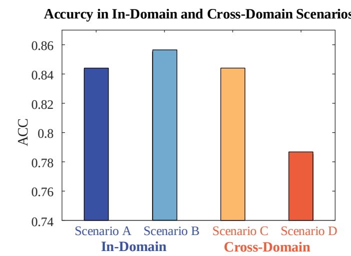
Prompt Transfer for Dual-Aspect Cross Domain Cognitive Diagnosis

Fei Liu[1,2], Yizhong Zhang[1,2], Shuochen Liu[3], Shengwei Ji[1,4], Kui Yu[1,2], and Le Wu *,[1,2]

[1]Key Laboratory of Knowledge Engineering with Big Data, the Ministry of Education of China, Hefei, China
[2]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China
[3]School of Computer Science and Technology, University of Science and Technology of China, Hefei, China
[4]School of Big Data And Artificial Intelligence, Hefei University, Hefei, China

***Abstract*—Cognitive Diagnosis (CD) aims to evaluate students' cognitive states based on their interaction data, enabling downstream applications such as exercise recommendation and personalized learning guidance. However, existing methods often struggle with accuracy drops in cross-domain cognitive diagnosis (CDCD), a practical yet challenging task. While some efforts have explored exercise-aspect CDCD, such as cross-subject scenarios, they fail to address the broader dual-aspect nature of CDCD, encompassing both student- and exercise-aspect variations. This diversity creates significant challenges in developing a scenario-agnostic framework. To address these gaps, we propose PromptCD, a simple yet effective framework that leverages soft prompt transfer for cognitive diagnosis. PromptCD is designed to adapt seamlessly across diverse CDCD scenarios, introducing PromptCD-S for student-aspect CDCD and PromptCD-E for exercise-aspect CDCD. Extensive experiments on real-world datasets demonstrate the robustness and effectiveness of PromptCD, consistently achieving superior performance across various CDCD scenarios. Our work offers a unified and generalizable approach to CDCD, advancing both theoretical and practical understanding in this critical domain. The implementation of our framework is publicly available at https://github.com/Publisher-PromptCD/PromptCD.**

***Index Terms*—Educational Data Mining, Cognitive Diagnosis, Cross-Domain, Prompt Transfer.**



Scenario Description

Scenarios	Source	Target
Scenario A	20% Mathematics	Mathematics
Scenario B	60% Mathematics	Mathematics
Scenario C	Physics 20% Mathematics	Mathematics
Scenario D	Chinese 20% Mathematics	Mathematics

Fig. 1. Performance comparison of MIRT in in-domain scenarios A and B versus cross-domain scenarios C and D.

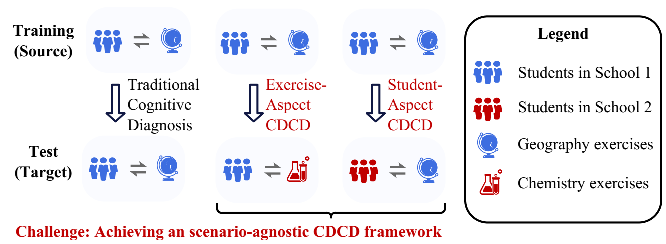


Fig. 2. Traditional cognitive diagnosis versus cross-domain cognitive diagnosis.

I. Introduction

Cognitive diagnosis aims to assess students' proficiency based on historical interactions [1]–[3]. It is a crucial task in educational data mining, supporting many downstream tasks like exercise recommendation [4], [5], learning guidance [6]–[10], and computerized adaptive testing [11].

In recent efforts, the primary focus has been on enhancing the accuracy of cognitive diagnosis models [12]. Specifically, these models aim to learn the characteristics of students and exercises from training data and utilize these learned representations to predict scores on test data [13]–[18]. Despite advancements in these models, they rely on the assumption that student and exercise characteristics are consistent across training and test data, which can be referred to as *in-domain cognitive diagnosis*. However, the assumption mentioned above is strict in practice. Students and exercises with diverse characteristics create diverse domains. For instance, students from different schools or countries, while exercises span various subjects. Thus, considering ***C**ross-**D**omain **C**ognitive **D**iagnosis (CDCD)* for students or exercises with various characteristics is more practical, however, posing many challenges to existing models.

First, representations of students or exercises learned from training data (source domains) cannot be directly applied to testing data (target domains), leading to a sharp decline in diagnostic accuracy. A straightforward approach to this limitation is to retrain the model. However, retraining solely on new domain data often results in overfitting and catastrophic forgetting, severely compromising the model's generalization capabilities. Alternatively, retraining with all existing and new data is computationally intensive and impractical for real-world applications. To illustrate these limitations, we conducted experiments using MIRT [19] on the SLP dataset [20] in in-domain (A and B) and cross-domain (C and D) scenarios. Figure 1 provides detailed scenario descriptions, and the results highlight the following: 1) Cross-Domain

* Corresponding author.

This work was supported in part by grants from the National Science and Technology Major Project (under Grant 2021ZD0111802), the National Natural Science Foundation of China (under Grants 62406096, 72188101, 62376086, U23B2031, 721881011, 62306100), the China Postdoctoral Science Foundation (under Grant No. 2024M760722), the Anhui Postdoctoral Scientific Research Program Foundation (under Grant No. 2024C934), the Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China) (under Grant No. BigKEOpen2025-01), and the Fundamental Research Funds for the Central Universities (under Grant No. JZ2024HGQB0093).

Challenges: C and D performed significantly worse than A and B, indicating the struggles of traditional models in CDCD scenarios. 2) Overfitting Risks: A performed worse than B, highlighting the impact of overfitting when retraining with limited data. 3) Sensitivity to Domain Differences: D significantly underperformed C due to greater distributional differences between Chinese and mathematics compared to physics and mathematics, underscoring the sensitivity of traditional models to source domains. These observations demonstrate that retraining cognitive diagnosis models is impractical, necessitating a better approach to mitigate accuracy declines in CDCD.

Second, CDCD scenarios are inherently complex due to the diverse characteristics of students and exercises, creating a variety of domains. As shown in Figure 2, CDCD can be categorized into two types: 1) **Student-Aspect CDCD**: Differences arise from varying student demographics, such as urban versus rural populations. 2) **Exercise-Aspect CDCD**: Variations occur across subject domains, such as mathematics and physics. Unfortunately, CDCD remains an underexplored area. Existing studies [21], [22] have primarily focused on specific CDCD aspects, proposing scenario-specific models with limited compatibility. Developing a generalizable framework capable of addressing both student- and exercise-aspect CDCD scenarios remains a significant challenge.

In this paper, we propose PromptCD, a simple yet generalizable framework designed to address the challenges of dual-aspect cross-domain cognitive diagnosis (CDCD). Dual-aspect CDCD introduces unique complexities, as knowledge transfer must consider both student-aspect and exercise-aspect scenarios. These scenarios involve distinct challenges: 1) Students and exercises across domains can be either overlapping or non-overlapping. Overlapping entities require personalized adaptation to maintain consistency, while non-overlapping entities necessitate a generalized representation to ensure effective knowledge transfer. 2) Diverse target domains vary significantly in their characteristics, making it difficult to adapt representations while preserving diagnostic accuracy and avoiding issues like overfitting or catastrophic forgetting. To address these challenges, PromptCD introduces a unified framework leveraging soft prompt transfer, a proven technique in cross-domain tasks across various fields [23]–[28]. Specifically, we design personalized prompts for overlapping entities and shared domain-adaptive prompts for non-overlapping entities. These prompts enhance representation learning and transfer, ensuring robustness in diverse CDCD scenarios. Additionally, PromptCD adopts a two-stage training strategy—pre-training on source domains and fine-tuning on target domains—for efficient and scalable adaptation. To demonstrate its versatility, we develop PromptCD-S and PromptCD-E, tailored to student-aspect and exercise-aspect CDCD scenarios, respectively. We summarize the contributions of this paper as follows:

- We propose the PromptCD framework, introducing soft prompt transfer and a two-stage training strategy to address dual-aspect CDCD challenges.
- We develop PromptCD-S and PromptCD-E, showcasing the framework's ability to generalize across student- and exercise-aspect scenarios.
- Extensive experiments on real-world datasets validate the effectiveness of PromptCD, achieving significant performance improvements over baselines.

II. Preliminaries

A. Cognitive Diagnosis

Cognitive diagnosis aims to evaluate students' proficiency in knowledge concepts based on their response records $\boldsymbol{L}$. This task involves modeling the interaction between student features $\boldsymbol{\alpha}$ and exercise features $\boldsymbol{\beta}$ to predict scores. Since students' proficiency is not directly observable, the model is trained to optimize predictive accuracy using the cross-entropy loss $\mathcal{L}_{CE}$. Below, we outline key interaction functions used in classic cognitive diagnosis models.

IRT [29] and MIRT [19] use the logistic function in a unidimensional and multidimensional manner, respectively. The detailed interaction functions are as follows: $y_{uv} = \frac{1}{1+e^{-C*\boldsymbol{D}_v(\boldsymbol{\alpha}_u-\boldsymbol{\beta}_v)}}$ and $y_{uv} = \frac{1}{1+e^{-\boldsymbol{\alpha}_u^T\boldsymbol{\beta}_v+D_v}}$, where D_v is discrimination of exercise v. C is a constant. NeuralCD [2], [15] utilizes neural networks to model the complex interactions between representations of students and exercises as follows: $\boldsymbol{x}_{uv} = \boldsymbol{Q}_v \circ (\boldsymbol{\alpha}_u - \boldsymbol{\beta}_v) * D_v, y_{uv} = f_1(f_2(f_3(\boldsymbol{x}_{uv})))$, where $\boldsymbol{Q}_v \in \{0,1\}^{1*K}$ indicates whether an exercise is associated with a knowledge concept. f_1, f_2, f_3 are the fully connected layers with positive weights to ensure monotonicity. KSCD [3] further explores the impact of potential associations between knowledge concepts on diagnostic results, shown as follows: $\boldsymbol{\alpha}'_{uc} = \phi\left(f_{sk}(\boldsymbol{\alpha}_u \oplus \mathbf{h}_c^K)\right), \boldsymbol{\beta}'_{vc} = \phi\left(f_{ek}(\boldsymbol{\beta}_v \oplus \mathbf{h}_c^K)\right), y_{uv} = \phi\left(\frac{1}{n_v}\sum_{c=1}^{C}\mathbf{Q}_{vc} \times fse(\boldsymbol{\alpha}'_{uc} - \boldsymbol{\beta}'_{vc})\right)$, where $\mathbf{h}_c^K$ represents the initialized embedding representations of knowledge concept c. $\mathbf{Q}_{vc}$ is the knowledge relevance vector $\mathbf{Q}_v$ of the concept c. n_v indicates the number of knowledge concepts contained in exercise e_v. ϕ is the activation function. f_{se}, f_{sk}, f_{ek} are linear transformation functions that correspond to different fully connected layers.

B. Cross-Domain Cognitive Diagnosis (CDCD)

Consider $|\boldsymbol{S}|$ source domains $\{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_{|\boldsymbol{S}|}\}$ and $|\boldsymbol{T}|$ target domains $\{\boldsymbol{T}_1, \boldsymbol{T}_2, \ldots, \boldsymbol{T}_{|\boldsymbol{T}|}\}$. Let $\boldsymbol{LS}_s$ and $\boldsymbol{LT}_t$ denote the interaction records for source domain $\boldsymbol{S}_s$ and target domain $\boldsymbol{T}_t$, respectively, where $s \in \{1, 2, \ldots, |\boldsymbol{S}|\}$ and $t \in \{1, 2, \ldots, |\boldsymbol{T}|\}$. The CDCD task aims to identify and leverage cognitive patterns and learning structures that generalize across domains, enhancing the model's performance in the new domain $\boldsymbol{T}_t$. By leveraging the abundant data $\boldsymbol{LS}_1, \boldsymbol{LS}_2, \ldots, \boldsymbol{LS}_{|S|}$ from source domains, CDCD can rapidly establish cognitive diagnosis models in target domain $\boldsymbol{T}_t$ with few-shot data $\boldsymbol{LT}_t^{few} \subset \boldsymbol{LT}_t$.

To facilitate the introduction of the subsequent framework, we define sets $\boldsymbol{O}$ and $\boldsymbol{D}$ to represent the overlapping and non-overlapping entities between the source and target domains, as depicted in Figure 3. The entity here refers to students or exercises. For instance, in the exercise-aspect CDCD scenarios, there exists non-overlapping groups of exercises, defined as $\boldsymbol{D}$. Conversely, we define the set of overlapping students as $\boldsymbol{O}$, which allows the transfer of cross-domain information.

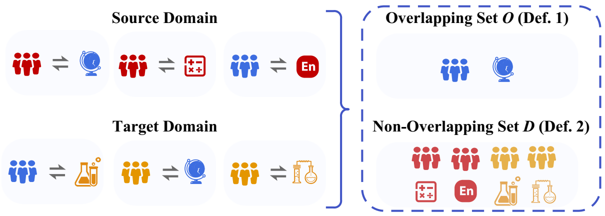


Fig. 3. Illustration of overlapping and non-overlapping sets in CDCD.

Definition 1 (Overlapping Set)**.** The overlapping set $\boldsymbol{O}$ represents the entities that exist in both the source and target domains, which is defined as:

$$\boldsymbol{O} = \bigcup_{s=1}^{|\boldsymbol{S}|} \boldsymbol{S}_s \cap \bigcup_{t=1}^{|\boldsymbol{T}|} \boldsymbol{T}_t \tag{1}$$

Definition 2 (Non-Overlapping Set)**.** Let $\boldsymbol{\Omega}$ be the universal set of all entities in both the source and target domains. The non-overlapping set $\boldsymbol{D}$ is defined as the complement of $\boldsymbol{O}$ with respect to the universal set $\boldsymbol{\Omega}$:

$$\boldsymbol{D} = \boldsymbol{\Omega} \setminus \boldsymbol{O} = \left(\bigcup_{s=1}^{|\boldsymbol{S}|} \boldsymbol{S}_s \cup \bigcup_{t=1}^{|\boldsymbol{T}|} \boldsymbol{T}_t \right) \setminus \boldsymbol{O} \tag{2}$$

III. Proposed Framework

In this section, we propose the scenario-agnostic PromptCD , applicable to both student- and exercise-aspect CDCD scenarios.

A. Overall Architecture

The overall architecture is shown in Figure 4. Our two-stage framework abstracts scenario-agnostic features and unified learning strategies, enabling rapid adaptation to new domains. The pseudo-code is presented in Algorithm 1.

Pre-training Stage. In Section III-B, we present an exposition of the personalized and shared prompts, as well as the processing strategies for entity representations. During the pre-training stage, the prompts are updated using the data $\boldsymbol{LS}_s$ ($s \in \{1, 2, \ldots, |S|\}$) from the source domains.

Fine-tuning Stage. In Section III-C, we outline the prompt transfer process and introduce a variant that enhances adaptation. After pre-training, we fine-tune the trainable parameters using few-shot data $\boldsymbol{LS}_t^{few}$ from the target domain $\boldsymbol{T}_t$, thereby transferring knowledge from the source domains and adapting to the target domain's distribution.

B. Source Prompt Enhancement

Considering the characteristics of overlapping and non-overlapping entities in cross-domain scenarios, relying solely on entity representations may not accommodate the diverse interactions across different domains. Consequently, we designed two types of learnable soft prompts—personalized prompts and shared prompts—to establish connection between the source and target domains by associating these prompts with different entity representations.

Algorithm 1 PromptCD

Input: cognitive diagnosis model $\mathcal{M}$, records $\boldsymbol{LS}$ for pre-training and $\boldsymbol{LT}_t^{few}$ in target doamin t for fine-tuning.
Output: fine-tuned model $\mathcal{M}$, the transfer prompts $\hat{\boldsymbol{p}}^o$ and $\hat{\boldsymbol{p}}_t^d$.
—**Pre-training Stage**—
while $e_1 \leqslant Epoch_{Pretrain}$ **do**
  for $\boldsymbol{LS}_s \in \{\boldsymbol{LS}_1, \boldsymbol{LS}_2, ..., \boldsymbol{LS}_{|\boldsymbol{S}|}\}$ **do**
    Initialize embeddings $\boldsymbol{o}_s^{\text{orig}}$, $\boldsymbol{d}_s^{\text{orig}}$, prompts $\boldsymbol{p}^o$, $\boldsymbol{p}^d$ and $\mathcal{M}$;
    Enhance the representation of entities in Eq.(3) and Eq.(4);
    Input $\boldsymbol{o}_s^{\text{out}}$ and $\boldsymbol{d}_s^{\text{out}}$ to $\mathcal{M}$ to predict scores $\boldsymbol{y}_s$;
    Calculate the loss using $\boldsymbol{LS}_s$ to update the model;
  end for
end while
—**Fine-Tuning Stage**—
while $e_2 \leqslant Epoch_{Finetune}$ **do**
  Initialize the entities $\boldsymbol{o}_t^{\text{orig}}$, $\boldsymbol{d}_t^{\text{orig}}$;
  Obtain the transfer prompts $\hat{\boldsymbol{p}}^o$ and $\hat{\boldsymbol{p}}_t^d$ in Eq.(5) and Eq.(6));
  Activate improvement policy in Eq.(7);
  Enhance the representations in a manner similar to pre-training;
  Input $\boldsymbol{o}_t^{\text{out}}$ and $\boldsymbol{d}_t^{\text{out}}$ to $\mathcal{M}$ to predict scores $\boldsymbol{y}_t$;
  Calculate the loss using $\boldsymbol{LT}_t^{few}$ to update the model.
end while

Specifically, personalized prompts are tailored for individual entity within the overlapping set $\boldsymbol{O}$. Since the overlapping entities are identical in both the source and target domains, each entity can be associated with a personalized prompt, allowing for the transfer of more information. In contrast, shared prompts are utilized by all entities in each domain within the non-overlapping set $\boldsymbol{D}$, ensuring a common representation for domain-specific knowledge. The resulting composite representations can be expressed as:

$$\boldsymbol{o}_{k,i}^{\text{cat}} = [\boldsymbol{p}_i^o, \boldsymbol{o}_{k,i}^{\text{orig}}], \quad \boldsymbol{d}_{k,j}^{\text{cat}} = [\boldsymbol{p}_k^d, \boldsymbol{d}_{k,j}^{\text{orig}}], \tag{3}$$

where $\boldsymbol{p}_i^o$ is the personalized prompt for each individual overlapping entity i, and $\boldsymbol{p}_k^d$ is the shared prompt for domain-specific entities. $\boldsymbol{o}_{k,i}^{\text{orig}}$ and $\boldsymbol{d}_{k,j}^{\text{orig}}$ respectively denote original embedding of single entity in source domain $\boldsymbol{S}_k (k \in 1, 2, ..., |\boldsymbol{S}|)$ through random initialization, while $\boldsymbol{o}_{k,i}^{\text{cat}}$ and $\boldsymbol{d}_{k,i}^{\text{cat}}$ denote their corresponding representations after concatenation with prompts.

To integrate the concatenated features and extract the joint information, we utilize a fully connected layer defined as the operator Linear, which has different trainable parameters depending on the types of entities:

$$\boldsymbol{o}_k^{\text{out}} = \text{Linear}_o(\boldsymbol{o}_k^{\text{cat}}), \quad \boldsymbol{d}_k^{\text{out}} = \text{Linear}_d(\boldsymbol{d}_k^{\text{cat}}). \tag{4}$$

$\boldsymbol{o}_k^{\text{out}}$ and $\boldsymbol{d}_k^{\text{out}}$ denote final source-domain representations, aligned with the original embeddings. The processed representations are then input into the cognitive diagnosis model to predict scores.

C. Source-To-Target Transfer

In the fine-tuning stage, learned prompts are adapted to the target domains. Personalized prompts, designed for overlapping entities across domains, are transferred on a one-to-one basis in Eq.(5) to maintain the integrity of cross-domain

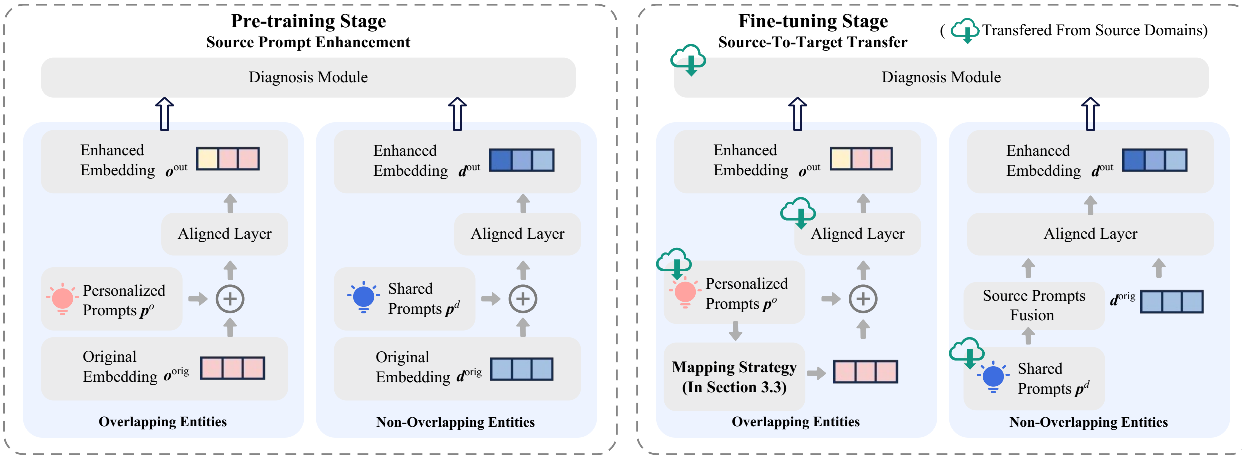


Fig. 4. Overall architecture of the proposed PromptCD framework, including the pre-training and fine-tuning stages.

connection information, as these prompts are also relevant to the target domains.

$$\hat{\boldsymbol{p}}_i^o = \boldsymbol{p}_i^o, \tag{5}$$

To effectively capture the commonalities across different source domains, the shared prompts should be concatenated and mapped into a unified representation for the target domain. Specifically, we apply a learnable linear transformation to project the concatenated prompts into a shared latent space, ensuring both domain adaptation and dimension consistency. Formally, this process is expressed as:

$$\hat{\boldsymbol{p}}_t^d = \text{Linear}_{s2t}(\boldsymbol{p}_1^d \oplus \boldsymbol{p}_2^d \oplus ... \oplus \boldsymbol{p}_{|\boldsymbol{S}|}^d), \tag{6}$$

where Linear_{s2t} is a trainable mapping function. More specifically, we define it as $\hat{\boldsymbol{P}}_t^d = \boldsymbol{W}\boldsymbol{P}_S + \boldsymbol{b}$, where $\boldsymbol{P}_S$ is the concatenated source-domain prompt matrix, $\boldsymbol{W}$ is the transformation matrix, and $\boldsymbol{b}$ is a bias term. This mapping serves to align feature distributions across domains while preserving essential shared knowledge, allowing the model to adaptively reweight different source-domain prompts for effective transfer.

The final representations $\boldsymbol{o}_t^{\text{out}}$ and $\boldsymbol{d}_t^{\text{out}}$ in target domain $\boldsymbol{T}_t$ are obtained by processing the original embeddings $\boldsymbol{o}_t^{\text{orig}}$ and $\boldsymbol{d}_t^{\text{orig}}$ through operations analogous to those described in Eq.(3) and Eq.(4), involving interactions with the transferred prompts.

Prompt-to-Representation Mapping: We propose a variant that leverages the cross-domain information characteristics of $\hat{\boldsymbol{p}}^o$, which are learned from the overlapping set $\boldsymbol{O}$. This approach is informed by knowledge transfer theory in cognitive science, suggesting that cognitive representations in one domain can be mapped to another through an intermediate transformation. Specifically, we hypothesize that a student's latent ability in one subject can be transferred to another via personalized prompts acting as intermediaries.

Personalized prompts, derived from subjects such as Physics or Biology, implicitly encode domain-general cognitive traits (e.g., logical reasoning, analytical skills). Our strategy employs a linear layer to learn the mapping relationship between these personalized prompts and shallow representations of entities in the target domains:

$$\boldsymbol{o}_t^{\text{orig}} = \text{Linear}_{init}(\hat{\boldsymbol{p}}^o) \tag{7}$$

where $\boldsymbol{o}_{\boldsymbol{t}}^{\text{orig}}$ is original representations from $\boldsymbol{O}$ in target domain $\boldsymbol{T}_t$. Linear_{init} comprises trainable parameters that are optimized using the interaction data in the target domains. This strategy supersedes random initialization, leveraging available data to access potential original information.

By decoding these domain-general cognitive traits, we can estimate the student's capability in a target subject, enabling predictions via a shared latent cognitive space. This integration of forward mapping and reverse inference enhances the effectiveness of our model in transferring knowledge across domains.

IV. Proposed Instantiations

Based on the unified framework above, we instantiate specific scenarios to illustrate its application in the following two scenarios. The pseudo-codes for the proposed PromptCD-S and PromptCD-E instantiations are detailed Algorithms 2 and 3 in the Supplements.

A. Student-Aspect CDCD: PromptCD-S

Student-aspect CDCD focuses on a cross-school scenario where $\boldsymbol{O}$ and $\boldsymbol{D}$ respectively represent a set of exercises and students. This indicates that the source and target domains have overlapping students. Each exercise item has a specific personalized prompt $\boldsymbol{p}_{exer}^o$. We use personalized prompts to uncover the basic requirements of exercise, which are the same for students from different schools.

In this scenario, each school has a corresponding shared prompt $\boldsymbol{p}_{sch}^d$, which reflects the collective performance of students from the common school. Then we employ the PromptCD to concatenate representations of students and exercises with their prompts in Eq.(3) and input them into the cognitive diagnosis model for interaction. By optimizing the cross-entropy loss derived from predicted scores and ground-truth from the source domain response records, PromptCD

TABLE I
DATASET STATISTICS IN THE EXPERIMENTS

Scenarios	**Exercise-aspect (Humanities)**			**Exercise-aspect (Sciences)**			**Student-aspect (Mathematics)**			
Domains	**Chinese**	**History**	**Geography**	**Mathematics**	**Physics**	**Biology**	**A-bin**	**B-bin**	**C-bin**	**D-bin**
Student Number	4,021	4,021	4,021	4,021	4,021	4,021	1,758	984	824	455
Exercise Number	92	164	117	137	115	120	137	137	137	137
Concept Number	14	12	24	31	34	16	31	31	31	31
Total Interactions	263,485	583,334	381,772	435,797	387,535	400,858	197,048	103,852	86,940	47,957
Interactions Per Student	66	145	95	108	96	100	112	106	106	105
Sparsity	0.29	0.12	0.19	0.21	0.16	0.17	0.18	0.23	0.23	0.23
Positive_Negative Ratio	7.04	3.03	1.69	2.80	1.49	1.96	4.46	2.72	1.87	1.35

updates the prompts to extract information across source domains.

To accomplish the source-to-target prompt transfer, we concatenate $\boldsymbol{p}^d_{sch}$ for different schools in the source domains and map them to the original dimension in Eq.(6) to capture commonalities. Finally, a few records from the target domains are used to fine-tune $\hat{\boldsymbol{p}}^d_{sch}$ and personalized prompts $\hat{\boldsymbol{p}}^o_{exer}$ trained from source domains, enhancing their accuracy for future predictions.

B. Exercise-Aspect CDCD: PromptCD-E

Exercise-aspect CDCD, on the other hand, addresses the cross-subject scenario where $\boldsymbol{O}$ and $\boldsymbol{D}$ respectively represent a set of students and exercises. Each student is associated with a personalized prompt $\boldsymbol{p}^o_{stu}$ to capture their basic capability across various subjects, such as mathematics or physics.

Similarly, each subject has a shared prompt $\boldsymbol{p}^d_{sub}$ for all exercise items to enhance the understanding of subject-specific knowledge. Employing the pre-training and fine-tuning methodology analogous to PromptCD-S, we derive the ultimate representations of prompts.

In different scenarios, the meaning and dimensions of entity representations often differ. PromptCD mitigates the sensitivity of existing studies [21], [22] to cross-domain data by employing specific prompts to transfer information across domains, thereby aiding the cognitive diagnosis model in predicting scores accurately.

V. EXPERIMENTS

To validate the effectiveness of the PromptCD in cross-domain scenarios, we conducted extensive experiments on real-world datasets, to address the following questions:

- **RQ1**: How does PromptCD perform in student- and exercise-aspect CDCD scenarios?
- **RQ2**: How efficient are the key components in PromptCD?
- **RQ3**: Can feature visualization demonstrate the effectiveness of prompts in enhancing cross-domain representations?
- **RQ4**: How to conduct personalized learning guidance using PromptCD?

A. Experimental Settings

We present the experimental setup, including the datasets, baselines, metrics, and implementation details.

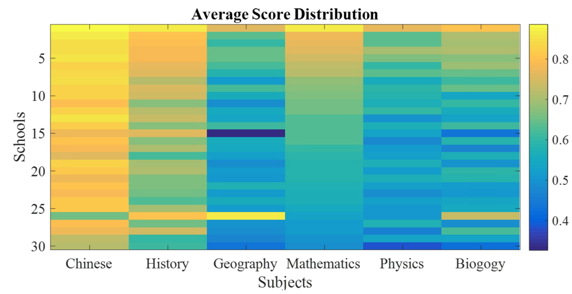


Fig. 5. The average score distribution across different schools and subjects in the SLP dataset

1) Datasets: SLP [20] is a real-world educational dataset that collects students from different schools' responses to multiple subjects in K-12 education. The average score distribution across different schools and subjects in the SLP dataset is illustrated in Figure 5. Research has identified variations in the interaction levels of students when answering exercises across different subjects and schools. For students within the same school, there are similar patterns in their interaction levels across different subjects. From the perspective of student interactions, different schools represent distinct domains in the aforementioned phenomenon. This observation precisely confirms the challenges described in Section I regarding CDCD task, as similar challenges also arise from the exercise-aspect perspective.

To validate the exercise-aspect CDCD, we extracted interaction data from the SLP dataset for three humanities subjects (Chinese, History, Geography) and three science subjects (Mathematics, Physics, and Biology). For the student-aspect CDCD, we focused on interaction data from students at 30 schools in a single subject(e.g. Mathematics). We addressed the challenge of varying average cognitive levels by categorizing the schools into four bins (A, B, C, and D) based on average scores. The dataset statistics for these scenarios are presented in Table I.

2) Baselines: We utilized four widely recognized CD models as the backbone diagnostic models: IRT [29], MIRT [19], NeuralCD [15], and KSCD [3]. We applied our framework to these models, denoted as [Backbone]-Ours. If the prompt-to-representation mapping (Section III-C) is incorporated, the model is denoted as [Backbone]-Ours+. The original backbone versions without cross-domain prompt transfer are used as baselines, denoted as [Backbone]-Origin. Additionally, we included state-of-the-art cross-domain cognitive diagnosis models, TechCD [21], ZeroCD [22], and CCLMF [30] as baselines, referred to as [Backbone]-Tech, [Backbone]-Zero, and [Backbone]-CCLMF, respectively. For student-aspect, TechCD and ZeroCD utilize source domain representations as initial representations for the target domains.

- **Origin**: The backbones without cross-domain prompt and parameter transfer.
- **Tech**: Apply TechCD [21] to the backbone. It requires the relations between knowledge concepts, constructed using a statistical method proposed in RCD [14].
- **Zero**: Apply ZeroCD [22] to the backbone.
- **CCLMF**: Apply CCLMF [30] to the backbone.
- **Ours**: Apply the proposed PromptCD to the backbone.
- **Ours+**: Using the prompt-to-representation mapping strategy on the basis of Ours.

3) Metrics: Cognitive diagnosis is to assess students' proficiency. However, since proficiency is an unobservable variable, researchers typically predict students' future responses (correct or incorrect) and evaluate the model's performance using the accuracy of these predictions. Therefore, we use classification evaluation metrics, namely AUC, ACC, RMSE, and F1.

4) Implementation Details: For both exercise- and student-aspect CDCD, we adopted a similar approach to obtain the data. In the exercise-aspect CDCD, we conducted separate diagnoses for the humanities and sciences, using any two subjects within each category as source domains and the remaining subject as the target domain. For the student-aspect CDCD, students from any three bins are treated as source domains, with the remaining bin as the target domain. In both cases, 20% of the interaction records in the target domain are randomly selected for fine-tuning, while the remaining records are used for testing. We determined the appropriate prompt dimensions for the backbone. Specifically, we set the dimensions for IRT and MIRT to 5 and 10, respectively, while NeuralCD and KSCD were set to 20.

B. Overall Comparison (RQ1)

We compare the overall performance of the models under both the student- and exercise-aspect CDCD.

Exercise-Aspect CDCD. Table II presents the performance under the exercise-aspect CDCD. The cross-subject scenario divides six subjects into two categories: humanities and sciences, with three subjects in each category. Two subjects are treated as source domains, and one as the target domain. Across all target domains, the proposed PromptCD consistently outperforms state-of-the-art baselines. Specifically, when comparing the baselines and our framework across the IRT, MIRT, NeuralCD, and KSCD backbones, our method demonstrates significant improvement in the AUC metric, with an average increase of up to nearly 20% compared to the Origin version, especially in the Biology ($0.667 \rightarrow \mathbf{0.798}$ in IRT) and Chinese subjects ($0.736 \rightarrow \mathbf{0.864}$ in IRT). Additionally, as a unified framework for cross-domain scenarios, our approach demonstrates notable improvements over TechCD [21], ZeroCD [22], and CCLMF [30]. Similarly, the RMSE metric shows a significant reduction in all scenarios. The Ours+ version, which incorporates an additional adaptation strategy, shows consistent improvements over the Ours version in most test scenarios, further validating the robustness of our approach.

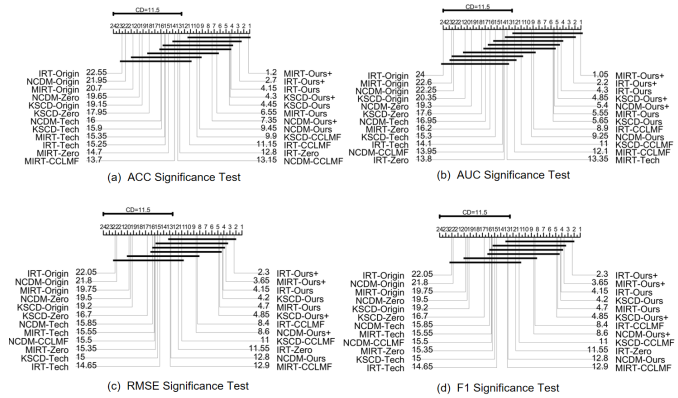


Fig. 6. Visualization of Significance Test Results for Evaluation Metrics. Models connected by the same horizontal line do not show statistically significant differences, while models located on different lines exhibit significant performance differences at the significance level of $\alpha = 0.05$.

Student-Aspect CDCD. Table III showcases the performance under the student-aspect CDCD. In this setup, schools are categorized into four bins (A, B, C, D) based on their average scores, with three bins designated as source domains and the remaining as the target domain. Specifically, in the A-bin target domain, the NeuralCD-Ours shows 27.9% improvement ($0.687 \rightarrow \mathbf{0.879}$) in the AUC metric relative to the origin, while KSCD-Ours achieves 14.5% increase ($0.764 \rightarrow \mathbf{0.875}$). Through the aforementioned experiments, we observed similar patterns across other metrics when applying PromptCD in various scenarios. Comparing the two tables, we see that our proposed framework shows a significant improvement over TechCD and ZeroCD in Table III than in Table II, primarily because the baseline algorithms are not specifically designed to address student-aspect CDCD.

Significance Analysis of Model Performance. We conducted Nemenyi tests on various baseline models used in different scenarios to report statistical significance for metrics such as AUC, ACC, RMSE and F1, as shown in Figure 6. In the figure, models connected by the same horizontal line do not show statistically significant differences, while models located on different lines exhibit significant performance differences at the significance level of $\alpha = 0.05$. The PromptCD model (especially the "Ours+" version) significantly outperforms other baseline models across multiple domains. This further demonstrates the effectiveness and robustness of cross-domain prompt transfer methods in cognitive diagnosis tasks.

C. Detailed Analysis (RQ2)

To address the role of PromptCD in key aspects, we analyze the effects of different modules in the proposed model.

Different Fine-tuning Ratios. We examine the performance of the PromptCD across varying proportions of fine-tuning data, addressing the challenge of data sparsity in cross-domain knowledge adaptation. Specifically, we evaluate the model's performance with fine-tuning data proportions of 0.1, 0.2,

TABLE II
COMPARISON RESULTS IN EXERCISE-ASPECT CDCD SCENARIOS

Target	**Biology**				**Mathematics**				**Physics**			
Metrics	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1
IRT-Origin	0.667	0.652	0.466	0.738	0.699	0.725	0.434	0.814	0.761	0.706	0.442	0.757
IRT-Tech	0.779	0.729	0.421	0.800	0.861	0.819	0.357	0.885	0.845	0.767	0.397	0.808
IRT-Zero	0.721	0.696	0.460	0.809	0.809	0.788	0.414	0.871	0.805	0.736	0.421	0.799
IRT-CCLMF	0.775	0.723	0.425	0.808	0.870	0.817	0.353	0.884	0.850	0.761	0.403	0.813
IRT-Ours	0.798	0.742	**0.411**	0.815	0.874	0.826	0.347	0.886	0.858	0.776	0.387	0.818
IRT-Ours+	**0.799**	**0.743**	**0.411**	**0.816**	**0.880**	**0.832**	**0.342**	**0.890**	**0.864**	**0.781**	**0.384**	**0.823**
MIRT-Origin	0.672	0.669	0.459	0.763	0.712	0.746	0.421	0.835	0.771	0.717	0.440	0.774
MIRT-Tech	0.779	0.727	0.422	0.797	0.864	0.817	0.355	0.878	0.847	0.766	0.395	0.807
MIRT-Zero	0.723	0.706	0.439	0.799	0.786	0.786	0.395	0.867	0.802	0.734	0.420	0.781
MIRT-CCLMF	0.771	0.726	0.426	0.805	0.867	0.810	0.368	0.879	0.843	0.768	0.398	0.813
MIRT-Ours	0.793	0.738	0.415	0.816	0.881	0.834	0.347	**0.893**	0.855	0.775	0.398	**0.822**
MIRT-Ours+	**0.801**	**0.743**	**0.411**	0.809	**0.886**	**0.838**	**0.343**	**0.893**	**0.865**	**0.785**	**0.389**	0.821
NCDM-Origin	0.706	0.655	0.456	0.724	0.755	0.775	0.403	0.856	0.790	0.725	0.432	0.771
NCDM-Tech	0.780	0.727	0.421	0.795	0.865	0.809	0.360	0.874	0.797	0.732	0.423	0.784
NCDM-Zero	0.734	0.697	0.437	0.792	0.824	0.788	0.373	0.871	0.791	0.714	0.438	0.746
NCDM-CCLMF	0.765	0.731	0.424	0.811	0.844	0.808	0.363	0.875	0.839	0.769	0.403	0.809
NCDM-Ours	0.785	**0.735**	**0.417**	**0.815**	0.852	0.813	0.359	**0.878**	0.848	0.764	0.397	0.796
NCDM-Ours+	**0.788**	0.731	0.418	0.812	**0.872**	**0.816**	**0.357**	0.874	**0.861**	**0.782**	**0.386**	**0.820**
KSCD-Origin	0.710	0.691	0.445	0.779	0.761	0.774	0.401	0.854	0.797	0.729	0.426	0.772
KSCD-Tech	0.778	0.729	0.422	0.799	0.859	0.818	0.356	0.882	0.842	0.765	0.398	0.807
KSCD-Zero	0.728	0.703	0.431	0.792	0.801	0.793	0.382	0.872	0.798	0.732	0.428	0.788
KSCD-CCLMF	0.782	0.732	0.420	0.799	0.861	0.815	0.357	0.879	0.843	0.765	0.397	0.813
KSCD-Ours	0.795	**0.741**	**0.413**	**0.818**	0.869	0.826	**0.349**	**0.888**	**0.855**	**0.777**	**0.389**	**0.817**
KSCD-Ours+	**0.796**	0.739	0.414	0.812	**0.870**	**0.828**	0.350	0.886	**0.855**	0.776	**0.389**	0.816

Target	**Chinese**				**History**				**Geography**			
Metrics	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1
IRT-Origin	0.736	0.872	0.319	0.931	0.707	0.752	0.415	0.843	0.687	0.659	0.465	0.731
IRT-Tech	0.830	0.862	0.319	0.922	0.799	0.735	0.417	0.810	0.754	0.707	0.438	0.782
IRT-Zero	0.821	0.878	0.306	0.934	0.790	0.774	0.391	0.859	0.760	0.705	0.437	0.780
IRT-CCLMF	0.855	0.872	0.296	0.934	0.803	0.778	0.381	0.865	0.776	0.711	0.433	0.786
IRT-Ours	0.864	**0.886**	0.290	0.937	0.811	0.791	0.377	0.870	0.783	0.722	0.425	0.790
IRT-Ours+	**0.865**	**0.886**	**0.289**	**0.938**	**0.814**	**0.791**	**0.376**	**0.871**	**0.787**	**0.724**	**0.424**	**0.793**
MIRT-Origin	0.744	0.873	0.319	0.932	0.719	0.756	0.410	0.846	0.689	0.671	0.461	0.763
MIRT-Tech	0.811	0.742	0.411	0.835	0.795	0.752	0.408	0.829	0.761	0.709	0.435	0.781
MIRT-Zero	0.822	0.879	0.302	0.935	0.782	0.776	0.392	0.856	0.724	0.662	0.454	0.724
MIRT-CCLMF	0.845	0.867	0.296	0.930	0.804	0.774	0.385	0.861	0.758	0.704	0.439	0.776
MIRT-Ours	0.863	0.885	0.289	0.937	0.818	0.791	0.376	**0.871**	0.778	0.714	0.431	**0.793**
MIRT-Ours+	**0.866**	**0.888**	**0.288**	**0.937**	**0.822**	**0.794**	**0.374**	0.870	**0.792**	**0.728**	**0.422**	0.787
NCDM-Origin	0.782	0.851	0.323	0.916	0.742	0.740	0.412	0.826	0.717	0.679	0.453	0.752
NCDM-Tech	0.809	0.875	0.308	0.933	0.740	0.769	0.405	0.866	0.721	0.678	0.451	0.749
NCDM-Zero	0.805	0.876	0.306	0.932	0.742	0.761	0.407	0.863	0.728	0.687	0.445	0.755
NCDM-CCLMF	0.838	0.865	0.305	0.930	0.789	0.772	0.397	0.850	0.766	0.711	0.438	0.775
NCDM-Ours	0.843	**0.881**	0.297	**0.934**	0.793	0.778	0.389	0.856	0.774	0.717	0.430	0.780
NCDM-Ours+	**0.852**	0.878	**0.296**	0.931	**0.807**	**0.788**	**0.379**	**0.868**	**0.786**	**0.720**	**0.426**	**0.783**
KSCD-Origin	0.787	0.869	0.318	0.929	0.744	0.771	0.402	0.857	0.722	0.688	0.448	0.771
KSCD-Tech	0.829	0.797	0.371	0.875	0.798	0.749	0.410	0.825	0.756	0.706	0.438	0.772
KSCD-Zero	0.825	0.851	0.328	0.914	0.789	0.769	0.406	0.853	0.749	0.702	0.443	0.778
KSCD-CCLMF	0.842	0.875	0.305	0.925	0.793	0.776	0.388	0.856	0.772	0.711	0.432	0.787
KSCD-Ours	**0.854**	0.883	**0.292**	**0.936**	**0.807**	**0.789**	**0.380**	0.868	0.785	**0.723**	**0.425**	0.794
KSCD-Ours+	**0.854**	**0.884**	**0.292**	**0.936**	0.804	**0.789**	**0.380**	**0.869**	**0.787**	0.719	0.426	**0.797**

and 0.3 on both the exercise-aspect CDCD for the biology target domain and the student-aspect for the A-bin target domain, shown in Figure. 7 (a) and (b). In most cases, the performance improves as the proportion increases, even at a small proportion of 0.1, the models can achieve satisfactory performance in dual-aspect scenarios by using PromptCD.

Different Prompt Dimensions. We demonstrate the influence of prompt dimensionality on PromptCD. We provide results using the NeuralCD, which employs horizontal concatenation, as an example. Specifically, we set the prompt dimensionality to 1, 5, 10, or 20. The comparison results are shown in Figure. 7 (c) and (d). Performance is lowest when the prompt dimension is 1. As the prompt dimension increases, performance improves, reflecting the enhanced information capacity of the prompts and their effectiveness in improving model performance.

Various Source Domains. We evaluate the impact of various source domains PromptCD. Specifically, for the exercise-aspect scenario with biology as the target domain, We consider three scenarios: using mathematics, physics, or both as source domains, respectively. Similarly, for the student-aspect scenario with A-bin as the target domain, we consider three types of combinations of source domains, which are illustrated in Table IV. The performance achieved using data from two source domains is superior to that obtained from a single source domain, which indicates that PromptCD can learn common knowledge across multiple source domains.

Various Cross-Domain Types. Table V presents the experimental results of the model in more scenarios, including from humanities to sciences and from sciences to humanities. The detailed experimental setups are as follows: Sciences-Humanities: Source: Biology, Mathematics $\rightarrow$ Target: Geography. Humanities-Humanities: Source: Chinese, History $\rightarrow$ Target: Geography. Humanities-Sciences: Source: Chinese, History $\rightarrow$ Target: Physics. Sciences-Sciences: Source: Biology, Mathematics $\rightarrow$ Target: Physics Our results indicate that in

TABLE III
COMPARISON RESULTS IN STUDENT-ASPECT CDCD SCENARIOS

Target	A-bin				B-bin				C-bin				D-bin			
Metrics	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1
IRT-Origin	0.670	0.800	0.393	0.884	0.548	0.728	0.451	0.841	0.678	0.679	0.477	0.777	0.519	0.537	0.512	0.630
IRT-Tech	0.821	0.847	0.336	0.912	0.837	0.811	0.365	0.877	0.769	0.741	0.425	0.830	0.809	0.733	0.419	0.767
IRT-Zero	0.855	0.854	0.324	0.917	0.854	0.823	0.360	0.884	0.847	0.784	0.390	0.835	0.831	0.753	0.406	0.792
IRT-CCLMF	0.851	0.854	0.321	0.920	0.853	0.822	0.362	0.883	0.855	0.790	0.384	0.847	0.839	0.757	0.403	0.799
IRT-Ours	0.871	0.867	0.316	0.922	0.870	0.826	0.351	0.884	0.877	0.806	0.368	**0.858**	0.857	**0.766**	0.397	**0.814**
IRT-Ours+	**0.881**	**0.872**	**0.308**	**0.925**	**0.881**	**0.834**	**0.344**	**0.89**	**0.881**	**0.811**	**0.363**	**0.858**	**0.858**	0.764	**0.395**	0.813
MIRT-Origin	0.718	0.820	0.372	0.896	0.742	0.762	0.413	0.848	0.715	0.706	0.461	0.798	0.729	0.682	0.470	0.749
MIRT-Tech	0.820	0.845	0.339	0.912	0.840	0.811	0.361	0.878	0.805	0.760	0.403	0.831	0.817	0.739	0.414	0.780
MIRT-Zero	0.841	0.841	0.347	0.902	0.849	0.814	0.366	0.880	0.845	0.799	0.367	0.854	0.804	0.735	0.433	0.754
MIRT-CCLMF	0.834	0.849	0.331	0.911	0.842	0.814	0.368	0.881	0.842	0.770	0.398	0.841	0.814	0.742	0.428	0.762
MIRT-Ours	0.861	0.859	0.324	0.917	0.863	0.816	0.365	0.883	0.862	0.778	0.395	0.844	0.844	0.766	0.409	0.806
MIRT-Ours+	**0.886**	**0.872**	**0.311**	**0.923**	**0.886**	**0.836**	**0.347**	**0.891**	**0.881**	**0.807**	**0.375**	**0.858**	**0.859**	**0.778**	**0.398**	**0.813**
NCDM-Origin	0.687	0.809	0.387	0.894	0.693	0.743	0.465	0.847	0.709	0.643	0.485	0.782	0.533	0.574	0.653	0.729
NCDM-Tech	0.818	0.845	0.340	0.910	0.838	0.805	0.364	0.873	0.816	0.766	0.401	0.835	0.796	0.697	0.429	0.714
NCDM-Zero	0.761	0.800	0.376	0.877	0.798	0.712	0.430	0.783	0.754	0.724	0.430	0.804	0.797	0.706	0.455	0.718
NCDM-CCLMF	0.844	0.851	0.332	0.911	0.846	0.813	0.368	0.871	0.838	0.775	0.396	0.820	0.813	0.734	0.414	0.757
NCDM-Ours	**0.879**	**0.870**	**0.308**	**0.924**	0.865	0.820	0.357	0.877	0.856	0.780	0.390	0.822	0.837	0.751	**0.408**	0.765
NCDM-Ours+	0.878	0.865	0.319	0.920	**0.878**	**0.833**	**0.352**	**0.891**	**0.864**	**0.796**	**0.380**	**0.841**	**0.840**	**0.752**	0.409	**0.785**
KSCD-Origin	0.764	0.831	0.368	0.901	0.756	0.775	0.414	0.860	0.766	0.726	0.448	0.807	0.769	0.706	0.449	0.737
KSCD-Tech	0.809	0.848	0.338	0.913	0.839	0.809	0.366	0.884	0.780	0.756	0.412	0.830	0.794	0.725	0.425	0.772
KSCD-Zero	0.778	0.806	0.346	0.877	0.808	0.798	0.390	0.879	0.785	0.763	0.408	0.839	0.797	0.737	0.423	0.782
KSCD-CCLMF	0.854	0.855	0.327	0.915	0.853	0.822	0.358	0.886	0.853	0.794	0.385	0.846	0.818	0.759	0.418	0.786
KSCD-Ours	0.875	0.865	0.312	0.921	0.876	0.834	0.346	**0.891**	**0.871**	**0.806**	**0.371**	**0.854**	0.844	0.768	**0.403**	**0.805**
KSCD-Ours+	**0.880**	**0.868**	**0.310**	**0.923**	**0.878**	**0.835**	**0.345**	**0.891**	0.870	0.804	**0.371**	0.851	**0.846**	**0.771**	0.404	0.795

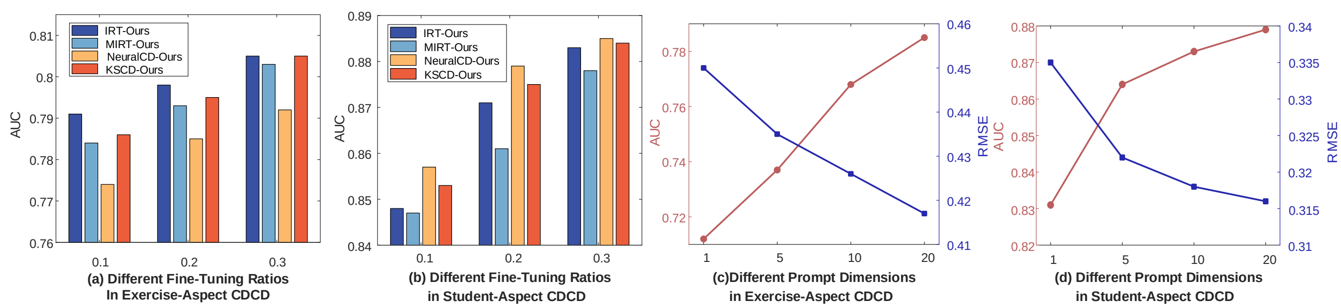


Fig. 7. Performance comparisons with (a-b) different tuning ratios and (c-d) different prompt dimensions

TABLE IV
PERFORMANCE COMPARISONS WITH DIFFERENT SOURCE DOMAINS IN EXERCISE AND STUDENT-ASPECT SCENARIOS

Backbone	Source Domain	Biology				Source Domain	A-bin			
		AUC	ACC	RMSE	F1		AUC	ACC	RMSE	F1
IRT	Mathematics	0.781	0.729	0.414	0.814	B	0.860	0.861	0.321	0.919
	Physics	0.782	0.729	0.413	0.814	B+C	0.870	0.865	**0.314**	0.921
	Mathematics+Physics	**0.798**	**0.742**	**0.411**	**0.815**	B+C+D	**0.871**	**0.867**	0.316	**0.922**
MIRT	Mathematics	0.781	0.730	0.421	0.814	B	0.853	0.856	0.324	0.917
	Physics	0.784	0.732	0.419	0.814	B+C	**0.862**	**0.861**	**0.320**	**0.919**
	Mathematics+Physics	**0.793**	**0.738**	**0.415**	**0.816**	B+C+D	0.861	0.859	0.321	**0.919**
NeuralCD	Mathematics	0.774	0.726	0.423	0.796	B	0.865	0.862	0.318	0.920
	Physics	0.773	0.726	0.423	0.800	B+C	0.874	0.868	0.311	0.922
	Mathematics+Physics	**0.785**	**0.735**	**0.417**	**0.815**	B+C+D	**0.879**	**0.870**	**0.308**	**0.924**
KSCD	Mathematics	0.782	0.735	0.418	0.812	B	0.868	0.864	0.316	0.921
	Physics	0.781	0.734	0.418	0.815	B+C	0.875	**0.868**	**0.311**	**0.923**
	Mathematics+Physics	**0.795**	**0.741**	**0.413**	**0.818**	B+C+D	**0.875**	0.865	0.312	0.921

both cases, PromptCD outperforms the comparison algorithms. Interestingly, its performance shows a slight decline compared to the “humanities to sciences” and “sciences to sciences” scenarios, which aligns with empirical expectations. Specifically, the transfer of student states is more effective between disciplines with similar characteristics, whereas greater differences between disciplines result in more significant deviations in the transferred student states.

Alternative Architectures for Prompt Fusion. To further validate our architectural choices, we conducted additional ablation experiments to explore alternative prompt fusion strategies beyond simple concatenation followed by a linear transformation. Specifically, we evaluate the following variants: MLP Fusion: Applying a multi-layer perceptron (MLP) instead of Linear_{s2t} to capture non-linear interactions between prompts. Element-wise Addition: The original embedding and the corresponding prompt features are added element-wise to generate the final individual representation. The new features

TABLE V
COMPARISON RESULTS OF VARIOUS CDCD SCENARIOS

Scenarios	**Sciences-Humanities**				**Humanities-Humanities**				**Humanities-Sciences**				**Sciences-Sciences**			
Metrics	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1	AUC	ACC	RMSE	F1
IRT-Origin	0.677	0.650	0.472	0.721	0.687	0.659	0.465	0.731	0.747	0.686	0.460	0.731	0.761	0.706	0.442	0.757
IRT-Tech	0.757	0.705	0.438	0.784	0.754	0.707	0.438	0.782	0.829	0.753	0.407	0.791	0.845	0.767	0.397	0.808
IRT-Zero	0.734	0.693	0.446	0.776	0.760	0.705	0.437	0.780	0.804	0.730	0.424	0.766	0.805	0.736	0.421	0.799
IRT-CCLMF	0.775	0.715	0.427	0.791	0.776	0.711	0.433	0.786	0.804	0.730	0.424	0.766	0.805	0.736	0.421	0.799
IRT-Ours	0.791	0.726	0.422	**0.795**	0.783	0.722	0.425	0.790	**0.854**	0.773	**0.390**	**0.815**	0.858	0.776	0.387	0.818
IRT-Ours+	**0.792**	**0.727**	**0.421**	0.792	**0.787**	**0.724**	**0.424**	**0.793**	**0.854**	**0.774**	**0.390**	0.814	**0.864**	**0.781**	**0.384**	**0.823**
MIRT-Origin	0.695	0.669	0.459	0.773	0.689	0.671	0.461	0.763	0.775	0.718	0.439	0.779	0.771	0.717	0.440	0.774
MIRT-Tech	0.765	0.709	0.435	0.787	0.761	0.709	0.435	0.781	0.828	0.744	0.416	0.807	0.847	0.766	0.395	0.807
MIRT-Zero	0.722	0.686	0.450	0.763	0.724	0.662	0.454	0.724	0.794	0.721	0.428	0.757	0.802	0.734	0.420	0.781
MIRT-CCLMF	0.767	0.710	0.434	0.775	0.758	0.704	0.439	0.776	0.838	0.756	0.399	0.810	0.843	0.768	0.398	0.813
MIRT-Ours	0.786	0.720	0.428	**0.794**	0.778	0.714	0.431	**0.793**	0.843	0.764	0.405	**0.813**	0.855	0.775	0.398	**0.822**
MIRT-Ours+	**0.791**	**0.728**	**0.423**	0.789	**0.792**	**0.728**	**0.422**	0.787	**0.852**	**0.771**	**0.397**	0.806	**0.865**	**0.785**	**0.389**	0.821
NCDM-Origin	0.714	0.683	0.460	0.765	0.717	0.679	0.453	0.752	0.782	0.721	0.435	0.764	0.790	0.725	0.432	0.771
NCDM-Tech	0.771	0.715	0.431	0.788	0.721	0.678	0.451	0.749	0.821	0.743	0.414	0.799	0.797	0.732	0.423	0.784
NCDM-Zero	0.718	0.688	0.451	0.777	0.728	0.687	0.445	0.755	0.798	0.729	0.426	0.787	0.791	0.714	0.438	0.746
NCDM-CCLMF	0.769	0.715	0.435	0.781	0.766	0.711	0.438	0.775	0.831	0.754	0.408	0.791	0.839	0.769	0.403	0.809
NCDM-Ours	0.779	0.720	0.427	**0.784**	0.774	0.717	0.430	0.780	0.832	0.757	0.407	0.806	0.848	0.764	0.397	0.796
NCDM-Ours+	**0.786**	**0.722**	**0.424**	0.783	**0.786**	**0.720**	**0.426**	**0.783**	**0.848**	**0.769**	**0.397**	**0.814**	**0.861**	**0.782**	**0.386**	**0.820**
KSCD-Origin	0.722	0.687	0.448	0.770	0.722	0.688	0.448	0.771	0.798	0.728	0.426	0.771	0.797	0.729	0.426	0.772
KSCD-Tech	0.761	0.709	0.435	0.785	0.756	0.706	0.438	0.772	0.826	0.753	0.409	0.805	0.842	0.765	0.398	0.807
KSCD-Zero	0.734	0.695	0.442	0.776	0.749	0.702	0.443	0.778	0.809	0.733	0.420	0.791	0.798	0.732	0.428	0.788
KSCD-CCLMF	0.776	0.719	0.428	0.796	0.772	0.711	0.432	0.787	0.836	0.761	0.404	0.806	0.843	0.765	0.397	0.813
KSCD-Ours	**0.788**	**0.726**	**0.424**	0.791	0.785	**0.723**	**0.425**	0.794	0.848	0.769	0.395	0.804	0.855	**0.777**	**0.389**	**0.817**
KSCD-Ours+	**0.788**	0.724	**0.424**	**0.794**	**0.787**	0.719	0.426	**0.797**	**0.849**	**0.772**	**0.393**	**0.815**	**0.855**	0.776	**0.389**	0.816

TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT PROMPT FUSION STRATEGIES

Target	**Mathematics**			
	AUC	ACC	RMSE	F1
MIRT-Ours+	**0.886**	**0.838**	**0.343**	**0.893**
MIRT-MLP	0.881	0.833	0.350	0.890
MIRT-Addition	0.882	0.833	0.347	0.890
MIRT-Multiplication	0.744	0.754	0.410	0.850
Target	**A-bin**			
	AUC	ACC	RMSE	F1
MIRT-Ours+	**0.886**	**0.872**	**0.311**	**0.923**
MIRT-MLP	0.877	0.865	0.318	0.920
MIRT-Addition	0.881	0.853	0.322	0.916
MIRT-Multiplication	0.769	0.816	0.389	0.898

TABLE VII
ABLATION STUDY ON THE IMPACT OF PROMPT COMPONENTS

Target	**Mathematics**			
	AUC	ACC	RMSE	F1
MIRT-Ours+	**0.886**	**0.838**	**0.343**	**0.893**
MIRT-1	0.728	0.758	0.410	0.852
MIRT-2	0.879	0.830	0.351	0.890
MIRT-3	0.712	0.746	0.421	0.835
Target	**A-bin**			
	AUC	ACC	RMSE	F1
MIRT-Ours+	**0.886**	**0.872**	**0.311**	**0.923**
MIRT-1	0.732	0.831	0.359	0.904
MIRT-2	0.877	0.865	0.317	0.920
MIRT-3	0.718	0.820	0.372	0.896

effectively integrate the local attributes of the node with the global information carried by the prompt. Element-wise Multiplication: The original embedding and the corresponding prompt features are multiplied element-wise to generate the final individual representation. By applying element-wise multiplication, the global semantic information of the prompt features can be enhanced while preserving important detailed information. These fusion strategies are evaluated under both student-aspect and exercise-aspect CDCD settings, with performance compared in terms of AUC, ACC, and generalization capability. The results in Table VI demonstrate that our method achieves superior performance across multiple metrics.

Different Types of Prompts. We analyze the impact of removing different prompt components by evaluating three variants: MIRT-1: Removes personalized prompts $\hat{\boldsymbol{p}}^o$ to assess their contribution. MIRT-2: Removes shared prompts $\hat{\boldsymbol{p}}_t^d$ to evaluate their role in cross-domain adaptation. MIRT-3: Removes both components to observe the overall effect. The results, shown in Table VII, confirm that both personalized and shared prompts contribute significantly to performance. Removing either component leads to a noticeable decline, highlighting their complementary roles in capturing individualized and domain-level knowledge.

D. Feature Visualization (RQ3)

In this section, we visualized the representations learned by the model in CDCD scenarios for both exercises and students to explain why the Prompts learned from the source domains are effective.

For the student-aspect CDCD scenario, Figure. 8 (a) displays the distribution of the original student representations after dimensionality reduction and reveals that the original representations of students from different bins do not display any discernible patterns. In contrast, Figure. 8 (b) displays the distribution of the final representations, obtained after the operations described in Section III-B. The clustering of students within the same bin suggests that our Prompts effectively capture the overall characteristics of the domain. Additionally, student representations from A-bin are more distant from those of D-bin and closer to those of B-bin, indicating that the prompts effectively distinguish between different levels of student groups. In the exercise-aspect CDCD, as shown

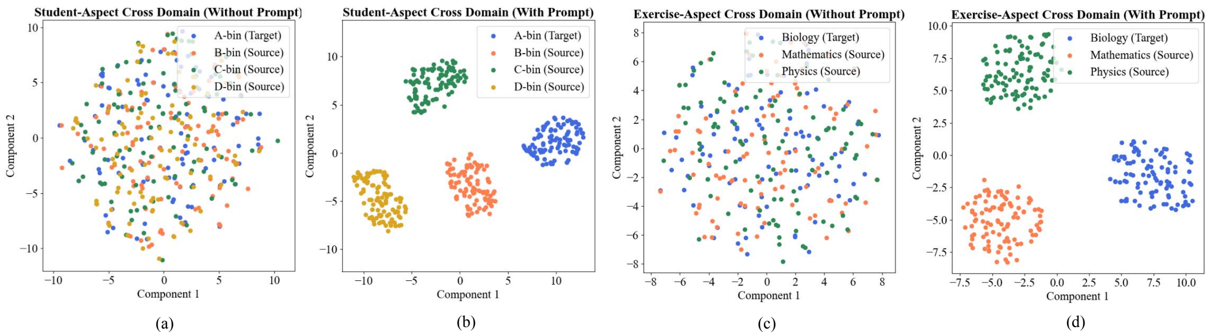


Fig. 8. Visualization of (a, c) origin representations without prompt and (b, d) our representations with prompt

TABLE VIII
CLUSTER ANALYSIS BEFORE AND AFTER PROMPTING

Status	Dimension	Intra-Dist	Inter-Dist
without Prompt	Exercise Embedding	6.6352	0.2911
without Prompt	Student Embedding	7.8856	0.6041
with Prompt	Exercise Embedding	2.8690	12.2945
with Prompt	Student Embedding	2.9120	13.0087

TABLE IX
RECOMMENDING EXAMPLE IN EXERCISE-ASPECT CDCD

Concept ID	11	14	19	25	26	28	33
Exercise ID	16	2	14	25	23	54	3
Student Mastery	0.491	0.472	0.484	0.489	0.490	0.481	0.485
Exercise Difficulty	0.550	0.481	0.520	0.428	0.530	0.487	0.469
True Performance	0	0	0	1	0	0	1

in Figure 8 (c) and (d), the transfer prompts also effectively capture both the internal characteristics of each subject and the distinctions between different subjects.

Additionally, we employed two quantitative metrics—inter-cluster distance and intra-cluster distance—to analyze the changes in student and exercise representations before and after introducing prompts. As shown in Table VIII, the inter-cluster and intra-cluster distances are significantly smaller after introducing prompts compared to before, providing a clearer demonstration of the effectiveness of our method and ensuring consistency in the presentation across both student and exercise dimensions.

E. Personalized Recommendation (RQ4)

In this section, we illustrate how the PromptCD facilitates personalized learning guidance for exercise recommendations. We employ a straightforward yet effective strategy to suggest exercises related to concepts that students have not yet mastered [21], ensuring they are of appropriate difficulty [5]. Based on a CD backbone, we first determine whether this is a student-side or exercise-side cross-domain recommendation scenario, then select the corresponding framework to be added to the CD model. After the model undergoes a two-stage training process, it produces a well-trained model $\mathcal{M}$. $\mathcal{M}$ can determine the student's mastery level of knowledge concepts through a diagnostic module. We select N exercises associated with the knowledge concepts that the student has not yet mastered. Furthermore, we aim for the exercises to be of moderate difficulty for the student, as exercises that are too difficult or too easy may hurt their learning interest. Therefore, from this set of K exercises, we ultimately choose K exercises of moderate difficulty to form the recommended list for the student. Let's take the example of applying the proposed framework to NeuralCD [2] to showcase the results of the personalized learning recommendation. In the exercise-aspect CDCD, mathematics and biology are treated as the source domains, and physics is the target domain. We recommend exercises to a randomly sampled student. The recommended concept ID, exercise ID, corresponding student mastery, exercise difficulty, and true performance are detailed in Table IX. The results demonstrate that the recommended exercises align with the requirements of practical applications, which are exercises that the student has not mastered yet and are of moderate difficulty.

VI. RELATED WORK

A. Cognitive Diagnosis

Cognitive diagnosis [1], [3], [14] is a form of student learning modeling that plays a vital role in educational recommendation tasks [4], [5]. Traditional cognitive diagnosis models, such as IRT [29] and MIRT [19], utilize unidimensional and multidimensional latent traits, respectively, to represent student and exercise features. The DINA model [31] incorporates guessing and slipping parameters but often relies on assumptions that oversimplify student interactions. These traditional models have established a foundation for cognitive diagnosis but struggle to capture the complexities of student behavior and learning patterns.

To address these limitations, various deep learning-based cognitive diagnosis models have been proposed. Wang et al. [15] introduced NeuralCD, which uses neural networks to enhance both accuracy and interpretability. Many works have expanded upon NeuralCD, such as KaNCD [15] and KSCD [3], which make full use of information from non-interactive knowledge concepts. Beyond these, recent studies have explored more advanced cognitive diagnosis frameworks. For example, ORCDF [32] introduces an oversmoothing-resistant model to enhance learning representations in online education systems, while SCD [33] combines symbolic reasoning with hybrid optimization to improve diagnosis performance. Moreover, FineCD [34] explores how foundation models can enhance derivative-free cognitive diagnosis, improving generalization across diverse student populations and subjects. These models that assume the training and testing data are from the same distribution will experience a significant decline in performance when confronted with non-identical distributions.

Cross-Domain Cognitive Diagnosis: The introduction of new domains in online education often leads to the unavailability of practice logs for many students, creating the CDCD issue [21]. Gao et al. [21] proposed TechCD for exercise-aspect CDCD, which uses transferable knowledge concept graphs to address the cold-start problem in new domains. This method embeds knowledge concepts and student behaviors into a graph, leveraging transferable knowledge to accurately assess cognitive abilities. ZeroCD [22] tackles the CDCD problem by utilizing early-participating student data to assess cognitive abilities with minimal data. In addition, Hu et al. [30] proposed CCLMF, which leverages a meta-learner to predict network parameters and enhances model performance on target courses using knowledge from source courses. However, most of them only focuses on just one aspect of the issue. LRCD [35] introduces a language representation-favored approach to zero-shot cross-domain cognitive diagnosis, demonstrating the potential of pre-trained language models in CDCD settings. In this paper, we focus on a scenario-agnostic CDCD framework, maintaining compatibility with both exercise-aspect and student-aspect scenarios.

B. Prompt Learning

Prompt learning [36] is a technique applied to pre-trained language models [23] and has demonstrated significant success in various applications, including recommendation tasks [37]–[40]. This approach guides the model's generation process using prompts. Prompts can be either hard (discrete words) or soft (continuous learnable embeddings) [41]. Soft prompts, in particular, offer greater flexibility as they can be optimized and adjusted during training, allowing them to better adapt to specific tasks and data requirements [42].

Prompt Learning for Cross-Domain Tasks. The prompt learning method, through the adjustment and optimization of prompts, can adapt to the language and characteristics of different domains. Consequently, it has shown promising results in cross-domain recommendation tasks [24]–[28]. In these tasks, shared knowledge from source domains is often transferred to the target domain via knowledge-enhanced prompts. Unlike hard prompts, which require extensive handcrafting and are highly specific to individual tasks, soft prompts offer greater flexibility and can be more easily optimized and adapted [27]. This adaptability reduces inefficiencies and improves robustness across various tasks and models. Hard prompts, on the other hand, often present significant challenges; poorly designed prompts can negatively impact model performance and may not transfer effectively across tasks [43]. Moreover, the need for prompt engineering and the difficulty of creating effective templates for each task further limit their efficiency [44], [45]. Therefore, this paper focuses on leveraging soft prompt learning for the CDCD task.

VII. Conclusion

In this paper, we proposed the PromptCD framework for cross-domain cognitive diagnosis tasks in intelligence education. Specifically, we designed the prompts to enhance the student and exercise representation across domains. The prompts follow a two-stage mode of pre-training and fine-tuning. Importantly, the proposed framework can be applied to both student-aspect and exercise-aspect cross-domain scenarios. Experimental results on the real-world datasets illustrated the effectiveness of the proposed framework.

References

[1] Yingjie Liu, Tiancheng Zhang, Xuecen Wang, Ge Yu, and Tao Li. New development of cognitive diagnosis models. FCS, 17(1):171604, 2023.

[2] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang. Neural cognitive diagnosis for intelligent education systems. In AAAI, pages 6153–6161, 2020.

[3] Haiping Ma, Manwei Li, Le Wu, Haifeng Zhang, Yunbo Cao, Xingyi Zhang, and Xuemin Zhao. Knowledge-sensed cognitive diagnosis for intelligent education platforms. In CIKM, pages 1451–1460, 2022.

[4] Fei Liu, Xuegang Hu, Shuochen Liu, Chenyang Bu, and Le Wu. Meta multi-agent exercise recommendation: A game application perspective. In SIGKDD, page 1441–1452, New York, NY, USA, 2023.

[5] Zhenya Huang, Qi Liu, Chengxiang Zhai, Yu Yin, Enhong Chen, Weibo Gao, and Guoping Hu. Exploring multi-objective exercise recommendations in online education systems. In CIKM, pages 1261–1270, 2019.

[6] Siyu Wu, Jun Wang, and Wei Zhang. Contrastive personalized exercise recommendation with reinforcement learning. IEEE TLT, 2023.

[7] Ju Fan, Yuanchun Jiang, Yezheng Liu, and Yonghang Zhou. Interpretable mooc recommendation: a multi-attention network for personalized learning behavior analysis. Internet Research, 32(2):588–605, 2022.

[8] Haoyang Bi, Haiping Ma, Zhenya Huang, Yu Yin, Qi Liu, Enhong Chen, Yu Su, and Shijin Wang. Quality meets diversity: A model-agnostic framework for computerized adaptive testing. In ICDM, pages 42–51. IEEE, 2020.

[9] Haw-Shiuan Chang, Hwai-Jung Hsu, and Kuan-Ta Chen. Modeling exercise relationships in e-learning: A unified approach. In EDM, pages 532–535, 2015.

[10] Zichao Wang, Angus Lamb, Evgeny Saveliev, Pashmina Cameron, Yordan Zaykov, José Miguel Hernández-Lobato, Richard E. Turner, Richard G. Baraniuk, Craig Barton, Simon Peyton Jones, Simon Woodhead, and Cheng Zhang. Instructions and guide for diagnostic questions: The neurips 2020 education challenge, 2021.

[11] Qi Liu, Yan Zhuang, Haoyang Bi, Zhenya Huang, Weizhe Huang, Jiatong Li, Junhao Yu, Zirui Liu, Zirui Hu, Yuting Hong, Zachary A. Pardos, Haiping Ma, Mengxiao Zhu, Shijin Wang, and Enhong Chen. Survey of computerized adaptive testing: A machine learning perspective. CoRR, abs/2404.00712, 2024.

[12] Fei Wang, Weibo Gao, Qi Liu, Jiatong Li, Guanhao Zhao, Zheng Zhang, Zhenya Huang, Mengxiao Zhu, Shijin Wang, Wei Tong, and Enhong Chen. A survey of models for cognitive diagnosis: New developments and future directions, 2024.

[13] Yuqiang Zhou, Qi Liu, Jinze Wu, Fei Wang, Zhenya Huang, Wei Tong, Hui Xiong, Enhong Chen, and Jianhui Ma. Modeling context-aware features for cognitive diagnosis in student learning. In SIGKDD, pages 2420–2428, 2021.

[14] Weibo Gao, Qi Liu, Zhenya Huang, Yu Yin, Haoyang Bi, Mu-Chun Wang, Jianhui Ma, Shijin Wang, and Yu Su. Rcd: Relation map driven cognitive diagnosis for intelligent education systems. In SIGIR, pages 501–510, 2021.

[15] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yu Yin, Shijin Wang, and Yu Su. Neuralcd: A general framework for cognitive diagnosis. TKDE, 2022.

[16] Xinping Wang, Caidie Huang, Jinfang Cai, and Liangyu Chen. Using knowledge concept aggregation towards accurate cognitive diagnosis. In CIKM, pages 2010–2019, 2021.

[17] Jiatong Li, Fei Wang, Qi Liu, Mengxiao Zhu, Wei Huang, Zhenya Huang, Enhong Chen, Yu Su, and Shijin Wang. Hiercdf: A bayesian network-based hierarchical cognitive diagnosis framework. In SIGKDD, pages 904–913, 2022.

[18] Shiwei Tong, Jiayu Liu, Yuting Hong, Zhenya Huang, Le Wu, Qi Liu, Wei Huang, Enhong Chen, and Dan Zhang. Incremental cognitive diagnosis for intelligent education. In SIGKDD, pages 1760–1770, 2022.

[19] Mark D Reckase and Mark D Reckase. Multidimensional item response theory models. Springer, 2009.

[20] Yu Lu, Yang Pian, Ziding Shen, et al. Slp: A multi-dimensional and consecutive dataset from k-12 education. In ICCE, volume 1, pages 261–266, 2021.
[21] Weibo Gao, Hao Wang, Qi Liu, Fei Wang, Xin Lin, Linan Yue, Zheng Zhang, Rui Lv, and Shijin Wang. Leveraging transferable knowledge concept graph embedding for cold-start cognitive diagnosis. In SIGIR, pages 983–992, 2023.
[22] Weibo Gao, Qi Liu, Hao Wang, Linan Yue, Haoyang Bi, Yin Gu, Fangzhou Yao, Zheng Zhangm Xin Li, and Yuanjing He. Zero-1-to-3: Domain-level zero-shot cognitive diagnosis via one batch of early-bird students towards three diagnostic objectives. In AAAI, 2024.
[23] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9):1–35, 2023.
[24] Chunjiang Ge, Rui Huang, Mixue Xie, Zihang Lai, Shiji Song, Shuang Li, and Gao Huang. Domain adaptation via prompt learning. IEEE TNNLS, 2023.
[25] Jinxing Liu, Junjin Xiao, Haokai Ma, Xiangxian Li, Zhuang Qi, Xiangxu Meng, and Lei Meng. Prompt learning with cross-modal feature alignment for visual domain adaptation. In CAAI International Conference on Artificial Intelligence, pages 416–428. Springer, 2022.
[26] Hui Wu and Xiaodong Shi. Adversarial soft prompt tuning for cross-domain sentiment analysis. In ACL, pages 2438–2447, 2022.
[27] Wenbo Zhao, Arpit Gupta, Tagyoung Chung, and Jing Huang. Spc: Soft prompt construction for cross domain generalization. In Workshop on RepL4NLP, pages 118–130, 2023.
[28] Lulu Zhao, Fujia Zheng, Weihao Zeng, Keqing He, Ruotong Geng, Huixing Jiang, Wei Wu, and Weiran Xu. Adpl: Adversarial prompt-based domain adaptation for dialogue summarization with knowledge disentanglement. In SIGIR, pages 245–255, 2022.
[29] Susan E Embretson and Steven P Reise. Item response theory. Psychology Press, 2013.
[30] Liya Hu, Zhiang Dong, Jingyuan Chen, Guifeng Wang, Zhihua Wang, Zhou Zhao, and Fei Wu. Ptadisc: a cross-course dataset supporting personalized learning in cold-start scenarios. In NeurIPS, NIPS '23, Red Hook, NY, USA, 2024. Curran Associates Inc.
[31] De La Torre and Jimmy. Dina model and parameter estimation: A didactic. Journal of Educational and Behavioral Statistics, 34(1):115–130, 2009.
[32] Hong Qian, Shuo Liu, Mingjia Li, Bingdong Li, Zhi Liu, and Aimin Zhou. Orcdf: An oversmoothing-resistant cognitive diagnosis framework for student learning in online education systems. In SIGKDD, page 2455–2466, 2024.
[33] Junhao Shen, Hong Qian, Wei Zhang, and Aimin Zhou. Symbolic cognitive diagnosis via hybrid optimization for intelligent education systems. AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024.
[34] Mingjia Li, Hong Qian, Jinglan Lv, Mengliang He, Wei Zhang, and Aimin Zhou. Foundation model enhanced derivative-free cognitive diagnosis. FCS, 19(1):191318, 2025.
[35] Shuo Liu, Zihan Zhou, Yuanhao Liu, Jing Zhang, and Hong Qian. Language representation favored zero-shot cross-domain cognitive diagnosis. In SIGKDD, pages 1–8, 2025.
[36] Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. Language models as knowledge bases? arXiv preprint arXiv:1909.01066, 2019.
[37] Zizhuo Zhang and Bang Wang. Prompt learning for news recommendation. arXiv preprint arXiv:2304.05263, 2023.
[38] Jinming Li, Wentao Zhang, Tian Wang, Guanglei Xiong, Alan Lu, and Gerard Medioni. Gpt4rec: A generative framework for personalized recommendation and user interests interpretation. arXiv preprint arXiv:2304.03879, 2023.
[39] Yupeng Hou, Junjie Zhang, Zihan Lin, Hongyu Lu, Ruobing Xie, Julian McAuley, and Wayne Xin Zhao. Large language models are zero-shot rankers for recommender systems. arXiv preprint arXiv:2305.08845, 2023.
[40] Xuansheng Wu, Huachi Zhou, Wenlin Yao, Xiao Huang, and Ninghao Liu. Towards personalized cold-start recommendation with prompts. arXiv preprint arXiv:2306.17256, 2023.
[41] Lei Li, Yongfeng Zhang, and Li Chen. Personalized prompt learning for explainable recommendation. ACM TOIS, 41(4):1–26, 2023.
[42] Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang. Ppt: Pre-trained prompt tuning for few-shot learning. arXiv preprint arXiv:2109.04332, 2021.
[43] Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are human-level prompt engineers. arXiv preprint arXiv:2211.01910, 2022.
[44] Vivian Liu and Lydia B Chilton. Design guidelines for prompt engineering text-to-image generative models. In CHI, pages 1–23, 2022.
[45] Victor Sanh, Albert Webson, Colin Raffel, Stephen H Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, et al. Multitask prompted training enables zero-shot task generalization. arXiv preprint arXiv:2110.08207, 2021.

SUPPLEMENTS

In this section, the experimental details, the pseudocodes of the proposed models, and more experiments and analysis are presented.

A. Experimental Details of Section I

The experiment was conducted on the SLP dataset using the MIRT model, focusing on three subjects: Mathematics, Physics, and Chinese. The Mathematics dataset was randomly divided into three subsets—Mathematics1, Mathematics2, and Mathematics3—with a ratio of 2:2:6. In this configuration, Mathematics2 served as the testing set, while the other two subsets were combined in various ways for training. This approach allowed for diverse training sets, enhancing the robustness of the results. The hyperparameter settings for the MIRT model included a latent trait dimension of 10, a learning rate of 0.001, and a batch size of 256.

B. Pseudocodes of PromptCD-S and PromptCD-E

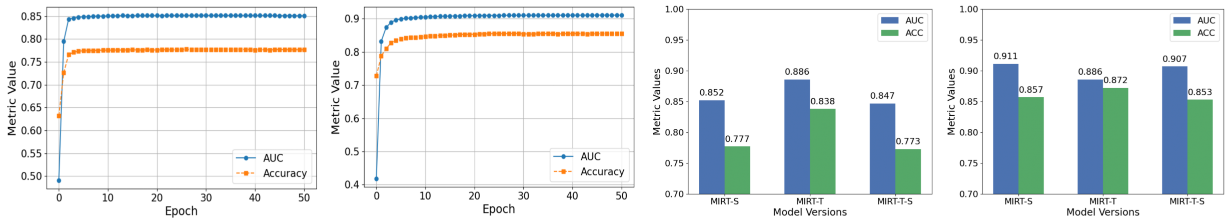
Pseudocodes of PromptCD-S and PromptCD-E applied to NeuralCD are shown in Algorithms 2 and 3.

Algorithm 2 PromptCD-S for NeuralCD

Input: The cognitive diagnosis model $\mathcal{M}$ based on the NeuralCD backbone. $\mathcal{M}$, records $\boldsymbol{LS}$ for pre-training and $\boldsymbol{LT}_t^{few}$ in target domain t for fine-tuning.
Output: fine-tuned model $\mathcal{M}$, the transfer prompts $\hat{\boldsymbol{p}}_{exer}^{o}$ and $\hat{\boldsymbol{p}}_{sch}^{d}$.
—**Pre-training Stage**—
while $e_1 \leqslant Epoch_{Pretrain}$ **do**
  for $\boldsymbol{LS}_s \in \{\boldsymbol{LS}_1, \boldsymbol{LS}_2, ..., \boldsymbol{LS}_{|\boldsymbol{S}|}\}$ **do**
    Initialize the students embedding $\boldsymbol{\alpha}_s^{\text{orig}}$, the exercises embedding $\boldsymbol{\beta}_s^{\text{orig}}$, prompts $\boldsymbol{p}_{exer}^{o}$, $\boldsymbol{p}_{sch}^{d}$ and $\mathcal{M}$;
    Enhance the representation of students and exercises in Eq.(3) and Eq.(4), connect $\boldsymbol{p}_{exer}^{o}$ to $\boldsymbol{\beta}_s^{\text{orig}}$, and $\boldsymbol{p}_{sch}^{d}$ to $\boldsymbol{\alpha}_s^{\text{orig}}$, obtaining $\boldsymbol{\alpha}_s^{\text{out}}$ and $\boldsymbol{\beta}_s^{\text{out}}$ after mapping;
    Input $\boldsymbol{\alpha}_t^{\text{out}}$ and $\boldsymbol{\beta}_t^{\text{out}}$ to $\mathcal{M}$. Specifically, subtract $\boldsymbol{\beta}_t^{\text{out}}$ from $\boldsymbol{\alpha}_t^{\text{out}}$, multiply by the exercise and knowledge vectors, and pass through NeuralCD to get the predicted score $\boldsymbol{y}_s$;
    Calculate the loss using $\boldsymbol{LS}_s$ to update the model;
  end for
end while
—**Fine-Tuning Stage**—
while $e_2 \leqslant Epoch_{Finetune}$ **do**
  Initialize the students embedding $\boldsymbol{\alpha}_t^{\text{orig}}$, the exercises embedding $\boldsymbol{\beta}_t^{\text{orig}}$;
  Obtain the transfer prompts $\hat{\boldsymbol{p}}_{exer}^{o}$ and $\hat{\boldsymbol{p}}_{sch}^{d}$ in Eq.(5) and Eq.(6));
  Activate improvement policy in Eq.(7);
  Enhance the representations in a manner similar to pre-training;
  Input $\boldsymbol{\alpha}_t^{\text{out}}$ and $\boldsymbol{\beta}_t^{\text{out}}$ to $\mathcal{M}$ to obtain the final predicted score $\boldsymbol{y}_t$, as in the pre-training process;
  Calculate the loss using $\boldsymbol{LT}_t^{few}$ to update the model.
end while

C. Hyperparameter Tuning Strategies and Training Epoch Ratios in Experiments

We first set the initial hyperparameters based on prior domain knowledge and experience, typically setting the learning rate to 0.001 and the batch size to 256. Then, we conduct a learning rate range test, gradually increasing the learning rate (e.g., doubling it each time). After each adjustment, we train



(a) Performance Trends During Pretraining

(b) Performance Trends During Fine-Tuning

(c) Comparison of AUC and ACC Across Models in Math

(d) Comparison of AUC and ACC Across Models in A-bin

Fig. 9. Performance trends (a-b) and knowledge retention across phases (c-d)

Algorithm 3 PromptCD-E for NeuralCD

Input: The cognitive diagnosis model $\mathcal{M}$ based on the NeuralCD backbone. $\mathcal{M}$, records $\boldsymbol{LS}$ for pre-training and $\boldsymbol{LT}_t^{few}$ in target doamin t for fine-tuning.
Output: fine-tuned model $\mathcal{M}$, the transfer prompts $\hat{\boldsymbol{p}}_{stu}^o$ and $\hat{\boldsymbol{p}}_{sub}^d$.
—**Pre-training Stage**—
while $e_1 \leqslant Epoch_{Pretrain}$ **do**
  for $\boldsymbol{LS}_s \in \{\boldsymbol{LS}_1, \boldsymbol{LS}_2, ..., \boldsymbol{LS}_{|S|}\}$ **do**
    Initialize the students embedding $\boldsymbol{\alpha}_s^{\text{orig}}$, the exercises embedding $\boldsymbol{\beta}_s^{\text{orig}}$, prompts $\boldsymbol{p}_{stu}^o$, $\boldsymbol{p}_{sub}^d$ and $\mathcal{M}$;
    Enhance the representation of student and exercise in Eq.(3) and Eq.(4); connect $\boldsymbol{p}_{stu}^o$ to $\boldsymbol{\alpha}_s^{\text{orig}}$, and $\boldsymbol{p}_{sub}^d$ to $\boldsymbol{\beta}_s^{\text{orig}}$, obtaining $\boldsymbol{\alpha}_s^{\text{out}}$ and $\boldsymbol{\beta}_s^{\text{out}}$ after mapping.
    Input $\boldsymbol{\alpha}_s^{\text{out}}$ and $\boldsymbol{\beta}_s^{\text{out}}$ into $\mathcal{M}$. Specifically, subtract $\boldsymbol{\beta}_t^{\text{out}}$ from $\boldsymbol{\alpha}_t^{\text{out}}$, multiply by the exercise and knowledge vectors, and pass through NeuralCD to get the predicted score $\boldsymbol{y}_s$;
    Calculate the loss using $\boldsymbol{LS}_s$ to update the model;
  end for
end while
—**Fine-Tuning Stage**—
while $e_2 \leqslant Epoch_{Finetune}$ **do**
  Initialize the students embedding $\boldsymbol{\alpha}_t^{\text{orig}}$, the exercises embedding $\boldsymbol{\beta}_t^{\text{orig}}$;
  Obtain the transfer prompts $\hat{\boldsymbol{p}}_{stu}^o$ and $\hat{\boldsymbol{p}}_{sub}^d$ in Eq.(5) and Eq.(6));
  Activate improvement policy in Eq.(7);
  Enhance the representations in a manner similar to pre-training;
  Input $\boldsymbol{\alpha}_t^{\text{out}}$ and $\boldsymbol{\beta}_t^{\text{out}}$ to $\mathcal{M}$ to obtain the final predicted score $\boldsymbol{y}_t$, as in the pre-training process;
  Calculate the loss using $\boldsymbol{LT}_t^{few}$ to update the model.
end while

for a fixed number of batches and record the corresponding loss values. By plotting the relationship between the learning rate and loss, we identify the range where the loss decreases most rapidly and remains stable. Additionally, for batch size adjustments, we follow a linear scaling rule: when the batch size increases by a factor of $\boldsymbol{k}$, the learning rate is also increased by a factor of $\boldsymbol{k}$ accordingly to ensure training stability and efficiency.

In practical experiments, we observed that datasets composed of different disciplines require varying numbers of iterations to reach convergence during the pre-training and fine-tuning stages. Therefore, using a fixed training epoch ratio may not be suitable for different scenarios. To address this issue, we introduce an Early Stopping mechanism. Specifically, during training, we continuously monitor performance metrics on the validation set (e.g., loss value or accuracy). If the performance does not improve for a certain number of consecutive iterations, training is terminated early. This strategy not only significantly reduces unnecessary computational costs but also effectively prevents overfitting, ensuring both training efficiency and model generalization capability.

D. Convergence Analysis of Pre-training and Fine-tuning

To analyze convergence, we take the target domain as the mathematics subject in a cross-disciplinary scenario and track the AUC and ACC curves over the two training stages. As shown in Fig. 9, the experimental results demonstrate that during the pre-training phase, the loss values stabilized after sufficient iterations, providing robust initialization for subsequent fine-tuning. The fine-tuning phase further optimized model performance, and after adequate iterations, the evaluation metrics showed no significant fluctuations, indicating that the model achieved stable adaptation to the target domain.

E. Trade-off between Knowledge Preservation and Adaptation

To address potential catastrophic forgetting during fine-tuning, we employ the following strategies: Prompt Transfer Mechanism: To ensure knowledge retention across domains, personalized prompts are directly transferred as $\hat{\boldsymbol{p}}_i^o = \boldsymbol{p}_i^o$, maintaining the original representations. Shared prompts undergo a transformation to align with the target domain's feature space, preserving adaptability. Few-shot Fine-tuning: Transferred prompts are further refined using a small set of target domain samples ($\boldsymbol{LT}_t^{few}$), enabling domain-specific adaptation while maintaining core knowledge integrity.

To evaluate the retention of source domain knowledge after fine-tuning, we selected the scenario where the target domain is Math in the Exercise-Aspect CDCD task (Fig. 9. (c)) and the scenario where the target domain is A-bin in the Student-Aspect CDCD task (Fig. 9. (d)). Based on this, we performed a reverse transfer of both personalized and shared prompts, adapted to the target domain, back to the source domain model. Specifically, MIRT-S represents the source domain model, MIRT-T denotes the target domain model that received transferred prompts from the source domain, and MIRT-T-S refers to the source domain model that received reverse-transferred prompts from the target domain. Results indicate that after adapting to the target domain, these prompts can still effectively retain key information from the source domain, without significant information loss or catastrophic forgetting. This demonstrates that during fine-tuning, the model can achieve a good balance between improving adaptability to the target domain and preserving knowledge from the source domain.