

Article

NeuralNCD: A Neural Network Cognitive Diagnosis Model Based on Multi-Dimensional Features

Guangquan Li ^{1,2}, Yuqing Hu ¹, Junkai Shuai ¹, Tonghua Yang ^{2,*}, Yonghong Zhang ^{3,*}, Shiming Dai ¹ and Naixue Xiong ⁴ 

¹ School of Computer and Information Engineering, Jiangxi Agriculture University, Nanchang 330045, China

² School of Vocational Teachers, Jiangxi Agriculture University, Nanchang 330045, China

³ School of Information Engineering, Jiangxi Vocational College of Mechanical & Electrical Technology, Nanchang 330013, China

⁴ Department of Computer, Mathematical and Physical Sciences, Sul Ross State University, Alpine, TX 79830, USA

* Correspondence: yangth883@jxau.edu.cn (T.Y.); wzfzyh@jxau.edu.cn (Y.Z.)

Abstract: One of the most critical functions of modern intelligent teaching technology is cognitive diagnostics. Traditional cognitive diagnostic models (CDMs) usually use designed functions to deal with the linear interaction between students and exercises, but it is difficult to adequately deal with the complex relationship of non-linear interaction between students and exercises; moreover, existing cognitive diagnostic models often lack the integrated consideration of multiple features of exercises. To address these issues, this paper proposes a neural network cognitive diagnosis model (NeuralNCD) that incorporates multiple features. The model obtains more accurate diagnostic results by using neural networks to handle the nonlinear interaction between students and exercises. First, the student vector and the exercise vector are obtained through the Q-matrix; second, the multi-dimensional features of the exercises (e.g., difficulty, discrimination, guess and slip) are obtained using the neural network; finally, item response theory and a neural network are employed to characterize the interaction between the student and the exercise in order to determine the student's cognitive state. At the same time, monotonicity assumptions and data preprocessing mechanisms are introduced into the neural network to improve the accuracy of the diagnostic results. Extensive experimental results on real world datasets present the effectiveness of NeuralNCD with regard to both accuracy and interpretability for diagnosing students' cognitive states. The prediction accuracy (ACC), root mean square error (RMSE), and area under the curve (AUC) were 0.734, 0.425, and 0.776, respectively, which were about 2–10% higher than the related works in these evaluation metrics.

Keywords: cognitive diagnosis; neural network; deep learning; intelligent educational system



Citation: Li, G.; Hu, Y.; Shuai, J.; Yang, T.; Zhang, Y.; Dai, S.; Xiong, N. NeuralNCD: A Neural Network Cognitive Diagnosis Model Based on Multi-Dimensional Features. *Appl. Sci.* **2022**, *12*, 9806. <https://doi.org/10.3390/app12199806>

Academic Editor: Amerigo Capria

Received: 23 August 2022

Accepted: 21 September 2022

Published: 29 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, with the fast proliferation of big data in education, intelligent education systems [1] and large-scale online education platforms [2] have become more and more widely accepted. Smart education systems and online education platforms enable learners to learn what they need faster and easier. However, the large amount of educational resources and the difference in the knowledge level of each learner make it a challenge for learners to learn what they need in a limited time [3]. The ways in which one can quickly and accurately understand learners' knowledge levels, help learners learn the concepts they have not yet mastered, or consolidate the concepts they have mastered have become an important research topic in the current field of intelligent education [4,5]. For this reason, scholars have introduced cognitive diagnostic models to track and diagnose students' knowledge states [6].

Recently, some scholars have put significant effort and extensive work into the subject of cognitive diagnostics, and it has resulted in significant progress [7]. For example, item re-

sponse theory (IRT) [8] modeled students into one-dimensional ability values by combining exercise characteristics, such as difficulty, discrimination, and students' knowledge states. Deterministic input, noisy "And" gate model (DINA) [9] combined the Q-matrix and the student's existing answer records to model the student as a multidimensional ability vector on the knowledge points. However, the interaction functions used in the above models still have some shortcomings in dealing with the relationship between students and exercises. For example, the logistic function [10] and the inner product [11].

With the rapid development of deep learning algorithms, the powerful processing of information by artificial neural networks has been noticed by scholars. For example, neural networks have been used to simulate energy conversion [12] and pedestrian detection [13], and good results have been obtained for both. In the field of cognitive diagnostics, neural cognitive diagnosis for intelligent education systems (NeuralCDM) was proposed by WANG et al. [14] and used neural networks to model the interaction between students and exercises, which enhances the use of information. However, these models still have some shortcomings. For example, there is a lack of exploration of integrated modeling of multiple exercise features and the further processing of the nonlinear connection between students and exercises that neural networks can provide.

To this end, this paper proposes a neural network cognitive diagnosis model based on multi-dimensional features, which is based on the NeuralCDM and uses a 3-parameter logistic model [8] to enable the incorporation of multiple exercise features into the cognitive diagnosis model. The complex interaction of students and exercise is modeled by a neural network. In addition, data preprocessing is performed on the input of each layer of the neural network [15], so that the neural network can better capture the connection between students and exercises and improve the accuracy of the model. Finally, the monotonicity assumption is applied to the neural network to enhance the interpretability of the model.

The main contributions in this paper are summarized as follows:

(1) In this paper, we propose a neural network cognitive diagnosis model based on multi-dimensional features. Based on a 3-parameter logistic model, the model can effectively utilize multiple exercise features (e.g., difficulty, discrimination, guess and slip factor), so that it can better diagnose students' knowledge states.

(2) The model introduces data preprocessing mechanisms and monotonicity assumptions to enhance the accuracy and interpretability of the model. A data batching layer is added to the neural network to enable it to better capture the interaction information between students and exercises. The monotonicity assumption is introduced to make the diagnosed student's states more interpretable.

The rest of the paper is organized as follows. Section 2 presents a short literature review on epoch-based developmental cognitive diagnostic models. Section 3 details the proposed methods for acquiring exercise features (e.g., difficulty, discrimination, guess and slip factor) and neural networks for modeling nonlinear interactions between students and exercises, respectively. Section 4 provides an in-depth analysis of the effectiveness of neural networks and exercise feature preferences and compares them with existing work. Finally, conclusions and recommendations for future work are drawn in Section 5.

2. Related Work

In smart education, cognitive diagnosis of student status is a vital and fundamental task in realistic scenarios, such as personalized student tutoring [16], early warning of academic status [17], and student preference prediction [18]. For this reason, some classical cognitive diagnostic models have been proposed to solve such problems [19]. For example, the IRT model is one of the most classic models that models students' diagnoses through their answer records, and characteristics such as the difficulty of the exercises. Multidimensional item response theory models (MIRT) proposed by Reckase M D et al. [20] is a type of modeling of competencies, item characteristics, and diagnostic performance. It transforms the previous unidimensional cognitive diagnostic model into a multidimensional cognitive diagnostic model. Although these cognitive diagnostic models can diagnose the student's

knowledge state, the description of the student's state is a one-dimensional competency value. The potential representation of student states by ability values makes the models poorly represented in terms of interpretability.

In order to enhance the understandability of the model, the DINA was proposed by De La Torre et al. [9], who utilized students' exercises records and the exercise knowledge concept association matrix (Q-matrix) to describe the students as a multidimensional vector of competencies on the knowledge points, so that the student competency vector corresponds to the knowledge concepts. However, the DINA also has some shortcomings. In response to the problems of the DINA model, researchers have carried out extensive research. For example, fuzzy cognitive diagnosis for modeling examinee performance (FuzzyCDF) was proposed by Wu et al. [21], who addressed the situation that the traditional DINA only diagnoses objective questions and does not consider the diagnosis of subjective questions. Student ability was represented as the affiliation of a fuzzy set (a real number in the range of the set $[0, 1]$) and two cognitive diagnosis models, fuzzy intersection and fuzzy concatenation, were used to diagnose objective and subjective questions separately. Higher-order latent trait models for cognitive diagnosis (HO-DINA) were proposed by De la Torre and Douglas [22] considered that students' knowledge acquisition is influenced by one or more higher-order features of the exercises. Therefore, the model combines a higher-order latent structure model and DINA to specify the joint distribution of binary attributes in the cognitive diagnostic model through higher-order latent features, improving the utilization of less informative, but cognitively specific, data. In addition, there are also DINA-based cognitive diagnostic methods, such as GDINA [23], GDI [24], ESVE-DINA [25] and other cognitive diagnostics [26]. All these methods further enhance the diagnostic capability of DINA. However, the use of information from the exercises is still limited by the manually designed functions, which make it difficult to adequately deal with the nonlinear relationship between students and the exercises.

With the rapid development of deep learning, the powerful data processing ability of neural networks has achieved great success in all major fields [27,28], and the introduction of neural networks into cognitive diagnosis models has become a hot research topic [29]. Deep learning enhanced item response theory for cognitive diagnosis (DIRT) was proposed by Song et al. [30], who used neural networks to enhance the extraction of textual information from exercises and improved the application of the model to sparse data. The NeuralCDM [14] proposed the use of neural networks instead of traditional logical functions to model the interaction between students and exercises. However, the processing of interaction information is not comprehensive enough and lacks the modeling of some of the exercise features (e.g., guess and slip factor) and the integration of multiple exercise features.

In this paper, we propose a neural network cognitive diagnosis model that incorporates multiple features. With the model in this paper, the features of the exercises are enriched by introducing a 3-parameter logistic model, while the student states are modeled interpretably using the Q-matrix, and the model uses a neural network for interaction modeling on the interaction between the student and the exercises, and data preprocessing is performed on the input of each layer of the neural network, so that the neural network can better capture the interaction information of the student and the exercises. Finally, the monotonicity assumption is used to further enhance the interpretability of the model.

3. Our Proposed NeuralNCD Framework

The traditional cognitive diagnostic model has difficulties in handling the nonlinear connection between students and exercises and in considering multiple exercise features (e.g., difficulty, discrimination, guess and slip factor). In this paper, we propose a new neural network cognitive diagnosis model (NeuralNCD).

The model's broad framework is depicted in Figure 1. The model contains three main modules, namely the input module, the exercise feature module and the IRT interaction module. The input module is used to initialize the students' mastery of each knowledge

concept and obtain the knowledge concepts that the exercises have through the Q-matrix, which is obtained through expert labeling. The IRT interaction module is based on the 3-parameter logistic model theory to obtain the interaction vectors between students and the exercises and uses the neural network to model the interaction vectors and finally predict the students' performance to obtain students' mastery of the knowledge concept.

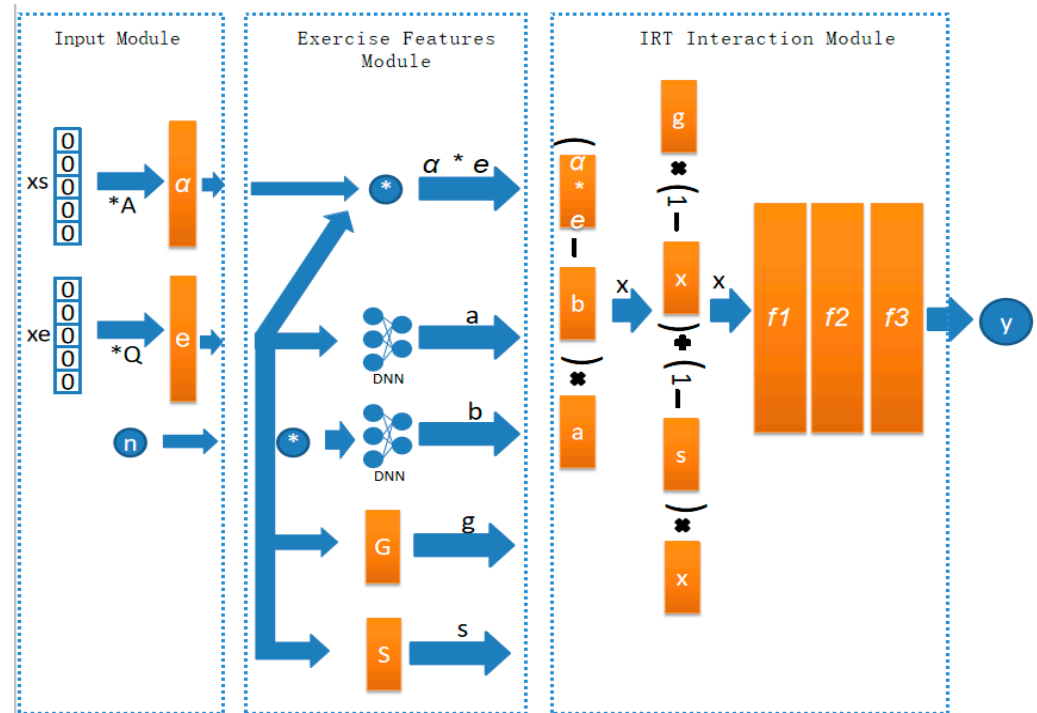


Figure 1. Framework diagram of NeuralNCD model. Where * is the multiplication of vectors.

3.1. Input Module

Assume there are N students, M exercises, and K knowledge ideas in the online education system. For students HS , first initialize the students' mastery of knowledge concepts, and use xs to denote the one-hot vector of students. Then, the student vector α is obtained by Equation (1).

$$\alpha = \text{sigmoid}(xs * A) \quad (1)$$

where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$, $\alpha_i \in [0, 1]$ denotes the student's mastery range of the knowledge concept k_i , $xs \in [0, 1]^{1 \times N}$; $A \in R^{N \times K}$ is a trainable matrix. Moreover, considering that the purpose of the model is to give students an idea of their knowledge level, the student vector is an interpretable vector similar to DINA for student proficiency assessment.

For exercise E , this paper constructs a Q-matrix with M rows and K columns through the expert-labeled exercise–knowledge concept relationship mapping, uses xe to represent the one-hot vector of the exercise, and multiplies xe with the Q-matrix by Equation (2) to obtain the exercise vector e .

$$e = xe * Q \quad (2)$$

where $e = (e_1, e_2, \dots, e_k)$, $e_i = 1$ means that exercise e has the knowledge concept k_i , $e_i = 0$ means that exercise e does not have the knowledge concept k_i , $xe \in [0, 1]^{1 \times M}$, e and α have the same dimensionality.

3.2. Exercise Features Module

The module uses a neural network to train a vector of exercises and obtains some of the features of the exercises (e.g., difficulty, discrimination, guess and slip factor) by computing the data of the exercises.

The discriminative power [30] identifies the degree of students' mastery of knowledge concepts. Since neural networks can automatically learn higher-order nonlinear features, the model in this paper uses a neural network to determine the discriminative power n of the exercises, which is implemented as shown in Equation (3), using DNN_a to train the exercise vector e and normalizing it to satisfy a in the range $[0, 10]$.

$$a = 10 * (\text{sigmoid}(\text{DNN}_a(e))) \quad (3)$$

Exercises in online education systems consist of multiple knowledge concepts, and exercises with the same knowledge concepts may have different difficulty levels depending on the depth and breadth of the concepts examined. In this paper, the initial difficulty of an exercise is determined by the correctness of the exercise, and the difficulty is converged by using a neural network for iteration. The specific implementation is shown in Equations (4) and (5). First, the initial difficulty n of the exercise is obtained by counting the number of times exercise E is answered N_j and the number of times it is answered correctly N_i . Then, it is iterated through the neural network DNN_b . Finally, the difficulty b of exercise E is obtained by multiplying the result of the iteration with the exercise vector e .

$$n = 1 - \frac{N_i}{N_j} \quad (4)$$

$$b = ((\text{sigmoid}(\text{DNN}_b(n * xe)) - 0.5) * e) * 2 \quad (5)$$

where the range of b is $[-1, 1]$ and DNN_a and DNN_b are both single-layer neural networks with the same structure, but the weights of the neural network are not the same. i th is a question with i correct answers and j th is a question with j questions.

Since students will guess the right answer to the exercise and accidentally answer the wrong exercise when they actually answer the questions, failure to model guesses and slips in the model [31] may result in the model not working as expected. Therefore, in order to more accurately predict students' mastery of knowledge concepts and to make the student vector more interpretable, the misleading features of exercise E with a guess factor g and a slip factor s are specified. Unlike the traditional random sampling method, two single-layer neural networks G and S are created in this paper to model the guess and slip factors by Equation (6).

$$g = G(e), s = S(e) \quad (6)$$

where $g \in [0, 1]^{1 \times K}$ and $s \in [0, 1]^{1 \times K}$.

3.3. IRT Interaction Module

The model parameters are represented as vectors in this module's interaction layer, which is based on the 3-parameter logistic model. The model is combined with a neural network and the powerful fitting ability of the neural network is used to deal with the nonlinear relationship between students and exercises. Meanwhile, to improve the speed and accuracy of the neural network and prevent overfitting, a data pre-processing layer is added before each neural network layer to preprocess the input data. Specifically, the module contains one interaction layer, two fully connected layers and one output layer of the neural network, and the data pre-processing layer is added before the activation functions of the fully connected layers and the output layer to normalize the input data. The expression of the interaction layer is given in Equations (7) and (8).

$$x = (1.7 * e * \alpha - b) * a \quad (7)$$

$$x = g * (1 - x) + (1 - s) * x \quad (8)$$

where x represents the interaction vector after students interact with the exercise features, 1.7 is an empirical parameter, $g * (1 - e * \alpha)$ represents the change in student status when students answer the exercise correctly without possessing the corresponding knowledge,

but by guessing, and $(1 - s) * (e * \alpha)$ represents the change in student status when students answer the exercise correctly with the corresponding level of knowledge.

After obtaining the interaction vector x , it is fitted by the fully connected layer and the output layer, as shown in Equations (9)–(11).

$$f1 = \Phi(BN(W_1 * x^T)) \quad (9)$$

$$f2 = \Phi(BN(W_2 * f1)) \quad (10)$$

$$y = \Phi(BN(W_3 * f2)) \quad (11)$$

where W_1, W_2, W_3 are the neural network weight parameters and φ is the activation function *sigmoid*. The *BN* in Equation (9) is the preprocessing of the input data for each layer of the network, and the processing is shown in Equations (12)–(15).

$$\mu_\beta = \frac{1}{m} \sum_{i=1}^m x_i \quad (12)$$

$$\sigma_\beta^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_\beta)^2 \quad (13)$$

$$\hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \theta}} \quad (14)$$

$$y_i = \gamma \hat{x}_i + \beta \quad (15)$$

where x_i is the value of the input data, μ_β is the mean of the data, σ_β^2 is the standard deviation of the data and \hat{x}_i is the value after normalizing the data. y_i is the output value after the reconstruction transformation of the normalized data. γ, β, θ are the learnable reconstruction parameters. Training γ, β helps the network to learn how to recover the distribution of features that the original network learned. However, the θ and γ, β parameters have opposite effects, and training θ helps the neural network to normalize the data better.

In the module, to mitigate the black-box property of the neural network, we use the activation function *sigmoid* to satisfy the monotonicity assumption and enhance the interpretability of the diagnostic results. In the neural network, each element of W_1, W_2, W_3 is restricted to be positive in order for the output of the network to increase monotonically.

3.4. Algorithm Description

The design and implementation of the NeuralNCD algorithm is presented in this section. For ease of understanding, we illustrate the k-means clustering method in Algorithm 1. Algorithm 1 utilizes the 3-parameter logistic model and the iterative computational power of neural networks to maximize the nonlinear relationship between students and exercises to obtain the student's knowledge state.

3.5. Learning Module

The loss function of the model is a cross-entropy function of the true grade r and the predicted grade y . Specifically, y is the student *HS* grade for exercise E predicted by the NeuralNCD model, and r is the true grade for exercise E . Therefore, the loss of the model is obtained from Equation (16).

$$Loss = - \sum_{i=1}^M (r_i \log y_i + (1 - r_i) \log(1 - y_i)) \quad (16)$$

Algorithm 1: NeuralNCD

Input: Dataset $D = \{(\alpha_i, e_i, k_i)\}; i = 0, \dots, N; j = 0, \dots, M; m = 0, \dots, K$; Learning Rate η .

Output: Judgment results y_i .

1: **Step 1.** Create trainable matrix $A, A \in R^{N \times K}, \alpha = \text{sigmoid}(xs * A)$ is received.

2: **Step 2.** Constructing Q-matrices by mapping exercise—knowledge concept relationships, $Q \in R^{M \times N}, e = xe * Q$ is received. // Feature parameter selection.

3: **Step 3.** initialize all connection weights and queues of the network in the (0, 1) range.

4: **Step 4.** $a = 10 * (\text{sigmoid}(\text{DNN}_a(e)))$ is received. // Obtain exercise feature differentiation a

5: **Step 5.** $b = ((\text{sigmoid}(\text{DNN}_b(n * xe)) - 0.5) * e) * 2$ is received. // Get the difficulty b

6: **Step 6.** $g = G(e), s = S(e)$ is received. // Obtain exercise feature guesses g and miss s

7: **Step 7.** Get the optimal network.

8: repeat

9: for all $(\alpha_i, e_i, k_i) \in D$ do

10: Obtain the interaction vector x .

11: Calculate the current sample output: $\bar{y}_i = \Phi(\text{BN}(W_1 * x^T))$

12: Calculate the gradient of the output neuron: $g_i = -\frac{\partial E_i}{\partial \bar{y}} * \frac{\partial \bar{y}}{\partial (w_1 * x^T)}$

13: Calculating the gradient of hidden layer neurons: $e_h = -\frac{\partial E_i}{\partial b_h} * \frac{\partial b_h}{\lambda_h}$

14: Update weights: $\Delta w_{hj} = \eta g_j b_h$

$\Delta v_{ih} = \eta e_h x$

15: $u_1 \leftarrow \tilde{u}_1, u_2 \leftarrow \tilde{u}_2$ // updating the mass center of each cluster

16: end for

17: until $y_i = 1$

18: **Step 8.** The optimal network is used to process the relationship between students and

exercises to obtain the decision result y_i , which is stored in the array L.

20: **Step 9.** Get the results of the recommended exercises and get the students' knowledge status.

4. Performance Analysis**4.1. Dataset Description**

The two real datasets used in the experiments are the public datasets ASSIST2009 and FrcSub. ASSIST2009 is an open dataset collected by the ASSISTment online education platform, and each dataset contains two different versions of the interaction model. The skill-builder version is used in this paper, which provides a log of student responses to exercises and marks each interaction with a knowledge concept number. FrcSub is a small-sample public dataset, widely used in cognitive modeling and consists of test takers' responses (right or wrong) to fraction-subtraction questions.

To make sure that each student has enough information to diagnose themselves, the experiment filtered out students with less than 15 answer records and questions that students did not answer [32]. After filtering, for ASSIST2009, 4163 student responses were obtained for 17,746 exercises, which contained a total of 123 knowledge concepts. For FrcSub, 536 student interactions on 20 questions were obtained, and the exercises contained a total of 8 different knowledge points. The relevant statistical information about the two data sets is shown in Table 1.

Table 1. Information about the dataset.

| Dataset | Students | Exercises | Knowledge Concept |
|------------|----------|-----------|-------------------|
| ASSIST2009 | 4163 | 17746 | 123 |
| FrcSub | 536 | 20 | 8 |

4.2. Evaluation Metrics

To measure whether NeuralNCD, the cognitive diagnostic model proposed in this paper, can more accurately provide students with a diagnosis that matches their own level of performance, the following steps were taken. In this paper, we use the prediction accuracy (ACC), root means square error (RMSE) and area under the curve (AUC) to evaluate the effectiveness of NeuralNCD and other cognitive diagnostic models in diagnosing students'

status [33]. The smaller the value of RMSE, which is used to assess the discrepancy between predicted and true values, the better. However, the bigger the value of ACC and AUC, the better it reflects the actual situation of the error between the expected and true values. Specifically, the prediction accuracy (ACC), root mean square error (RMSE) and area under the curve (AUC) are defined as shown in Equations (17)–(19).

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (17)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - y_i)^2} \quad (18)$$

$$AUC = \frac{\sum_{insi \in positiveclass} rankinsi - \frac{M*(M+1)}{2}}{M * N} \quad (19)$$

where TP stands for the number of properly predicted positive instances, FP for mistakenly forecasted negative cases, TN for correctly predicted negative cases, FN for wrongly predicted positive cases, and M, N for the number of positive and negative samples. The summation $= \sum_{i=1}^M r_i$ in Equation (19), r_i is the ranking position of the i th positive class sample, and its computational complexity is $O(n \log(n))$ [34].

4.3. Baselines

The NeuralNCD is a typical cognitive diagnostic model with the ultimate goal of obtaining the student's knowledge state, but it is difficult to assess the performance of the model because the real knowledge state of the student is not available. As a result, the model's success is determined indirectly by anticipating the results of the recommended exercises. In order to test the model's efficacy, the NeuralNCD is compared with the following experimental methods in this paper:

1. IRT [8]: IRT is a typical continuum type of cognitive diagnostic model. It diagnoses the test parameters, as well as the students' potential ability, by jointly modeling the test questions and the students based on their answers.
2. MIRT [19]: the MIRT involves modeling of the multiple competencies required for a recommendation, the relationship between item characteristics and recommendation performance. The model can more accurately respond to the complexity of the interaction between students and exercises.
3. DINA [9]: the DINA is one of the discrete cognitive diagnostic models, which describes the student as a multidimensional knowledge mastery vector and diagnoses students' knowledge states from the student's existing answer records. Its framework is simple and the interpretability of the results is good.
4. PMF [35]: probability matrix decomposition is a traditional recommendation method, and the model is used for exercise recommendation by relating students and exercises to users and goods, and students' answer records correspond to users and goods. The performance of the exercise recommendation is obtained by decomposing the student's answer records to obtain the low-dimensional potential vectors of students and exercises.
5. NeuralCDM [14]: the NeuralCDM obtains the difficulty and differentiation characteristics of the exercises by using the textual information of the exercises and the labeled knowledge concepts, and then combines the IRT and the neural network to diagnose the students.

4.4. Experimental Setup

This paper uses Python to implement the algorithms in this paper. A 7:1:2 division of the training set, validation set and test set is used for each data set. All data are cross-

validated with a 5-fold validation, where the validation set is utilized to determine the best model parameters.

In both datasets, the fully connected layer widths in Equations (9)–(11) are set to 256, 128, and 1, respectively, and the input data are preprocessed using Equations (12)–(15) in this paper. The Adam optimization algorithm was used to train the model. The rate of learning η is 0.002. The Mini-batch size of the dataset has been set to 32.

4.5. Experimental Results

This experiment aimed to analyze the following:

1. Comparison experiments on whether there is an improvement in the effectiveness of exercise recommendation compared to other traditional cognitive diagnostic models.
2. Can the performance of the model be improved by multiple exercise features?
3. The effect of optimization of neural networks on model performance.
4. Is the recommendation of exercises for students a true reflection of their knowledge through case studies?

4.5.1. Comparison Experiments with Other Cognitive Diagnostic Models

In this paper, the proposed model and the comparison algorithm are compared on the two datasets and the results are shown in Table 2, which shows that the model outperforms all the comparison algorithms on the ASSIST2009 and FrcSub datasets for all three evaluation metrics.

Table 2. Evaluation results of the model on ASSIST2009 and FrcSub datasets.

| Model | ASSIST2009 | | | FrcSub | | |
|-----------|------------|-------|-------|--------|-------|-------|
| | ACC | REMS | AUC | ACC | REMS | AUC |
| IRT | 0.664 | 0.459 | 0.684 | 0.802 | 0.385 | 0.867 |
| MIRT | 0.695 | 0.451 | 0.709 | 0.819 | 0.378 | 0.885 |
| PMF | 0.671 | 0.478 | 0.729 | 0.810 | 0.386 | 0.877 |
| DINA | 0.649 | 0.467 | 0.675 | 0.825 | 0.373 | 0.890 |
| NeuralCDM | 0.720 | 0.436 | 0.749 | 0.824 | 0.375 | 0.888 |
| NeuralNCD | 0.734 | 0.425 | 0.776 | 0.836 | 0.364 | 0.905 |

Among them, in the ASSIST 2009 dataset, the NeuralNCD model showed a large improvement (5–10%) in individual evaluation metrics compared to the CDMS without neural networks (IRT, MIRT, PMF, DINA), indicating that the neural networks used in this paper can better capture the relationships between students compared to traditional cognitive diagnostic models. There is also a 2.5% improvement in the evaluation metric AUC, compared to the latest NeuralCDM model using neural networks, verifying the effectiveness of the model in modeling guess and slip features and the ability of the data preprocessing mechanism to improve the neural network's ability to capture student and exercise relationships. In addition, the NeuralNCD has a relatively low AUC metric improvement relative to each comparison algorithm on the FrcSub dataset, e.g., only 1.1% improvement compared to DINA, and 13.3% improvement on the ASSIST2009 dataset. This may be due to the following two reasons: (1) the sample size of the FrcSub dataset is small; (2) the variety of exercises on the FrcSub dataset is small, with a fixed number of 20 exercises. In addition, all models performed better on the FrcSub dataset than on the ASSIST2009 dataset, which may be due to the fact that the exercises on the FrcSub dataset have fewer knowledge points, only eight, and students can learn the knowledge points more easily.

Although the performance of the model proposed in this paper on the ASSIST2009 dataset demonstrated a relatively small improvement compared to that on the FrcSub dataset, the overall results are all better than other models for the following three main reasons: (1) the neural network used in this paper can better capture the relationship between students and exercises compared to traditional cognitive diagnostic models that only use interaction func-

tions (e.g., logistic functions, inner products). (2) Integrating multiple exercise features into the cognitive diagnostic model can obtain more accurate and reasonable diagnostic results. (3) The data preprocessing and monotonicity assumptions added in this paper enhance the performance of the neural network and improve the accuracy and interpretability of the model diagnosis results.

4.5.2. Comparative Experiments with Multiple Exercise Features

In this paper, we compare the model performance of Neu-p1, Neu-p2 and NeuralNCD on two datasets, ASSIST2009 and FrcSub, and the comparison results are shown in Figure 2.

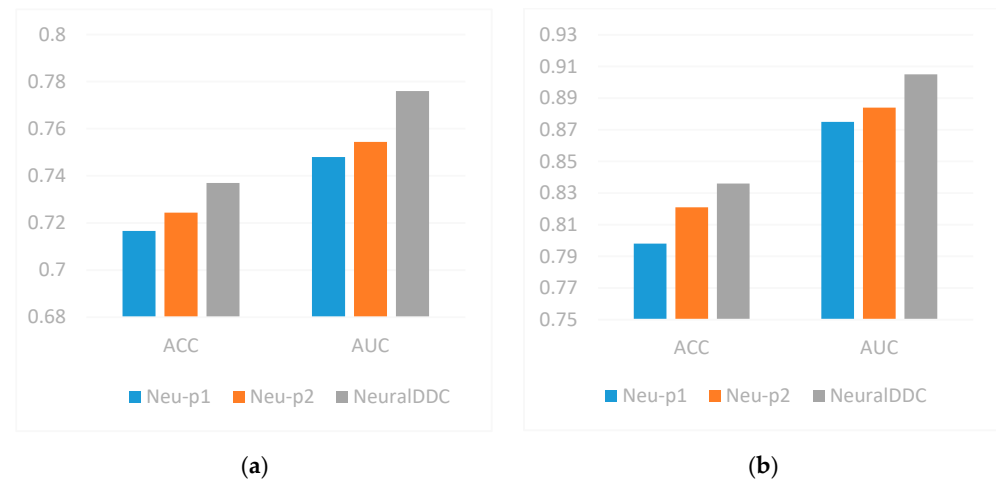


Figure 2. Model performance comparison of Neu-p1, Neu-p2 and NeuralNCD on datasets (a) ASSIST2009, (b) FrcSub.

To investigate how different exercise characteristics affected model performance, the interaction layers on the NeuralNCD are replaced with interaction functions with different exercise feature parameters and named Neu-p1 (single-feature neurocognitive diagnostic model) and Neu-p2 (two-feature neurocognitive diagnostic model) in this paper. The results in Figure 2 show that the data from the NeuralNCD are optimal for the evaluation metrics of ACC and AUC on both datasets, and the results of ACC and AUC on the ASSIST2009 dataset are improved by 1.6% and 2.4% on average. The results of ACC and AUC on the FrcSub dataset were improved by 2.6% and 2.5% on average. Thus, the experimental results mentioned above indicate that the model proposed in this paper that uses exercise answer records with multiple exercise features has a positive impact on diagnosing student status and serves to improve the accuracy of the model.

4.5.3. Influence of Neural Network Optimization on the Model

It is known from previous work [11] that using neural networks instead of the interaction functions (e.g., logistic functions, inner products) used in previous cognitive diagnostic models can be helpful in diagnosing student states. As a result, this study goes on to investigate whether strengthening the neural network's functions (by incorporating data preprocessing and monotonicity requirements) may enhance the model's performance in this part. In this paper, the data preprocessing mechanism of the NeuralNCD is removed and the weights of the fully connected layers are randomly valued. It is named NeuralNCDa (neurocognitive diagnostic model with missing neural network optimization), and it is compared with the NeuralNCD on two data sets.

As can be observed in Figure 3, the AUC evaluation metrics of the NeuralNCD with data preprocessing and monotonicity assumptions are significantly better than those of the NeuralNCDa with missing neural network optimization. This indicates that the data preprocessing and monotonicity assumptions are important for obtaining interpretable diagnostic results. It is also evident that the convergence speed of the NeuralNCD is

substantially faster, indicating that the optimized neural network is able to handle the relationship between students and exercises more effectively.

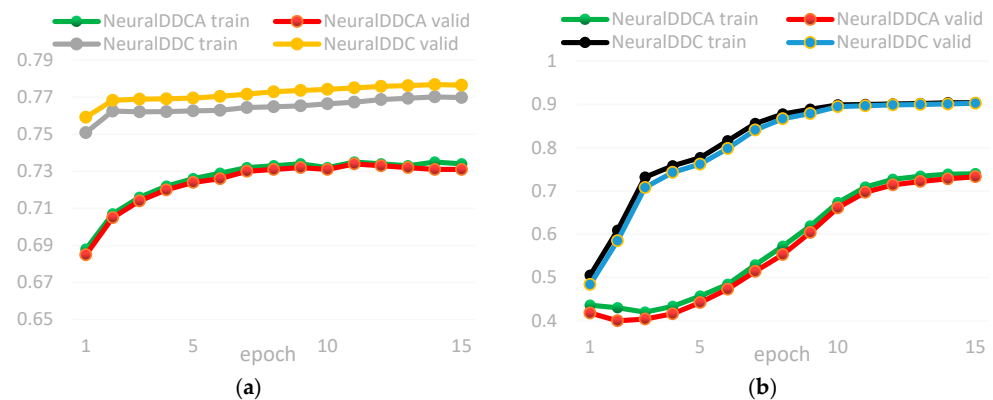


Figure 3. AUC scores for NeuralNCD and NeuralDDCA training and validation on datasets (a) ASSIST2009, (b) FrcSub.

4.5.4. Case Study

In this paper, we show the diagnosis of students' knowledge level by a cognitive diagnostic model through a practical case study. Table 3 shows the student's prediction of the exercises by IRT and NeuralNCD. Exercise 1 has three knowledge points, K4, K6, and K7, and Exercise 2 and Exercise 3 have two knowledge points, K4 and K6. For example, the difficulty of the third problem is 0.533, and the students' mastery of the relevant knowledge points is 0.686 and 0.897, which is higher than the required difficulty. Figure 4 shows the level of knowledge diagnosed by the student HS through IRT and NeuralNCD. Since IRT [8] only diagnoses students' ability and no knowledge concepts are subdivided, the diagnostic results of IRT are positive polygons. Therefore, the NeuralNCD proposed in this paper can diagnose students' cognition by analyzing the exercise answer records with multiple exercise characteristics, and the diagnostic results have good accuracy and interpretability.

Table 3. Case study table.

| Q | K | Real | Pre | | Difficulty b | |
|---|------------|------|-----|-----------|--------------|-----------|
| | | | IRT | NeuralNCD | IRT | NeuralNCD |
| 1 | k4, k6, k7 | 1 | 0 | 1 | 0.753 | 0.633 |
| 2 | k6, k7 | 0 | 0 | 0 | 0.801 | 0.851 |
| 3 | k6, k7 | 1 | 1 | 1 | 0.465 | 0.533 |



Figure 4. Case study diagram.

5. Conclusions and the Future Work

To address the problems of traditional cognitive diagnostic models in terms of insufficient consideration of multiple exercise features and difficulties in handling the nonlinear interaction between students and exercises, this paper proposes a neural network cognitive diagnosis model based on multi-dimensional features (NeuralNCD). First, the model uses a neural network to extract features of the exercises (e.g., difficulty, discrimination, guess and slip factor) for modeling the student learning process. Second, the model uses the neural network to model the nonlinear interaction between the student and the exercises, allowing the model to diagnose the student's state more accurately. Finally, combining data preprocessing and monotonicity assumption mechanisms in the neural network not only improves the accuracy and interpretability of the diagnostic results, but also effectively increases the convergence speed of the model. The experimental results show that the NeuralNCD performs better in both real data sets ASSIST2009 and FrcSub compared with other classical models, which validates the effectiveness of the multiple exercise features (e.g., difficulty, discrimination, guess and slip factor) and the neural network with data preprocessing and monotonicity assumption mechanisms was proposed in this paper for diagnosing students' knowledge states.

The NeuralNCD still has the following shortcomings: (1) since the model uses deep learning methods, it requires a large amount of data to train the model, and the practical application may not guarantee a large enough data size for training the model. (2) The model is based on a 3-parameter logistic model, and although the influence of several exercise features such as difficulty, discrimination, guess and slip coefficients on diagnosing students' knowledge level is considered, there may still be features that are not considered for real application scenarios. Therefore, to address these shortcomings, the model needs to be further enhanced in future work.

Author Contributions: Conceptualization, G.L. and Y.H.; methodology, G.L. and Y.H.; validation, J.S., T.Y. and Y.Z.; writing—original draft preparation, G.L. and Y.H.; writing—review and editing, G.L., Y.H., J.S., T.Y., Y.Z., S.D. and N.X.; supervision, S.D. and N.X.; funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Key Project of Education Science Planning in Jiangxi Province (Grant No. 19ZD024).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data supporting the results of this study can be obtained from the corresponding corresponding author(s) upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Burns, H.; Luckhardt, C.A.; Parlett, J.W.; Redfield, C.L. *Intelligent Tutoring Systems: Evolutions in Design*; Psychology Press: London, UK, 2014.
2. Anderson, A.; Huttenlocher, D.; Kleinberg, J.; Leskovec, J. Engaging with massive online courses. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 687–698.
3. Yang, P.; Xiong, N.; Ren, J. Data security and privacy protection for cloud storage: A survey. *IEEE Access* **2020**, *8*, 131723–131740. [[CrossRef](#)]
4. Hengyu, L.; Tiancheng, Z.; Peiwen, W.; Ge, Y. A review of knowledge tracking. *J. East China Norm. Univ. Nat. Sci.* **2019**, *2019*, 1–15.
5. Wang, C.; Liu, Q.; Chen, E.H.; Huang, Z.Y. The rapid calculation method of DINA model for large scale cognitive diagnosis. *Acta Electronica Sin.* **2018**, *46*, 1047.
6. DiBello, L.V.; Roussos, L.A.; Stout, W. 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handb. Stat.* **2006**, *26*, 979–1030.
7. Liu, Q. Towards a New Generation of Cognitive Diagnosis. In Proceedings of the 30th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 19–26 August 2021; pp. 4961–4962.
8. Janssen, R.; Tuerlinckx, F.; Meulders, M.; De Boeck, P. A hierarchical IRT model for criterion-referenced measurement. *J. Educ. Behav. Stat.* **2000**, *25*, 285–306. [[CrossRef](#)]

9. De La Torre, J. DINA model and parameter estimation: A didactic. *J. Educ. Behav. Stat.* **2009**, *34*, 115–130. [\[CrossRef\]](#)
10. Embretson, S.E.; Reise, S.P. *Item Response Theory*; Psychology Press: London, UK, 2013.
11. Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, *42*, 30–37. [\[CrossRef\]](#)
12. Sciuto, G.L.; Susi, G.; Cammarata, G.; Capizzi, G. A spiking neural network-based model for anaerobic digestion process. In Proceedings of the 2016 International Symposium on Power Electronics, Electrical Drives, Automation and Motion, Capri, Italy, 22–24 June 2016; pp. 996–1003.
13. Jiang, Y.; Tong, G.; Yin, H.; Xiong, N. A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. *IEEE Access* **2019**, *7*, 118310–118321. [\[CrossRef\]](#)
14. Wang, F.; Liu, Q.; Chen, E.; Huang, Z.; Chen, Y.; Yin, Y.; Wang, S. Neural cognitive diagnosis for intelligent education systems. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 6153–6161.
15. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 448–456.
16. Wan, S.; Niu, Z. A hybrid e-learning recommendation approach based on learners' influence propagation. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 827–840. [\[CrossRef\]](#)
17. Bernacki, M.L.; Chavez, M.M.; Uesbeck, P.M. Predicting achievement and providing support before STEM majors begin to fail. *Comput. Educ.* **2020**, *158*, 103999. [\[CrossRef\]](#)
18. Jiang, P.; Wang, X. Preference cognitive diagnosis for student performance prediction. *IEEE Access* **2020**, *8*, 219775–219787. [\[CrossRef\]](#)
19. Zhan, P.; Jiao, H.; Man, K.; Wang, L. Using JAGS for Bayesian Cognitive Diagnosis Modeling: A Tutorial. *J. Educ. Behav. Stat.* **2019**, *44*, 473–503. [\[CrossRef\]](#)
20. Reckase, M.D. Multidimensional item response theory models. In *Multidimensional Item Response Theory*; Springer: New York, NY, USA, 2009; pp. 79–112.
21. Liu, Q.; Wu, R.; Chen, E.; Xu, G.; Su, Y.; Chen, Z.; Hu, G. Fuzzy cognitive diagnosis for modelling examinee performance. *ACM Trans. Intell. Syst. Technol. TIST* **2018**, *9*, 1–26. [\[CrossRef\]](#)
22. De La Torre, J.; Douglas, J.A. Higher-order latent trait models for cognitive diagnosis. *Psychometrika* **2004**, *69*, 333–353. [\[CrossRef\]](#)
23. Ma, W.; de la Torre, J. GDINA: An R package for cognitive diagnosis modeling. *J. Stat. Softw.* **2020**, *93*, 1–26. [\[CrossRef\]](#)
24. Wang, D.; Cai, Y.; Tu, D. Q-matrix estimation methods for cognitive diagnosis models: Based on partial known Q-matrix. *Multivar. Behav. Res.* **2020**, 1–13. [\[CrossRef\]](#)
25. Dong, L.; Ling, Z.; Ling, Q.; Lai, Z. Cognitive Diagnosis with Explicit Student Vector Estimation and Unsupervised Question Matrix Learning. *arXiv* **2022**, arXiv:2203.03722.
26. Zhan, P.; Man, K.; Wind, S.A.; Malone, J. Cognitive Diagnosis Modeling Incorporating Response Times and Fixation Counts: Providing Comprehensive Feedback and Accurate Diagnosis. *J. Educ. Behav. Stat.* **2022**. [\[CrossRef\]](#)
27. Li, H.; Liu, J.; Wu, K.; Yang, Z.; Liu, R.W.; Xiong, N. Spatio-temporal vessel trajectory clustering based on data mapping and density. *IEEE Access* **2018**, *6*, 58939–58954. [\[CrossRef\]](#)
28. Huang, S.; Liu, A.; Zhang, S.; Wang, T.; Xiong, N.N. BD-VTE: A novel baseline data based verifiable trust evaluation scheme for smart network systems. *IEEE Trans. Netw. Sci. Eng.* **2020**, *8*, 2087–2105. [\[CrossRef\]](#)
29. Lu, Y.; Wu, S.; Fang, Z.; Xiong, N.; Yoon, S.; Park, D.S. Exploring finger vein based personal authentication for secure IoT. *Future Gener. Comput. Syst.* **2017**, *77*, 149–160. [\[CrossRef\]](#)
30. Cheng, S.; Liu, Q.; Chen, E.; Huang, Z.; Huang, Z.; Chen, Y.; Hu, G. DIRT: Deep learning enhanced item response theory for cognitive diagnosis. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2397–2400.
31. Cheng, S.; Liu, Q.; Chen, E. Domain adaption for knowledge tracing. *arXiv* **2020**, arXiv:2001.04841.
32. Wu, M.; Tan, L.; Xiong, N. A structure fidelity approach for big data collection in wireless sensor networks. *Sensors* **2014**, *15*, 248–273. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Wu, S.; Flach, P. *A Scored AUC Metric for Classifier Evaluation and Selection*; Second Workshop on ROC analysis in ML: Bonn, Germany, 2005; pp. 247–262.
34. Hand, D.J.; Till, R.J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **2001**, *45*, 171–186. [\[CrossRef\]](#)
35. Mnih, A.; Salakhutdinov, R.R. Probabilistic matrix factorization. In Proceedings of the NIPS'07: 20th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 12 December 2007; Volume 20.