Abstract. Learning tests play an important role in both traditional learning and online learning. The traditional education test can only report students’ scores or abilities, but not their knowledge level, which is no longer satisfied with people’s requirements. In recent years, DINA model of cognitive diagnosis model has been widely used to diagnose students’ knowledge mastery. DINA model can dig out the students’ knowledge points and give feedback to the teachers, so that the teachers can make remedial plans for the students’ deficiencies in time. This paper first introduces the basic principle of DINA model and the improvement of DINA model in the field of education in recent years. Secondly, we introduce the development of Dina model under the trend of online education, and prove the availability of DINA model in online platform with experimental data. Finally, we predict and analyze the research direction of DINA algorithm in the future.

Keywords: Cognitive diagnosis · DINA model · Online education

1 Introduction

As an important part of students’ learning, test plays an important role in traditional education and online learning. With the maturity of testing theory and its wide application, people have a higher pursuit of testing. In traditional education tests, only the test scores or ability values of the testes can be reported and taken as the evaluation index. In fact, the subjects with the same score have different cognitive structures and different cognitive processes [1]. In 2001, the United States passed the bill “No Child Left Behind Act of 2001”. The act requires that each student’s diagnostic information be added to the educational test report. Research shows that education with only one test result but
no diagnosis and remedy is meaningless. Now people need not only the superficial evaluation of ability, but also the internal cognition [2]. Therefore, modern psychological researchers combined cognitive psychology and psychometrics to develop a new psychometric model with cognitive diagnosis function, namely cognitive diagnosis model (CDM) [3].

In cognitive psychology, cognitive diagnosis model can be used to model students’ cognitive state from the knowledge point level. After the theory came into being, it attracted wide attention. After the model of cognitive diagnosis came into being, it has attracted a lot of attention. Common cognitive diagnostic models include: Linear logistic trait model proposed by Fisher [4]; Rule space model proposed by Tatsuoka et al. [5]; Hartz’s fusion model [6]; attribute hierarchy mode by Leighton [7]; the deterministic input noise and gate model (DINA) proposed by Torre et al. [8]; a neural network-based PSP method proposed by Qian et al. [9] and so on. According to its theoretical basis, cognitive diagnosis models are divided into four categories in literature [1]: (1) potential trait models; (2) methods based on non-parametric artificial intelligence; (3) potential classification models; (4) evidence center design. Among the cognitive diagnostic models, Dina model is one of the most popular. The model has simple parameters and wide application range. It can dig out the students’ knowledge points and give feedback to the teachers, so that the teachers can make remedial plans for the students’ deficiencies in time. This article studies the DINA model, focusing on the application of the DINA model in the context of new information technology.

The logic structure of this paper is shown in Fig. 1, and the basic structure is as follows: The second section introduces the principle of DINA model; the third section introduces the application of DINA model under online education, including fuzzy diagnosis, common diagnosis and large sample diagnosis; the fourth section makes experimental analysis on the availability of DINA model under online education, and compares it with the traditional algorithm; the fifth section puts forward the prospect.
2 Model Specification

The DINA model is a typical discrete cognitive diagnostic model. This model describes the students as a multi-dimensional knowledge point grasping vector, and diagnoses from the actual answer results of the students. It can determine whether the test individual has mastered each skill required to correctly answer the test questions, help students understand their deficiencies. And let teachers teach purposefully [10].

In the DINA model, a student’s knowledge vector and the Q-matrix constitute a latent response variable for student m to question v:

\[ \eta_{mv} = \prod_{k=1}^{K} \alpha_{mk}^{q_{vk}}. \]

Here, \( \eta_{mv} \) is a potential response indicator, \( \eta_{mv} = 1 \) means the student I_m mastered the test question J_v. If not, \( \eta_{mv} = 0 \). K is the number of attributes. Let \( \alpha_{mk} \) be the binary variable for student i on knowledge k \((k = 1, \ldots, K)\), where a value of 1 indicates that student m shows mastery of knowledge k and a value of 0 indicates non-mastery, and let \( \alpha_{mk} \) be the vector of knowledge of student m. Let Q be a V by K matrix, with element \( q_{vk} \) indicating whether knowledge k is required to answer question j correctly.

The probability that student m correctly answers question v(\( Y_{mv} = 1 \)) is expressed in the DINA model as follows:

\[ P(Y_{mv} | \alpha_m) = g_v^{1-\eta_{mv}} (1 - s_v)^{\eta_{mv}}, \]

Where the slipping parameter \( s_j \) is the probability of an incorrect response to question m if all of the required knowledge of question j have been mastered, that is, \( s_j = P(Y_{mv} = 0 | \eta_{mv} = 1) \), and the guessing parameter \( g_j \) is the probability of a correct response to question m if a student lacks at least one of the required knowledge for question m, that is, \( g_j = P(Y_{mv} = 1 | \eta_{mv} = 0) \).

The parameter estimation methods commonly used in cognitive diagnostic models are the maximum expectation algorithm (Expectation Maximization Algorithm, EM) [8] and Markov Chain Monte Carlo Algorithm (MCMC) [11]. MCMC algorithm is more suitable for complex multi-parameter estimation. In the DINA basic model, only the \( s_j \) and \( g_j \) needs to be calculated by parameter estimation. We use the EM algorithm for parameter estimation. EM algorithm is an iterative algorithm for finding maximum likelihood estimates of parameters in a probability model. It has two steps: the first step is Expectation Step, which is called the E step, and the second step is the maximization step, which is called the M step. Each iteration can ensure that the value of the likelihood function increases and converges to a local maximum. For detailed steps, please refer to the literature [8]. The DINA model uses the EM algorithm to maximize the edge likelihood of (2) to obtain parameter estimates. The likelihood function is expressed as:

\[ L(s, g | \alpha) = \prod_{m=1}^{I} \prod_{v=1}^{J} P_v(\alpha_m)^{Y_{mv}} [1 - P_v(\alpha_m)]^{1-Y_{mv}}. \]

The diagnostic process of the DINA model is shown in Fig. 2. As can be seen from Fig. 2, the diagnostic process mainly includes the input of candidates’ answers, the diagnosis of mastery, and the output of score prediction. Before diagnosis, the Q matrix is defined.
by an expert and is used as a known condition input with the student’s answer matrix $Y$ to diagnose and obtain the student’s potential master set $S$. Then, the master can be used to diagnose and predict the student’s response to the unanswered questions.

In the education field, improvements have been made by many scholars to the DINA model, mainly divided into the following three categories:

1) Improvement of Q matrix. The Q matrix in the traditional DINA model is a subjective process and model parameter estimates may be inaccurate in cases where researchers employ an unspecified Q. Some studies assume part of the knowledge in the Q matrix, and have developed refining programs and algorithms. For example, de la Torre and Chiu provide a validation process that first estimates the project parameters in a temporary Q matrix, and then updates the elements of Q with an algorithm [12]. A fundamentally different approach to estimation of Q is introduced in Chen et al. [13]. This technique provides a prior specification of Q, which can be extracted from the posterior distribution by MCMC algorithm.

2) Multi-strategy improvements. When examinees can apply multiple strategies to solve problems, single strategy model (such as SS-DINA model) cannot fully capture the complexity of multiple strategy features [14]. For test tasks with multiple strategies, different strategies may be applied by different students [15]. Multi Strategies diagnosis model is an inevitable choice. For example, De la Torre and Douglas proposed MS-DINA model to provide multiple solutions to problems [15]. The model establishes multiple q-matrices, corresponding to different strategies, and students only need to use one of the strategies to solve the questions, which is more suitable for the actual situation of students.

3) Multi-level scoring improvements. Almost all cognitive diagnostic models can only adapt to binary data, which limits the application and development of cognitive diagnostics and cannot meet the needs of practical work [3]. Tu, Cai, et al. combined the idea of cumulative category response function in the hierarchical response model proposed by Samejima, and extended the dichotomy DINA model to a multi-model
called P-DINA model [3]. In addition, Torre relaxed the constraints and proposed a generalized Dina model [16].

With the development of online education, cognitive diagnosis is no longer confined to traditional classrooms, and researchers have begun to develop online platforms.

3 Development of the DINA Model in Online Education

With the continuous development of information science, online education platforms such as MOOC and NetEase cloud classroom are springing up rapidly. As an important supplement and extension of the traditional education mode, online education can provide a novel teaching method. However, at present, existing algorithms for analyzing students’ learning situation are basically based on line test questions, which are not based on students’ knowledge points and are prone to produce large deviation. In traditional learning, Torre puts forward a DINA model, which can get students’ knowledge points by question answering. If DINA model is applied to online education platforms to diagnose students’ learning situation, it will be helpful for online education. Therefore, scholars have developed a variety of application methods of DINA model on online education platforms. These methods have great practical value in using DINA model from different entry points.

3.1 Fuzzy Diagnosis

The diagnosis of students’ learning situation in online learning includes two parts: objective problem diagnosis and subjective problem diagnosis. Although the traditional DINA model for the diagnosis of objective questions (only two kinds of answers to the wrong questions) has been gradually matured, but it is not very good to diagnose the subjective questions. In order to apply DINA model to online education platforms, it is first necessary to break the limitation that DINA model can only diagnose objective questions, so that it can diagnose subjective questions. After mastering the learning situation of students, online education platform can carry out personalized teaching for students.

Most people’s understanding of knowledge is based on vague concepts. In the data source, it is very difficult to accurately find the relationship between different value points or value ranges of each attribute, and people are more concerned about the high level of abstraction [17]. In 1965, Zadeh established a method, fuzzy theory, to describe fuzzy phenomena on the basis of mathematics [18]. At present, fuzzy set theory has been successfully applied in many fields such as pattern recognition, intelligent control, machine learning, and artificial intelligence [19]. Like the traditional set operation, in the fuzzy theory, fuzzy sets also have fuzzy intersection and fuzzy union calculation - fuzzy Union takes the larger one, and fuzzy intersection takes the smaller one. The formula is:

\[ A \cap B(x) = \min(A(x), B(x)). \]  \hspace{1cm} (4)

\[ A \cup B(x) = \max(A(x), B(x)). \]  \hspace{1cm} (5)
Wu et al. propose a fuzzy cognitive diagnosis framework (FuzzyCDF) for examinees' cognitive modelling with both objective and subjective problems [20]. In this model, the probability expressions of students answering subjective and subjective questions are:

$$P(X_{ij} = 1 | \eta_{ij}, s_j, g_j) = (1 - s_j) \eta_{ij} + g_j (1 - \eta_{ij}). \quad (6)$$

$$P(X_{ij} = 1 | \eta_{ij}, s_j, g_j) = N(X_{ij}|[(1 - s_j) \eta_{ij} + g_j (1 - \eta_{ij})], \sigma^2). \quad (7)$$

Here, (6) is the diagnosis of objective problems. There are only two answers for students, either right or wrong, so $X_{ij}$ takes the value 0 or 1. In (7), students have many possibilities to answer subjective questions. So $X_{ij}$ is transformed into a continuous value of [0, 1]. $\sigma$ is the standardized variance of the subjective question score. The degree $\alpha_{ik}$ of the knowledge point in the model, that is, the degree of membership $i(k)$ between the student $i$ and the knowledge point $k$, is determined by the degree of membership of the fuzzy set related to the knowledge point.

Formally, $\alpha_{ik}$ is defined as:

$$\alpha_{ik} = i(k) = \frac{1}{1 + \exp[D \star \alpha_{ik} (\theta_i - b_{ik})]}, \quad (8)$$

the meaning of this definition is that the candidate’s proficiency in a particular skill($\alpha_{ik}$) depends on the difference between the candidate’s advanced trait ($\theta_i$) and the attributes of the skill: difficulty ($b_{ik}$) and discrimination ($\alpha_{ik}$) of skill $k$ for student $j$. Here, the coefficient $D$ is set to 1.7, which is an empirical scale constant in a log-cognitive model. $\eta_{ij}$ is student $i$’s mastery of test $j$. In objective questions, it is defined as the fuzzy intersection of student $i$’s grasp of all the knowledge points examined in test question $j$, expressed as $\eta_{ij} = i \cap k \leq K, q_k = 1(k)$; in subjective questions, it is the fuzzy union of student $i$’s grasp of all the knowledge points examined in test $j$, which is expressed as $\eta_{ij} = i \cup k \leq K, q_k = 1(k)$.

Fuzzy-CDF can extract information from objective and subjective issues, obtain more accurate and interpretable cognitive analysis, and analyze the characteristics of each student, which has great practical effects.

In 2016, Liu et al. proposed the SDINA model [21] (soft deterministic inputs, noisy “and” gate model Model), which considers the posterior probability of each knowledge point, improves the parameter estimation process of Dina model, and changes the knowledge point mastery of students into a continuous value between 0 and 1, making the diagnosis result more accurate. Later, some researchers proposed R-FuzzyCDF model based on FuzzyCDF [22]. This model redefines the mastery degree of knowledge points in the FuzzyCDF model and the mastery degree of subjective test questions by students. R-FuzzyCDF model relates the importance of knowledge points with the number of subsequent knowledge points and the number of related questions, that is, the more a knowledge point is used by other knowledge points, the more important the knowledge point is. In addition, with the increase of the number of knowledge points mastered by students, the probability of a correct answer will increase. To some extent, R-FuzzyCDF model makes up for the deficiency of FuzzyCDF, which further improves the accuracy of diagnosis model. But the parameters to be estimated also become more, which to some extent destroys the feature of simple parameters of DINA model.
3.2 General Character Diagnosis

The DINA model can diagnose students’ knowledge mastery according to the students’ questions, and then contact the students’ knowledge mastery to evaluate the students. In this way, students’ weak learning parts can be found and consolidated. However, the DINA model can only obtain diagnostic results based on its own conditions, without considering the commonality between students. It is too one-sided, and the results may have large errors, so it has not been promoted and has not played its true role.

In the development of e-commerce, how to recommend favorite products to users of online malls is a hot topic. To accurately recommend products to users, it is necessary to first analyze and diagnose the users, extract the characteristics of the users, and then judge the users based on the characteristics. Collaborative filtering algorithms are the most popular type of recommendation system in applied research. It takes into account the performance of similar users to model the target user. When the collaborative filtering algorithm is applied to student diagnosis, it takes into account the commonalities among the students, instead of unilaterally considering the target students. However, because collaborative filtering algorithms cannot analyze the students themselves, they are not interpretable. Researchers combine the DINA model with collaborative filtering algorithms and use their respective strengths to make up for each other’s shortcomings to achieve the effect of common diagnosis.

Zhu [23] and others combined the DINA model and the probability matrix decomposition method in collaborative filtering to propose a PMF-CD model. In the calculation process, the knowledge level of the students given by the previous DINA model was only 1 and 0. The PMF-CD model considers all posterior probabilities of the degree of mastery of the knowledge points, and converts the degree of mastery of the knowledge points into continuous values between 0 and 1. Then, they modeled how well they mastered the test questions based on their level of knowledge. Finally, the student’s true answer level is used as the prior information of the probability matrix decomposition algorithm, and the student’s score matrix is decomposed into the student’s feature matrix and the test question’s feature matrix to calculate the student’s potential correct answer probability. Compared with a single DINA model, the PMD-CD model considers the connection between students and is more interpretable than the DINA model. It takes into account the learning personality between students and the commonality of student learning, and can more effectively analyze the learning of students.

Based on the probability matrix decomposition method, researchers have successively combined other algorithms of collaborative filtering with cognitive diagnosis to obtain the best diagnostic results. Qi B proposed a test recommendation method based on collaborative filtering and cognitive diagnosis in an adaptive testing environment, extended the application rules of the DINA model on multi-level attribute scoring, and expanded the application scenarios of test recommendation [24]. Reference [25] takes into consideration the knowledge of students’ knowledge points and recommends secondary collaborative filtering test questions based on knowledge points, and then uses item response theory and deep self-encoder to predict students’ fullness of recommended knowledge points on recommended test questions and Comprehensive and wonderful, and finally generate a list of recommended test questions based on the prediction results.
Shan R T combines the SVD algorithm with the DINA model [26]. When analyzing target students, the algorithm adds similar user performance to the target student diagnosis. Regarding the wrong question raised by the target user, it is believed that the target user has not grasped the relevant knowledge points of the problem. For the question raised by the target user, if similar users made mistakes, it will be also thought that the target user may not be able to grasp the relevant knowledge points of the problem. Diagnostic results can further reduce calculation errors. Li modeled the students through the cognitive diagnosis model [27], and then integrated the knowledge points of the test questions and the students’ knowledge points into the matrix according to the joint probability matrix decomposition method. This method is a good description of the performance of students and test questions, ensuring the interpretability and rationality of scores. Moreover, through time complexity analysis, the method can be extended to large-scale data sets and applied to online education platforms for big data.

3.3 Large Sample Diagnosis

With the development of Internet education, educational methods have also undergone major changes with changes in computer technology and people’s teaching concepts [28]. A large number of students have begun to use online education platforms for learning. Faced with such huge and complicated data on online education platforms, the research and application of big data in the field of online education are more important [29]. The DINA model is a good diagnostic model. In recent years, researchers have applied it to online education. However, traditional DINA models are diagnosed by students with less sample data. When faced with an online education platform for large-scale student groups, the diagnostic efficiency of the DINA model will be greatly reduced [30], which makes the application of the DINA model very limited. Therefore, how to optimize the diagnostic efficiency of the DINA model has always been the focus of scholars’ research.

In 2004, in order to reduce the time complexity of the DINA model and improve its efficiency, Torre proposed the HO-DINA model [11]. This model assumes that the control of knowledge points is controlled by a hyperparameter to reduce the dimensions of knowledge points. Although this method can improve the calculation efficiency to a certain extent, it destroys the good interpretability of the original DINA model. In the DINA model, there are hidden variables that cannot be directly observed, that is, the degree of mastery of the students’ knowledge points, so the EM algorithm (see the introduction in Sect. 2) is needed for parameter estimation. However, when the amount of missing information is large, the calculation speed of the EM algorithm becomes slower [31], which will reduce the diagnostic efficiency of the DINA model.

In order to improve efficiency while maintaining the characteristics of the DINA model, Wang C et al. started from the EM algorithm used for parameter estimation in the DINA model and indirectly achieved the effect of diagnostic acceleration by improving the calculation efficiency of the EM algorithm [32]. The acceleration of the DINA model is achieved by increasing the parameter estimation speed through the following three different data set partitioning methods: 1) Incremental DINA model (I-DINA). The model mainly divides the student data, and accesses only one student block during each iteration to update the likelihood function. Other students who have not yet visited retain the results of the last iteration. 2) Maximum Entropy DINA model
(ME-DINA). The ME-DINA model filters out data sets that have little effect on each iteration to form a lazy set. The changed data set is called a change set. During the next iteration, the change set will be iterated and the lazy set will not be iterated. The final result will be retained. 3) A hybrid model based on the former two—Incremental Maximum Entropy DINA (IME-DINA). In the change set of the ME-DINA model, the division of student blocks was added. Experiments show that all three methods can achieve acceleration effects of several to several tens of times while ensuring the validity of the DINA model, which effectively improves the calculation efficiency of the DINA model. I-DINA calculation is fast, and ME-DINA calculation is stable while IME-DINA model combines the advantages of both and can achieve stable and fast results. Moreover, the above three methods are not limited to the DINA model and can also be continuously applied to other cognitive diagnostic models based on the EM algorithm, which have good generality and promotion value.

4 Usability Analysis of DINA Model

In online education, it is important to predict the next answer of a student based on their learning and recommend personalized test questions for the student. This article takes the test question recommendation algorithm as an example. The traditional test recommendation algorithm is compared with the recommendation algorithm added to the DINA model, and the availability of DINA is proved by experimental results.

4.1 Dataset Introduction

The data set comes from reference [20], which are FrcSub data set, Math1 data set and Math2 data set. Table 1 shows the statistics of the three data sets. The three data sets contain two data, matrix Q and scoring matrix X. Each column of matrix Q represents a knowledge point, and each row represents a test problem. 0 means the test question does not contain the knowledge point, and 1 means it does. Each column of score matrix X represents the score of a test question, each row represents a test question, 0 indicates that the student’s score on the question is 0, and 1 indicates that the student’s score is 1.

| Student | Questions | Knowledge |
|---------|------------|-----------|
| FrcSub  | 536        | 20        | 8         |
| Math1   | 4209       | 20        | 11        |
| Math12  | 3911       | 20        | 16        |
4.2 Experimental Evaluation Indicators

This article uses the precision, recall and F1 indicators commonly used in recommendation fields [24]. The specific definitions are as follows:

\[
\text{precision} = \frac{TP}{TP + FP} \quad \text{(9)}
\]

\[
\text{recall} = \frac{TP}{TP + FN} \quad \text{(10)}
\]

\[
F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad \text{(11)}
\]

Here, TP indicates the number of students who really answered correctly in the recommended test questions, FP indicates the number of students who incorrectly answered the recommended test questions, and FN indicates the number of students who really answered the recommended test questions.

4.3 Comparative Test Methods

(1) DINA model [8]. Calculate students’ knowledge of mastery according to the DINA model, and select students’ weak knowledge points from test questions for recommendation.

(2) User-based collaborative filtering [33]. This method selects similar students to do wrong test questions to recommend to users.

(3) Collaborative filtering algorithm based on the DINA model. Comprehensive consideration of user-based collaborative filtering algorithm and DINA model. Choosing test questions with weak students’ knowledge points and similar students’ mistakes and recommend them.

4.4 Analysis of Experimental Results

Select 30%, 40% and 50% of the test questions as the test set, recommend three test questions to the students of the test set, and calculate the accuracy, recall, and F1 values of the three methods, respectively. The experimental results are shown in Table 2.

From this figure, we have the following conclusions:

(1) Observed from the data, the collaborative filtering algorithm that takes into account the degree of knowledge of the students’ knowledge is better than the traditional cognitive diagnostic model, and recommends to the students based on the knowledge of the similar users of the target user Title, the algorithm is more reliable and persuasive.

(2) When the data is larger, the performance of the improved algorithm based on cognitive diagnosis collaborative filtering algorithm is superior, which shows that the algorithm is more suitable after adding DINA model, and can be applied to real life.
Table 2. Accuracy, recall and F1 of different recommendation methods in the same dataset

| Method | FrcSub | Math1 | Math2 |
|--------|--------|-------|-------|
|        | 30%    | 40%   | 50%   | 30%    | 40%   | 50%   | 30%    | 40%   | 50%   |
| Accuracy (1) | 0.4264 | 0.4336 | 0.4421 | 0.3535 | 0.3323 | 0.3110 | 0.4208 | 0.3833 | 0.3693 |
| (2)     | 0.8723 | 0.8729 | 0.8535 | 0.6735 | 0.6479 | 0.6470 | 0.6017 | 0.6183 | 0.6052 |
| (3)     | 0.9067 | 0.8852 | 0.8744 | 0.7132 | 0.7200 | 0.7151 | 0.8042 | 0.7698 | 0.7348 |
| Recall (1) | 0.4617 | 0.3365 | 0.2799 | 0.3679 | 0.2592 | 0.1920 | 0.4117 | 0.2816 | 0.2161 |
| (2)     | 0.3365 | 0.3157 | 0.2937 | 0.1657 | 0.1674 | 0.1084 | 0.2777 | 0.2786 | 0.2692 |
| (3)     | 0.3659 | 0.2611 | 0.2165 | 0.2766 | 0.3097 | 0.2848 | 0.2469 | 0.2048 | 0.1828 |
| F1 (1)  | 0.4433 | 0.3789 | 0.3428 | 0.3606 | 0.2913 | 0.2381 | 0.4162 | 0.3246 | 0.2727 |
| (2)     | 0.4546 | 0.4362 | 0.4029 | 0.2660 | 0.1843 | 0.1392 | 0.3021 | 0.3147 | 0.2882 |
| (3)     | 0.5214 | 0.4032 | 0.3471 | 0.3773 | 0.3834 | 0.3549 | 0.3778 | 0.3359 | 0.2927 |

5 Conclusion

Cognitive diagnosis models are bridges between internal attribute and external response, which plays an important role in cognitive diagnosis and evaluation [34]. It can accurately measure and diagnose the cognitive attribute structure of students, provide exact basis for improving or remediating teaching, and point out the direction for improving teaching quality [35]. Compared with other models, DINA model is an excellent model with simple parameters and easy to understand [34], which has been widely used in the field of education. Researchers have also constantly improved the model. In recent years, with the development of social information, education has been transforming towards educational information. The rapid rise of various online education platforms will be a major development trend of education in the future. In this development trend, the application of DINA model to online education platforms can help to analyze the status of students, thus contributing to the development of online education. Therefore, researchers continue to integrate DINA model and information technology, so that DINA model can get rid of the limitations of traditional education. Although DINA model has developed for many years, it is still in its infancy in information education. In the future, our research on DINA model can be carried out from the following aspects:

(1) How to show the diagnosis process of students?
    Hidden attributes of students are difficult to measure in DINA model. If the diagnosis process is shown in the diagnosis, the reliability of the diagnosis will be greatly improved.

(2) How to explore other learning factors of students?
    In the DINA model, only two factors of students’ guesses and errors are considered. In fact, students are influenced by many factors when they complete a set of test questions. Fully mining the potential factors of students can help teachers better understand students’ learning state and help improve students’ academic level.
(3) How to make cognitive diagnosis of multi-disciplinary integration?
Current cognitive diagnosis work is all around the test of a single subject for students. However, students learn multiple subjects at the same time, and the subjects affect each other, actually. Through the combination of data mining technology and cognitive diagnosis, the learning situation of students can be directly understood through students’ multi-disciplinary integration for diagnosis.

(4) How to overcome the data redundancy brought by big data in the diagnosis process?
With the development of the era of big data, how to overcome the data redundancy and recommend the students’ performance prediction quickly and stably in the massive data is the next issue to be considered.

References

1. Chen, Q.M., Zhang, M.Q.: Development of cognitive diagnosis models and their application methods. Adv. Psychol. Sci. 18(03), 522–529 (2010)
2. Dongbo, T., Cai, Y., Qi, S., Dai, H.: A review on cognitive diagnostic models under modern test theory. Psychol. Explor. 28(2), 64–68 (2008)
3. Tu, D.B., Cai, Y., Dai, H.-Q.D.: A polytomous cognitive diagnosis model: P-DINA model. Acta Psychologica Sinica 42(10), 1011–1020 (2010)
4. Fischer, G.H.: The linear logistic test model as an instrument in educational research. Acta Physiol. (Oxf) 7(6), 359–374 (1973)
5. Tatsuoka, K.K.: Rule space: an approach for dealing with misconceptions based on item response theory. J. Educ. Meas. 20(4), 345–354 (1983)
6. Hartz, M.C.: A Bayesian framework for the unified model for assessing cognitive abilities: blending theory with practicality. Am. J. Gastroenterol. 95(4), 906–909 (2002)
7. Leighton, J.P., Gierl, M.J., Hunka, S.M.: The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka’s rule-space approach. J. Educ. Meas. 41(3), 205–237 (2004)
8. De La Torre, J.: DINA model and parameter estimation: a didactic. J. Educ. Behav. Stat. 34(1), 115–130 (2009)
9. Qian, J.X., Yu, J.Y.: The neural network-based PSP method in cognitive diagnosis. Psychol. Sci. 33(4), 915–917 (2010)
10. Ozaki, K.: DINA models for multiple-choice items with few parameters: considering incorrect answers. Appl. Psychol. Meas. 39(6), 431–447 (2015)
11. De La Torre, J., Douglas, J.A.: Higher-order latent trait models for cognitive diagnosis. Psychometrika 69(3), 333–353 (2004)
12. de la Torre, J., Chiu, C.Y.: A general method of empirical Q-matrix validation. Psychometrika 81(2), 253–273 (2016)
13. Chen, Y., Culpepper, S.A., Chen, Y., Douglas, J.: Bayesian estimation of the DINA Q matrix. Psychometrika 83(1), 89–108 (2018)
14. Liu, T.H., Zhao, Y., Dai, H.: Mix-DINA model based on mixture distribution for multiple strategy cognitive diagnosis. Psychol. Explor. 36(5), 464–471 (2016)
15. Huo, Y., de la Torre, J.: Estimating a cognitive diagnostic model for multiple strategies via the EM algorithm. Appl. Psychol. Meas. 38(6), 464–485 (2014)
16. De La Torre, J.: The generalized DINA model framework. Psychometrika 76(2), 179–199 (2011)
17. Xiong, X.H., Yao, J.C.: Research and application of data mining based on fuzzy sets. Comput. Eng. Appl. 01, 203–205 (2002)
18. Zadeh, L.A.: Fuzzy sets. Inf. Control 8(3), 338–353 (1965)
19. Li, G., Yang, B., Liu, Y., et al.: Survey of data mining based on fuzzy set theory. Comput. Eng. Des. 32(12), 4064–4067 (2011)
20. Wu, R., Liu, Q., Liu, Y., et al.: Cognitive modelling for predicting examinee performance. In: International Conference on Artificial Intelligence, pp. 1017–1024. AAAI Press (2015)
21. Liu, Y., et al.: Collaborative learning team formation: a cognitive modeling perspective. In: Navathe, S.B., Wu, W., Shekhar, S., Du, X., Wang, X.S., Xiong, H. (eds.) DASFAA 2016. LNCS, vol. 9643, pp. 383–400. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32049-6_24
22. Li, Y.X., Wen, Y.M., Yi, X.H., et al.: Revised model of fuzzy cognitive diagnosis framework. J. Data Acquisition Process. 05, 110–121 (2017)
23. Tianyu, Z., Zhenya, H., Enhong, C., et al.: Personalized test item recommendation method based on cognitive diagnosis. Chin. J. Comput. 01, 178–193 (2017)
24. Qi, B., Zou, H.X., Wang, Y., et al.: Question recommendation based on collaborative filtering and cognitive diagnosis. Comput. Sci. 235–240 (2019)
25. Xiong, H.J., Song, Y.F., Zhang, P., et al.: Personalized question recommendation based on autoencoder and two-step collaborative filtering. Comput. Sci. 172–177 (2019)
26. Shan, R.T., Luo, Y.C., Sun, Y.: Collaborative filtering algorithm based on cognitive diagnosis. Comput. Syst. Appl. 27(3), 136–142 (2018)
27. Li, Q., Liu, X.H., Xu, X.H., et al.: Personalized test question recommendation method based on unified probabilistic matrix factorization. J. Comput. Appl. 38(3), 639–643 (2018)
28. Aparicio, M., Bacao, F.: E-learning concept trends. In: Proceedings of the 2013 International Conference on Information Systems and Design of Communication, pp. 81–86. ACM, New York (2013)
29. Chen, C., Wang, Y.P., Li, C., et al.: The research and application of big data in the field of online education. J. Comput. Res. Dev. S1, 67–74 (2014)
30. Torre, J., Minchen, N.: Cognitively diagnostic assessments and the cognitive diagnosis model framework. Educ. Psychol. 20(2), 89–97 (2014)
31. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc.: Ser. B (Methodol.) 39(1), 1–22 (1977)
32. Wang, C., Liu, Q., Chen, E.H., et al.: The rapid calculation method of DINA model for large scale cognitive diagnosis. Acta Electronica Sinica 46(5), 1047–1055 (2018)
33. Goldberg, D., Nichols, D., Oki, B.M., et al.: Using collaborative filtering to weave an information tapestry. Commun. ACM 35(12), 61–70 (1992)
34. Xiao, Z., Ruxue, S.: Research advance in DINA model of cognitive diagnosis. China Examinations 1, 32–37 (2013)
35. Zhang, M.Q., Fan, X.Z., Guo, K.Y., et al.: The evolution and development of education quality assessment under the concept of big data. Mod. Educ. J. 3, 2–6 (2016)