of iterations of Algorithm 1 is set to 5000 and the parameters are estimated based on the last 2,500 samples to guarantee the convergence of the Markov chain. Moreover, the setting of w will be discussed in Subject. 7.2.

7.1.3 Evaluation metrics

The root-mean-square error (RMSE) and the mean absolute error (MAE) are used as the metrics as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_{U_i} - X_{U_i})^2} \tag{22}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |F_{U_i} - X_{U_i}| \tag{23}$$

where N is the number of learners in the testing dataset; F_{U_i} denotes the predicted performance of learner U_i ; and X_{U_i} represents the real performance of learner U_i .

The learners' performance in items is predicted by Eqs. (14)–(18) after obtaining the slip and guess factors of each item through the model training. Note that the predicted results are continuous, for objective items, they are supposed to be discretized by

$$F_{ij}^* = \begin{cases} 1 & \text{if } F_{ij} \geq TH_F \\ 0 & \text{otherwise} \end{cases}$$

where F_{ij}^* is the discrete predicted score of learner $\#i$ on objective item $\#j$; TH_F is the threshold, which is usually set as 0.5.

7.2 Value of w (RQ1)

In FC-CDF, w is a trade-off parameter to balance the influence of the cognitive hyper-entropy on learners' skill

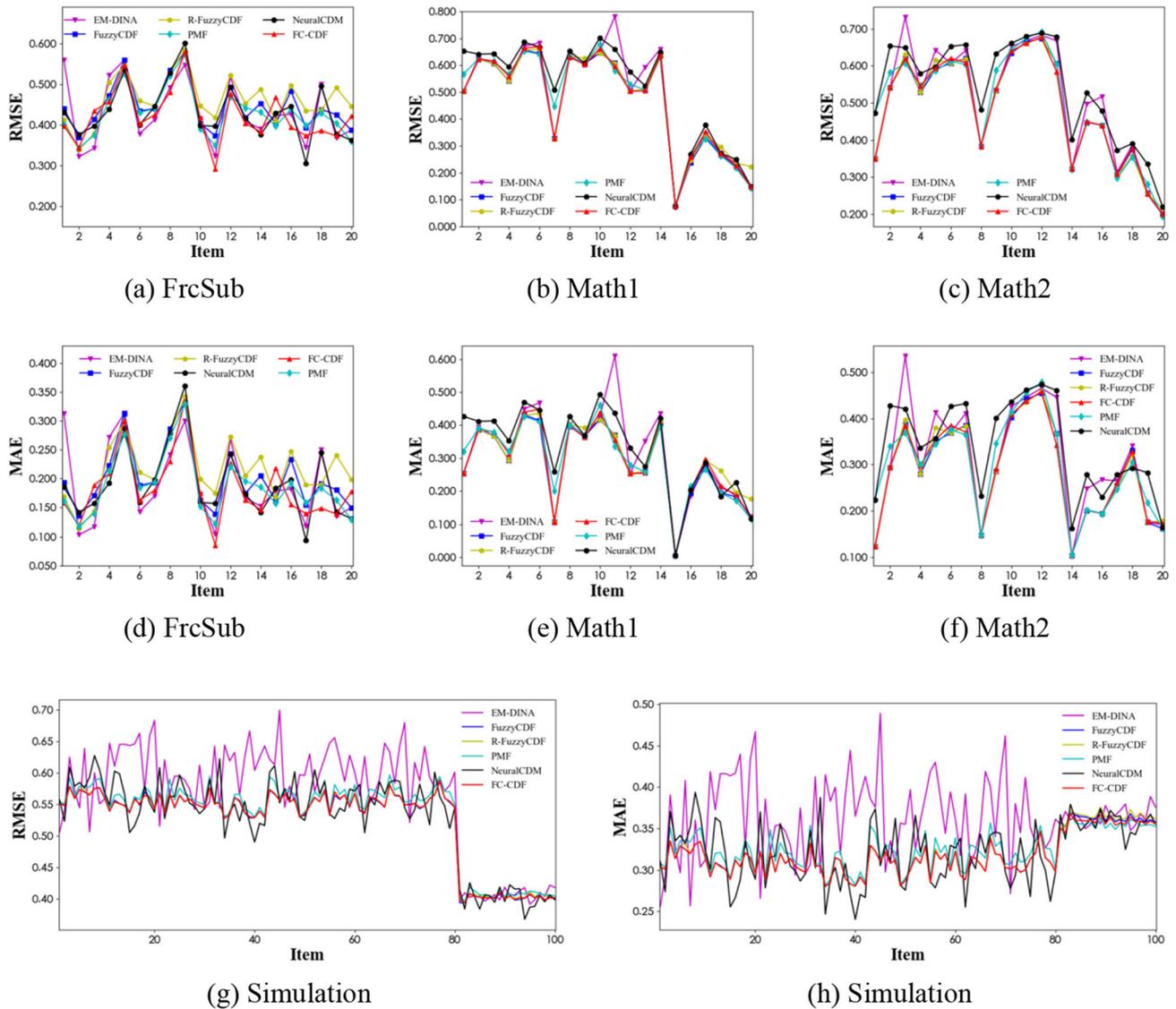


Fig. 6 Performance comparison on each item

Table 3 Performance comparison

Approaches	FrcSub		Math1		Math2		Simulation	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
PMF	0.4320	0.1870	0.5130	0.2970	0.5170	0.2980	0.5361	0.3248
MCMC-DINA	0.4470	0.2000	0.5250	0.3100	0.5300	0.3100	0.5472	0.3386
EM-DINA	0.4400	0.1940	0.5270	0.3110	0.5300	0.3100	0.5484	0.3414
FuzzyCDF	0.4540	0.2060	0.5050	0.2910	0.5070	0.2860	0.5282	0.3189
R-FuzzyCDF	0.4650	0.2170	0.5080	0.2940	0.5090	0.2880	0.5291	0.3191
NeuralCDM	0.4630	0.2150	0.5310	0.3140	0.5260	0.3060	0.5296	0.3197
FC-CDF	0.4247	0.1815	0.5060	0.2904	0.5064	0.2861	0.5278	0.3185
<i>p</i> value	1.98e-2	4.71e-2	8.30e-1	7.77e-1	8.72e-1	9.38e-1	6.85e-1	5.91e-1

proficiency. The experiments use FrcSub and simulation dataset to find a reasonable value of w , which is set from the range [0.0, 1.0], and the performance of FC-CDF with

different w is verified via a fivefold cross-validation method. The results are shown in Fig. 5.

As shown in Fig. 5, FC-CDF performs the best in FrcSub and the simulation dataset when w is set as 0.1. Moreover, with the increase of w , the performance of FC-CDF decreases, especially in the small-scale dataset (i.e., FrcSub). The reason is that large w would amplify the fuzziness and uncertainty of cognitive cloud, which leads to inevitable errors in cognitive diagnosis. Based on the above analysis, 0.1 is used as the suggested value of w .

7.3 Prediction performance analysis (RQ2)

7.3.1 Performance analysis in real-world and simulation datasets

All approaches are verified by a fivefold cross-validation method for a fair comparison, i.e., 80% of the score profiles of each learner are selected randomly as the training set, and the rest 20% for testing. Moreover, we also conduct a one-sample t -test to compare our FC-CDF with the strongest baseline approaches. The results are shown in Fig. 6 and Table 3. Considering that the performance of MCMC-DINA is like that of EM-DINA, only EM-DINA is displayed in Fig. 6 for easy reading.

From Fig. 6 and Table 3, the analysis is given as follows:

- (1) Compared with the classic approaches, i.e., PMF and DINA, FC-CDF performs optimally in all cases. By introducing cognitive cloud and interval numbers, FC-CDF could depict learners' cognitive status comprehensively and accurately predict their performance on the items.
- (2) Compared with the fuzzy theory-based approaches, i.e., FuzzyCDF and R-FuzzyCDF, our FC-CDF outperforms them in the symmetric datasets (i.e., FrcSub and simulation datasets). For the asymmetric datasets (i.e., Math1 and Math2), the cognitive clouds in FC-CDF suffer from data deficiency problems on some items because some skills are checked by only 1 test item. Thus, the performance of FC-CDF is close to that of FuzzyCDF. However, as mentioned before, since FuzzyCDF needs to estimate more parameters than FC-CDF, the subsequent experiments confirm that the execution time of FuzzyCDF is one hundred times more than that of FC-CDF.
- (3) Compared with the neural networks-based approach, i.e., NeuralCDM, there is a gap between NeuralCDM and FC-CDF in the real-world datasets. The performance of NeuralCDM declines because that neural networks need more training data, while the small

scale of a dataset (e.g., FrcSub) may cause the overfitting for NeuralCDM. It should be noted that, the cross-entropy was chosen as the loss function in the original paper of NeuralCDM. Obviously, it is not a perfect choice to train a model for performance prediction tasks when dealing with the datasets consisting of both objective and subjective items. NeuralCDM might perform better if tailored loss functions are adopted.

- (4) In the FrcSub dataset, the p value < 0.05 indicates the improvement of our FC-CDF is statistically significant. While in other datasets, the differences between FC-CDF and FuzzyCDF are not statistically significant. However, as mentioned before, FuzzyCDF consuming a lot of time to estimate parameters sacrifices efficiency for performance improvement. It is difficult to apply FuzzyCDF to real e-learning applications, while our FC-CDF could provide both competitive prediction accuracy and satisfactory diagnosis efficiency. The efficiency analysis is discussed in Subsect. 7.4 in detail.

7.3.2 Performance analysis in extended datasets

The following experiments use Vir_FrcSub, Vir_Math1 and Vir_Math2 datasets, and FC-CDF is compared with six approaches including PMF, MCMC-DINA, EM-DINA, FuzzyCDF, R-FuzzyCDF and NeuralCDM. Different percentages of the simulated data are mixed with the corresponding real-world data. The performance of all the approaches is verified by a fivefold cross-validation method to guarantee the fairness of the comparison. The experimental result is shown in Fig. 7, where the x -axis represents the percentage of simulated data, and the y -axis denotes the average RMSE or MAE values of the prediction performance in the test dataset.

Obviously, Fig. 7 demonstrates similar results to Fig. 6. FC-CDF outperforms other approaches in the symmetric datasets (i.e., Vir_FrcSub), while for the asymmetric datasets (i.e., Vir_Math1 and Vir_Math2), the predicted results of FuzzyCDF and R-FuzzyCDF are a bit better than that of FC-CDF. However, FuzzyCDF and R-FuzzyCDF need more training time than FC-CDF. For instance, compared with FuzzyCDF and R-FuzzyCDF on Vir_Math1, when the percent of added simulation data reaches 90%, the RMSE value of FC-CDF declines by 1.14% and 0.79%, respectively, and the training time of FuzzyCDF and R-FuzzyCDF are 144.29 and 223.99 times longer than FC-CDF's, and more analysis about the execution time will be mentioned in the next section.

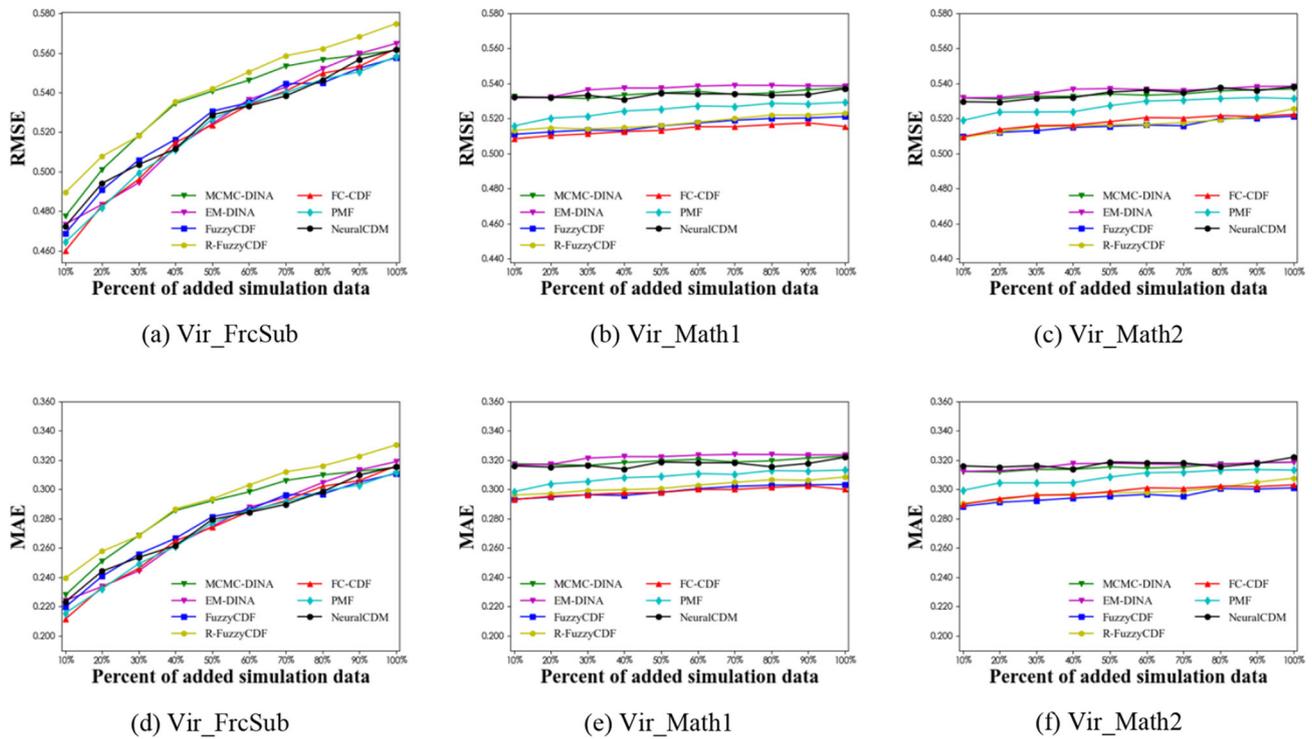


Fig. 7 Performance comparison in extended datasets

Table 4 Execution time analysis in real-world and simulation datasets (seconds)

Approaches	FrcSub	Math1	Math2	Simulation
PMF	10.80	83.88	74.94	91.14
MCMC-DINA	24.58	5,578.56	6,068.75	3,052.18
EM-DINA	0.11	5.83	72.83	1.15
FuzzyCDF	598.54	7,902.17	9,729.13	3,233.95
R-FuzzyCDF	763.44	12,583.80	15,442.89	4,379.25
NeuralCDM	234.36	1,560.53	1,384.33	1,844.08
FC-CDF	11.93	76.75	70.42	89.99

7.4 Execution time analysis (RQ3)

To support the learners’ autonomic learning online in their fragmented time, the e-learning platform is required to diagnose the learners’ cognitive status quickly for recommending suitable exercises to them. In this section, the experiments are used to analyze the execution efficiency of FC-CDF. The experiments are executed on a Linux server with a 2.3 GHz Intel Xeon CPU and 16 GB memory. This server only runs our experimental program.

7.4.1 Execution time in real-world and simulation datasets

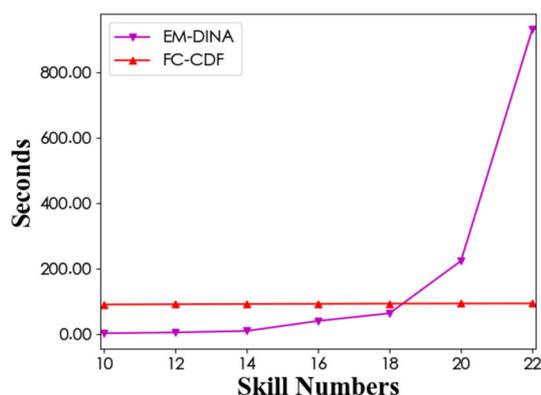
FC-CDF is compared with the six approaches including PMF, MCMC-DINA, EM-DINA, FuzzyCDF, R-FuzzyCDF and NeuralCDM. The training time of every approach in three real-world datasets and a simulation dataset is recorded, and the number of iterations of MCMC sampling is set to 5000. Moreover, we introduce a Tesla-T4 16G GPU to boost the training speed of NeuralCDM, the number of iterations of the training process is set as 5 for NeuralCDM. Table 4 shows the results.

From Table 4, the analysis is given as follows:

- (1) FC-CDF outperforms the other MCMC-based baseline approaches (e.g., R-FuzzyCDF, MCMC-DINA). More importantly, FC-CDF performs much better than FuzzyCDF. In the asymmetric datasets (i.e., Math1, Math2), the execution time of FuzzyCDF is 102.96 and 138.16 times as long as FC-CDF’s, respectively, and in the symmetric datasets (i.e., FrcSub, Simulation), FC-CDF still beats FuzzyCDF. The reason is that FC-CDF only uses the MCMC algorithm to estimate the item parameters (i.e., the slip and guess factors), while for other approaches, MCMC is utilized to estimate more parameters. For example, R-FuzzyCDF estimates the learners’ traits and skills’ difficulties via MCMC, and MCMC-

Table 5 Execution time analysis in simulation datasets (seconds)

Skill numbers	10	12	14	16	18	20	22
EM-DINA	1.29	3.81	8.31	38.97	62.33	222.47	931.36
FC-CDF	89.30	90.12	90.77	91.30	91.70	92.30	92.40

**Fig. 8** Execution time comparison in simulation datasets

DINA estimates the learners' skills proficiency via MCMC.

- (2) Although we used a GPU to boost the training speed for NeuralCDM, the execution time of NeuralCDM is 20.33 and 19.65 times as long as FC-CDF's in Math1 and Math2 datasets, respectively.
- (3) In terms of execution time, although the gap between PMF and our FC-CDF is small, the lack of interpretability and limited performance on score prediction make PMF hard to compete with our FC-CDF.
- (4) EM-DINA is obviously better than FC-CDF in FrcSub, Math1 and simulation datasets, while FC-CDF outperforms EM-DINA in Math2 dataset. The reason is that the EM-DINA's complexity is exponential with the number of skills. In the FrcSub, Math1 and simulation datasets, the number of skills is 8, 11 and 10, respectively. Among these datasets, EM-DINA could enumerate at most 2^{11} possible patterns of skill proficiency for one learner and find the most suitable skill proficiency for one learner. However, the number of skills is 16 in the Math2 dataset. In this case, EM-DINA should enumerate 2^{16} possible patterns of skill proficiency for each learner, which is more time-consuming than the performance on the former datasets.

Taking EM-DINA's performance in Math1 and Math2 as an example, when the number of skills increases from 11 to 16, the execution time of EM-DINA in Math2 is 12.49 times as long as that in Math1. Even there are fewer learners in Math2 than Math1. As for FC-CDF, its execution time does not grow fast with the increasing skill

numbers like EM-DINA. It could be foreseen that FC-CDF would outperform EM-DINA in terms of efficiency when the number of skills is more than a certain number. To verify this, a series of simulation datasets are constructed. In these datasets, the number of learners and items are set as 1000 and 100, respectively; the number of skills is set from the range [10, 22]; and the training time of FC-CDF and EM-DINA in every simulation dataset is recorded. Considering that the specific values of the score matrix and Q -matrix make no difference to the training time, their values are randomly generated in these datasets. The results are shown in Table 5 and Fig. 8.

From Table 5 and Fig. 8, it could be observed that there is a surge in EM-DINA when the number of skills increases from 18 to 22, while our FC-CDF performs stably. In practice, a course (e.g., "Java Programming") usually covers more than 20 skills, even 30 skills. In this case, the execution time of EM-DINA would be unacceptable. Compared with the baseline approaches, FC-CDF is a preferable cognitive diagnosis framework for practical e-learning scenarios.

7.4.2 Execution time in extended datasets

In the following experiments, the extended datasets (i.e., Vir_FrcSub, Vir_Math1 and Vir_Math2) are used to verify the execution efficiency of FC-CDF in larger-scale data. Different percentages from 10 to 100% of the simulation dataset are added. The execution time of training each approach in each extended dataset is recorded, and the number of iterations of MCMC sampling is set to 5000. The results are shown in Table 6.

Obviously, compared with other MCMC-based approaches (i.e., MCMC-DINA, FuzzyCDF and R-FuzzyCDF), FC-CDF significantly improves the execution efficiency of the training process. For example, in Vir_Math2, when the percent of added simulation data reaches 100%, the execution time of MCMC-DINA, FuzzyCDF and R-FuzzyCDF is 47.59, 144.25 and 229.90 times longer than that of FC-CDF, respectively. More importantly, the gap between FC-CDF and EM-DINA is getting smaller with the increase in learners in the Vir_Math2 dataset. When the percent of added extension data is more than 60%, FC-CDF beats EM-DINA because the EM-DINA's complexity is exponential with the number of skills.

Table 6 Execution time analysis in extended datasets (seconds)

Percent	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
(a) Vir_FrcSub										
PMF	13.09	11.93	13.74	14.79	15.60	15.83	19.33	18.62	19.49	19.78
MCMC-DINA	26.20	28.46	30.13	32.40	34.09	36.02	37.86	39.69	41.78	43.69
EM-DINA	0.12	0.20	0.25	0.34	0.31	0.32	0.33	0.33	0.33	0.34
NeuralCDM	237.08	240.49	249.37	259.14	277.09	303.45	326.56	341.21	372.47	377.62
FuzzyCDF	664.12	699.67	755.41	813.14	849.78	910.92	971.32	1,043.18	1,099.23	1,133.44
R-FuzzyCDF	815.24	920.32	940.24	1,035.15	1,104.92	1,174.75	1,260.56	1,307.55	1,389.67	1,462.13
FC-CDF	13.21	13.87	15.04	15.27	16.29	17.09	17.78	19.09	19.34	19.98
(b) Vir_Math1										
PMF	87.04	94.34	101.11	109.08	116.38	126.07	134.25	145.35	150.98	160.86
MCMC-DINA	5,610.23	5,648.21	5,759.16	5,820.00	5,908.04	5,945.24	6,032.65	6,287.09	6,393.37	6,452.42
EM-DINA	6.08	6.73	7.19	7.91	8.16	9.03	9.87	10.89	11.21	12.51
NeuralCDM	1778.23	1826.44	2172.09	2383.69	2520.75	2689.18	2776.05	2889.14	2916.09	3013.36
FuzzyCDF	8,433.56	9,129.17	9,963.76	10,722.78	11,509.29	12,225.45	13,000.66	13,544.20	14,231.62	14,967.56
R-FuzzyCDF	13,584.67	14,785.93	16,078.49	17,235.10	18,635.56	19,426.89	20,844.10	21,747.14	22,836.09	24,110.45
FC-CDF	82.87	89.28	96.15	104.08	110.90	122.01	131.18	137.00	142.41	153.65
(c) Vir_Math2										
PMF	83.92	92.06	99.27	106.41	115.22	120.53	132.50	137.07	143.15	151.10
MCMC-DINA	6,168.56	6,189.65	6,239.09	6,303.54	6,379.24	6,409.43	6,448.27	6,517.89	6,613.67	6,720.90
EM-DINA	76.19	82.87	89.05	94.22	108.81	128.98	135.41	140.47	145.78	150.70
NeuralCDM	1,491.12	1,650.71	1,785.98	1,880.76	2,024.66	2,200.01	2,298.45	2,414.66	2,507.09	2,634.03
FuzzyCDF	11,245.00	12,288.67	13,251.61	14,213.42	15,260.87	16,204.95	17,407.16	18,344.19	19,371.07	20,369.43
R-FuzzyCDF	17,625.45	19,187.76	20,859.00	22,300.08	24,245.91	25,884.54	27,340.67	28,834.23	30,070.94	32,464.96
FC-CDF	77.96	83.52	91.30	98.92	110.29	112.48	120.90	126.33	134.25	141.21

7.5 Experimental summary

In experiments, the prediction performance and execution efficiency of FC-CDF are compared with those of the other six approaches. By introducing cognitive cloud and fuzzy interval numbers, FC-CDF could improve the prediction accuracy in the symmetric datasets (i.e., FrcSub and simulation dataset) and obtains satisfactory prediction accuracy in the asymmetric datasets (i.e., Math1, Math2, Vir_Math1 and Vir_Math2). Most importantly, FC-CDF greatly reduces the training time of the cognitive diagnosis by employing an efficient RCG algorithm, which quickly recommends suitable exercises to learners in a new round of practice according to their latest score profiles. Although all the real-world data used in the experiments are only related to mathematics, it is worth mentioning that the proposed approach is easily generalized to other subjects or courses. In summary, FC-CDF provides good prediction accuracy and decreases the execution time in a large-scale data environment involving in many learners, test items and skills, which strongly supports the learners' autonomic learning online in their fragmented time.

8 Conclusion and discussion

The current CDMs employ a simple measurement mechanism to measure the learners' skill proficiency with low computational efficiency in a large-scale data environment. This paper proposed a new cognitive diagnosis framework FC-CDF, which takes advantage of both the cloud model theory and the fuzzy interval number. The learners' skills proficiency is modeled as cognitive clouds via RCG, and the learners' skills proficiency is measured by transforming each cognitive cloud into fuzzy interval numbers. Then, the mastery of learners on the items is simulated by mapping conjunctive and compensatory interactions into the fuzzy set operations. Based on it, the learners' performance could be predicted by considering the slip and guess factors of each item. Finally, aiming at the data deficiency problem, PMF is integrated into FC-CDF to deal with the examinee performance prediction task. The experimental results demonstrate that the proposed approach via FC-CDF significantly reduces the execution time of cognitive diagnosis and provides good accuracy of learners' performance prediction.

There are still some open issues, which can be explored in the future: (1) In a big data environment, the neural networks could be utilized to mine the latent attributes of learners and test items for more accurate cognitive modeling. Based on the neural networks, it is possible to incorporate the cloud model theory and fuzzy interval numbers to measure the fuzziness and uncertainty of cognitive status more exactly. (2) Based on the cognitive status and predicted scores via FC-CDF, the personalized learning resources recommendation, collaborative learning team formation and learning path recommendation could be achieved.

Acknowledgements This work was partially supported by National Natural Science Foundation of China (Grant Numbers 62077014 and 71971221) and Natural Science Foundation of Hunan Province (Grant Number 2021JJ30886). The authors thank Prof. Liu Qi from University of Science and Technology of China and his team for providing the MATLAB codes of FuzzyCDF and the Python codes of NeuralCDM, which are of great help to the authors.

Author contribution HM contributed to conceptualization, methodology, investigation, writing—review and editing, funding acquisition, project administration and supervision. ZH contributed to conceptualization, methodology, investigation, formal analysis, writing—original draft, and validation. HZ contributed to conceptualization and writing—review and editing. WT contributed to conceptualization and writing—review and editing. HZ contributed to conceptualization and writing—review and editing. KL contributed to writing—review and editing.

Funding This work was partially supported by National Natural Science Foundation of China (G. No. 62077014 and 71971221) and Natural Science Foundation of Hunan Province (G. No. 2021JJ30886).

Data availability All data and material used by us are available.

Code availability The custom codes will be available after this manuscript is published.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

- Bernacki ML, Chavez MM, Uesbeck PM (2020) Predicting achievement and providing support before STEM majors begin to fail. *Comput Educ* 158:103999
- Cai Y, Zhao Y, Liu S, Zhang S, Tu D (2017) An extended polytomous cognitive diagnosis model. *J Psychol Sci* 40(6):1491–1497
- Chakraborty A (2002) Markov chain monte carlo. *Resonance* 7(5):66–75
- Cheng S, Liu Q, Chen E, Huang Z, Huang Z, Chen Y, Ma H, Hu G (2019) DIRT: deep learning enhanced item response theory for cognitive diagnosis. In: Proceedings of the 28th ACM international conference on information and knowledge management, pp 2397–2400
- DeCarlo LT (2011) On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-Matrix. *Appl Psychol Meas* 35(1):8–26
- DiBello LV, Roussos LA, Stout W (2006) A review of cognitively diagnostic assessment and a summary of psychometric models. *Handb Stat* 26:979–1030
- Gao W, Liu Q, Huang Z, Yin Y, Bi H, Wang M-C, Ma J, Wang S, Su Y (2021) RCD: relation map driven cognitive diagnosis for intelligent education systems. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, pp 501–510
- Gong X, Yu C, Min L (2021) A cloud theory-based multi-objective portfolio selection model with variable risk appetite. *Expert Syst Appl* 176:114911
- Gregory C (1994) Teacher's corner: Origin of the scaling constant $d=1.7$ in item response theory. *J Educ Behav Stat* 19(3):293–295
- Janssen R, Tuerlinckx F, Meulders M, Boeck PD (2000) A hierarchical IRT model for criterion-referenced measurement. *J Educ Behav Stat* 25(3):285–306
- Jiang B, Li X, Yang S, Kong Y, Cheng W, Hao C, Lin Q (2022) Data-driven personalized learning path planning based on cognitive diagnostic assessments in MOOCs. *Appl Sci* 12(8):3982
- Klir G, Yuan B (1995) Fuzzy sets and fuzzy logic. Prentice Hall
- Li D, Meng H, Shi X (1995) Membership clouds and membership cloud generators. *Comput Res Dev* 32(6):15–20
- Li D, Liu C, Gan W (2009) A new cognitive model: cloud model. *Int J Intell Syst* 24(3):357–375
- Li Y, Wen Y, Yi X, Xu Z (2017) Revised model of fuzzy cognitive diagnosis framework. *J Data Acquis Process* 32(5):110–121
- Liu H, Li X (2020) Learning path combination recommendation based on the learning networks. *Soft Comput* 24(6):4427–4439
- Liu Q, Wu R, Chen E, Xu G, Su Y, Chen Z, Hu G (2018) Fuzzy cognitive diagnosis for modelling examinee performance. *ACM Trans Intell Syst Technol* 9(4):296–321
- Liu H, Wang L, Li Z, Hu Y (2019a) Improving risk evaluation in FMEA with cloud model and hierarchical TOPSIS method. *IEEE Trans Fuzzy Syst* 27(1):84–95
- Liu Q, Tong S, Liu C, Zhao H, Chen E, Ma H, Wang S (2019b) Exploiting cognitive structure for adaptive learning. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 627–635
- Liu Q (2021) Towards a new generation of cognitive diagnosis. In: Proceedings of the 30th international joint conference on artificial intelligence, pp 4961–4964
- Ma H, Hu Z, Li K, Zhang H (2016) Toward trustworthy cloud service selection: a time-aware approach using interval neutrosophic set. *J Parallel Distrib Comput* 96:75–94
- Ma Y, Cui C, Yu J, Guo J, Yang G, Yin Y (2020) Multi-task MIML learning for pre-course student performance prediction. *Front Comput Sci* 14(5):145313
- Ma H, Tang W, Zhu H, Zhang H (2021) Resource utilization-aware collaborative optimization of IaaS cloud service composition for data-intensive applications. *IEEE Trans Syst Man Cybern Syst* 51(2):1322–1333
- Ma H, Li J, Zhu H, Tang W, Huang Z, Tang Y (2022) Collaborative optimization of learning team formation based on multidimensional characteristics and constraints modeling: a team leader-centered approach via E-CARGO. *IEEE Trans Comput Soc Syst*. <https://doi.org/10.1109/TCSS.2022.3224762>
- Ma H, Li J, Tang Y, Zhu H, Huang Z, Tang W (2023) A universal optimization framework of leader-centered learning team formation based on fuzzy evaluations of learners and E-CARGO. *IEEE Syst Man Cybern Mag*. <https://doi.org/10.1109/MSMC.2022.3231698>
- Maatuk AM, Elberkawi EK, Aljawarneh S, Rashaideh H, Alharbi H (2022) The COVID-19 pandemic and e-learning: challenges and

- opportunities from the perspective of students and instructors. *J Comput High Educ* 34(1):21–38
- Mai T, Bezbradica M, Crane M (2021) Learning behaviours data in programming education: community analysis and outcome prediction with cleaned data. *Futur Gener Comput Syst* 127:42–55
- Nichols PD, Joldersma K (2008) Cognitive diagnostic assessment for education: theory and applications. *J Educ Meas* 45(4):407–411
- Niyogisubizo J, Liao L, Nziyumva E, Murwanashyaka E, Nshimyumukiza PC (2022) Predicting student's dropout in university classes using two-layer ensemble machine learning approach: a novel stacked generalization. *Comput Educ Artif Intell*. <https://doi.org/10.1016/j.caeai.2022.100066>
- Peng H, Zhang H, Wang J (2018) Cloud decision support model for selecting hotels on TripAdvisor.com with probabilistic linguistic information. *Int J Hosp Manag* 68:124–138
- Salakhutdinov R, Mnih A (2007) Probabilistic matrix factorization. In: *Proceedings of the 19th international conference in neural information processing systems*, pp 1257–1264
- Sessoms J, Henson RA (2018) Applications of diagnostic classification models: a literature review and critical commentary. *Meas Interdiscip Res Perspectives* 16(1):1–17
- Song X, Li J, Lei Q, Zhao W, Chen Y, Mian A (2022) Bi-CLKT: Bi-graph contrastive learning based knowledge tracing. *Knowl Based Syst*. <https://doi.org/10.1016/j.knosys.2022.108274>
- Tatsuoka KK (1983) Rule space: an approach for dealing with misconceptions based on item response theory. *J Educ Meas* 20(4):345–354
- Thai-Nghe N, Drumond L, Krohn-Grimberghe A, Schmidt-Thieme L (2010) Recommender system for predicting student performance. *Procedia Comput Sci* 1(2):2811–2819
- Thai-Nghe N, Schmidt-Thieme L (2015) Multi-relational factorization models for student modeling in intelligent tutoring systems. In: *Proceedings of the 7th international conference on knowledge and systems engineering*, pp 61–66
- Torre J (2008) DINA model and parameter estimation: a didactic. *J Educ Behav Stat* 34(1):115–130
- Torre J, Douglas J (2004) Higher-order latent trait models for cognitive diagnosis. *Psychometrika* 69(3):333–353
- Tscher A (2010) Collaborative filtering applied to educational data mining. In: *Proceedings of the 16th ACM SIGKDD conference on knowledge discovery and data mining*, pp 13–23
- Tu D, Cai Y, Dai H, Ding S (2010) A polytomous cognitive diagnosis model: P-DINA model. *Acta Psychol Sin* 42(10):1011–1020
- Wan S, Niu Z (2020) A hybrid e-learning recommendation approach based on learners' influence propagation. *IEEE Trans Knowl Data Eng* 32(5):827–840
- Wang G, Xu C, Li D (2014) Generic normal cloud model. *Inf Sci* 41:724–744
- Wang C, Liu Q, Chen E, Huang Z, Zhu T, Ss Y, Hu G (2018) The rapid calculation method of DINA model for large scale cognitive diagnosis. *Acta Electron Sin* 46(5):26–34
- Wang P, Huang S, Cai C (2021) Dual linguistic term set and its application based on the normal cloud model. *IEEE Trans Fuzzy Syst* 29:2180–2194
- Wang F, Liu Q, Chen E, Huang Z, Yin Y, Wang S, Su Y (2022) NeuralCD: a general framework for cognitive diagnosis. *IEEE Trans Knowl Data Eng*. <https://doi.org/10.1109/TKDE.2022.3201037>
- Wang F, Liu Q, Chen E, Huang Z (2020) Neural cognitive diagnosis for intelligent education systems. In: *Proceedings of the AAAI conference on artificial intelligence*, pp 6153–6161
- Wu Z, He T, Mao C, Huang C (2020) Exam paper generation based on performance prediction of student group. *Inf Sci* 532:72–90
- Yang J, Wang G, Liu Q, Guo Y, Liu Y, Gan W, Liu Y (2018) Retrospect and prospect of research of normal cloud model. *Chin J Comput* 41(3):724–744
- Yang S, Wei H, Ma H, Tian Y, Zhang X, Cao Y, Jin Y (2023) Cognitive diagnosis-based personalized exercise group assembly via a multi-objective evolutionary algorithm. *IEEE Trans Emerg Top Comput Intell*. <https://doi.org/10.1109/TETCI.2022.3220812>
- Yao H, Lian D, Cao Y, Wu Y, Zhou T (2019) Predicting academic performance for college students: a campus behavior perspective. *ACM Trans Intell Syst Technol* 10(3):1–21
- Yun Y, Dai H, Zhang Y, Wei S, Shang X (2022) Interpretable educational recommendation: an open framework based on Bayesian principal component analysis. In: *Proceedings of the 2022 IEEE international conference on systems, man, and cybernetics (SMC)*, pp 3409–3414
- Zhan P, Yu Z, Li F, Wang L (2019) Using a multi-order cognitive diagnosis model to assess scientific literacy. *Acta Psychol Sin* 51(6):734–746
- Zhang G, Li D, Li P, Kang J, Chen G (2007) A collaborative filtering recommendation algorithm based on cloud model. *J Softw* 18(1):2403–2411

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.