RCD: Relation Map Driven Cognitive Diagnosis for Intelligent Education Systems

Weibo Gao[1], Qi Liu[1,*], Zhenya Huang[1], Yu Yin[1], Haoyang Bi[1],
Mu-Chun Wang[1], Jianhui Ma[1], Shijin Wang[2], Yu Su[2]
[1]Anhui Province Key Laboratory of Big Data Analysis and Application,
School of Computer Science and Technology, University of Science and Technology of China,
{weibogao,yxonic,bhy0521,amoswang2000}@mail.ustc.edu.cn;{qiliuql,huangzhy,jianhui}@ustc.edu.cn;
[2]IFLYTEK, {sjwang3, yusu}@iflytek.com

ABSTRACT

Cognitive diagnosis (CD) is a fundamental issue in intelligent educational settings, which aims to discover the mastery levels of students on different knowledge concepts. In general, most previous works consider it as an inter-layer interaction modeling problem, e.g., student-exercise interactions in IRT or student-concept interactions in DINA, while the inner-layer structural relations, such as educational interdependencies among concepts, are still underexplored. Furthermore, there is a lack of comprehensive modeling for the student-exercise-concept hierarchical relations in CD systems. To this end, in this paper, we present a novel **R**elation map driven **C**ognitive **D**iagnosis (RCD) framework, uniformly modeling the interactive and structural relations via a multi-layer student-exercise-concept relation map. Specifically, we first represent students, exercises and concepts as individual nodes in a hierarchical layout, and construct three well-defined local relation maps to incorporate inter- and inner-layer relations, including a student-exercise interaction map, a concept-exercise correlation map and a concept dependency map. Then, we leverage a multi-level attention network to integrate node-level relation aggregation inside each local map and balance map-level aggregation across different maps. Finally, we design an extendable diagnosis function to predict students' performance and jointly train the networks. Extensive experimental results on real-world datasets clearly show the effectiveness and extendibility of our RCD in both diagnosis accuracy improvement and relation-aware representation learning.

CCS CONCEPTS

• **Applied computing → E-learning**.

KEYWORDS

Cognitive diagnosis; Student performance prediction; Graph neural network

* Corresponding Author.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3462932

ACM Reference Format:
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 *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3404835.3462932

1 INTRODUCTION

Cognitive Diagnosis (CD) has been recognized as a crucial task in intelligent educational applications [3, 6, 20], such as student assessment [4, 22] and resource recommendation [25]. In principle, a CD system consists of three components: students, exercises and knowledge concepts. Specifically, given the students' exercising records (e.g., answer correctly or not), the goal of CD is discovering their actual mastery levels on specific knowledge concepts [27].

As shown in Figure 1 (a), a CD system can be abstracted as a three-layer student-exercise-concept hierarchy. Existing CD studies can be basically summarized as modeling interactions between different layers from a higher perspective. For example, Item Response Theory (IRT) [13], Multidimensional IRT (MIRT) [1, 36] and Matrix Factorization (MF) [24, 41] try to model the interactions between students and exercises with linear functions, while Neural Cognitive Diagnosis (NCD) [44] leverages higher-order student-exercise interactions with neural networks. There are also some researches (e.g., Deterministic Inputs, Noisy-And gate (DINA) [10]) that model the student-concept interaction through mapping exercises to corresponding concepts directly with a so-called Q-matrix.

Although great efforts have been made, rich information contained in the student-exercise-concept hierarchy is not effectively utilized yet. Specifically, there exist two problems along this line. First, the multi-layer interactions among the hierarchy should be fully modeled instead of partially modeled (i.e., student-exercise or student-concept). As illustrated in Figure 1 (b, c), in a learning system, students (e.g., s_1) interacts with exercises (e.g., q_2) by responsing them, meanwhile exercises relate to concepts (e.g., c_1 and c_3) with knowledge correlations. Since entities of each layer have particular features and unique inter-layer interactions with other layers, a complete and comprehensive modeling will benefit the CD task best. Second, in addition to the inter-layer interactions (e.g., between students and exercises), there also exist complex inner-layer structures, especially among concepts. Sufficient pedagogical researches [12, 48] have proved the existence of interdependencies between knowledge concepts, which are usually represented by a concept dependency map [30]. Without loss of generality, Figure 1

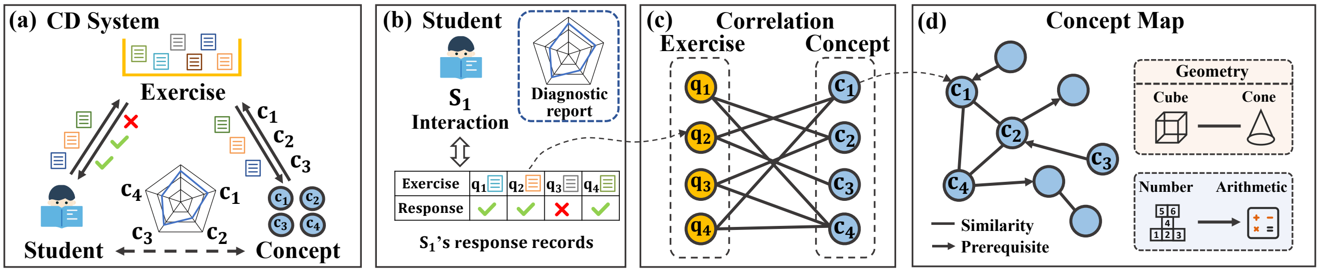


Figure 1: The illustrative examples of (a) the cognitive diagnosis system; (b) students exercising interactions; (c) correlations between exercises and concepts; (d) a concept dependency map showing educational relations between concepts.

(d) demonstrates a concept dependency map with two typical relations: similarity and prerequisite. Specifically, similarity relations involve a pair of concepts belonging to the same topic (e.g., concept *Cone* and concept *Cube* of topic *Geometry*). Prerequisite relations involve a pair of concepts which are assumed that students should logically learn one before another (e.g., concept *Number* before concept *Arithmetic*). Such structural information can obviously help the CD task. Taking Figure 1 (d) as an example, if student s_1 is known to have mastered concept c_2 , she has probably mastered the similar concepts c_1 and c_4, and the prerequisite concept c_3. In summary, an ideal cognitive diagnosis system should exploit the complete inter-layer student-exercise-concept interactions as well as the inner-layer concept dependency structures.

However, difficulties lie ahead to achieve the ideal CD system due to the complexity of the student-exercise-concept hierarchy. On the one hand, the system contains different types of relations that make it highly heterogeneous. On the other hand, these relations are not thoroughly unorganized but in hierarchical layouts. To address the challenges of modeling the co-existing heterogeneity and organization, we propose a general CD framework, namely **R**elation map driven **C**ognitive **D**iagnosis (RCD), which uniformly models the interactive and structural relations from the multi-layer student-exercise-concept relation map. Specifically, different from previous work, RCD explicitly represents all the entities in the student-exercise-concept hierarchy with an embedding layer. In order to effectively learn relation-aware representations, we introduce three well-defined local maps, i.e., student-exercise interaction map, exercise-concept correlation map and concept dependency map, (described in section 3), on which the nodes recursively aggregate information from neighbors. Since each node is shared by multiple maps, a fusion layer is designed to apply sophisticated node-level and map-level aggregation for each node and leverage an attention network to automatically balance the multi-level information. Thereafter, given the relation-aware representations, RCD predicts the students' performance with a carefully designed diagnosis function. Furthermore, as a general framework, RCD is expected to have the capacity of extending the existing methods. In other words, RCD is well-designed so that it can naturally blend additional interactions and structural information into current methods and improve their performance. Finally, extensive experiments on real-world datasets clearly demonstrate the effectiveness and extendibility of RCD. In summary, our key contributions are listed as follows:

- We comprehensively model the hierarchical student-exercise-concept relations for CD and integrate interactive and structural information with multiple relations explicitly. Besides, we design a multi-level attention network to attentively learn how the model prefers different map sources.
- Our proposed RCD has high extendibility with existing CD methods, which easily improves previous CD studies that only model student-exercise interaction in a natural way.
- We conduct extensive experiments on real-world datasets to validate the effectiveness and extendibility of RCD, including quantitative comparison, qualitative analysis and interpretable visualization.

2 RELATED WORK

Cognitive Diagnosis. The task of cognitive diagnosis (CD) originates from the pedagogical assumption that the cognitive state of each student is stable in a static scene (e.g., an exam) thus can be diagnosed through their interactive behaviors (i.e., historical exercising records). Therefore, existing research of CD can be generally categorized into two genres which model student-exercise interaction and student-concept interaction respectively. The first genre of work characterizes the student feature (e.g., ability level) and the exercise feature (e.g., difficulty level) and models the performance of exercising behavior with a interaction function. As typical representatives, traditional IRT [13], MIRT [1, 36] and MF [24, 41] use linear functions, while recent NCD [44] leverages neural networks to learn a high-order non-linear function. Another genre of work directly builds student-concept interaction by simply ignoring the feature of individual exercises and replacing them with their corresponding concepts. For example, DINA [10] characterizes each student with a binary vector which indicates whether or not the student has mastered the knowledge concepts. Although existing methods have made great progress, none of them models students, exercises and concepts thoroughly, thus leaving abundant information underexploited.

Knowledge Concept Structure. According to educational theories [12, 34, 35], the knowledge concepts usually do not exist alone. There are many conceptual relations which specify the implicit interdependencies between concepts. Without loss of generality, these relations can be categorized into the directed and the undirected. The directed relations generally indicate knowledge influence propagation among concepts, such as prerequisite relation [8] and remedial relation [38]. On the contrary, the undirected relations usually indicate knowledge overlapping or interplay among concepts, such as similarity relation [47] and collaboration relation [19]. The rich structural information between concepts has been proved to be fairly helpful for many educational tasks [8, 29, 39]. Along this way, our RCD properly incorporates knowledge concept structure into cognitive diagnosis via modeling the relation map.

Graph Structure Modeling. In recent years, graph structure modeling [9] has been one of the most popular topics due to the remarkable performance for modeling many graph-structured data [7, 52]. The goal of graph based models [15, 33, 45] is learning structural representation for nodes or edges to further support various graph mining tasks. As the state of the art technique, graph neural network (GNN) [14, 17, 37] has shown huge success for modeling graph structures. The key idea behind GNN is to recursively aggregate feature information from neighboring nodes via neural networks. However, a lot of previous work in GNNs, such as GAT [43] and GraphSage [16], are designed solely for homogeneous graphs, which are not enough for many real-world problems with various types of nodes and edges, e.g., the student-exercise-concept relations mentioned above. Recently, much efforts have been made to address graph heterogeneity which mainly focuses on preserving the meta-path based structural information. For example. metapath2vec [11] and GTN [50] capture hierarchy with manually designed or automatically learned meta-paths. HAN [46] and HetG [51] consider the multi-level information and the attention mechanism to further improve heterogeneous graph learning.

3 PRELIMINARIES

In this section, we first introduce the student-exercise-concept relation map, which is illustrated in Figure 2 (a). Then we formally define the problem of relation map driven cognitive diagnosis.

3.1 Relation Map

The relation map presents as three local maps, from top to bottom are a concept dependency map, an exercise-concept correlation map and a student-exercise interaction map, respectively.

3.1.1 Concept dependency map. The concept dependency map with educational dependencies labeled by experts could be represented as a graph $\mathcal{G}_c\ (C, \mathcal{R}_c)$. In this graph, C is a set of concepts and $\mathcal{R}_c = \left\{\mathcal{R}_c^r, r = 1, \dots, R\right\}$ is a set of multiple educational dependency relations, where r stands for a certain type of educational dependency relations (e.g., prerequisite and similarity).

3.1.2 Exercise-concept Correlation Map. An exercise-concept correlation map $\mathcal{G}_{qc}\ \left(Q \cup C, \mathcal{R}_{qc}\right)$, could be denoted as a bidirectional exercise-concept bipartite graph, where Q is a set of exercises, C is a set of concepts, and $\mathcal{R}_{qc}$ refers to the set of correlations between exercises and concepts. In addition, the relation $r_{q_i \leftrightarrow c_j} = 1$ refers to that exercise q_i includes the main concept c_j. For example, exercise q_1 has main concepts c_2 and c_4 in Figure 1 (c).

3.1.3 Student-Exercise Interaction Map. The student-exercise interaction data could be represented as a bidirectional student-exercise bipartite graph $\mathcal{G}_{sq}\ \left(\mathcal{S} \cup Q, \mathcal{R}_{sq}\right)$, where $\mathcal{S}$ is a set of students, Q is a set of exercises, and $\mathcal{R}_{sq}$ refers to the set of interactions from training data. If relation $r_{s_i \leftrightarrow q_j} = 1$, student s_i has chosen exercise q_j and answered it. Please note that, here we do not reveal students' exercising response data (i.e., answer correctly or not) in fact.

3.1.4 Relation Map in RCD. After introducing the above three local maps, here we define the relation map as $\mathcal{G}(\mathcal{G}_c \cup \mathcal{G}_{qc} \cup \mathcal{G}_{sq}, \mathcal{R}_c \cup \mathcal{R}_{qc} \cup \mathcal{R}_{sq})$, which contains educational dependencies among concepts, exercise-concept correlations and student exercising response interactions as illustrated in Figure 2 (a). For the hierarchical relation map, it naturally presents a multi-level structure. From the map-level perspective, the map consists of three cross-layer local maps. From the node-level perspective, each local map has a relatively independent map source (i.e., nodes and relations).

3.2 Problem Statement

Let $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$ be the set of N students, $Q = \{q_1, q_2, \dots, q_M\}$ be the set of M exercises and $C = \{c_1, c_2, \dots, c_K\}$ be the set of K knowledge concepts. $\mathcal{R}_{qc} = \left\{r_{q_i \leftrightarrow c_j} \mid q_i \in Q, c_j \in C\right\}$ is the set of correlations between exercises and concepts. Suppose each student individually chooses some exercises to practice. We record the response records of a certain student as a set of triplet (s, q, r_{sq}) where $s \in \mathcal{S}$, $q \in Q$ and r_{sq} is the score that student s got on exercise e. Let $\mathcal{L}$ be the set of response records. As described above, we have the hierarchical relation map $\mathcal{G}$. Then we give a clear formulation of the cognitive diagnosis driven by relation map as:

Given: students' response records $\mathcal{L}$, exercises and concepts' correlations $\mathcal{R}_{qc}$ and relation map $\mathcal{G}$;

Goal: diagnosing students' cognitive states (i.e., proficiency on specific knowledge concepts) by modeling the student performance prediction process.

4 RCD: RELATION MAP DRIVEN COGNITIVE DIAGNOSIS

In this section, we first briefly present a general description of our model. Then we introduce each part of the model in detail. Finally, we demonstrate the expandability of the model.

Overview. Our RCD model can jointly learn the relation-aware representations of students, exercises and concepts based on the heterogeneous relation map, and utilize the learned representations to infer students' proficiency on concepts. As shown in Figure 2, RCD contains three main parts: an embedding layer, a fusion layer and an extensible diagnosis layer. Specifically, by taking response records, the embedding layer outputs the vectorized embedding representations of concepts, exercises and students. In the fusion layer, we implement a multi-level attention structure to automatically balance interactive information among the multi-layer relation map, for effectively embedding learning. After fusing interactive information, the diagnosis layer utilizes the relation-aware embeddings to infer students' cognitive level through predicting the response score of each pair of student-exercise. Particularly, our model has superior expandability since it can naturally blend additional interactions and structural information into current methods and improve their performance.

4.1 Embedding Layer

It encodes students, exercises and knowledge concepts with d-dimensional trainable matrices $\mathrm{S} \in \mathbb{R}^{N \times d}$, $\mathrm{Q} \in \mathbb{R}^{M \times d}$ and $\mathrm{C} \in \mathbb{R}^{K \times d}$ respectively. Here N, M and K are the numbers of students, exercises and concepts respectively. For student s_z, her factor $\boldsymbol{s}_z^0$ aforementioned is obtained by multiplying his one-hot representation vector $\boldsymbol{x}_z$ with the trainable matrix S, i.e., the transpose of z-th row from student embedding matrix S. Similarly, exercise q_d's embedding $\boldsymbol{q}_d^0$ is the transpose of d-th row of matrix Q and each related concept c_k's embedding $\boldsymbol{c}_k^0$ is the transpose of k-th row of matrix C. We model the embedding layer as:

$$\boldsymbol{s}_z^0 = \boldsymbol{x}_z^T \mathrm{S}, \quad \boldsymbol{q}_d^0 = \boldsymbol{x}_d^T \mathrm{Q}, \quad \boldsymbol{c}_k^0 = \boldsymbol{x}_k^T \mathrm{C}, \tag{1}$$

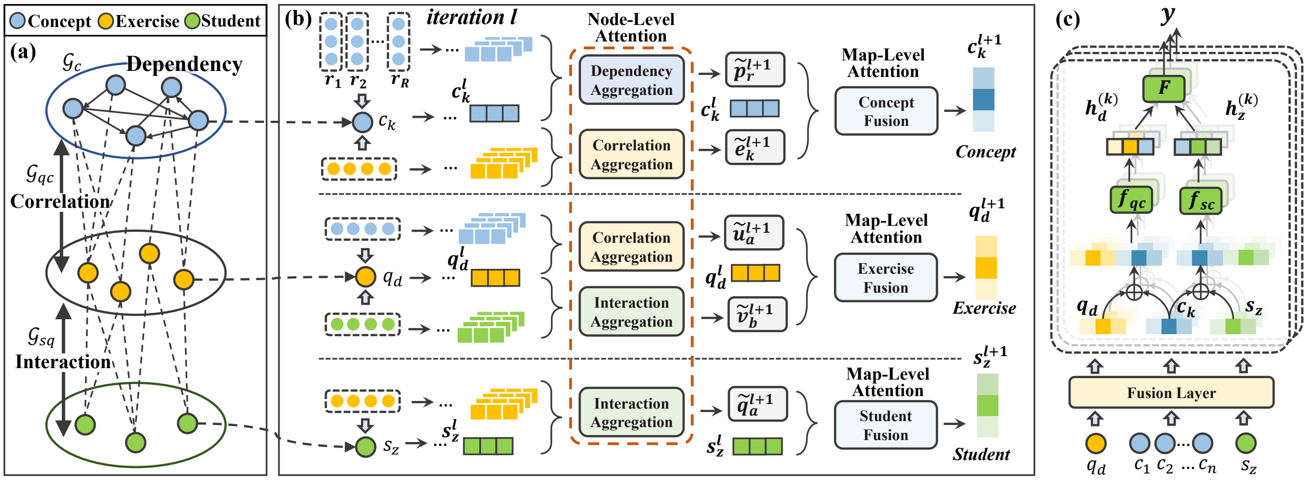


Figure 2: The overview architecture of RCD: (a) The relation map with multiple relations; (b) The embedding layer and fusion layer (illustrate the l-th iteration fusing process); (c) The diagnosis layer for performance prediction.

where $\boldsymbol{s}_z, \boldsymbol{q}_d, \boldsymbol{c}_k \in \mathbb{R}^d$, $\boldsymbol{x}_z \in \{0,1\}^N$, $\boldsymbol{x}_d \in \{0,1\}^M$ and $\boldsymbol{x}_k \in \{0,1\}^K$.

4.2 Fusion Layer

After obtaining the vectorized representations of each student s_z's embedding $\boldsymbol{s}_z^0$, each exercise q_d's embedding $\boldsymbol{q}_d^0$ and each concept c_k's embedding $\boldsymbol{c}_k^0$ from the previous layer, to address the first challenge of modeling the co-existing heterogeneity and organization among student-exercise-concept, the fusion layer uniformly models the interactive and structural relations with iteration-wise multi-level fusion. Specifically, at each iteration $l+1$, our fusion operation updates student s_z's embedding $\boldsymbol{s}_z^{l+1}$, exercise q_d's embedding $\boldsymbol{q}_d^{l+1}$ and concept c_k's embedding $\boldsymbol{c}_k^{l+1}$ from the previous iteration l, as illustrated in Figure 2 (b). In the following, we would respectively introduce how to update concept, exercise and student embeddings.

4.2.1 ***Knowledge Concept Fusion.*** For each knowledge concept c_k, let $\boldsymbol{c}_k^l$ denote its embedding at the l-th iteration. Each concept appears in the concept dependency map $\mathcal{G}_c$, and also plays an important role in the exercise-concept correlation map $\mathcal{G}_{qc}$.

In the $\mathcal{G}_c$, there exist multiple educational dependency relations $\mathcal{R}_c$ between knowledge concepts. For those directed dependencies, such as prerequisite relation [8] and remedial relation [38], the proficiency of the former one is expected to be higher than the latter [8, 29], where the influence propagates unidirectionally from predecessor concepts to successor concepts among directed dependencies. Therefore, we define this type of dependencies as the directed relations in $\mathcal{G}_c$, where concepts can interact with their predecessor concepts, not with successor concepts. While for those undirected dependencies, such as similarity relation [47] and collaboration relation [19], the promotion of the proficiency of a certain concept brought the promotion to its neighbor concepts and vice versa, which can be further explained as the influence is bidirectionally propagated between neighbor concepts. Inspired by these observations, we define this type of dependencies as the bidirectional relations in $\mathcal{G}_c$, where neighbor concepts can interact each other. Besides, in the $\mathcal{G}_{qc}$, concept c_k relates to a set of exercises, where concepts can interact with related exercises.

Thus, c_k is influenced by the concept dependency aggregation from $\mathcal{G}_c$ and the exercise correlation aggregation from $\mathcal{G}_{qc}$. Let $\tilde{\boldsymbol{p}}_r^{l+1}$ denote the aggregated embedding of concept dependency and $\tilde{\boldsymbol{e}}_k^{l+1}$ denoted the aggregated embedding of exercise correlation. Then, each concept's updated embedding $\boldsymbol{c}_k^{l+1}$ is modeled as:

$$\boldsymbol{c}_k^{l+1} = \boldsymbol{c}_k^l + \sum\nolimits_{r\in\mathcal{R}_c} \eta_{kr}^{l+1}\tilde{\boldsymbol{p}}_r^{l+1} + \eta_{k1}^{l+1}\tilde{\boldsymbol{e}}_k^{l+1}, \tag{2}$$

$$\tilde{\boldsymbol{p}}_r^{l+1} = \sum\nolimits_{a\in N_k^r} \alpha_{ka}^{l+1}\mathbf{W}_c{}^l\boldsymbol{c}_a^l, \tag{3}$$

$$\tilde{\boldsymbol{e}}_k^{l+1} = \sum\nolimits_{b\in N_k^q} \beta_{kb}^{l+1}\mathbf{W}_q{}^l\boldsymbol{q}_b^l, \tag{4}$$

where Eq. (2) models that how each concept c_k updates its embedding by fusing concept dependency aggregation $\tilde{\boldsymbol{p}}_r^{l+1}$ with certain educational relation r and exercise correlation aggregation $\tilde{\boldsymbol{e}}_k^{l+1}$, as well as its own embedding $\boldsymbol{c}_k^l$ from the previous iteration. Since each concept appears in both the concept dependency map and the exercise correlation map, Eq. (3, 4) model the concept dependency aggregation for each relation r and exercise correlation aggregation from the two local maps respectively. Specifically, N_k^r and N_k^q are the concept sets that contains c_k's neighbor concepts with relation r and exercise set that contains c_k's neighbor exercises respectively. $\mathbf{W}^l \in \mathbb{R}^{d\times d}$ is a trainable matrix for linear transformation.

In addition, there are three groups of weights (i.e., α_{ka}^{l+1}, β_{kb}^{l+1}, η_{kr}^{l+1} and η_{k1}^{l+1}) in the above three equations. They naturally present a multi-level structure. Specially, α_{ka}^{l+1} denotes the concept dependency strength between concept c_k and c_a, where c_k and c_a maintain the education dependency relation r in the concept dependency map. β_{kb}^{l+1} denotes the correlation strength between exercise q_b and concept c_k in the exercise-concept correlation map. These two weights can be seen as the node-level attentive weights, which model how each concept balances different relations in each local map. For the Eq. (2), η_{kr}^{l+1} and η_{k1}^{l+1} are the map-level weights which learn the contribution of each aspect with different maps. The map-level weights balance the educational dependency relation and the exercise correlation strength from different maps. To model the multi-level structure, we design a multi-level attention network to learn the attentive weights. Specifically, we first calculate the node-level weights as:

$$\alpha_{ka}^{l+1} = \boldsymbol{F}_{cc}([\mathbf{W_c}{}^l\boldsymbol{c}_k^l, \mathbf{W_c}{}^l\boldsymbol{c}_a^l]), \tag{5}$$

$$\beta_{kb}^{l+1} = \boldsymbol{F}_{cq}([\mathbf{W_q}{}^l\boldsymbol{c}_k^l, \mathbf{W_q}{}^l\boldsymbol{q}_b^l]), \tag{6}$$

where a full connection layer $\boldsymbol{F}(\cdot)$ is used to learn the node attention weights with concatenation $[\cdot]$, where bias is optional. After that, we normalize the attention weights and take α_{ka}^{l+1} as example:

$$\alpha_{ka}^{l+1} = \frac{\exp\left(\alpha_{ka}^{l+1}\right)}{\sum_{j\in N_k^r}\exp\left(\alpha_{kj}^{l+1}\right)}. \tag{7}$$

As all of them share the similar form as shown in Eq. (7), we omit the normalization step of all node- and map-level attention modeling in the following without confusion. After obtaining the node-level attention weights α_{ka}^{l+1} and β_{kb}^{l+1}, by taking the aggregation information $\tilde{\boldsymbol{p}}_r^{l+1}$ and $\tilde{\boldsymbol{e}}_k^{l+1}$ into Eq. (2), the map level attention weights η_{kr}^{l+1} and η_{k1}^{l+1} can be calculated as:

$$\eta_{kr}^{l+1} = \boldsymbol{F}_{cp}([\boldsymbol{c}_k^l, \tilde{\boldsymbol{p}}_r^{l+1}]), \quad \eta_{k1}^{l+1} = \boldsymbol{F}_{ce}([\boldsymbol{c}_k^l, \tilde{\boldsymbol{e}}_k^{l+1}]). \tag{8}$$

4.2.2 ***Exercise Fusion***. For each exercise q_d, let $\boldsymbol{q}_d^l$ denote its embedding at the l-th layer. As each exercise appears in both the exercise-concept correlation map G_{qc} and the student-exercise interaction map G_{sq}. Exercise q_d is influenced by the concept correlation aggregation from G_{qc} and student exercising interaction aggregation from G_{sq}. Specifically, in the G_{qc}, each exercise q_d has main concept set N_d^c and exercises can interact with contained concepts. In the G_{sq}, each exercise will be practiced by students and interacts with students by exercising behaviors and responses. Let $\tilde{\boldsymbol{u}}_d^{l+1}$ and $\tilde{\boldsymbol{v}}_d^{l+1}$ denote the aggregated embeddings of correlation and interaction respectively, similar to concept fusion, each exercise q_d's updated embedding $\boldsymbol{q}_d^{l+1}$ can be calculated as:

$$\boldsymbol{q}_d^{l+1} = \boldsymbol{q}_d^l + \gamma_{d1}^{l+1}\tilde{\boldsymbol{u}}_a^{l+1} + \gamma_{d2}^{l+1}\tilde{\boldsymbol{v}}_b^{l+1}, \tag{9}$$

$$\tilde{\boldsymbol{u}}_d^{l+1} = \sum\nolimits_{a\in N_d^c}\mu_{da}^{l+1}\mathbf{W_{cu}}{}^l\boldsymbol{c}_a^l, \tag{10}$$

$$\tilde{\boldsymbol{v}}_d^{l+1} = \sum\nolimits_{b\in N_d^s}\nu_{db}^{l+1}\mathbf{W_s}{}^l\boldsymbol{s}_b^l, \tag{11}$$

where Eq. (9) shows how each exercise q_d fuses concept correlation aggregation $\tilde{\boldsymbol{u}}_a^{l+1}$ and student interaction aggregation $\tilde{\boldsymbol{v}}_b^{l+1}$ from different maps, as well as its own embedding $\boldsymbol{q}_d^l$ from the previous iteration. N_d^c and N_d^s denotes the set of exercise q_d's related concepts and students respectively, and γ_{d1}^{l+1} and γ_{d2}^{l+1} are the map-level weights to balance the relation aggregation from G_{qc} and G_{sq}. μ_{ka}^{l+1} and ν_{kb}^{l+1} are the node-level weights to model how each exercise balances correlation aggregation and interaction aggregation in each local map. The above three groups of weights can be modeled similar to the forms as shown in Eq. (5 - 8).

4.2.3 ***Student Fusion***. Generally, at an online learning system, each student will choose some exercises for practicing, and she can obtain the responses (e.g., answer correctly or not). Thus, students directly interact with exercises in G_{sq}. Given each student s_z's l-th iteration embedding $\boldsymbol{s}_z^l$, then we model the updated student embedding $\boldsymbol{s}_z^{l+1}$ at the $(l+1)$-th iteration as:

$$\boldsymbol{s}_z^{l+1} = \boldsymbol{s}_z^l + \tilde{\boldsymbol{q}}_a^{l+1}, \quad \tilde{\boldsymbol{q}}_a^{l+1} = \sum\nolimits_{a\in N_z^q}\rho_{za}^{l+1}\mathbf{W_{qs}}{}^l\boldsymbol{q}_a^l, \tag{12}$$

where attention weight ρ_{za}^{l+1} balances the exercise aggregation for each student s_z and N_z^q is the exercise set practiced.

Generally, the fusion layer uniformly models the interactive and structural relations among student-exercise-concept after the iterative fusion process with L times.

4.3 Extendable Diagnosis Layer

With the fused relation-aware representations of students, exercises and concepts (i.e., $\boldsymbol{s}_z$, $\boldsymbol{q}_d$ and $\boldsymbol{c}_k$ respectively), we finally predict the student exercising performance and jointly train RCD via student response records in the diagnosis layer. In order to keep RCD as a general framework, the diagnosis layer is carefully designed to be compatible with a uniform diagnosis pattern followed by most previous works, which can be formally described as follows. For each record of student s_z practicing exercise q_d that contains n concepts $\{c_k\}_1^n$, the diagnosis layer predicts the probability of s_z answers q_d correctly as:

$$y_{q_d} = \mathcal{A}(\mathcal{S}(\boldsymbol{h}_z, \boldsymbol{h}_d)), where \quad \mathcal{S}(\boldsymbol{h}_z, \boldsymbol{h}_d) = \mathcal{T}(\mathcal{U}(\boldsymbol{h}_z - \boldsymbol{h}_d)). \tag{13}$$

$\boldsymbol{h}_z$ represents student factor (e.g., ability level) and $\boldsymbol{h}_d$ represents exercise factor (e.g. difficulty level). $\mathcal{S}(\cdot)$ is a scoring function which evaluates the student's performance upon each concept of the exercise, and $\mathcal{A}(\cdot)$ is an accumulation function, such as average operation, that selectively accumulates the utility according to the concepts contained in the exercise. Specifically, the score function is comprised of two parts. Firstly, a utility function $\mathcal{U}(\cdot)$ quantifies the advantage utility between the student and the exercise represented by $(\boldsymbol{h}_z - \boldsymbol{h}_d)$, e.g., comparing the student ability to the exercise difficulty. Then, a transform function $\mathcal{T}(\cdot)$ normalizes the score into a probabilistic prediction.

In the typical implementation of RCD, $\mathcal{A}(\cdot)$ is an average operation over the related concepts, $\mathcal{U}(\cdot)$ is a full connection layer denoted as $\boldsymbol{F}(\cdot)$ which outputs scalar, and $\mathcal{T}(\cdot)$ is the commonly used Sigmoid $\sigma(\cdot)$.

To generate the student factor $\boldsymbol{h}_z$ and exercise factor $\boldsymbol{h}_d$ from the relation-aware representations, we incorporate the concept embedding into the student and exercise embeddings by neural network $\boldsymbol{f}_{sc}$ and $\boldsymbol{f}_{qc}$. Formally, the diagnosis layer of RCD is:

$$y_{q_d} = \frac{1}{n}\sum\nolimits_{c_k}\sigma(\boldsymbol{F}(\boldsymbol{h}_z^{(k)} - \boldsymbol{h}_d^{(k)})), \tag{14}$$

$$\boldsymbol{h}_z^{(k)} = \sigma(\boldsymbol{f}_{sc}(\boldsymbol{s}_z \oplus \boldsymbol{c}_k)), \tag{15}$$

$$\boldsymbol{h}_d^{(k)} = \sigma(\boldsymbol{f}_{qc}(\boldsymbol{q}_d \oplus \boldsymbol{c}_k)). \tag{16}$$

We can jointly train RCD with the cross entropy loss between each diagnosis prediction y and its corresponding ground truth r:

$$\text{loss} = -\sum\nolimits_i \left(r_i \log y_i + (1 - r_i)\log(1 - y_i)\right). \tag{17}$$

By optimizing with the above loss, the vector of $\boldsymbol{h}_z^{(k)}$ in Eq. (15) denotes student s_z's knowledge proficiency on the concept c_k. If we just want to estimate her mastery value, inspired by [26] we mask the exercise factor $\boldsymbol{h}_d^{(k)}$ to zero as $\mathbf{0} = (0, 0, \ldots, 0)$, and obtain the mastery level estimation from Eq. (14):

Table 1: The statistics of the dataset.

Statistics	Junyi	ASSIST
#Students	10,000	2,493
#Exercises	835	17,746
#Knowledge concepts	835	123
#Response records	353,835	267,415
#Knowledge concepts per exercise	1	1.20
#Response records per student	35.38	107.27
#Prerequisite relations among concepts	988	1,164
#Similarity relations among concepts	1,040	1,256

$$m_z^{(k)} = \sigma(\boldsymbol{F}(\boldsymbol{h}_z^{(k)})). \tag{18}$$

Finally, we demonstrate how the diagnosis layer of RCD Eq. (14, 15, 16) also adapts to the general diagnosis pattern Eq. (13). In fact, we can utilize the outputs from the fusion layer to construct arbitrary form of CD methods. Here we give the general methods as examples. Generally, to extend an existing CD model, we replace the student factor and exercise factor with the ones output by RCD (i.e., $\boldsymbol{h}_z$ and $\boldsymbol{h}_d$), and then specialize the utility and accumulation functions. The extendibility over representative models are illustrated as follows:

IRT. Take the typical formation of IRT: $y = \sigma((h_{pro} - h_{diff}) \times h_{disc})$ as example, where h_{pro}, h_{diff} and h_{disc} demote student proficiency, exercise difficult and discrimination respectively. To extend from IRT, in the utility function $\mathcal{U}(\cdot)$, we project $\boldsymbol{h}_z$ and $\boldsymbol{h}_d$ to unidimensional h_{pro} and h_{diff} respectively, and set h_{disc} as additional trainable parameters; as for the accumulation function $\mathcal{A}(\cdot)$, we simply set it as identity.

MIRT. MIRT is a multidimensional extension of IRT [2]: $y = \sigma(\boldsymbol{Q_e} \cdot (\boldsymbol{h}^s - \boldsymbol{h}^e))$, where $\boldsymbol{h}^s$ and $\boldsymbol{h}^e$ are latent trait vectors of students and exercises, $\boldsymbol{Q_e}$ is the one-hot index vector of related concepts for exercise e. To extend from MIRT, we just replace $\boldsymbol{h}^s$ and $\boldsymbol{h}^e$ with our $\boldsymbol{h}_z$ and $\boldsymbol{h}_d$, set the utility function $\mathcal{U}(\cdot)$ as identity, and change the accumulation function $\mathcal{A}(\cdot)$ from the average operation to the sum operation.

NCD. NCD models the student-exercise interaction more generally as: $y = F(\boldsymbol{Q_e} \circ \left(\boldsymbol{h}^s - \boldsymbol{h}^{diff}\right) \times h^{disc})$, where F denotes multi-layer full connection layer. Similar to IRT and MIRT, we replace $\boldsymbol{h}^s$ and $\boldsymbol{h}^{diff}$ with our $\boldsymbol{h}_z$ and $\boldsymbol{h}_d$, treat h^{disc} as trainable parameters in the neural network based utility function $\mathcal{U}(\cdot)$, and set the accumulation function $\mathcal{A}(\cdot)$ as the sum operation.

MF. According to [44], MF can be treated as a special case of the NCD model where $\boldsymbol{h}^{diff} \equiv \boldsymbol{0}$ and $h^{disc} \equiv \boldsymbol{1}$. Therefore, extending from MF is straightforward. In addition, we can also implement MF by factoring score matrix to $\boldsymbol{h}_z$ and $\boldsymbol{h}_d$.

5 EXPERIMENTS

As the key contribution of this work is the design of relation map with interactive and structural information for accurate and explainable cognitive diagnosis (CD), we conduct experiments to answer the following research questions:

RQ1: Comparing with existing CD models, how does our proposed RCD perform in terms of accuracy?

RQ2: How about the model's performance when utilizing RCD blend interactions and structural information into current methods?

RQ3: How about the effectiveness of modeling the multi-level student-exercise-concept interaction by RCD?

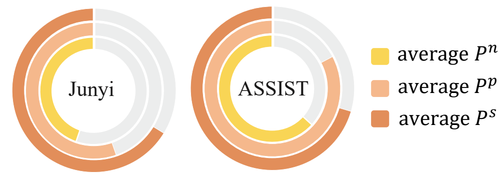


Figure 3: Correct rate comparison.

RQ4: How about the ability of RCD to capture student-exercise-concept interaction in relation map?

RQ5: How about the interpretation of RCD on diagnosing student knowledge states for CD?

5.1 Dataset Description

We conduct experiments on two real-world datasets, i.e., Junyi[1] and ASSIST[2], which both contain learners' exercising performance records and exercise-concept correlations. Besides, Junyi provides the concept dependency relations labeled by experts. In these two learning systems, students are allowed to resubmit their answers until they passed the assessment. Since these original datasets are not suitable for static diagnosis, we first preprocess each dataset by only employing the first submission for multiple submissions per student inspired by [21]. To guarantee that each student has enough data for diagnosis, we remove the students who attempted fewer than 15 response records for each dataset similar to [31, 44]. The complete statistical information for datasets is depicted in Table 1.

5.1.1 *Junyi.*

For Junyi dataset, the concept dependency relations contain the prerequisite and similarity. In prerequisite relation, each concept pair (c_i, c_j) indicates the former is the prerequisite of the latter, e.g., (*one_digit_addition, one_digit_multiplication*). In similarity relation, as the original data format is like $(c_i, c_j, value)$ where *value* indicates the strength of similarity between two concepts and $1 \le value \le 9$, e.g., (*vertical_angles, angle_types, 5.22*), we set the thresholds as 5.0 to get the similarity relations, i.e., c_i and c_j are similar if $value \ge 5.0$. In addition, we randomly select 10,000 learners' response records to yield the dataset similar to [49].

Furthermore, we conduct analysis based on response records to verify the existence of the educational dependency. Inspired by [29, 40], we define the following probability equation for concept pairs (c_i, c_j): $P_{ij} = n_c(c_j|c_i)/n(c_j|c_i)$ where $n(c_j|c_i)$ and $n_c(c_j|c_i)$ respectively are the counts that concept c_j is answered and concept c_j is answered correctly immediately after c_i is correctly answered. We respectively calculate the probability for prerequisite and similarity relations and denote them as P_{ij}^p and P_{ij}^s. Then, let $P_j^n = n_c(c_j)/n(c_j)$ denote the non-conditional correctness probability, where $n_c(c_j)$ and $n(c_j)$ respectively are the count that c_j is correctly answered and c_j is answered. As shown in Figure 3, we can obtain that the probability of student correctly answer will be improved after mastering related knowledge concepts. This result shows that there are educational dependencies between concepts.

5.1.2 *ASSIST.*

For ASSIST dataset, the educational dependencies are not provided in original data explicitly. To construct the concept structure, we provide an implementation method based on certain

[1] https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=1198
[2] https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data/skill-builder-data-2009-2010

statistics inspired by [29, 40]. Here, we define two types of educational relations (i.e., the prerequisite relation and the similarity relation) similar to Junyi, and introduce the following matrices to assist in constructing the concept dependency map.

Correct matrix is a probability matrix denoted as C. We calculate the matrix C based on exercising records, where $C_{i,j}$ is $\frac{n_{i,j}}{\sum_k n_{i,k}}$ if $i \neq j$; else, it is 0. Here, $n_{i,j}$ represents the count that concept j is answered correctly and immediately after i is answered correctly.

Transition matrix is a binary transition matrix denoted as T, where $T_{i,j} = 1$ indicates concept i has an edge pointing to j. To obtain the transition matrix, we first calculate $\tilde{T}$: $\tilde{T}_{ij} = \frac{C_{ij} - min(C)}{max(C) - min(C)}$. $\tilde{T}_{ij}$ is the probability that the existence of certain educational relation between concept i and concept j. Then, we determine the relations by $T_{i,j} = 1$ if $\tilde{T}_{ij} > threshold$. In this paper, we set $threshold$ as third power of the average value of matrix T.

Concept dependency map is the concept dependent matrix. Specially, the dependency r_{ij} between concept c_i and concept c_j denotes the prerequisite, if $T_{i,j} = 1$ but $T_{j,i} \neq 1$, which denotes that c_j is a successor of c_i. And r_{ij} denotes the similarity, if $T_{i,j} \times T_{j,i} = 1$.

5.2 Experimental Settings

5.2.1 Baselines and Evaluation Metrics. To verify the effectiveness of our proposed RCD model, we compare it with some baselines. Specifically, the compared baselines are selected from two aspects. One is the representative methods in CD field like IRT, MIRT, MF and NCD. The details are illustrated as follows:

- **IRT** [13] as the most popular CD method, models unidimensional students and exercises' feature by a linear function.
- **MIRT** [1] is a multidimensional extension of IRT, modeling multiple knowledge proficiency of students and exercises.
- **MF** [28, 41] predicts student performance by factoring score matrix and get students and exercises' latent trait vectors.
- **NCD** [44] is one of the most recent deep learning CD models. It models high-order and complex student-exercise interaction functions with neural networks.

The other aspect is competitive heterogeneous graph representation learning methods as illustrated in follows:

- **metapath2vec** [11] designs a meta-path based random walk and utilizes skip-gram to learn heterogeneous graph.
- **GTN** [50] is a semi-supervised method which can automatically learn useful meta-paths. We generate the adjacency matrix of the meta-paths based on local maps.
- **HAN** [46] leverages structure and node features for network embedding. We set multiple meta-paths for better HAN.
- **HetG** [51] samples node-level information based random walk with restart and fuses these nodes by LSTM [18] aggregation. We implement HetG by setting the depth of each path as 5 and the number of sampling for each node as 20.

To measure the performance of our model, we adopt different metrics from the perspectives of both regression and classification. From the regression perspective, we select *Root Mean Square Error* (RMSE) [32] to quantify the distance between predicted scores (i.e., continuous variable ranges from 0 and 1) and actual ones. From the classification perspective, we consider that students answer incorrectly or correctly can be represented 0 and 1 respectively. Thus, we use *Prediction Accuracy* (ACC) [44] and *Area Under an ROC Curve* (AUC) [5] for model evaluation.

5.2.2 Parameter Settings. To set up training process, we initialize all network parameters with Xavier initialization [29]. Each parameter is sampled from $U(-\sqrt{2/(n_{in} + n_{out})}, \sqrt{2/(n_{in} + n_{out})})$, where n_{in} and n_{out} denote the numbers of neurons feeding in and feeding out, respectively. We use the Adam algorithm [23] for optimization and set mini-batch size as 256. To be fair, we set the embedding dimension of student, exercise and concept as the number of concepts for all relation map driven methods as well as NCD Besides, we just set the iterations of fusion operation as 2 to explore the high-order complex student-exercise-concept interaction.

We implement all models with PyTorch by Python and conduct our experiments on a Linux server with four 2.0GHz Intel Xeon E5-2620 CPUs and a Tesla K2/0m GPU. All models were tuned to have the best performance to ensure the fairness[3].

5.3 Performance Comparison (RQ1 and RQ2)

To answer the RQ1 and RQ2, we compare the performance of RCD with several baselines on student performance prediction. Besides, we implement some relation-based methods (i.e., R-IRT, R-MIRT, R-MF and R-NCD) by blending structural information into current methods based on RCD. Specially, we additionally fuse concept factors into student and exercise factors similar to Eq. (15, 16) by neural networks for improving performance, and other settings are appropriately reinforced by neural networks based on the description in subsection 4.3. Table 2 reports the overall performance of all models for student performance prediction and the best scores are denoted in bold. There are several observations. Firstly, RCD consistently performs the best at all data sparsities, which indicates that modeling the multi-level student-exercise-concept interaction for the CD task can achieve outstanding performance, which can answer RQ1. Secondly, almost all relation map based methods perform better than baselines which do not consider such the complex interaction. This observation demonstrates that it is necessary to model complete and comprehensive interactions among students, exercises and concepts for the CD and answers RQ2. Lastly, NCD performs best in baselines, probably because NCD models high-order student-exercise relationships which have a satisfactory approximation ability than those linearly modeling methods.

These evidence demonstrates that integrating the multi-level interaction relations can improve accuracy for student performance prediction. In the following, we further conduct experiments on the training data to verify the modeling ability of RCD.

5.4 The Effectiveness of Multi-level Relations Modeling (RQ3)

To answer the RQ3, we compare RCD with several competitive heterogeneous graph representation learning methods on the performance prediction task. To be specific, we employ different graph learning methods (i.e., HetG, HAN, GTN and metapath2vec) in fusion layer and set the same diagnosis layer. As listed in Figure 4, RCD achieves the best scores. This result demonstrates that the multi-level fusion modeling is suitable to model multi-layer student-exercise-concept interaction. Meanwhile, HAN and HetG

[3]The code is available at https://github.com/bigdata-ustc/RCD

Table 2: Experimental results on student performance prediction.

(a) Junyi

Train/Test radio		50%/50%			60%/40%			70%/30%			80%/20%		
Methods		ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE
Baseline	IRT	67.56%	75.12%	44.53%	67.46%	76.09%	43.65%	67.45%	76.66%	43.13%	67.60%	77.50%	42.68%
	MIRT	74.70%	79.16%	41.43%	74.83%	79.32%	41.36%	74.81%	79.47%	41.35%	75.13%	79.89%	41.17%
	MF	66.87%	74.43%	45.20%	67.35%	74.98%	44.79%	67.81%	75.46%	44.37%	68.34%	76.44%	43.73%
	NCD	74.21%	78.68%	42.01%	74.40%	79.05%	41.67%	74.30%	78.95%	41.78%	74.43%	79.09%	41.72%
Relation-based	R-IRT	74.64%	79.12%	41.51%	74.67%	79.18%	41.44%	74.71%	79.35%	41.39%	74.68%	79.43%	41.36%
	R-MIRT	76.60%	81.82%	40.14%	76.59%	81.71%	40.15%	76.67&	81.87%	40.15%	76.48%	82.23%	40.11%
	R-MF	74.38%	78.79%	41.63%	74.33%	78.75%	41.66%	74.54%	78.85%	41.60%	74.43%	79.06%	41.59%
	R-NCD	74.48%	78.94%	41.90%	74.56%	78.86%	42.00%	74.61%	79.07%	41.85%	74.66%	79.31%	41.59%
	RCD	**76.61%**	**81.84%**	**40.05%**	**76.66%**	**81.88%**	**40.10%**	**77.01%**	**82.26%**	**39.82%**	**77.16%**	**82.62%**	**39.63%**

(b) ASSIST

Train/Test radio		50%/50%			60%/40%			70%/30%			80%/20%		
Methods		ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE	ACC	AUC	RMSE
Baseline	IRT	63.46%	66.46%	49.10%	63.76%	67.81%	48.13%	64.00%	68.79%	47.36%	64.26%	69.83%	46.59%
	MIRT	67.41%	67.85%	47.98%	68.53%	71.18%	48.39%	70.39%	72.83%	46.03%	71.70%	74.94%	45.17%
	MF	61.84%	70.09%	49.34%	64.23%	72.85%	47.13%	65.67%	74.62%	45.86%	67.12%	76.45%	44.51%
	NCD	71.51%	73.77%	43.90%	71.85%	74.21%	43.81%	72.17%	74.92%	43.78%	73.14%	75.94%	43.08%
Relation-based	R-IRT	69.53%	69.43%	45.15%	70.44%	72.59%	44.11%	71.43%	74.08%	43.58%	69.53%	69.43%	45.15%
	R-MIRT	72.82%	76.13%	42.59	72.50%	76.18%	42.73%	72.90%	76.55%	42.49%	73.40%	77.20%	42.22%
	R-MF	66.99%	66.90%	46.23%	67.54%	67.53%	45.98%	71.96%	74.85%	44.45%	71.61%	74.35%	43.70%
	R-NCD	71.46%	74.42%	43.61%	71.24%	73.84%	43.73%	71.78%	74.62%	43.75%	72.39%	75.32%	43.25%
	RCD	**72.87%**	**76.24%**	**42.50%**	**72.82%**	**76.19%**	**42.54%**	**73.11%**	**76.63%**	**42.41%**	**73.55%**	**77.21%**	**42.13%**

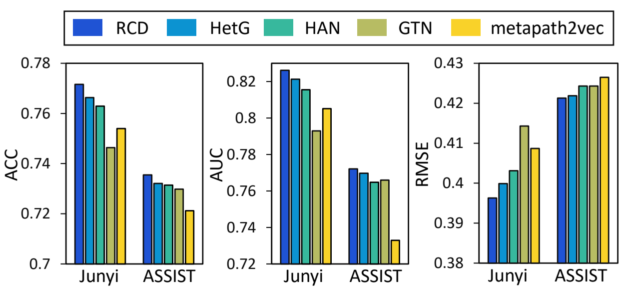
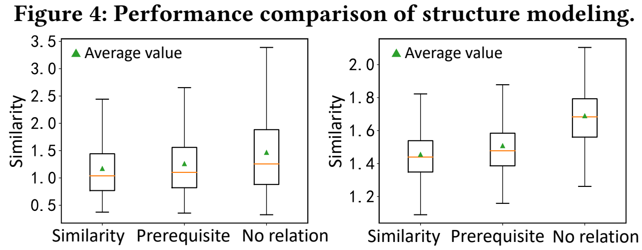


Figure 4: Performance comparison of structure modeling.



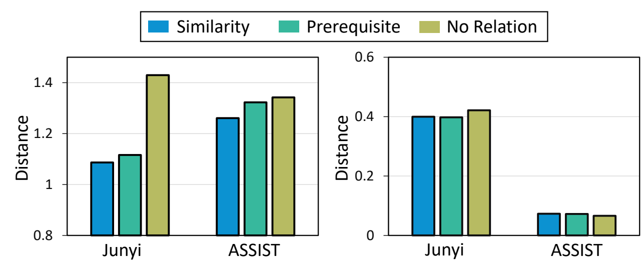
(a) Junyi (b) ASSIST

Figure 5: Results comparison of concept similarity.

also achieve good performance as their carefully designed hierarchical aggregation functions. These results can prove that our proposed RCD model has outstanding modeling abilities.

5.5 Relation Map Analysis (RQ4)

To answer RQ4, we conduct some interesting visualization experiments on concept dependency map, exercise-concept correlation map and student-exercise interaction map respectively, to demonstrate the ability of modeling interaction relations of RCD.



(a) RCD Model (b) Model Without Relation Fusion

Figure 6: Comparison of exercise-concept correlation study.

5.5.1 Concept Similarity. In concept dependency map, we explore the similarity between knowledge concepts. Specifically, after obtaining each concept embedding by trained model, we measure the similarity by Euclidean Distance between each pair of concepts. We respectively calculate the similarity between concepts with similarity and prerequisite relations as well as no explicit relations. Here we define that concepts have no relations if they exceed two hops. Figure 5 illustrates the results of concept pairs in the forms of box figures. From the figure, we can see that concepts with similar relations have the closest similarity on each dataset. Meanwhile, concepts with similar or prerequisite relations are closer than concepts without explicit relations. These results clearly demonstrate that RCD can capture the relations between concepts well.

5.5.2 Exercise-Concept Correlation. In the exercise-concept correlation map, we aim to analyze the relations between exercises and concepts learned from RCD. Specifically, between each pair of exercise (q_i, q_j), there exist three types of relations: (1) q_i's main concept is similar to q_j; (2) q_i's main concept is the prerequisite

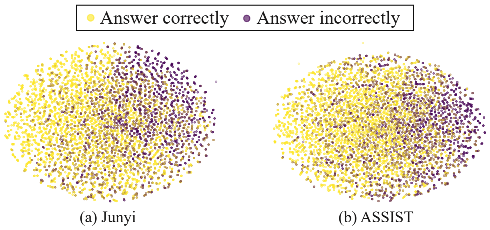


Figure 7: The visualization of interactive response relations between students and exercises.

of q_j and vice versa; (3) there is no explicit relation between q_i and q_j's concepts. In this part, we use all exercises in Junyi and randomly sample 2,000 exercises in ASSIST. We partitioned exercise pairs into three groups according to these three relations and calculate the Euclidean Distance for each pair of exercises. Besides, we also measure distances between exercises learned from the model without relation fusing process for comparison. Figure 6 illustrates the results which demonstrate that: (1) the distances between exercises learned by RCD, with similar, prerequisite, and no relation show a clear increasing trend; (2) for model without relation fusion, it is difficult to distinguish different relations between exercises. These findings demonstrate that RCD has a good ability to learn correlations between exercises and knowledge concepts.

5.5.3 Student Exercising Interaction. In the student-exercise interaction map, we aim to explore whether RCD can capture interactions between students and exercises. Specifically, inspired by [15], we first define each student-exercise interaction relation r_{sq}'s embedding as: $f(r_{sq}) = f(s) - f(q)$, where $f(s)$ and $f(q)$ denote student s and exercise q's embedding representations learned from RCD respectively. As exercising interaction relations can reflect students' performance, we explicitly label each relation r_{sq} with "*correctly*" or "*incorrectly*" based on the true performance data. Then we project all the student exercising responses provided from datasets into vector $f(r_{sq})$ and reduce the vector dimension to two-dimension space utilizing the T-SNE method [42]. We randomly select 10,000 interactions for visualization. As shown in Figure. 7, the interactions in the same color have the same label. We can see that the interactions that answer correctly are mainly clustered on the left side of the figure, and the interactions that answer incorrectly are mainly clustered on the other side. This shows that our relation map can capture the interactive relations between students and exercises. Besides, there are also some messy interaction points in the figure. This is because the student exercising response is a complex process and it is not accurate enough to estimate students' performance only based on their states on concepts. For example, even if a student does not master some certain concepts, she may answer related exercises correctly by coincidence. We will continue to explore the deeper interactions in the future.

5.6 Interpretation of the Diagnosis (RQ5)

We visualize the diagnostic reports and evaluate the interpretation of RCD. This visualization helps students and teachers recognize the former's knowledge state, efficiently and intuitively. Here, we evaluate the interpretability based on whether the diagnostic proficiency of students is reasonable with the given educational dependencies in concept dependency map. We randomly sampled a student and depict her proficiency of a subset of concepts in Figure 8.

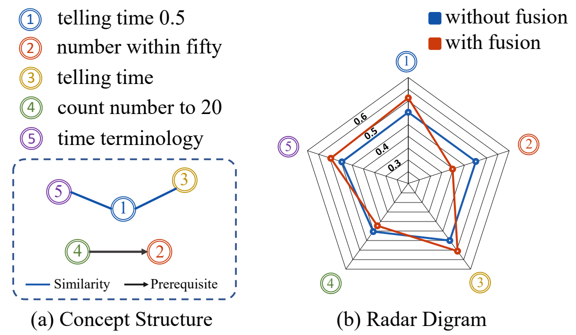


Figure 8: The part of concept structure and the visualization of student diagnostic report.

Figure 8 (a) shows the dependency structure of a subset of concepts, which contains similarity and prerequisite relations. Figure 8 (b) shows the diagnostic result and each point on the radar diagram represents the mastery level of the certain knowledge concept. Specially, the red line (*with fusion*) denotes the diagnostic proficiency by integrating the relation map, provided by a trained model, and the blue line (*without fusion*) denotes the proficiency which ignores the fusing process. From the figure, we can obtain that the proficiency of concept ① (*alternate exterior angles*), ③ (*parallel lines 1*) and ⑤ (*segment addition*) gets promoted after fusing concept dependency relations, where concept ③ and concept ⑤ serve as the similar neighbor of ①. Meanwhile, concept ④ (*same side exterior angles*) is stable while concept ② (*corresponding angles*) significantly reduces, where ② is a successor of ④. These observations indicate that: (1) the proficiency of similar knowledge concepts can affect each other; (2) the influence along prerequisite relations is only unidirectionally propagated in RCD. From the results, we can see that, owing to the ability of modeling the concept dependencies, RCD is able to provide a better interpretable insight on diagnosing student knowledge states for CD.

6 CONCLUSION

In this paper, we proposed a general cognitive diagnosis (CD) framework, namely **R**elation map driven **C**ognitive **D**iagnosis (RCD), which comprehensively considers the student-exercise-concept relations for cognitive diagnosis. Specifically, we first represented students, exercises and concepts as individual nodes in a hierarchical layout, and constructed three well-defined local relation maps to incorporate inter- and inner-layer relations. Then, we designed a multi-level attention network to integrate node-level relation aggregation in each local map and balance map-level aggregation from different maps. After that, we designed an extendable diagnosis function to predict students' performance. Finally, extensive experiments on real-world datasets clearly showed the effectiveness and extendibility of our RCD framework. We hope this work could lead to further studies.

Acknowledgements. This research was partially supported by grants from the National Natural Science Foundation of China (Grants No. 61922073, U20A20229 and 61727809), the Foundation of State Key Laboratory of Cognitive Intelligence (No. iED2020-M004), and the Iflytek joint research program.

REFERENCES

[1] Terry A Ackerman, Mark J Gierl, and Cindy M Walker. 2003. Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice* 22, 3 (2003), 37–51.

[2] Raymond J Adams, Mark Wilson, and Wen-chung Wang. 1997. The multidimensional random coefficients multinomial logit model. *Applied psychological measurement* 21, 1 (1997), 1–23.

[3] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web*. 687–698.

[4] Haoyang Bi, Haiping Ma, Zhenya Huang, Yu Yin, Qi Liu, Enhong Chen, Yu Su, and Shijin Wang. 2020. Quality meets Diversity: A Model-Agnostic Framework for Computerized Adaptive Testing. In *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 42–51.

[5] 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.

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

[7] Lei Chen, Le Wu, Richang Hong, Kun Zhang, and Meng Wang. 2020. Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 27–34.

[8] 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)*. IEEE, 39–48.

[9] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. 2018. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering* 31, 5 (2018), 833–852.

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

[11] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 135–144.

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

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

[14] Hongyang Gao, Zhengyang Wang, and Shuiwang Ji. 2018. Large-scale learnable graph convolutional networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1416–1424.

[15] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.

[16] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*. 1024–1034.

[17] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 639–648.

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[19] Xiaoqing Huang, Qi Liu, Chao Wang, Haoyu Han, Jianhui Ma, Enhong Chen, Yu Su, and Shijin Wang. 2019. Constructing Educational Concept Maps with Multiple Relationships from Multi-Source Data. In *2019 IEEE ICDM*. IEEE, 1108–1113.

[20] Zhenya Huang, Qi Liu, Yuying Chen, Le Wu, Keli Xiao, Enhong Chen, Haiping Ma, and Guoping Hu. 2020. Learning or forgetting? A dynamic approach for tracking the knowledge proficiency of students. *ACM Transactions on Information Systems (TOIS)* 38, 2 (2020), 1–33.

[21] Zhenya Huang, Qi Liu, Chengxiang Zhai, Yu Yin, Enhong Chen, Weibo Gao, and Guoping Hu. 2019. Exploring multi-objective exercise recommendations in online education systems. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1261–1270.

[22] Mohammad Khajah, Rowan Wing, Robert Lindsey, and Michael Mozer. 2014. Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In *Educational Data Mining 2014*. Citeseer.

[23] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[24] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[25] George D Kuh, Jillian Kinzie, Jennifer A Buckley, Brian K Bridges, and John C Hayek. 2011. *Piecing together the student success puzzle: Research, propositions, and recommendations: ASHE higher education report*. Vol. 116. John Wiley & Sons.

[26] 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.

[27] 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.

[28] Andriy Mnih and Russ R Salakhutdinov. 2007. Probabilistic matrix factorization. *Advances in neural information processing systems* 20 (2007), 1257–1264.

[29] 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)*. IEEE, 156–163.

[30] Joseph D Novak. 2010. *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Routledge.

[31] Shalini Pandey and Jaideep Srivastava. 2020. RKT: Relation-Aware Self-Attention for Knowledge Tracing. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1205–1214.

[32] Hongbin Pei, Bo Yang, Jiming Liu, and Lei Dong. 2017. Group sparse bayesian learning for active surveillance on epidemic dynamics. *arXiv preprint arXiv:1712.00328* (2017).

[33] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

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

[35] William F Pinar, William M Reynolds, Patrick Slattery, Peter M Taubman, et al. 1995. *Understanding curriculum: An introduction to the study of historical and contemporary curriculum discourses*. Vol. 17. Peter Lang.

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

[37] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE Transactions on Neural Networks* 20, 1 (2008), 61–80.

[38] Yi Shang, Hongchi Shi, and Su-Shing Chen. 2001. An intelligent distributed environment for active learning. *Journal on Educational Resources in Computing (JERIC)* 1, 2es (2001), 4–es.

[39] Shan-Yun Teng, Jundong Li, Lo Pang-Yun Ting, Kun-Ta Chuang, and Huan Liu. 2018. Interactive unknowns recommendation in e-learning systems. In *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 497–506.

[40] 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.

[41] Andreas Toscher and Michael Jahrer. 2010. Collaborative filtering applied to educational data mining. *KDD cup* (2010).

[42] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[43] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[44] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang. 2020. Neural cognitive diagnosis for intelligent education systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 6153–6161.

[45] Hao Wang, Tong Xu, Qi Liu, Defu Lian, Enhong Chen, Dongfang Du, Han Wu, and Wen Su. 2019. MCNE: An end-to-end framework for learning multiple conditional network representations of social network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1064–1072.

[46] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The World Wide Web Conference*. 2022–2032.

[47] Zhiwei Wang, Xiaoqin Feng, Jiliang Tang, Gale Yan Huang, and Zitao Liu. 2019. Deep knowledge tracing with side information. In *International Conference on AIEd*. Springer, 303–308.

[48] Robert S Woodworth and EL Thorndike. 1901. The influence of improvement in one mental function upon the efficiency of other functions.(I). *Psychological review* 8, 3 (1901), 247.

[49] Haiqin Yang and Lap Pong Cheung. 2018. Implicit heterogeneous features embedding in deep knowledge tracing. *Cognitive Computation* 10, 1 (2018), 3–14.

[50] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. 2019. Graph transformer networks. In *Advances in Neural Information Processing Systems*. 11983–11993.

[51] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 793–803.

[52] Yuhao Zhang, Peng Qi, and Christopher D Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. *arXiv preprint arXiv:1809.10185* (2018).