Knowledge-Sensed Cognitive Diagnosis for Intelligent Education Platforms

Haiping Ma
Information Materials and Intelligent Sensing Laboratory of Anhui Province, Institutes of Physical Science and Information Technology, Anhui University
Hefei, China
hpma@ahu.edu.cn

Manwei Li
Information Materials and Intelligent Sensing Laboratory of Anhui Province, Institutes of Physical Science and Information Technology, Anhui University
Hefei, China
lmanwei@163.com

Le Wu
Hefei University of Technology Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
Hefei, China
lewu.ustc@gmail.com

Haifeng Zhang
The Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Mathematical Science, Anhui University
Hefei, China
haifengzhang1978@gmail.com

Yunbo Cao
Tencent Cloud Xiaowei
Beijing, China
yunbocao@tencent.com

Xingyi Zhang*
The Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Artificial Intelligence, Anhui University
Hefei, China
xyzhanghust@gmail.com

Xuemin Zhao
Tencent Cloud Xiaowei
Chengdu, China
xueminzhao@tencent.com

ABSTRACT

Cognitive diagnosis is a fundamental issue of intelligent education platforms, whose goal is to reveal the mastery of students on knowledge concepts. Recently, certain efforts have been made to improve the diagnosis precision, by designing deep neural networks-based diagnostic functions or incorporating more rich context features to enhance the representation of students and exercises. However, how to interpretably infer the student's mastery over non-interactive knowledge concepts (i.e., knowledge concepts not related to his/her exercising records) still remains challenging, especially when not giving relations between knowledge concepts. To this end, we propose a Knowledge-Sensed Cognitive Diagnosis (KSCD) framework, aiming at learning intrinsic relations among knowledge concepts from student response logs and incorporating them for inferring students' mastery over all knowledge concepts in an end-to-end manner. Specifically, we firstly project students, exercises and knowledge concepts into embedding representation matrices, where the intrinsic relations among knowledge concepts are reflected in the knowledge embedding representation matrix. Then, the knowledge-sensed student knowledge mastery vector and exercise factor vectors are obtained by the multiply product of their embedding representations and the knowledge embedding representation matrix, which make the student's mastery of non-interactive knowledge concepts be interpretably inferred. Finally, we can utilize classical student-exercise interaction functions to predict student's exercising performance and jointly train the model. In additional, we also design a new function to better model the student-exercise interactions. Extensive experimental results on two real-world datasets clearly show the significant performance gain of our KSCD framework, especially in predicting students' mastery over non-interactive knowledge concepts, by comparing to state-of-the-art cognitive diagnosis models (CDMs).

CCS CONCEPTS

• **Information systems** → *Data mining*.

KEYWORDS

Cognitive diagnosis, Intelligent education, Relation learning among knowledge concepts

ACM Reference Format:
Haiping Ma, Manwei Li, Le Wu, Haifeng Zhang, Yunbo Cao, Xingyi Zhang, and Xuemin Zhao. 2022. Knowledge-Sensed Cognitive Diagnosis for Intelligent Education Platforms. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3511808.3557372

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557372

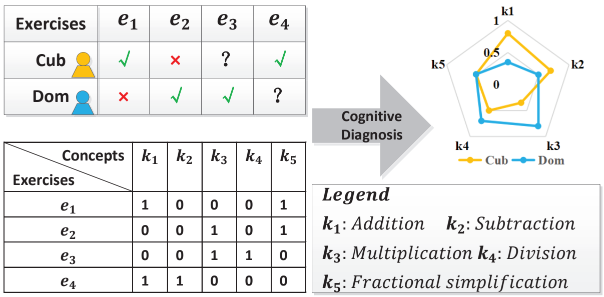


Figure 1: Illustration of cognitive diagnosis. The left part presents the students' response logs and the exercise-concept relational matrix named Q-matrix, while the upper right part shows the corresponding diagnosis results.

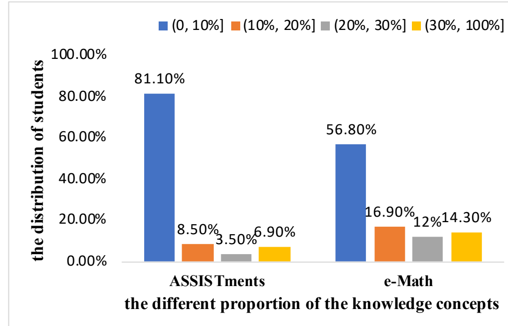


Figure 2: The distribution of students with different proportion interactive knowledge concepts in two real-world datasets. We can see that more than 80% of students interact with less than 10% knowledge concepts in the ASSISTments, and more than 55% students interact with less than 10% knowledge concepts in the e-Math.

1 INTRODUCTION

The aim of cognitive diagnosis in the field of intelligent education is to diagnose students' knowledge mastery (i.e., students' mastery of knowledge concepts) based on their historical records of answering exercises and the exercise-concept relational matrix called Q-matrix [1, 4, 23, 30]. Taking Figure 1 as an example, students *Cub* and *Dom* have practiced a series of exercises (i.e., $\{e_1, e_2, e_4\}$ and $\{e_1, e_2, e_3\}$), respectively, and got responses (e.g., right or wrong), which is shown at the top left; the Q-matrix is shown at the bottom left. Through the process of cognitive diagnosis, their knowledge mastery is obtained, which is shown in the right of the figure. Indeed, such diagnosis results can benefit a wide range of intelligent educational applications, such as helping teachers make decisions regarding remedial instruction [18, 28, 35] or targeted interventions [2, 21].

With the development of a large number of intelligent education platforms, cognitive diagnosis has become one fundamental technology for enhancing the core competitiveness of the platforms and received a great deal of attention [15, 21]. The classic cognitive diagnosis models (CDMs) in the field of educational psychology [5, 19], such as Item Response Theory (IRT) [12], Deterministic Input, Noisy 'And' gate (DINA) [9], rely on manually designed simple diagnostic functions, which may not be sufficient for modeling the complex interaction relationships between students and exercises in the intelligent education platforms. Therefore, considerable efforts have been undertaken to improve the precision of cognitive diagnosis by the researchers of artificial intelligence and data science domains [16, 17, 33, 34].

These efforts mainly focus on designing neural network-based diagnostic functions to better fit the complex student-exercise interactions [31], or attempts to incorporate more rich context features and prior relations between knowledge concepts for enhancing the representation learning of students and exercises [13, 32, 36]. However, how to interpretably infer the student's mastery over non-interactive knowledge concepts still remains challenging, especially when not giving the prior relations between knowledge concepts (e.g., prerequisite relation and similarity relation). The non-interactive knowledge concepts of a student refers to these knowledge concepts not related to his/her exercising records (e.g., k_4 is the non-interactive knowledge concept of the student *cub* in Figure 1). Therefore, these existing CDMs have major limitations in most of the intelligent education platforms due to the following two reasons. On the one hand, most of the intelligent education platforms are usually with highly sparse student-concept interactions as shown in Figure 2, which indicates that the vast majority of knowledge concepts are not interacted by a student. On the other hand, it is difficult for most developing intelligent education platforms to obtain this prior relations between knowledge concepts since annotating them is labor-intensive and costly.

To this end, we propose a Knowledge-Sensed Cognitive Diagnosis (KSCD) framework, whose idea is that the intrinsic knowledge relations can be learned from response logs and Q-matrix (e.g., the similarity of knowledge correctness in response logs can reflect the similarity of knowledge difficulty), and benefit the performance of cognitive diagnosis. To be specific, we first encode students, knowledge concepts and exercises into the same hidden space. Then, we obtain the student knowledge mastery vector and exercise factor vectors in the knowledge space by the multiply product of their embedding representation vectors and knowledge embedding representation matrix. Finally, we can use previous CDMs (such as neural cognitive diagnosis model (NCD) [31]) to model the interactions between the student and exercise parameters in the knowledge space for predicting student exercising performance. In this paper, we also design a new student-exercise interaction function to better capture the information among students, exercises and knowledge concepts. Through model training, the student's mastery levels over interactive and non-interactive knowledge concepts can be diagnosed and interpretably inferred. Moreover, the intrinsic knowledge relations can be learned, which is reflected in the embedding representation matrix of knowledge concepts. Extensive experiments in real-world datasets demonstrate the effectiveness of the proposed KSCD.

2 RELATED WORK

2.1 Cognitive Diagnosis

In the past decades, many CDMs have been proposed by researches of educational psychology domain. IRT [12] and DINA [9] are the two most fundamental but classic CDMs. IRT portrays the student and exercise with the unidimensional and continuous latent traits (i.e., the student latent ability, exercise difficulty, and exercise discrimination), and predicts the probability of a student answering an exercise correctly via the logistic function or cumulative distribution function of the normal distribution. By extending the student latent trait and exercise parameters into multidimensional space, Reckase et al. proposed Multidimensional IRT (MIRT) [27]. DINA [9] represents the student's knowledge mastery with a binary vector indicating whether he/she has mastered each knowledge concept. In DINA, one of the exercise parameters comes directly from Q-matrix, which is used to guarantee the interpretation of the diagnosis results, while the dot product is used to predict student exercising performance.

In recent years, with the guidance of the educational psychology theories, CDMs based on neural networks have been explored to further improve both precision and interpretability. These works can be divided into the following three categories. The first category focuses on designing neural network-based diagnostic functions, so as to better fit the complex student-exercise interactions [31]. For example, NCD [31] models the interaction between students and exercises with one shadow layer inspired by MIRT models and two deep full connection layers. The second category focuses on utilizing more rich context features for enhancing the representation learning of students and exercises [8, 36]. For example, educational context-aware cognitive diagnosis (ECD) [36] focuses on modeling the effect of a student's educational contexts (e.g., parents' education, school resource, personal interest) on the cognitive state. These models of the two above groups cannot diagnose the student's mastery of knowledge concepts that are not correlated to her exercising records. The third category attempts to address this issue by incorporating prior relations between knowledge concepts [13, 32]. For example, relation map driven cognitive diagnosis (RCD) [13] models the student-exercise interactions and prior structural relations among knowledge concepts via a multi-layer relation map. However, the annotation of the relations is labor-intensive and costly [10], thus it is still difficult for most intelligent education platforms to obtain this information. Inspired by these facts, this paper attempts to learn the intrinsic knowledge relations from response logs and Q-matrix and incorporates them for enhancing the performance of cognitive through the multiply product. The multiply product can make it easy to explain students' mastery of non-interactive knowledge concepts based on the intrinsic knowledge relations. Moreover, the proposed model supports the fusion of prior knowledge structure in the embedding learning stage.

2.2 Knowledge-Sensed Student Modeling

Recently, several studies [11, 25, 26] have demonstrated that there usually exist interdependencies among knowledge concepts. The relations of knowledge concepts are fairly helpful for many educational tasks [7]. For example, in knowledge tracing, Chen et al. [7] incorporated the prior prerequisite relation among knowledge concepts into the knowledge tracing model by considering this property as a constraint of the model. Tong et al. [29] used prior knowledge concepts relations (e.g., prerequisite, similarity, etc.) to model the propagation of influence between concepts (e.g., unidirectional, bidirectional, etc.). Nakagawa et al. [22] formulated the knowledge tracking task as a time-series node-level classification problem by incorporating the prior knowledge structure as a graph, and then solved it by graph neural networks. These models mentioned above mainly study the time-aware knowledge relational effect for student modeling. Obviously, they cannot be directly applied for cognitive diagnosis. Moreover, these works are mainly based on prior knowledge concepts relations obtained by manual annotation, which are not easy to obtain in practice.

3 PROBLEM STATEMENT

For the cognitive diagnosis task in an intelligent education platform, there are three sets of entities: a set of students $\mathcal{S} = \{s_1, s_2, ..., s_N\}$, a set of exercises $\mathcal{E} = \{e_1, e_2, ..., e_M\}$, the associated knowledge concepts $\mathcal{K} = \{k_1, k_2, ..., k_C\}$. The exercise-knowledge relations are usually provided by domain experts and denoted as Q-matrix $\mathbf{Q} \in \mathbb{R}^{M\times C}$, where $\mathbf{Q}_j$ is a C-dimensional binary vector that indicates which knowledge concepts are correlated to the exercise e_j. Let $\mathcal{R}$ denote students' exercising response logs, which are represented by a set of triplet (s_i, e_j, r_{ij}), where $s_i \in \mathcal{S}$, $e_j \in \mathcal{E}$, and $r_{ij} \in \{0, 1\}$ represents the response score of student s_i got on exercise e_j (0 represents fault and 1 represents true).

Problem Definition. *Given student entity $\mathcal{S}$, exercise entity $\mathcal{E}$, and knowledge concept entity $\mathcal{K}$, the students' exercising response logs $\mathcal{R}$ and the exercise-knowledge relational matrix $\mathbf{Q}$, let the student' knowledge mastery vector $\boldsymbol{\alpha}_i$ be modeled into the knowledge space, that is, the c-th element of $\boldsymbol{\alpha}_i$ denotes the mastery of student s_i on knowledge concept k_c. Our goal is to obtain the value of $\boldsymbol{\alpha}_i$ through the student's exercising performance prediction $P(r_{ij} = 1) = f(\boldsymbol{\alpha}_i, \boldsymbol{\Phi}_j)$, where $\boldsymbol{\Phi}_j$ denotes the parameter set of exercise e_j (e.g., the knowledge difficulty $\boldsymbol{\beta}_j$, discrimination η_j).*

In the process of modeling the student's knowledge mastery vector $\boldsymbol{\alpha}_i$ and the exercise parameters $\boldsymbol{\Phi}_j$, how to learn the intrinsic relations among $\mathcal{K}$ and interpretably incorporate such information to model these parameters in the knowledge space in an end-to-end manner is the key purpose of this paper. Next, we introduce the parameter modeling methods of some representative CDMs, which are closely related to our work. Note that here we do not discuss the related work that relies on prior relations between knowledge concepts, annotated by domain experts.

Existing Representative Parameter Modeling Methods. The representative traditional method IRT [12] models the process of predicting the probability that student s_i correctly answering exercise e_j as follows:

$$P(r_{ij} = 1) = f_{IRT}((\alpha_i - \beta_j) \times \eta_j), \tag{1}$$

where the student's ability α_i and the exercise parameters including exercise's knowledge difficulty β_j and exercise's discrimination η_j are directly defined as unidimensional and continuous variables.

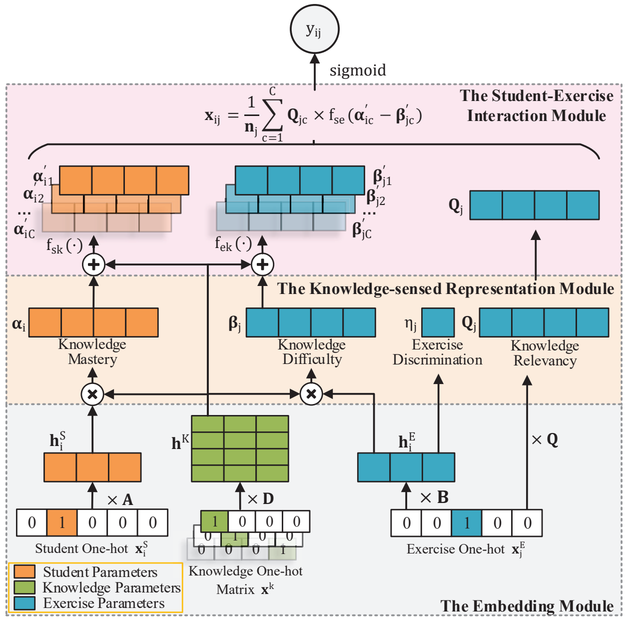


Figure 3: The KSCD framework. h_i^S is the embedding representation vector of student s_i, h^K is the knowledge representation matrix, and h_j^E is the embedding representation vector of exercise e_j. In the student-exercise interaction module, C is the number of knowledge concepts and n_j denotes the number of knowledge concepts contained in exercise e_j.

The representative neural network based method NCD [31] models the student's exercising performance as:

$$P(r_{ij} = 1) = f_{NCD}(\mathbf{Q}_j \circ (\boldsymbol{\alpha}_i - \boldsymbol{\beta}_j) \times \eta_j), \tag{2}$$

where the student's knowledge mastery vector $\boldsymbol{\alpha}_i$ is directly modeled by multiplying the student's one-hot representation vector with a trainable matrix. Similarly, the exercise's knowledge difficulty $\boldsymbol{\beta}_j$ and discrimination η_j are directly obtained by multiplying the exercise's one-hot representation vector with a trainable matrix and a trainable vector, respectively. Here, each dimension of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ can correspond to a specific knowledge concept with the help of the vector $\mathbf{Q}_j$ through element-wise product (i.e., "$\circ$").

In short, these parameter modeling methods directly project students and exercises into knowledge mastery vector and exercise factor vectors, and still seldom consider the intrinsic relations between concepts for learning the representations of these vectors. Therefore, they cannot diagnose the student's mastery on concepts that are not correlated to his/her exercising records. To address this issue, we propose a Knowledge-Sensed Cognitive Diagnosis (KSCD) framework, which will be presented in the next section.

4 THE PROPOSED KSCD FRAMEWORK

Figure 3 illustrates the framework of KSCD, which consists of three modules: (1) the embedding module; (2) the knowledge-sensed representation module; and (3) the student-exercise interaction module. Specifically, in the embedding module, we map the student s_i, exercise e_j, and knowledge concept k_c to a unified hidden space, respectively, to obtain the embedding representations $\mathbf{h}_i^S$, $\mathbf{h}_j^E$ and $\mathbf{h}^K$ for student s_i, exercise e_j, and knowledge concept set $\mathcal{K}$. In the knowledge-sensed representation module, we obtain the knowledge mastery vector $\boldsymbol{\alpha}_i$ and knowledge difficulty vector $\boldsymbol{\beta}_j$ for student s_i and exercise e_j by multiplying knowledge concept representation matrix $\mathbf{h}^K$ with student's embedding representation $\mathbf{h}_i^S$ and exercise's embedding representation $\mathbf{h}_j^E$, respectively. The exercise discrimination variable η_j is acquired from exercise's embedding representation $\mathbf{h}_j^E$ through non-linear transformation. The above two modules are the purpose of the proposed framework. Finally, with the obtained interpretable parameters of students and exercises, the student-exercise interaction module is designed with a specially designed function to predict the student's exercising performance. Actually, under the KSCD framework, these existing student-exercise interaction functions can also be used to define the last module.

4.1 The Embedding Module

Firstly, we encode the students, exercises, and knowledge concepts into a d-dimensional hidden space and obtain their initialized embedding representation vectors, by multiplying their one-hot representation vectors with the trainable matrices respectively, as shown in Eq. (3).

$$\mathbf{h}_i^S = \mathbf{x}_i^S \times \mathbf{A}, \mathbf{h}_j^E = \mathbf{x}_j^E \times \mathbf{B}, \mathbf{h}_c^K = \mathbf{x}_c^K \times \mathbf{D}, \tag{3}$$

where $\mathbf{h}_i^S$, $\mathbf{h}_j^E$, $\mathbf{h}_c^K \in \mathbb{R}^{1\times d}$ denote the initialized embedding representations of student s_i, exercise e_j and knowledge concept k_c. $\mathbf{x}_i^S$, $\mathbf{x}_j^E$, $\mathbf{x}_c^K$ denote the one-hot representation vectors of student s_i, exercise e_j, and concept k_c respectively. $\mathbf{A} \in \mathbb{R}^{N\times d}$, $\mathbf{B} \in \mathbb{R}^{M\times d}$ and $\mathbf{D} \in \mathbb{R}^{C\times d}$ are three trainable matrices (N, M and C are the number of students, exercises and knowledge concepts respectively).

Each knowledge concept is represented by a d-dimensional vector, so we can obtain the representation matrix $\mathbf{h}^K \in \mathbb{R}^{C\times d}$ about knowledge concept set $\mathcal{K}$.

4.2 The Knowledge-Sensed Representation Module

The main purpose of this module is to obtain the interpretable student and exercise parameters (i.e., the student's knowledge mastery vector $\boldsymbol{\alpha}_i$, exercise's knowledge difficulty vector $\boldsymbol{\beta}_j$, and exercise's discrimination η_j) based on the representations obtained from the embedding module. These parameters are universally needed by classical student-exercise interaction functions for student's exercising performance prediction.

The knowledge mastery vector $\boldsymbol{\alpha}_i$ of student s_i is obtained by multiplying the student's embedding vector $\mathbf{h}_i^S$ and the knowledge embedding matrix $\mathbf{h}^K$. The multiply product can make the diagnosis results of different knowledge concepts explainable by the intrinsic knowledge relations. We formulate it as:

$$\boldsymbol{\alpha}_i = \mathbf{h}_i^S \times (\mathbf{h}^K)^T, \tag{4}$$

where $\boldsymbol{\alpha}_i \in \mathbb{R}^{1\times C}$. Similarly, the knowledge difficulty vector $\boldsymbol{\beta}_j \in \mathbb{R}^{1\times C}$ of exercise e_j is obtained by multiplying the exercise embedding representation vector $\mathbf{h}_j^E$ and the knowledge embedding representation matrix $\mathbf{h}^K$.

$$\boldsymbol{\beta}_j = \mathbf{h}_j^E \times (\mathbf{h}^K)^T. \tag{5}$$

The unidimensional discrimination η_j of exercise e_j is obtained from the exercise embedding vector $\mathbf{h}_j^E$ through non-linear transformation, as shown in Eq. (6). Here, ϕ denotes the activation function.

$$\eta_j = \phi(\mathbf{W} \times \mathbf{h}_j^E + \mathbf{b}). \tag{6}$$

Besides, to ensure each dimension of the student's knowledge mastery $\boldsymbol{\alpha}_i$ and the exercise's knowledge difficulty $\boldsymbol{\beta}_j$ corresponding to the specific concept, the knowledge relevance vector $\mathbf{Q}_j \in \mathbb{R}^{1\times C}$ of exercise e_j is also needed. $\mathbf{Q}_j$ is calculated by multiplying the exercise one-hot representation vector $\mathbf{x}_j^E$ with matrix $\mathbf{Q}$ as:

$$\mathbf{Q}_j = \mathbf{x}_j^E \times \mathbf{Q}. \tag{7}$$

4.3 The Student-Exercise Interaction Module

With the obtained interpretable student and exercise parameters (i.e, the knowledge mastery vector $\boldsymbol{\alpha}_i$ of student s_i, knowledge difficulty $\boldsymbol{\beta}_j$ and discrimination η_j of exercise e_j, and Q_j), we can use classical student-exercise interaction functions to predict the student's exercising performance.

Here, we can resemble NCD because it is a representative neural network approach to define the complex interactions between students and exercises. Specifically, with the obtained parameters, the KSCD firstly adopts the MIRT-like function as the first layer of student-exercise interaction, which is shown in Eq. (8). Then, several fully connected layers satisfying the monotonicity assumption of student knowledge mastery and exercising response are used to predict the exercising performance, as shown in Eq. (9).

$$\mathbf{x}_{ij} = \mathbf{Q}_j \circ (\boldsymbol{\alpha}_i - \boldsymbol{\beta}_j) \times \eta_j, \tag{8}$$

$$y_{ij} = \phi(\mathbf{W}_3 \times \phi(\mathbf{W}_2 \times \phi(\mathbf{W}_1 \times \mathbf{x}_{ij}^T + b_1) + b_2) + b_3), \tag{9}$$

where ϕ denotes the activation function. W_1, W_2, W_3 are the weights of different neural network layers, respectively.

In this paper, we also design a new student-exercise interaction function as shown in Eq. (10), which can further incorporate the intrinsic relations among knowledge concepts. Specifically, we splice the student's knowledge mastery vector $\boldsymbol{\alpha}_i$ with the knowledge concept representation $\mathbf{h}_c^K$, respectively, and then pass it through the neural network layer to obtain the higher-order student's knowledge mastery vector $\boldsymbol{\alpha}'_{ic}$ of each knowledge concept k_c. A similar operation is adopted for the exercise parameters to obtain higher-order knowledge difficulty $\boldsymbol{\beta}'_{jc}$ of each knowledge concept k_c. A utility function $f_{se}(\boldsymbol{\alpha}'_{ic} - \boldsymbol{\beta}'_{jc})$ is then used to measure the student's ability advantage over exercise's knowledge difficulty on each knowledge concept k_c, and the student's ability advantage on exercise e_j is selectively accumulated through the knowledge relevance vector Q_j of the exercise e_j. The specific formulas are defined as follows:

$$y_{ij} = \phi(\frac{1}{n_j}\sum_{c=1}^{C}\mathbf{Q}_{jc} \times f_{se}(\boldsymbol{\alpha}'_{ic} - \boldsymbol{\beta}'_{jc})), \tag{10}$$

$$\boldsymbol{\alpha}'_{ic} = \phi(f_{sk}(\boldsymbol{\alpha}_i \oplus \mathbf{h}_c^K)), \tag{11}$$

$$\boldsymbol{\beta}'_{jc} = \phi(f_{ek}(\boldsymbol{\beta}_j \oplus \mathbf{h}_c^K)), \tag{12}$$

where n_j denotes the number of knowledge concepts contained in exercise e_j; ϕ denotes the activation function, and here we adopt *sigmoid* and $\oplus$ is the splice operation; $y_{ij} \in (0,1)$ is the predicted probability that student s_i correctly answers exercise e_j; f_{se}, f_{sk}, f_{ek} are linear transformation functions that represent different fully connection neural layers.

Table 1: Dataset statistics.

Statistics	JunYi	e-Math
# Students	1000	517
# Exercises	712	1582
# Knowledge Concepts	39	61
# Response logs	203,945	62,412
# Avg logs per student	203.94	120.71
# Avg concepts per exercise	1.00	1.21
# Avg logs per Knowledge Concept	8.09	6.87

4.4 Model Optimization

The final knowledge mastery vector $\boldsymbol{\alpha}_i$ of each student s_i can be obtained through the student's exercising performance prediction task. Loss function is defined with the cross entropy between the output y_{ij} of the student-exercise interaction module and true exercising response r_{ij} as follows:

$$loss = -\sum_{(s_i,e_j,r_{ij})\in\mathcal{R}}(r_{ij}log y_{ij}) + (1-r_{ij})log(1-y_{ij})). \tag{13}$$

5 EXPERIMENTS

Our experiments are designed to address the following research questions:

- **RQ1:** How does our framework perform when comparing with state-of-the-art CDMs?
- **RQ2:** How does our framework perform in predicting students' mastery over non-interactive knowledge concepts?
- **RQ3:** Whether the intrinsic relations among knowledge concepts learned by our model are reliable?
- **RQ4:** Whether our model is sensitive to hyperparameters?
- **RQ5:** Whether the students' mastery over non-interactive knowledge concepts predicted by our model is reasonable?

5.1 Experimental Settings

5.1.1 Dataset Description. To verify the superiority of the proposed KSCD framework, we conduct experiments on two real-world datasets: JunYi [6] and e-Math. JunYi is a public dataset collected by the JunYi Education Platform in Taiwan, containing nearly $20M$ responses from $1,000$ students. E-Math is a private dataset collected by a well-known electronic educational product, mainly containing math exercises and test logs of primary and secondary school students. We filter out students with less than 15 response logs for all datasets to guarantee that there are enough data for modeling each student. After preprocessing, for the JunYi dataset, we obtain $1,000$ students, $203,945$ response logs, and 39 knowledge concepts. Similarly in the e-Math dataset, we obtain 517 students, $62,412$ response logs, and 61 knowledge concepts. To demonstrate

Table 2: Overall experimental results, where 80%/20%, 70%/30%, 60%/40% and 50%/50% denote different train/test split ratios. The line named *Imp.(%)* provides the relative performance improvement of the KSCD model compared to the best baseline model. The best results are highlighted in bold and the best results of the baselines are underlined. It is noted that both our KSCD and KSCD_NCD significantly outperform the best baseline with p <0.01.

(a) JunYi

Methods	ACC				RMSE				AUC			
	50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%
DINA	0.6872	0.6892	0.6912	0.7021	0.4687	0.4717	0.4751	0.4612	0.7130	0.7181	0.7185	0.7364
MIRT	0.7409	0.7397	0.7434	0.7337	0.4211	0.4218	0.4192	0.4330	0.7741	0.7712	0.7752	0.7639
NCD	0.7508	0.7502	0.7543	0.7510	0.4122	0.4113	0.4091	0.4102	0.7890	0.7884	0.7903	0.7951
CDGK	0.7502	0.7522	0.7557	0.7601	0.4123	0.4107	0.4109	0.4059	0.7890	0.7903	0.7928	0.7983
ECD	0.7527	0.7512	0.7554	0.7591	0.4115	0.4119	0.4102	0.4065	0.7891	0.7887	0.7904	0.7909
KSCD_NCD	0.7649	0.7625	0.7693	0.7722	0.4039	0.4031	0.4001	0.3967	0.8065	0.8059	0.8124	0.8158
KSCD	**0.7728**	**0.7715**	**0.7758**	**0.7783**	**0.3951**	**0.3964**	**0.3929**	**0.3912**	**0.8178**	**0.8152**	**0.8191**	**0.8223**
Imp.(%)	2.670%	2.566%	2.660%	2.394%	3.985%	3.482%	4.217%	3.622%	3.637%	3.151%	3.317%	3.006%

(b) e_Math

Methods	ACC				RMSE				AUC			
	50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%
DINA	0.6508	0.6551	0.6429	0.6547	0.4921	0.4827	0.4819	0.4857	0.6881	0.6914	0.6930	0.6853
MIRT	0.6602	0.6635	0.6634	0.6613	0.4837	0.4792	0.4788	0.4770	0.6956	0.7021	0.7030	0.7018
NCD	0.6908	0.6903	0.6943	0.6975	0.4449	0.4434	0.4483	0.4486	0.7316	0.7422	0.7425	0.7440
CDGK	0.6844	0.6946	0.6906	0.6992	0.4457	0.4484	0.4473	0.4443	0.7289	0.7422	0.7438	0.7443
ECD	0.6910	0.6905	0.6953	0.6981	0.4442	0.4420	0.4424	0.4416	0.7317	0.7415	0.7418	0.7431
KSCD_NCD	0.7049	0.7038	0.7013	0.7085	0.4397	0.4380	0.4372	0.4345	0.7456	0.7534	0.7537	0.7579
KSCD	**0.7073**	**0.7098**	**0.7099**	**0.7135**	**0.4331**	**0.4314**	**0.4312**	**0.4298**	**0.7609**	**0.7647**	**0.7677**	**0.7685**
Imp.(%)	2.359%	2.188%	2.100%	2.045%	2.499%	2.398%	2.532%	2.672%	3.991%	3.032%	3.213%	3.251%

the effectiveness of our framework for different data sparsity, we apply 80%/20%, 70%/30%, 60%/40% and 50%/50% training/test splits for each student's response logs in all datasets, respectively.

The basic statistics of the datasets after preprocessed are summarized in Table 1. The "# Avg logs per Knowledge Concept" represents the average amount of logs that each student interacted with each concept, and the formula is expressed as shown in Eq. (14).

$$AVG_{log} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{C} log(i,j)}{\sum_{i=1}^{N}\sum_{j=1}^{C} F_J(log(i,j) > 0)}, \tag{14}$$

where $log(i, j)$ denotes the number of exercises containing knowledge concept k_j answered by student s_i; N, C are the number of students and knowledge concepts, respectively; $F_J(\cdot)$ is the indicator function, where $F_J(\cdot) = 1$ if $log(i, j) > 0$, otherwise $F_J(\cdot) = 0$.

5.1.2 Baseline Models and Evaluation Metrics. For our KSCD framework, we provide two implementations (named KSCD and KSCD_NCD) with different student-exercise interaction functions. Specifically, KSCD is with the specially designed interaction function, while KSCD_NCD is with the same interaction function as the NCD. We compare the two implementations with several baselines to verify the effectiveness of KSCD, including DINA [9], MIRT [27], NCD [31] CDGK [32] and ECD [36]. The details are described below:

- **DINA** as the classical CDM, utilizes a binary variable to characterize whether students have mastered specific concepts.
- **MIRT** is an extension of the classical IRT model, which uses multidimensional vectors to characterize the parameters of students and exercises, using linear functions for interaction.
- **NCD** attempts to use neural networks to automatically learn the complex interaction between students and exercises.
- **CDGK** is an artificial neural network-based CDM, which adds a guess adjustment layer to take into account guessing factors and performs the aggregation of concepts into CDM.
- **ECD** aggregates the cognitive states of students reflected in the educational environment and students' historical response records, so as to achieve diagnostic enhancement. Here, we characterize the cognitive state reflected by the educational contexts using a random initialization vector due to the absence of educational contexts.

Considering that there is no true knowledge mastery of students, we indirectly evaluate the effectiveness of our model by using the knowledge mastery vector obtained to predict the student's exercising performance according to recent literature [20, 31]. Corresponding, the ACC (Accuracy) [31], RMSE (root mean square error) [24], and AUC (area under the curve) [3] are used as metrics to assess the performance of CDMs in predicting students' exercising performance.

5.1.3 Parameters Setting. We initialize the network parameters with Xavier initialization [14]. Each parameter is sampled from *N(0,* μ^2*)*, where $\mu = -\sqrt{(2/(n_{in} + n_{out}))}$. n_{in}, n_{out} denote the number of neurons input and output, respectively. The relevant parameters are set as follows through hyperparameter experimental analysis.

Table 3: Experimental results over different categories of test sets. The line named *Imp.(%)* provides the relative performance gain of the KSCD model compared to the best baseline model. The best results are highlighted in bold and the best results of the baselines are underlined.

(a) JunYi

Methods		ACC				RMSE				AUC			
		50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%
Warm	DINA	0.6775	0.6762	0.6833	0.6891	0.4589	0.4617	0.4581	0.4542	0.7435	0.7414	0.7473	0.7540
	MIRT	0.7455	0.7430	0.7457	0.7354	0.4180	0.4196	0.4176	0.4317	0.7705	0.7672	0.7702	0.7605
	NCD	0.7560	0.7545	0.7574	0.7529	0.4076	0.4084	0.4070	0.4087	0.7860	0.7856	0.7901	0.7921
	CDGK	0.7562	0.7566	0.7594	0.7632	0.4084	0.4071	0.4085	0.4041	0.7861	0.7868	0.7897	0.7952
	ECD	0.7583	0.7558	0.7586	0.7616	0.4072	0.4090	0.4069	0.4048	0.7863	0.7858	0.7903	0.7959
	KSCD_NCD	0.7688	0.7653	0.7713	0.7736	0.4008	0.4009	0.3985	0.3955	0.8031	0.8025	0.8078	0.8124
	KSCD	**0.7767**	**0.7735**	**0.7775**	**0.7798**	**0.3923**	**0.3944**	**0.3916**	**0.3902**	**0.8142**	**0.8116**	**0.8153**	**0.8189**
	Imp.(%)	2.426%	2.234%	2.383%	2.175%	3.659%	3.120%	3.760%	3.440%	3.548%	3.152%	3.163%	2.890%
Cold	DINA	0.5228	0.5330	0.5697	0.5731	0.5212	0.5531	0.5403	0.5340	0.6152	0.6104	0.6194	0.6272
	MIRT	0.6783	0.6841	0.6909	0.6883	0.4603	0.4591	0.4515	0.4679	0.7485	0.7492	0.7485	0.7451
	NCD	0.6788	0.6778	0.6880	0.6997	0.4577	0.4577	0.4521	0.4474	0.7403	0.7322	0.7437	0.7443
	CDGK	0.6720	0.6835	0.6744	0.6954	0.4614	0.4570	0.4598	0.4498	0.7388	0.7368	0.7417	0.7512
	ECD	0.6759	0.6725	0.6854	0.6937	0.4598	0.4595	0.4542	0.4493	0.7409	0.7320	0.7436	0.7519
	KSCD_NCD	0.7125	0.7142	0.7250	0.7371	0.4435	0.4398	0.4340	0.4283	0.7786	0.7754	0.7907	0.7969
	KSCD	**0.7190**	**0.7208**	**0.7402**	**0.7386**	**0.4307**	**0.4284**	**0.4186**	**0.4187**	**0.7948**	**0.7959**	**0.8125**	**0.8097**
	Imp.(%)	5.922%	5.365%	7.136%	5.560%	5.899%	6.258%	7.287%	6.415%	6.186%	6.233%	8.550%	7.687%

(b) e-Math

Methods		ACC				RMSE				AUC			
		50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%	50%/50%	60%/40%	70%/30%	80%/20%
Warm	DINA	0.6003	0.6037	0.6076	0.6126	0.5197	0.5241	0.5223	0.5192	0.6321	0.6278	0.6297	0.6381
	MIRT	0.6588	0.6613	0.6615	0.6594	0.4844	0.4803	0.4804	0.4783	0.6942	0.7002	0.6998	0.6991
	NCD	0.6864	0.6960	0.6955	0.6976	0.4443	0.4485	0.4482	0.4384	0.7323	0.7448	0.7442	0.7456
	CDGK	0.6920	0.6906	0.6914	0.7013	0.4437	0.4484	0.4480	0.4405	0.7329	0.7405	0.7456	0.7543
	ECD	0.6832	0.6961	0.6965	0.6989	0.4448	0.4414	0.4470	0.4415	0.7288	0.7450	0.7437	0.7439
	KSCD_NCD	0.7029	0.7032	0.7035	0.7115	0.4356	0.4345	0.4377	0.4303	0.7471	0.7534	0.7539	0.7637
	KSCD	**0.7080**	**0.7093**	**0.7086**	**0.7125**	**0.4331**	**0.4320**	**0.4321**	**0.4301**	**0.7609**	**0.7643**	**0.7656**	**0.7671**
	Imp.(%)	2.312%	1.896%	1.737%	1.597%	2.630%	2.130%	3.333%	2.361%	3.820%	2.591%	2.682%	1.697%
Cold	DINA	0.5061	0.5118	0.5147	0.4983	0.5521	0.5482	0.5551	0.5482	0.6125	0.6117	0.6073	0.6129
	MIRT	0.6717	0.6873	0.6891	0.6892	0.4785	0.4662	0.4581	0.4581	0.7056	0.7225	0.7422	0.7384
	NCD	0.6809	0.6897	0.6959	0.6960	0.4493	0.4436	0.4426	0.4419	0.7084	0.7278	0.7342	0.7400
	CDGK	0.6701	0.6988	0.6834	0.7045	0.4519	0.4407	0.4401	0.4331	0.7039	0.7293	0.7529	0.7533
	ECD	0.6828	0.6934	0.6929	0.7041	0.4492	0.4429	0.4439	0.4405	0.7088	0.7280	0.7333	0.7403
	KSCD_NCD	0.6982	0.7133	0.7124	0.7254	0.4393	0.4308	0.4299	0.4228	0.7391	0.7540	0.7718	0.7797
	KSCD	**0.7052**	**0.7187**	**0.7267**	**0.7336**	**0.4311**	**0.4233**	**0.4182**	**0.4189**	**0.7623**	**0.7723**	**0.7962**	**0.7884**
	Imp.(%)	3.281%	2.848%	4.426%	4.131%	4.029%	3.948%	4.976%	3.279%	7.548%	5.896%	5.751%	4.659%

We uniformly set the embedding dimension of the first embedding layer to 20-dimension (i.e., $d = 20$ in Eq. (3)). The dimensions of the full connection layers (Eq. (9)) in the student-exercise interaction module are 512, 256, 1 respectively, and all layers use *sigmoid* as the activation function. We use the *Adam* optimizer to train all models with a batch size of 32 and the learning rate of 0.002. Finally, the experimental results on all models are obtained by performing standard 5-fold cross-validation. All models are implemented with *Pytorch*, and all experiments are conducted with *Tesla V*100.

5.2 Experimental Results

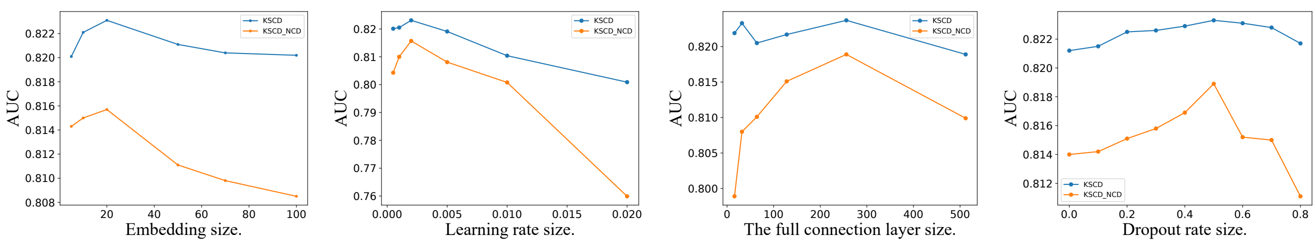
5.2.1 The Comparison Results between our models and Baselines. To verify the effectiveness of the proposed KSCD framework, Table 2 shows the performance of KSCD, KSCD_NCD and all baselines in predicting students' future exercising responses on all datasets with several train/test split ratios. Here, we report the average value of five evaluation runs. The best results are highlighted in bold and the best baselines are underlined. The last line shows the relative improvement of KSCD compared with the best baseline models.

From Table 2, we can obtain the following observations. (1) The performance gains of our proposed models (i.e., KSCD and KSCD_NCD) on different train/test ratios for all datasets are significantly outperforms all baselines, which demonstrates the effectiveness of our proposed framework and answers **RQ1**. (2) KSCD outperforms KSCD_NCD on all datasets, which confirms the validity of the specially designed student-exercise interaction function.

To further demonstrate the superiority of our model in predicting students' mastery on non-interactive concepts, we divide each test record (s_i, e_i, r_{ij}) in the test set into three categories (i.e., cold test set, normal test set and warm test set) according to the percentage of non-interactive knowledge concepts in e_j. Here, the non-interactive knowledge concepts in e_j means they are not correlated to the train records of student s_i. Specifically, let $\mathbf{K}^{s_i}$ and $\mathbf{K}^{e_j}$ be the associated concept set of student s_i from his/her training data

Table 4: The top ten relevant and irrelevant knowledge pairs based on the learned knowledge embedding representation matrix. The knowledge pairs whose relevance is consistent with the ground truth are highlighted in bold.

Relevant		Irrelevant	
Knowledge Concept	Knowledge Concept	Knowledge Concept	Knowledge Concept
Application of two-digit add and sub	**Application of three-digit add and sub**	**Recognize perimeter**	**Estimate distance**
Addition of two-digit	**Subtraction of three-digit**	**Addition of two-digit**	**Recognize fractions**
Calculate perimeter	**Application of calculating perimeter**	**Application of multi-digit multiplication**	**Distance problem**
Recognize perimeter	**Calculate perimeter**	**Conversion of mass units**	**Recognize seconds**
Recognize mass units	**Application of mass units**	**Conversion of time units**	**Calculation of add and sub**
Calculation within one hundred	**Calculation of two-digit**	**Multiplication of two-digit**	**Recognize rectangles**
Calculation of add and sub	**Subtraction of three-digit**	Recognize kilometers	Distance problem
Recognize fractions	Multiply with zero	**Conversion of mass units**	**Application of three-digit add and sub**
Addition of three-digit (rounded)	**Addition of three-digit (non-rounded)**	**Recognize mass units**	**Multiply with zero**
Addition of two-digit	**Subtraction of two-digit**	Subtraction of three-digit	Multiplication of three-digit



(a) Embedding size. (b) Learning rate size. (c) The full connection layer size. (d) Dropout rate size

Figure 4: Impact of each hyperparameter on the performance of the KSCD_NCD and KSCD in terms of AUC.

and the associated concept set of exercise e_j, respectively. A test record (s_i, e_j, r_{ij}) is divided into cold test set if $|\bigcap(\mathbf{K}^{e_j}, \mathbf{K}^{s_i})| = 0$, is divided into warm test set when $|\bigcap(\mathbf{K}^{e_j}, \mathbf{K}^{s_i})| = |\mathbf{K}^{e_j}|$. Otherwise it is put into normal test set. Here, $\bigcap$ is the intersection operation between sets and $|.|$ means the size of the set.

Table 3 shows the experimental results of all methods on warm test sets and cold test sets, where the train/test ratio is 80%/20%, 70%/30%, 60%/40% and 50%/50%. The experimental results for the normal test set are not given here since it is too small (less than 200) in both datasets. Then, the relative performance improvement of the KSCD model compared to the best baseline model, i.e., Imp.(%), is given in the last line of the corresponding category. From this table, we can obtain the following observations. (1) Our KSCD and KSCD_NCD perform the best and the second-best in both warm test sets and cold test sets, which demonstrates the effectiveness and robustness of our KSCD framework. (2) The relative performance gain of KSCD is more significant in the cold test set than that in the warm test set. Taking the JunYi dataset as an example, when train/test ratio is 70%/30%, our KSCD improves over the best baseline on the ACC, RMSE, and AUC by 7.136%, 7.287%, and 8.550% in the cold test set, respectively, while the improvements are 2.383%, 3.760%, and 3.163% in the warm test set. These observations indicate the superiority of our models in predicting students' mastery of non-interactive concepts and answer **RQ2**.

5.2.2 Evaluation of Learned Intrinsic Knowledge Relations. This experiment is based on the e-Math dataset that contains a true relation between knowledge concepts (i.e., prerequisite relation between knowledge concepts). We conduct this analysis experiment as follows. Firstly, we calculate the Pearson correlation coefficient between knowledge concepts based on the learned knowledge embedding matrix (i.e., the matrix $\mathbf{h}^K$) by KSCD model. Then, we obtain the top relevant and irrelevant knowledge concept pairs according to the Pearson correlation coefficient. The results for the top ten relevant and irrelevant knowledge pairs are shown in the "relevant" column and "irrelevant" column of Table 4 respectively. In the "relevant" column, we bold the knowledge pairs if they have true first- or second-order neighborhood relationships, while we bold the knowledge pairs if they do not have relationships in the "irrelevant" column. From Table 4, we find that more than 80% of the top ten relevant knowledge pairs have true first- or second-order neighborhood relations, and 80% of the top ten irrelevant knowledge pairs have no relation. This experiment validates the higher accuracy of the learned knowledge relations and answers **RQ3**.

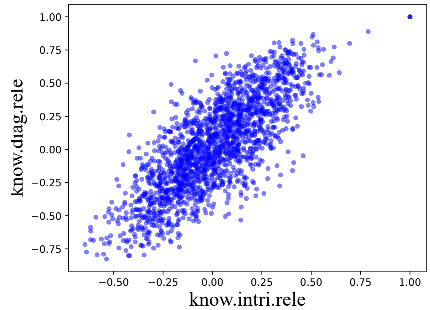
5.2.3 Sensitivity Analysis of Hyperparameters. For our models, we conduct the parameter sensitivity analysis for the embedding dimension, size of the full connection layer, dropout rate and learning rate, which are as: $\{5, 10, 20, 50, 70, 100\}$, $\{16, 32, 64, 128, 256, 512\}$, $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ and $\{0.0005, 0.001, 0.002, 0.005, 0.01, 0.02\}$. Figure 4(a), (b), (c) and (d) show the impacts of embedding dimension, learning rate size, the full connection layer size and dropout rate size on the KSCD_NCD and KSCD models in predicting student performance tasks, respectively. It is worth reminding that there are the same trends in terms of the other two metrics (i.e., ACC and RMSE), although we only provide the AUC values in Figure 4. From the results, we can obtain the following observations. Firstly, the selection of learning rate size has a more obvious impact on the model when comparing with other hyperparameters, such as dropout rate size. Second, compared with KSCD_NCD, the KSCD

Concept	Concept	Relevance
C	A	0.718
C	B	0.716
F	D	0.748
F	E	0.723

(a)

Concept	Response Logs	Mastery
A	✓✓✓✗✓✓	0.732
B	✓✗✓✓✓✗✓✓✓✓	0.819
C		0.719
D	✗✓✗✗	0.219
E	✗	0.268
F		0.380

(b)



(c)

Figure 5: (a) Knowledge relevance calculated from the learned knowledge representation matrix. (b) The visualization of student diagnosis report. (c) Relation between know. intri. rele. (i.e., knowledge intrinsic relevance) and know. diag. rele. (i.e., knowledge diagnosis relevance).

with a specially designed student-exercise interaction function is more stable and insensitive to hyperparameters (**RQ4**).

5.2.4 Rationality Analysis of Diagnosis Results. Here, we analyze the rationality of diagnosis results by KSCD through a case study (**RQ5**). The case study can make it convenient for teachers or students to understand how to use the output of the model. Specifically, we randomly select one student in the e-Math dataset and visualize his/her mastery of part knowledge concepts in Figure 5(b). Each line contains the knowledge concept, the exercising records on the knowledge concept, and the mastery of the corresponding knowledge concept, diagnosed by our model. Figure 5(a) represents the relevance between knowledge pairs calculated based on the learned knowledge concept representation. From Figure 5(b), we can obtain the following observations: (1) There are higher mastery values of the student on knowledge concepts Ⓐ and Ⓑ, whereas lower mastery values over knowledge concepts Ⓓ and Ⓔ, which is consistent with the true exercising records. (2) Even though the student has no exercising records on knowledge concepts Ⓒ and Ⓕ, the model infers that the student masters well on knowledge concept Ⓒ whereas worse over knowledge concept Ⓕ, which can be explained with the learned intrinsic relations among concepts shown in Figure 5(a). That is, the student has a better mastery on concept Ⓒ, since Ⓒ is highly relevant to concepts Ⓐ and Ⓑ. The two observations can prove that diagnosis results over both interactive and non-interactive concepts are intuitively reasonable.

To further verify that the diagnosis results of knowledge concepts can be explained by intrinsic knowledge relations, we analyze the relations between knowledge intrinsic relevance and knowledge diagnosis relevance. Specifically, we first define the Pearson coefficient of concepts according to their embedding representations as knowledge intrinsic relevance and the Pearson coefficient of concepts according to their diagnosis results of all students as knowledge diagnosis relevance. Figure 5(c) shows the experimental result, where each blue point represents a pair of concepts, and its values in the x-axis and y-axis denote intrinsic relevance and diagnosis relevance respectively. We can find that the diagnosis result of concepts can be explained by the intrinsic relations since they have a consistent positive correlation.

6 CONCLUSION

In this paper, we proposed a Knowledge-Sensed Cognitive Diagnosis (KSCD) framework, where the intrinsic relations of knowledge concepts can be learned and benefit the performance of cognitive diagnosis in an end-to-end manner. Specifically, we first leveraged neural networks to encode the embedding representations of students, exercises, and knowledge concepts. Then, we designed a knowledge-sensed representation module to map the students and exercises into the knowledge space through the multiply product of their embedding vectors and the knowledge embedding matrix. The multiply product can make that the diagnosis results of different concepts can be explained by the intrinsic knowledge relations reflected in the knowledge embedding representation matrix. Finally, a specially designed student-exercise interaction function is used to predict the student's exercising performance. Experimental results clearly demonstrate the superiority of our proposed KSCD framework.

7 ACKNOWLEDGMENTS

This research was partially supported by grants from the National Key Research and Development Project (NO. 2018AAA0100105), the National Natural Science Foundation of China (NO. 62107001), the Anhui Provincial Natural Science Foundation (NO. 2108085QF272), the Key Program of Natural Science Project of Educational Commission of Anhui Province (NO. KJ2020A0036) and the CCF-Tencent Open Fund.

REFERENCES

[1] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. Engaging with massive online courses. In *The 23rd International Conference on World Wide Web (WWW)*. 687–698.

[2] Joseph Beck. 2007. Difficulties in inferring student knowledge from observations (and why you should care). In *Educational Data Mining: Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education*. 21–30.

[3] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159.

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

[5] Jaime R Carbonell. 1970. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems* 11, 4 (1970), 190–202.

[6] Haw-Shiuan Chang, Hwai-Jung Hsu, and Kuan-Ta Chen. 2015. Modeling Exercise Relationships in E-Learning: A Unified Approach. In *International Educational Data Mining Society (EDM)*. 532–535.

[7] Penghe Chen, Yu Lu, Vincent W Zheng, and Yang Pian. 2018. Prerequisite-Driven Deep Knowledge Tracing. In *2018 IEEE International Conference on Data Mining (ICDM)*. 39–48.

[8] Song Cheng, Qi Liu, Enhong Chen, Zai Huang, Zhenya Huang, Yiying Chen, Haiping Ma, and Guoping Hu. 2019. DIRT: Deep Learning Enhanced Item Response Theory for Cognitive Diagnosis. In *The 28th ACM International Conference on Information and Knowledge Management (CIKM)*. 2397–2400.

[9] Jimmy De La Torre. 2009. DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics* 34, 1 (2009), 115–130.

[10] Louis V DiBello, Louis A Roussos, and William Stout. 2006. 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of statistics* 26 (2006), 979–1030.

[11] Henry C Ellis. 1965. The transfer of learning. (1965).

[12] Susan E Embretson and Steven P Reise. 2013. *Item response theory*. Psychology Press.

[13] Weibo Gao, Qi Liu, Zhenya Huang, Yu Yin, Haoyang Bi, Mu-Chun Wang, Jianhui Ma, Shijin Wang, and Yu Su. 2021. RCD: Relation Map Driven Cognitive Diagnosis for Intelligent Education Systems. In *The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 501–510.

[14] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 249–256.

[15] Xuan Guo, Rui Li, Qi Yu, and Anne R Haake. 2017. Modeling Physicians' Utterances to Explore Diagnostic Decision-making.. In *International Joint Conferences on Artificial Intelligence Organization (IJCAI)*. 3700–3706.

[16] Zhenya Huang, Qi Liu, Enhong Chen, Hongke Zhao, Mingyong Gao, Si Wei, Yu Su, and Guoping Hu. 2017. Question Difficulty Prediction for READING Problems in Standard Tests. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[17] Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su, and Guoping Hu. 2019. Ekt: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering* 33, 1 (2019), 100–115.

[18] Qi Liu, Shiwei Tong, Chuanren Liu, Hongke Zhao, Enhong Chen, Haiping Ma, and Shijin Wang. 2019. Exploiting cognitive structure for adaptive learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 627–635.

[19] Qi Liu, Runze Wu, Enhong Chen, Guandong Xu, Yu Su, Zhigang Chen, and Guoping Hu. 2018. Fuzzy cognitive diagnosis for modelling examinee performance. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 4 (2018), 1–26.

[20] Qi Liu, Runze Wu, Enhong Chen, Guandong Xu, Yu Su, Zhigang Chen, and Guoping Hu. 2018. Fuzzy Cognitive Diagnosis for Modelling Examinee Performance. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9 (2018), 1–26.

[21] Andrew J Martin and Goran Lazendic. 2018. Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of educational psychology* 110, 1 (2018), 27.

[22] Hiromi Nakagawa, Yusuke Iwasawa, and Yutaka Matsuo. 2019. Graph-based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network. In *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. 156–163.

[23] Zachary A Pardos and Neil T Heffernan. 2010. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International conference on user modeling, adaptation, and personalization*. Springer, 255–266.

[24] Hongbin Pei, Bo Yang, Jiming Liu, and Lei Dong. 2018. Group Sparse Bayesian Learning for Active Surveillance on Epidemic Dynamics. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*. 800–807.

[25] Jean Piaget. 1985. *The equilibration of cognitive structures: The central problem of intellectual development*. University of Chicago press.

[26] William F Pinar, William M Reynolds, Peter M Taubman, Patrick Slattery, et al. 1995. *Understanding Curriculum: An Introduction to the Study of Historical and Contemporary Curriculum Discourses*. Vol. 17. Peter Lang.

[27] Mark D Reckase. 2009. Multidimensional item response theory models. In *Multidimensional item response theory*. 79–112.

[28] Xueying Tang, Yunxiao Chen, Xiaoou Li, Jingchen Liu, and Zhiliang Ying. 2019. A reinforcement learning approach to personalized learning recommendation systems. *Brit. J. Math. Statist. Psych.* 72, 1 (2019), 108–135.

[29] Shiwei Tong, Qi Liu, Wei Huang, Zhenya Huang, Enhong Chen, Chuanren Liu, Haiping Ma, and Shijin Wang. 2020. Structure-based Knowledge Tracing: An Influence Propagation View. In *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 541–550.

[30] Kurt VanLehn. 1988. Student modeling. *Foundations of intelligent tutoring systems* 55 (1988), 78.

[31] F. Wang, Q. Liu, E. Chen, Z. Huang, Y. Chen, Y. Yin, Z. Huang, and S. Wang. 2020. Neural Cognitive Diagnosis for Intelligent Education Systems. In *The AAAI Conference on Artificial Intelligence (AAAI)*. 6153–6161.

[32] Xinping Wang, Caidie Huang, Jinfang Cai, and Liangyu Chen. 2021. Using Knowledge Concept Aggregation towards Accurate Cognitive Diagnosis. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2010–2019.

[33] Mike Wu, Milan Mosse, Noah Goodman, and Chris Piech. 2019. Zero shot learning for code education: Rubric sampling with deep learning inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 782–790.

[34] Yu Yin, Zhenya Huang, Enhong Chen, Qi Liu, Fuzheng Zhang, Xing Xie, and Guoping Hu. 2018. Transcribing content from structural images with spotlight mechanism. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2643–2652.

[35] Yuwen Zhou, Changqin Huang, Qintai Hu, Jia Zhu, and Yong Tang. 2018. Personalized learning full-path recommendation model based on LSTM neural networks. *Information Sciences* 444 (2018), 135–152.

[36] Yuqiang Zhou, Qi Liu, Jinze Wu, Fei Wang, Zhenya Huang, Wei Tong, Hui Xiong, Enhong Chen, and Jianhui Ma. 2021. Modeling context-aware features for cognitive diagnosis in student learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2420–2428.