Multi-Relational Cognitive Diagnosis for Intelligent Education

Kaifang Wu[1], Yonghui Yang[1], Kun Zhang[1], Le Wu[1,4](✉), Jing Liu[3], and Xin Li[2]

[1] Hefei University of Technology, Hefei, China
lewu.ustc@gmail.com
[2] University of Science and Technology of China, Hefei, China
leexin@ustc.edu.cn
[3] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China
jliu@nlpr.ia.ac.cn
[4] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China

Abstract. In intelligent education, cognitive diagnosis is a fundamental but important task, which aims to discover students' mastery of different knowledge concepts. Plenty of methods have been proposed to exploit student-exercise interactions, especially graph-based methods. However, most of them treat student behaviors to exercises as a binary interaction (i.e., interacted or not), neglecting diverse behavior patterns (i.e., correct and incorrect interactions). Moreover, the number of concepts is much smaller than exercises, presenting a challenge for measuring student proficiency. Therefore, in this paper, we propose a novel *Multi-Relational Cognitive Diagnosis* (*MRCD*) framework. Specifically, we first divide students' answer behaviors into correct and incorrect interactions with exercises, and form the corresponding two student-exercise relation graphs. We then leverage Graph Convolutional Network to learn exercise-level representations of students and exercises based on different relation graphs. Since dividing operation exacerbate the data sparsity problem, we employ graph contrastive learning to enhance *MRCD* on representation learning. Moreover, considering the relatively small number of concepts, we directly employ attention mechanism to generate student and exercise representations based on relevant concepts. After that, we fuse exercise-level and concept-level representations, and send them to a cognitive diagnosis model to predict student performance. Extensive experiments over two real-world datasets demonstrate the effectiveness of our proposed model.

Keywords: Cognitive diagnosis · Graph convolutional network · Graph contrastive learning

L. Fang et al. (Eds.): CICAI 2022, LNAI 13605, pp. 425–437, 2022.
https://doi.org/10.1007/978-3-031-20500-2_35

1 Introduction

Intelligent Tutoring Systems (ITS) [4,21] have been widely applied in recent years, such as Santa and ASSISTments online education platform. These platforms provide rich exercise resources and personalized exercise suggestions [3,9] for students. The crucial task of ITS is to obtain the proficiency of students on different concepts [20], which has drawn plenty of attention.

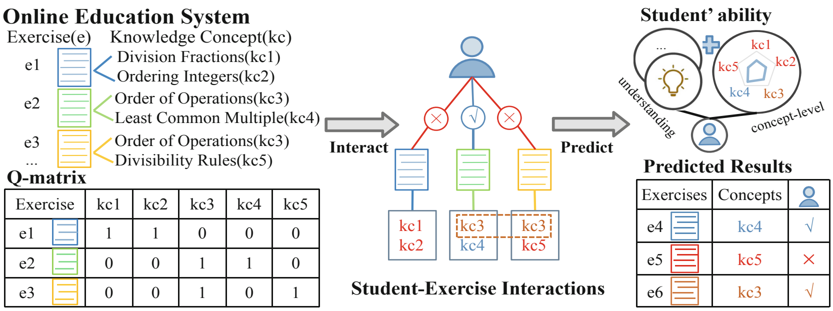


Fig. 1. The process of the cognitive diagnosis task.

Based on student-exercise historical interactions and exercise-concept correlation matrix (i.e., Q-matrix annotated by educational experts), early diagnosis models predicted student performance on exercises by handcrafted functions, such as Item Response Theory (IRT) [13,25], Deterministic Inputs, Noisy-And gate (DINA) [11], and Multidimensional IRT (MIRT) [1,26]. However, most of them only exploit shallow interactions, ignoring complex relationships among students, exercises, and concepts, which cause the performance of student prediction far from satisfaction.

With the rapid development of neural networks, many neural network-based methods have been proposed For example, [30] introduced neural network for cognitive diagnosis on the basis of traditional diagnosis methods. RCD [14] designed a relation-driven framework to learn representations by constructing the student-exercise-concept hierarchical graph, where student-exercise interactions are reduced to binary interactions. Despite the great progress, complex relations among students, exercises, and concepts are still under-exploited. One of main problems is the simplified binary processing of student-exercise interactions. As shown in Fig. 1, student-exercise interactions can be divided into correct and incorrect patterns. For the correct behavior pattern, it does not only reveal understanding to exercises, but also reflects association [37], guessing [11], and so forth. The incorrect behavior also contains different meanings, such as misunderstanding and carelessness. These phenomena demonstrate that it is useful to measure student-exercise interactions in terms of different relations. Moreover, one exercise usually contains multiple concepts, (e.g., Q-matrix in Fig. 1). However, most existing methods handle this phenomenon by treating the performance of a student in an exercise as the average capability of contained concepts [14], which ignores the potential information of student-concept interactions and is too coarse to measure the ability level of students on specific concepts.

To this end, we propose a novel *Multi-Relational Cognitive Diagnosis* (*MRCD*) framework. Specifically, to fully exploit student-exercise interactions, we divide answer records into correct and wrong patterns. Then, we employ Graph Convolutional Network (GCN) to learn exercise-level node representations from student-exercise relation graphs based on different behavior patterns. Meanwhile, the dividing operation will exacerbate data sparsity problem. Thus, we design a contrastive learning scheme with relational data augmentation to alleviate this problem. Moreover, considering that knowledge concepts are much less than exercises, and that student-exercise interactions are too coarse to describe the proficiency of students to concepts, we develop a self-attention module to model the relations among students, exercises, and concepts directly. Along this line, concept-level student ability and exercise characteristics can be better modeled. Next, we fuse exercise-level and concept-level representations of students and exercises, and send them into the neural diagnosis model to predict student performance. Extensive experiments over two real-world datasets demonstrate the superiority of our proposed *MRCD*.

2 Related Work

Cognitive Diagnosis Models. Cognitive diagnosis [12] focuses on assessing strengths and weaknesses regarding the student abilities. Traditionally, to overcome the limitation of Classical Test Theory [2] that evaluates student ability only by actual scores, IRT [13] leveraged a logical function to study the linear relationship between student ability and exercise characteristics (e.g., difficulty and discrimination). DINA [11] adopted discrete binary vectors to represent exercises and students. In addition, it also focused on the effect of the slip and guessing behaviors of students. MIRT [26] expanded the dimension of IRT to study the mastery levels of students in a fine-grained manner. In recent years, neural networks have been introduced for better diagnosis. For example, [30] combined neural network with the monotonicity assumptions for adaptive learning the representations of students and exercises, while ensuring the interpretability. And [10] used semantic information of exercises to obtain the difficulty and discrimination. Besides, there are also some literatures focused on the impact of educational context on students' implicit cognitive states [37], and the diagnostic task by considering both objective and subjective exercises [35].

Graph Structure Modeling. Due to the great success, graph-based modeling have attracted wide attention in many areas, which effectively improve the quality of learned representations by aggregating features from neighbors [7,16,32]. In intelligent education, graph-based methods also become one of the hot topics [14,22,29,36]. For example, [36] explored high-order relevance between exercises and concepts by aggregating node representations from exercise-concept graph. RCD [14] built a hierarchical graph consisting of three local graphs: student-exercise graph, exercise-concept graph, and concept dependency graph to learn better representations. However, these methods just mapped student-exercise interactions into binary values, ignoring the rich information hidden in different behavior patterns, which limited the performance of student prediction.

Meanwhile, Graph Contrastive learning (GCL) is one of the popular technologies on representation learning. Its core idea is to pull closer an anchor and positive samples while pushing away the anchor from negative samples in the representation space [17,18,24,31]. For alleviating the problem of data sparsity, [38] considered the original structure of graphs and node features by adopting an importance-driven approach. [33] introduced self-supervised learning as an auxiliary task to alleviate problems such as the long-tail and the robustness in recommendation. And [18] used structural neighbors and semantic neighbors to construct sample pairs. Considering two special subgraphs formed when constructing the relational graph, we construct positive and negative sample pairs from correct and wrong perspectives.

3 Problem Formulation

There are three entity sets: student set S ($|S| = M$), exercise set E ($|E| = N$), and knowledge concept set C ($|C| = K$), where M, N, K are sizes of each set. $\mathbf{Q} \in \mathbb{R}^{N\times K}$ matrix labeled by experts, describes correlations between exercise E and concept C. $q_{jk} = 1$ is that exercise e_j contains concept c_k, otherwise $q_{jk} = 0$. Besides, $\mathbf{R} \in \mathbb{R}^{M\times N}$ denotes student-exercise interactions, where $r_{ij} \in \{-1, 1, 0\}$ means that student s_i makes $\{wrong, correct, no\}$ answer to exercise e_j.

Considering the interaction type, we propose to divide $\mathbf{R}$ into two relational interactions $\mathbf{R}^+$ and $\mathbf{R}^-$, where $\mathbf{R}^+$ ($\mathbf{R}^-$) denotes correct (wrong) interactions. Taking $\mathbf{R}^+$ as an example, student-exercise interactions (i.e., correct answer and others) naturally form a bipartite graph $\mathcal{G}^+ = \{\mathbf{S} \cup \mathbf{E}, \mathbf{R}^+\}$, where the graph adjacent matrix is constructed as follows:

$$\mathbf{A}^+ = \begin{vmatrix} \mathbf{0}^{N\times M} & \mathbf{R}^+ \\ \mathbf{R}^{+T} & \mathbf{0}^{M\times N} \end{vmatrix}. \tag{1}$$

Meanwhile, the wrong relational graph $\mathcal{G}^- = \{\mathbf{S} \cup \mathbf{E}, \mathbf{R}^-\}$ is constructed similarly. The goal of cognitive diagnosis is to predict student performance on exercises, and diagnose cognitive states of students on specific knowledge concepts.

4 Multi-Relational Cognitive Diagnosis

Figure 2 shows the overall framework of *MRCD*, which consists of three main modules: *Exercise-Level Learning Module*, *Concept-Level Learning Module*, and *Diagnosis Module*. Next, we will introduce each of them in details.

4.1 Exercise-Level Learning Module

As mentioned before, complex student-exercise interactions imply a wealth of information, which requires a better utilization method than binary processing. Therefore, we propose to divide answer records into correct and wrong records, and leverage graph structure to describe student-exercise interactions.

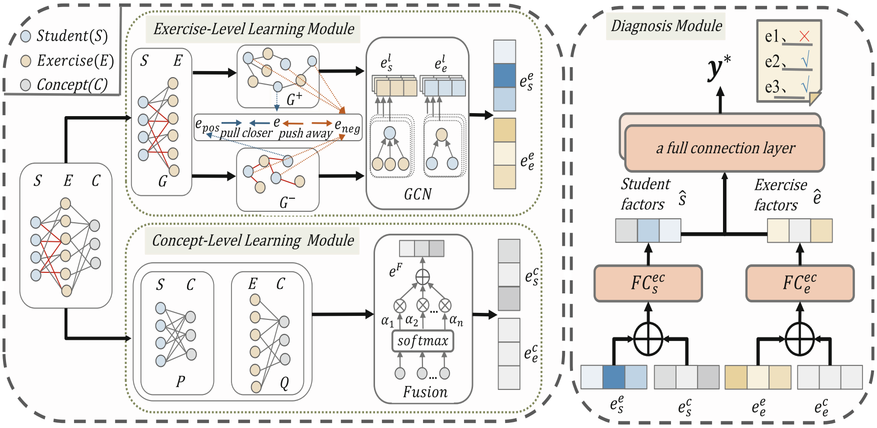


Fig. 2. The overall structure of our proposed *MRCD*. Note that the red lines indicate that students answer exercises wrong. (Color figure online)

For simplicity, in following parts, we take correct answer and other interactions $\mathbf{R}^+$ as an example to report technical details. Specifically, we leverage $\mathbf{S} \in \mathbb{R}^{M\times d}$ and $\mathbf{E} \in \mathbb{R}^{N\times d}$ to denote free embeddings of students and exercises, which are also initial values of the first layer in the graph. To obtain node embeddings at the $(t+1)^{th}$ layer based on its neighbors and its own embedding at the t^{th} layer, we utilize graph propagation and pooling operation to update each node embedding. $A_i^+ = \{j|r_{ij}^+ = 1\}$ and $A_j^+ = \{i|r_{ij}^+ = 1\}$ denote the exercise set that student s_i has answered correctly and the student set who has answered exercise e_j correctly, the updating process can be formulated as follows:

$$\mathbf{s}_i^{t+1} = \mathbf{s}_i^t + \sum_{j\in A_i^+} \frac{\mathbf{e}_j^t}{|A_i^+|}, \quad \mathbf{e}_j^{t+1} = \mathbf{e}_j^t + \sum_{i\in A_j^+} \frac{\mathbf{s}_i^t}{|A_j^+|}. \tag{2}$$

Moreover, we formulate this process in matrix norm. Let $\mathbf{S}^t$ and $\mathbf{E}^t$ denote embedding matrices of students and exercises after the t^{th} propagation, the updated embedding matrices at the $(t+1)^{th}$ propagation are calculated as follows:

$$\begin{vmatrix}\mathbf{S}^{t+1}\\ \mathbf{E}^{t+1}\end{vmatrix} = \begin{vmatrix}\mathbf{S}^{t}\\ \mathbf{E}^{t}\end{vmatrix} + (\mathbf{D}^{-1}\mathbf{A}^+) \times \begin{vmatrix}\mathbf{S}^{t}\\ \mathbf{E}^{t}\end{vmatrix}, \tag{3}$$

where $\mathbf{D}$ is a degree matrix of the $\mathbf{A}^+$, which efficiently transfers neighbor embeddings and updates fusion matrices.

Furthermore, we have noticed that dividing operation will exacerbate the data sparsity problem, which is harmful for student proficiency modeling. Thus, we employ GCL for better representation learning. Since we focus on different interactions among students and exercises, it is natural to treat node embeddings from different relational graphs as augmented embeddings. In $\mathcal{G}^+$ and $\mathcal{G}^-$, the nodes representing the same student or exercise are positive pairs of each other, while the other student(exercise) nodes in the graphs are negative samples. Then, we employ InfoNCE [8,23] to constrain *MRCD* to pull positive pairs closer and push negative pairs away as follows:

$$\begin{aligned} L_s &= \sum_{s_i \in S} -log \frac{exp(sim(\mathbf{s}_i^+, \mathbf{s}_i^-)/\tau)}{\sum_{s_k \in S'} exp(sim(\mathbf{s}_i^+, \mathbf{s}_k^+)/\tau)}, \\ L_e &= \sum_{e_j \in E} -log \frac{exp(sim(\mathbf{e}_j^+, \mathbf{e}_j^-)/\tau)}{\sum_{e_k \in E'} exp(sim(\mathbf{e}_j^+, \mathbf{e}_k^+)/\tau)}, \end{aligned} \tag{4}$$

where $\{\mathbf{s}_i^+, \mathbf{s}_i^-, \mathbf{e}_j^+, \mathbf{e}_j^-\}$ denote student and exercise embeddings from graph $\mathcal{G}^+$ and $\mathcal{G}^-$ separately. $S^{'}$ and $E^{'}$ are the batch data that exclude anchor example of student s_i and exercise e_j. τ is temperature. With these two optimizations, *MRCD* is able to make full use of sparse data to generate better representations.

4.2 Concept-Level Learning Module

Different from the situation that students and exercises have explicit interactions (i.e., correct, wrong, and no answers), we only obtain implicit student-concept interaction according to exercises. Meanwhile, each exercise often includes multiple concepts, and making wrong answer does not mean not mastering all included concepts in the exercise. Thus, we develop a concept-level learning module to model interactions at a fine-grained level [34].

Specifically, we first obtain student-concept interaction matrix $\mathbf{P} \in \mathbb{R}^{M \times K}$ based on $\mathbf{R}$ and $\mathbf{Q}$ matrix. $p_{ij} = 1$ means that student s_i has done exercises that contains concept c_j, Otherwise, $p_{ij} = 0$. Meanwhile, knowledge concepts are represented by $\mathbf{C} \in \mathbb{R}^{K \times d}$, where $\mathbf{c}_i$ denotes the i^{th} concept embedding. Considering the quantitative relationship between exercises and concepts, the student-concept interaction is relatively dense. It is easy to lead to oversmoothing by using multi-layer convolution to propagate information [19]. Thus, we employ attention mechanism to measure their connections as follows:

$$\mathbf{s}_i^c = \textstyle\sum_{c_k \in C_{s_i}} \alpha_{ik} \mathbf{c}_k, \quad \alpha_{ik} = \frac{\mathrm{sim}(\mathbf{s}_i, \mathbf{c}_k)}{\sum_{c_k \in C_{s_i}} \mathrm{sim}(\mathbf{s}_i, \mathbf{c}_k)}, \tag{5}$$

where $C_{s_i} = \{c_k | p_{ik} = 1\}$ denotes a set of concepts that student s_i has interacted with. $\mathrm{sim}(\cdot)$ is cosine similarity. $\mathbf{s}_i^c$ is the concept-level representation of student s_i. Meanwhile, concept-level exercise representations are obtained similarly:

$$\mathbf{e}_j^c = \sum_{c_l \in C_{e_j}} \beta_{jl} \mathbf{c}_l, \quad \beta_{jl} = \frac{\text{sim}(\mathbf{e}_j, \mathbf{c}_l)}{\sum_{c_l \in C_{e_j}} \text{sim}(\mathbf{e}_j, \mathbf{c}_l)}, \tag{6}$$

where $C_{e_j} = \{c_l | q_{jl} = 1\}$ denotes a set of concepts that exercise e_j contains. $\mathbf{e}_j^c$ is the concept-level representation of exercise e_j. By using this module, *MRCD* is able to measure the different impact of concepts on students or exercises, which is helpful for better modeling grasp level of students to concepts.

4.3 Diagnosis Module

After obtaining exercise-level and concept-level representations, what we do next is to fuse these representations and predict the student performance. First of all, we leverage a Fully Connected (FC) layer with a non-negative activation function (i.e., Sigmoid function) to fuse representations as follows:

$$\hat{\mathbf{s}}_i = \sigma(FC_s([\mathbf{s}_i; \mathbf{s}_i^c])), \quad \hat{\mathbf{e}}_j = \sigma(FC_e([\mathbf{e}_j; \mathbf{e}_j^c])), \tag{7}$$

where $[;]$ denotes concatenation operation. $\sigma(\cdot)$ is Sigmoid activation function. $\{\hat{\mathbf{s}}_i, \hat{\mathbf{e}}_j\}$ are the final representations of student s_i and exercise e_j.

Secondly, similar to existing neural diagnosis models [10], we also employ expert-designed diagnosis function MIRT [26] to finish the task. Specifically, we leverage concept-level exercise representation $\mathbf{e}_j^c$ to obtain the value of exercise's discrimination with another FC layer. Then, the diagnosis result of student s_i is calculated by predicting whether exercise e_j is answered correctly as follows:

$$y_{ij}^* = \sigma(F(\mathbf{k} \times (\hat{\mathbf{s}}_i - \hat{\mathbf{e}}_j) \times \mathbf{e}_j^{disc})), \quad \mathbf{e}_j^{disc} = \sigma(FC_1(\mathbf{e}_j^c)), \tag{8}$$

where $F(\cdot)$ is the neural network with two full connected layers in which each element of the weight is restricted to be positive [30]. $\mathbf{k}$ is a one-hot vector that denotes concepts contained in exercise e_j (i.e., the j-th row in the Q matrix). $\mathbf{y}_{ij}^*$ is the predicted probability that student s_i answers exercise e_j correctly.

Loss Function. Since cognitive diagnosis is formulated as predicting whether a student do exercises correctly, Cross-Entropy is employed as the optimization:

$$L_{se} = -\sum_i \sum_j \left(y_{ij} \log y_{ij}^* + (1 - y_{ij}) \log \left(1 - y_{ij}^*\right)\right), \tag{9}$$

where y_{ij} is the ground truth whether student s_i answers exercise e_j correctly. Meanwhile, we have employed GCL to constrain *MRCD* to learn better exercise-level representations with Eq. (4). Finally, the overall optimization target of *MRCD* is formulated with a hyper-parameter λ as follows:

$$Loss = L_{se} + \lambda(L_s + L_e). \tag{10}$$

5 Experiments

5.1 Experimental Settings

Table 1. The statistics of the datasets.

Datasets	Students	Exercises	Concepts	Correct logs	Wrong logs	KCs per exercise
ASSISTMents0910	4,163	17,751	123	183,356	95,520	1.97
EDNET	1,000	11,760	189	734,564	353,589	2.25

Datasets. We select ASSISTMents0910[1] and EdNet[2], as the evaluation datasets. These datasets count the interaction records between students and online tutoring systems, which mainly record the data of several attributes and results of students in the exercise process. Similar to existing work [14,28,30,36], we filter out exercises without concepts annotation and students who have less than 15 records. And we randomly select 1,000 students for cognitive diagnosis on EdNet dataset. The statistics of these two datasets are reported in Table 1.

Evaluation Metrics. Considering that the student ability level cannot be directly measured in cognitive diagnosis, we adopt some common indicators to evaluate the performance of the model, such as, *Root Mean Squared Error (RMSE)* [6], *Accuracy (Acc)* and *Area Under Curve (AUC)* [5].

Baselines. We compare *MRCD* with the following baselines:

- **IRT** [13]: IRT is a widely used probabilistic model based on a one-dimensional linear relationship between student ability and exercise characteristics.
- **MIRT** [26]: MIRT extends traditional IRT to model student-exercise interactions from multidimensional knowledge concepts.
- **DINA** [11]: DINA considers whether students have mastered the fine-grained knowledge concepts in a discrete manner. And the factors of student guessing and sliding are also concerned.
- **NeuralCD** [30]: NeuralCD introduces neural network to model the complex interaction relationship between students and exercises, and ensures the interpretability of student factors and exercise factors with help of the monotonicity assumption in the traditional diagnostic model.
- **RCD** [14]: RCD comprehensively models the student-exercise-concept relationship based on graph structure, especially concept dependency.

[1] https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data/skill-builder-data-2009-2010.

[2] http://ednet-leaderboard.s3-website-ap-northeast-1.amazonaws.com/.

Parameter Setting. To obtain the best performance, we tune the hyper-parameters over validation set and employ Early-Stop with the patience of 6 epochs to prevent overfitting. Some common hyper-parameters are set as follows. Learning rate is set as $lr = 0.0001$. Embedding size of students, exercises, and knowledge concepts are set as the number of concepts K. The number of layer of GCN in exercise-level learning module are selected from $\{1, 2, 3, 4\}$. The temperature τ in GCL is selected from $\{0.05, 0.1, 0.15, 0.2\}$. λ in Eq. (10) is selected from $\{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}\}$. We initialize the network parameters with Xavier initialization [15].

Table 2. Overall results on student performance prediction.

Models	ASSISTMents0910			EdNet		
	ACC↑	RMSE↓	AUC↑	ACC↑	RMSE↓	AUC↑
IRT	0.67946	0.45477	0.68273	0.71707	0.43444	0.72514
DINA	0.67635	0.48847	0.71167	0.70049	0.46659	0.69111
MIRT	0.72136	0.45105	0.74283	0.72931	0.42891	0.74826
NeuralCD	0.72782	0.43374	0.75469	0.72819	0.42469	0.75792
RCD	0.73264	0.42268	0.77013	0.73246	0.42311	0.76280
MRCD	**0.74130**	**0.41862**	**0.78208**	**0.73749**	**0.41962**	**0.77040**

5.2 Experimental Results

Overall Performance. Table 2 reports the overall performance of models on student performance prediction. From the table, we observe that *MRCD* achieves the best performance on all evaluation metrics compared with other baselines, especially the state-of-the-art graph-based RCD model. This phenomenon demonstrates the effectiveness of considering different type of interactions among students and exercises and detailed measurement of students and concepts. Moreover, *MRCD* has stable performance when data becomes sparser (i.e., ASSISTMents0910 dataset), indicating the usefulness of CL framework employed in exercise-level learning module. Among all baselines, we observe that neural network-based methods have better performance than traditional cognitive diagnosis models, indicating that neural network-based methods are more powerful to measure the complex relationships among students, exercises, and concepts.

Parameter Sensitive Test. There are two important hyper-parameters to control the impact of different modules, GCN layer number D and the weight λ of CL loss. Therefore, we conduct additional experiments to verify their impacts. Results are reported in Table 3 and Fig. 3. From the results, we observe that with the increasing of GCN layers, model performance becomes better, supporting that high-order interactions are very useful for student performance prediction. Moreover, the performance increasing on ASSISTMents0910 dataset is larger

Table 3. Performance comparisons of different propagation depth D.

Depth	ASSISTMents0910			EdNet		
	ACC↑	RMSE↓	AUC↑	ACC↑	RMSE↓	AUC↑
D = 0	0.73004	0.43093	0.76606	0.73560	0.42052	0.76869
D = 1	0.73802	0.42100	0.77237	0.73512	0.42205	0.76674
D = 2	0.73931	0.41932	0.77855	0.73609	0.42062	0.76916
D = 3	**0.74130**	**0.41862**	**0.78208**	**0.73749**	0.41962	0.77040
D = 4	0.73890	0.41894	0.78047	0.73667	**0.41953**	**0.77058**

than EdNet. The possible reason is that more GCN layers will help *MRCD* obtain more information for student proficiency modeling.

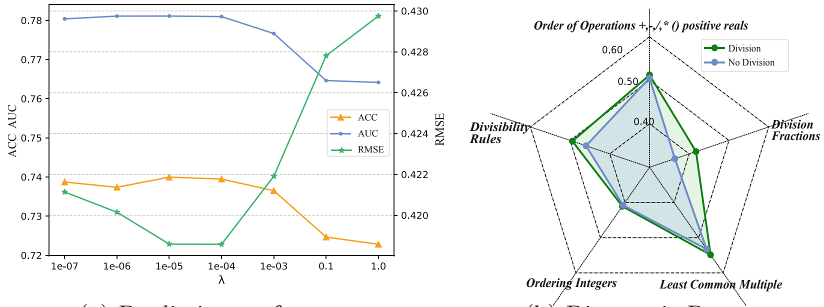
For the impact of CL framework, we observe from Fig. 3(a) that with the increasing of λ, the performance of *MRCD* is first increasing and then decreasing. When λ is bigger than 0.001, the performance will have a big drop. We speculate the possible reason is that when λ is too big, *MRCD* will be constrained to learn similar representations for all students, which will do harm to model performance. Based on three different evaluation metrics, we finally select $\lambda = 5e-5$ and $\lambda = 1e-6$ as best setting for two datasets separately.

Ablation Study. In this part, we make several ablation studies to verify the effectiveness of each component in *MRCD*, including *exercise-level learning module (MRCD-graph)*, *concept-level learning module (MRCD-kc)*, the consideration of *different student-exercise relations* by using RGCN [27] (*MRCD-graph-kc(rgcn)*) and *MRCD-graph-kc(t)* without the division operation. Results are illustrated in Table 4. We have to note that *Division* means whether dividing different relations. From the table, we observe that exercise-level learning module plays the most important role. Since this module focuses on high-order interactions among students and exercises, as well as employs CL framework to alleviate the sparsity problem that dividing operation introduced, it is natural that this module is very critical in our *MRCD* model. Moreover, when treating correct and wrong answer behaviors as the same, we observe that model performance also declined somewhat, demonstrating the necessity of processing correct and wrong answer behaviors differently.

Case Study. Here we visualize the diagnostic result of a student selected from the Assistment0910 dataset as shown in Fig. 3(b). And we record the student's historical responses and exercise-concept correlation relationship in Fig. 1. Note that *Order of Operations +,−,/,* () positive reals* is abbreviated to *Order of Operations*. From Fig. 3(b), we could clearly observe that different answer behaviors have an impact on the student' mastery of corresponding knowledge concepts. And considering students' different answer behavior patterns is useful to learn the ability of students, especially in the exercises with wrong answers.

Table 4. Results of ablation experiment.

Models	Division	ASSISTMents0910			EdNet		
		ACC↑	RMSE↓	AUC↑	ACC↑	RMSE↓	AUC↑
MRCD-kc	×	0.72500	0.43450	0.75752	0.71679	0.42992	0.74397
MRCD-graph	✓	0.73840	0.41975	0.77812	0.73059	0.42227	0.76466
MRCD-grph-kc(rgcn)	✓	0.73614	0.42274	0.77334	0.73587	0.42035	0.76909
MRCD-graph-kc(t)	×	0.73849	0.42076	0.77773	0.73597	0.41980	0.76964
MRCD	✓	**0.74130**	**0.41862**	**0.78208**	**0.73749**	**0.41962**	**0.77040**



(a) Prediction performance (b) Diagnostic Report

Fig. 3. Results of *MRCD* on ASSISTMents0910.

6 Conclusion

In this paper, we argued that different student-exercise interaction behaviors revealed different features of students, which should be considered explicitly. Thus, we proposed a novel *Multi-Relational Cognitive Diagnosis* (*MRCD*) framework to model student proficiency over exercises from exercise-level and concept-level perspectives simultaneously. Specifically, we first divided student-exercise interactions into correct and incorrect answer interactions and built graphs based on them. Then, we designed an exercise-level learning module, in which GCN and GCL framework are employed to learn better representations. Moreover, we developed a concept-level learning module to measure student-concept interactions and exercise-concept relationship directly. Then, we fused these two-level representations and sent them to a commonly used diagnosis model to predict student performance over exercises. Extensive experiments over two real-world datasets showed the superiority of *MRCD*. In the future, we will consider the dynamic change of student ability over time for better cognitive diagnosis.

Acknowledgments. This work was supported in part by grants from the National Key Research and Development Program of China (Grant No. 2021ZD0111802), the Fundamental Research Funds for the Central Universities (JZ2021HGTB0075), and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).

References

1. Adams, R.J., Wilson, M., Wang, W.C.: The multidimensional random coefficients multinomial logit model. Appl. Psychol. Meas. **21**(1), 1–23 (1997)
2. Allen, M.J., Yen, W.M.: Introduction to Measurement Theory. Waveland Press, Long Grove (2001)
3. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Engaging with massive online courses. In: WWW, pp. 687–698 (2014)
4. Anderson, J.R., Boyle, C.F., Reiser, B.J.: Intelligent tutoring systems. Science **228**(4698), 456–462 (1985)
5. Bradley, A.P.: The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recogn. **30**(7), 1145–1159 (1997)
6. Chai, T., Draxler, R.R.: Root mean square error (RMSE) or mean absolute error (MAE)? - arguments against avoiding RMSE in the literature. Geosci. Model Dev. **7**(3), 1247–1250 (2014)
7. Chen, L., Wu, L., Hong, R., Zhang, K., Wang, M.: Revisiting graph based collaborative filtering: a linear residual graph convolutional network approach. In: AAAI, vol. 34, pp. 27–34 (2020)
8. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: ICML, pp. 1597–1607. PMLR (2020)
9. Chen, Y., Li, X., Liu, J., Ying, Z.: Recommendation system for adaptive learning. Appl. Psychol. Meas. **42**(1), 24–41 (2018)
10. Cheng, S., et al.: Dirt: deep learning enhanced item response theory for cognitive diagnosis. In: CIKM, pp. 2397–2400 (2019)
11. De La Torre, J.: Dina model and parameter estimation: a didactic. J. Educ. Behav. Stat. **34**(1), 115–130 (2009)
12. DiBello, L.V., Roussos, L.A., Stout, W.: 31A review of cognitively diagnostic assessment and a summary of psychometric models. Handb. Stat. **26**, 979–1030 (2006)
13. Embretson, S.E., Reise, S.P.: Item Response Theory. Psychology Press, Hove (2013)
14. Gao, W., et al.: RCD: relation map driven cognitive diagnosis for intelligent education systems. In: SIGIR, pp. 501–510 (2021)
15. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
16. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: LightGCN: simplifying and powering graph convolution network for recommendation. In: SIGIR, pp. 639–648 (2020)
17. Khosla, P., et al.: Supervised contrastive learning. In: NeurIPS, vol. 33, pp. 18661–18673 (2020)
18. Lin, Z., Tian, C., Hou, Y., Zhao, W.X.: Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In: WWW, pp. 2320–2329 (2022)
19. Liu, M., Shao, P., Zhang, K.: Graph-based exercise-and knowledge-aware learning network for student performance prediction. In: Fang, L., Chen, Y., Zhai, G., Wang, J., Wang, R., Dong, W. (eds.) CICAI 2021. LNCS, vol. 13069, pp. 27–38. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-93046-2_3
20. Liu, Q., et al.: EKT: exercise-aware knowledge tracing for student performance prediction. IEEE Trans. Knowl. Data Eng. **33**(1), 100–115 (2019)
21. Liu, Q., et al.: Exploiting cognitive structure for adaptive learning. In: KDD, pp. 627–635 (2019)

22. Nakagawa, H., Iwasawa, Y., Matsuo, Y.: Graph-based knowledge tracing: modeling student proficiency using graph neural network. In: 2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 156–163. IEEE (2019)
23. Van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv e-prints arXiv-1807 (2018)
24. Qiu, J., et al.: GCC: graph contrastive coding for graph neural network pre-training. In: KDD, pp. 1150–1160 (2020)
25. Rasch, G.: On general laws and the meaning of measurement in psychology. In: Berkeley Symposium on Mathematical Statistics, vol. 4, pp. 321–333 (1961)
26. Reckase, M.D.: Multidimensional item response theory models. In: Reckase, M.D. (ed.) Multidimensional Item Response Theory. SSBS, pp. 79–112. Springer, New York (2009). https://doi.org/10.1007/978-0-387-89976-3_4
27. Schlichtkrull, M., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: Gangemi, A., et al. (eds.) ESWC 2018. LNCS, vol. 10843, pp. 593–607. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93417-4_38
28. Thai-Nghe, N., Schmidt-Thieme, L.: Multi-relational factorization models for student modeling in intelligent tutoring systems. In: 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE), pp. 61–66. IEEE (2015)
29. Tong, S., et al.: Structure-based knowledge tracing: an influence propagation view. In: 2020 IEEE International Conference on Data Mining (ICDM), pp. 541–550. IEEE (2020)
30. Wang, F., et al.: Neural cognitive diagnosis for intelligent education systems. In: AAAI, vol. 34, pp. 6153–6161 (2020)
31. Wang, T., Isola, P.: Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In: ICML, pp. 9929–9939. PMLR (2020)
32. Wang, X., He, X., Wang, M., Feng, F., Chua, T.S.: Neural graph collaborative filtering. In: SIGIR, pp. 165–174 (2019)
33. Wu, J., et al.: Self-supervised graph learning for recommendation. In: SIGIR, pp. 726–735 (2021)
34. Wu, L., He, X., Wang, X., Zhang, K., Wang, M.: A survey on accuracy-oriented neural recommendation: from collaborative filtering to information-rich recommendation. IEEE Trans. Knowl. Data Eng. (2022)
35. Wu, R., et al.: Cognitive modelling for predicting examinee performance. In: IJCAI (2015)
36. Yang, Y., et al.: GIKT: a graph-based interaction model for knowledge tracing. In: Hutter, F., Kersting, K., Lijffijt, J., Valera, I. (eds.) ECML PKDD 2020. LNCS (LNAI), vol. 12457, pp. 299–315. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67658-2_18
37. Zhou, Y., et al.: Modeling context-aware features for cognitive diagnosis in student learning. In: KDD, pp. 2420–2428 (2021)
38. Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., Wang, L.: Graph contrastive learning with adaptive augmentation. In: WWW, pp. 2069–2080 (2021)