DUAL BEHAVIOR SEQUENCE ENHANCEMENT AND CONDITIONAL DIFFUSION FOR COGNITIVE DIAGNOSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Cognitive diagnosis aims to infer students' latent cognitive states from their historical learning behaviors and predict their knowledge mastery on target exercises. Although existing methods have achieved promising results in cognitive state modeling, they still inadequately exploit temporal dynamics in students' learning processes. Moreover, ability representations derived from limited historical interaction statistics are susceptible to fluctuations in responses. To address these limitations, this paper proposes a Dual Behavior Sequence Enhancement and Conditional Diffusion method for Cognitive Diagnosis (DBSCD). DBSCD models students' response sequences and time-interval sequences from historical interactions, while incorporating corresponding exercise and knowledge concept information to construct a more comprehensive representation of students' historical cognitive states. This enhanced state then conditions the conditional diffusion process to progressively refine explicit ability representations, yielding more robust student cognitive states. Finally, the model integrates target exercise-related information to perform fine-grained cognitive diagnosis at the knowledge concept level. Experimental results on two real-world educational datasets demonstrate that DBSCD achieves favorable diagnostic performance, validating the effectiveness of both dual behavior sequence modeling and the conditional diffusion-based ability refinement mechanism. The source code is available at `https://anonymous.4open.science/r/DBSCD`.

1 INTRODUCTION

Cognitive diagnosis aims to infer students' mastery of knowledge concepts from their response records and exercise–concept associations, supporting personalized learning. Early cognitive diagnosis methods, such as IRT (Lord, 1980) and DINA (De La Torre, 2009), model student responses through psychometric parameters or patterns of knowledge mastery. With the development of deep learning, neural cognitive diagnosis models (Wang et al., 2020) further capture nonlinear interactions among students, exercises, and knowledge concepts. Subsequently, existing studies have further enhanced cognitive-state modeling by incorporating knowledge structures, affective states (Wang et al., 2024), cognitive processes (Gu et al., 2025), and conditional diffusion based on multi-view information.

However, existing methods still suffer from two limitations. First, most approaches focus mainly on response outcomes and interaction relationships, while insufficiently exploiting the learning dynamics within time intervals. Second, explicit ability representations estimated from limited historical interactions are vulnerable to response fluctuations and insufficient knowledge coverage, which may reduce the reliability of cognitive-state estimation. Therefore, effectively leveraging temporal behavioral information to refine explicit ability representations remains a key challenge in cognitive diagnosis.

To address these issues, we propose Dual Behavior Sequence Enhancement and Conditional Diffusion for Cognitive Diagnosis (DBSCD). The proposed method jointly models students' response outcomes and time-interval variations to enhance historical cognitive representations and capture students' learning dynamics. Meanwhile, DBSCD treats explicit ability estimated from historical

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

statistics as a statistical prior and conditionally refines it using behavioral information, improving the stability of ability representations. Furthermore, DBSCD integrates exercise semantics and knowledge-concept structures to establish fine-grained associations between student states and target exercises.

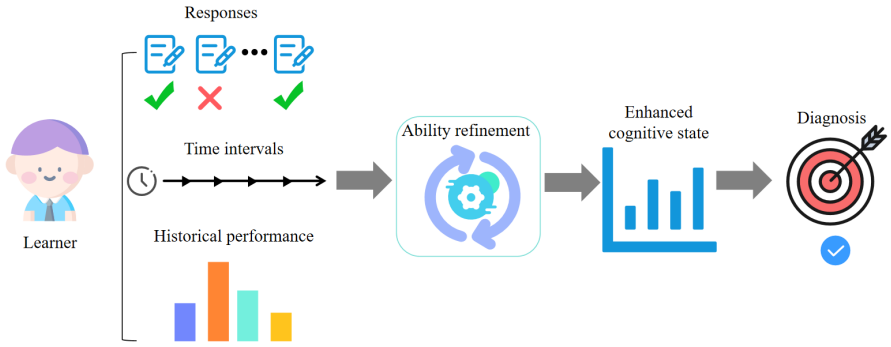

Figure 1: Motivation for DBSCD.

The main contributions of this paper are summarized as follows:

(1) DBSCD jointly models students' response outcomes and time intervals to construct historical cognitive representations that capture both response performance and temporal learning dynamics.

(2) We propose a conditional diffusion module that uses historical behavioral states to refine explicit ability representations estimated from response statistics.

(3) We evaluate the diagnostic effectiveness and accuracy of the proposed DBSCD through extensive experiments on three real-world datasets with multiple baseline models.

(4) A real-world offline examination dataset containing 87,614 student–exercise interaction records is constructed from an introductory computer science course at a university. Standardized question information, student responses, and exercise–concept associations support cognitive diagnosis in offline educational assessment settings.

2 Related Work

2.1 Cognitive Diagnosis Models

Existing cognitive diagnosis methods can be broadly categorized into psychometric theory-based, neural network-based, and relation-aware approaches. Early models, such as MIRT (Reckase, 2006) and DINA (De La Torre, 2009), were primarily grounded in educational measurement theory and explained response behavior through student ability, exercise parameters, or knowledge mastery patterns. With the development of deep learning, NCDM (Wang et al., 2020) and KaNCD (Wang et al., 2023b) employed neural networks to capture nonlinear interactions among students, exercises, and knowledge concepts. RCD (Gao et al., 2021) and HCD (Sun et al., 2025) further incorporated knowledge relationships, graph structures, or hierarchical information to enhance cognitive-state representations. Recent studies have also explored multidimensional behavioral information embedded in historical interactions and learning processes (Wang et al., 2023a). Although these methods have enriched the information used for cognitive modeling, the behavioral dynamics reflected by time intervals remain insufficiently explored. To address this issue, DBSCD jointly models response outcomes and time intervals to better represent students' historical cognitive states.

2.2 Diffusion Models

Diffusion models learn complex data distributions through progressive noising and reverse denoising (Ho et al., 2020) and have recently been introduced into student-state modeling (Zhao et al., 2024). DMC-CDM (Zhao et al., 2025) employs conditional diffusion to consolidate multi-perspective observations and enhance latent cognitive-state representations. MSKT (Zhang et al.,

2024) uses multi-state diffusion to characterize the transition of students' knowledge states from resting states to task-execution states, whereas DiffKT (Liu et al., 2026) applies diffusion-based probabilistic modeling to fine-grained knowledge states. Unlike existing diffusion-based cognitive diagnosis methods that consolidate latent observations, DBSCD treats concept-level explicit ability as a statistical prior and uses target-aware historical behavioral states to condition its refinement.

3 PRELIMINARY

Preliminaries of Cognitive Diagnosis. Let the sets of students, exercises and knowledge concepts be denoted by $S = \{s_1, s_2, \ldots, s_M\}$,$E = \{e_1, e_2, \ldots, e_N\}$, $C = \{c_1, c_2, \ldots, c_K\}$, respectively, where M, N, and K denote the numbers of students, exercises, and knowledge concepts. The binary matrix $\mathbf{Q} \in \{0,1\}^{N\times K}$ encodes exercise–concept associations, where $Q_{j,k} = 1$ iff exercise e_j involves concept c_k, and $\mathbf{q}_j = \mathbf{Q}_{j,:}$.

Definition 1. For a fixed student, the historical interaction sequence is $H = (I_1, \ldots, I_L)$, where $I_\ell = (e_\ell, \mathbf{q}_\ell, r_\ell, \Delta t_\ell)$, $\mathbf{q}_\ell \in \{0,1\}^K$, $r_\ell \in \{0,1\}$, and $\Delta t_\ell \geq 0$. The corresponding response and time-interval sequences are $\mathbf{R} = (r_1, \ldots, r_L)$ and $\Delta\mathbf{T} = (\Delta t_1, \ldots, \Delta t_L)$. We omit the student index and current time step superscript for brevity. Specifically, the response sequence R describes the student's answering performance during historical interactions, while the time interval sequence ΔT captures temporal variation patterns in the student's learning process.

Definition 2. The explicit ability of a student on knowledge concept c_k is defined as the accuracy of responses in historical interactions associated with this knowledge concept, which is formulated as follows

$$a_{j,k} = Q_{j,k}\frac{\sum_{\ell=1}^{L}\mathbb{I}(c_k \in \mathcal{C}(e_\ell))\, r_\ell}{\sum_{\ell=1}^{L}\mathbb{I}(c_k \in \mathcal{C}(e_\ell)) + \epsilon}, \qquad k = 1, \ldots, K. \tag{1}$$

where $C(e_\ell)$ denotes the set of knowledge concepts associated with historical exercise e_ℓ, and $\mathbb{I}(\cdot)$ is the indicator function, which equals 1 when the condition is satisfied and 0 otherwise. $\varepsilon > 0$ is a small constant introduced to avoid division by zero.

The vector $\mathbf{a_j} = (a_1, \ldots, a_K) \in [0,1]^K$ is utilized as a statistical prior derived from historical response performance and the refinement target of the ability enhancement module.

Problem Definition. Given the student's historical interaction sequence H, explicit ability vector a_j, target exercise e_j, and its associated knowledge concept vector q_j, the cognitive diagnosis task aims to learn a parameterized mapping F_Θ to estimate the student's current knowledge mastery state and predict the probability of correctly answering the target exercise e_j: $(\widehat{m}_j, \widehat{y}_j) = F_\Theta\left(H, a_j, e_j, q_j\right)$, where $\widehat{m}$ denotes the predicted mastery level vector over K knowledge concepts, and $\widehat{y}_j$ represents the predicted probability that the student correctly answers the target exercise e_j.

4 METHOD

Model Overview. As illustrated in Figure 2, DBSCD models historical behavioral information to construct a target-conditioned state, which guides the diffusion-based refinement of explicit ability representations. In parallel, target exercise information is encoded to obtain an exercise representation. The resulting student and exercise representations are then used for cognitive diagnosis and response prediction.

4.1 DUAL BEHAVIOR HISTORICAL ENHANCED STUDENT REPRESENTATION

Although students' response outcomes can directly reflect their historical answering performance, relying solely on binary correctness information is insufficient to capture the dynamic changes throughout the learning process. To address this limitation, we jointly model the response sequence and time interval sequence, incorporating associated exercise and knowledge concept information to construct a comprehensive representation of students' historical behavior.

Global Historical State. For the ℓ-th historical interaction, the exercise embedding, knowledge concept embedding, response embedding, and time interval embedding are denoted as follows e_ℓ^{ex}, e_ℓ^c, e_ℓ^r, $e_\ell^\Delta \in \mathbb{R}^d$, where d denotes the dimensionality of the hidden representation.For exercises

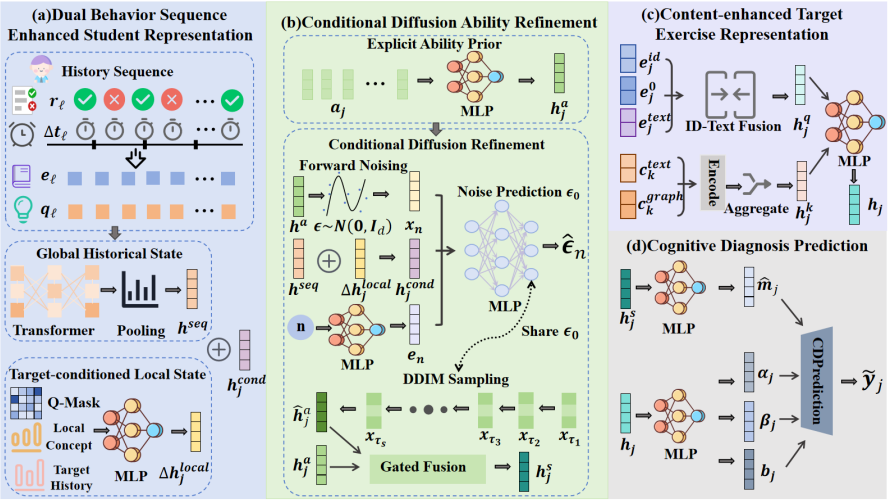

Figure 2: Overview of DBSCD. The framework consists of dual behavior historical representation, conditional diffusion ability refinement, content-enhanced target exercise representation, and cognitive diagnosis prediction.

associated with multiple knowledge concepts, the concept embedding is obtained by averaging the embeddings of the associated concepts; see Appendix A for the formulation.

For the time interval Δt_ℓ, we first apply a logarithmic transformation and then project it into the hidden representation space $e_\ell^{\Delta} = f_\Delta\left(\log(1+\Delta t_\ell)\right) \in \mathbb{R}^d$. Subsequently, the four types of information are integrated to obtain the representation of the ℓ-th historical interaction

$$z_\ell = \mathrm{LN}\left(e_\ell^{ex} + e_\ell^{c} + e_\ell^{r} + e_\ell^{\Delta}\right) \in \mathbb{R}^d, \tag{2}$$

where $\mathrm{LN}(\cdot)$ denotes the layer normalization operation. The sequence of historical interaction representations is further fed into a Transformer encoder (Vaswani et al., 2017) followed by a pooling operation to obtain the student's global historical behavioral state

$$U = \mathrm{SeqTransformer}(z_1, z_2, \ldots, z_L) \in \mathbb{R}^{L\times d}, \quad h^{seq} = \mathrm{Pooling}(U) \in \mathbb{R}^d. \tag{3}$$

Target-conditioned Local State. In addition to the global historical behavioral state, we further consider the knowledge relevance between historical interactions and the target exercise, and construct target-aware statistical features by incorporating corresponding historical performance and temporal information $r_j^{loc} \in \mathbb{R}^7$. The detailed computation process of r_j^{loc} is provided in Appendix B. Their projection is added to the global state to form the diffusion condition

$$\Delta h_j^{local} = \mathrm{MLP}_{loc}\left(r_j^{loc}\right) \in \mathbb{R}^d, \quad h_j^{cond} = h^{seq} + \Delta h_j^{local}. \tag{4}$$

4.2 Conditional Diffusion Ability Refinement Module

Explicit abilities estimated from historical response accuracy provide direct statistical estimates of students' knowledge states, but are sensitive to sparse interactions and response fluctuations.

First, the explicit ability vector $\mathbf{a}_j$ defined in the preliminary section is mapped into the diffusion latent space $\mathbf{h}_j^a = \mathrm{MLP}_a(\mathbf{a}_j) \in \mathbb{R}^d$, and $x_0 = \mathbf{h}_j^a \in \mathbb{R}^d$ is initialized. Assume that the diffusion process contains N_d timesteps, and $n \in \{1, 2, \ldots, N_d\}$ denotes the diffusion time step.

Let $\alpha_n = 1 - \beta_n$, $\bar{\alpha}_n = \prod_{u=1}^{n} \alpha_u$, where β_n denotes the noise-schedule coefficient at timesteps n, and $\bar{\alpha}_n$ represents the cumulative signal-retention coefficient (Ho et al., 2020).

Then the forward diffusion process is performed

$$\epsilon \sim \mathcal{N}(0, I_d), \quad \epsilon \in \mathbb{R}^d, \qquad x_n = \sqrt{\bar{\alpha}_n} h^a + \sqrt{1-\bar{\alpha}_n}\epsilon \in \mathbb{R}^d, \tag{5}$$

which perturbs the explicit ability representation at different noise levels.

To preserve student-specific historical behavioral information during denoising, the noisy ability representation, target-aware conditional representation, and diffusion timestep representation are jointly fed into the denoising network.

We encode diffusion step n as $e_n = \phi(n) = \mathrm{MLP}_t(\mathrm{PE}(n)) \in \mathbb{R}^d$, where PE is sinusoidal encoding. The denoiser takes $z_n = [x_n \| h_j^{\mathrm{cond}} \| e_n] \in \mathbb{R}^{3d}$ and predicts the injected noise

$$z_n = \left[x_n \parallel h_j^{cond} \parallel e_n\right] \in \mathbb{R}^{3d}. \tag{6}$$

The conditional denoising network predicts the Gaussian noise introduced during the forward process

$$\widehat{\epsilon}_n = \epsilon_\theta\left(x_n, h_j^{cond}, n\right) = \mathrm{MLP}_\epsilon\left(z_n\right) \in \mathbb{R}^d. \tag{7}$$

The denoising network is trained to predict the Gaussian noise introduced during the forward diffusion process

$$\mathcal{L}_{diff} = \mathbb{E}_{n,\epsilon}\left[\frac{1}{d}\left\|\epsilon - \epsilon_\theta\left(x_n, h_j^{cond}, n\right)\right\|_2^2\right]. \tag{8}$$

During the inference stage, we employ DDIM to perform few-step reverse sampling (Song et al., 2021). Let $\tau_1 > \tau_2 > \cdots > \tau_S = 0$ denote the selected DDIM timesteps. We first construct the initial noisy ability state based on the latent representation of explicit ability. Then, the corresponding denoised ability representation is estimated during the DDIM reverse process, followed by the computation of the next reverse state

$$x_{\tau_1} = \sqrt{\bar{\alpha}_{\tau_1}} h^a + \sqrt{1-\bar{\alpha}_{\tau_1}}\epsilon, \tag{9}$$

$$\hat{x}_0^{(m)} = \frac{x_{\tau_m} - \sqrt{1-\bar{\alpha}_{\tau_m}}\,\hat{\epsilon}_{\tau_m}}{\sqrt{\bar{\alpha}_{\tau_m}}}, \qquad x_{\tau_{m+1}} = \sqrt{\bar{\alpha}_{\tau_{m+1}}}\,\hat{x}_0^{(m)} + \sqrt{1-\bar{\alpha}_{\tau_{m+1}}}\,\hat{\epsilon}_{\tau_m}. \tag{10}$$

The final diffusion-enhanced ability is $\hat{h}_j^a = \hat{x}_0^{(S)} \in \mathbb{R}^d$.

Since the original explicit ability retains direct statistical information of knowledge concepts, we further employ a gating mechanism to fuse the original ability and the diffusion-enhanced ability

$$g_j = \sigma\Big(\mathrm{MLP}_g([h^a \| \hat{h}_j^a])\Big), \qquad h_j^s = \mathrm{MLP}_f\Big(g_j h^a + (1-g_j)\hat{h}_j^a\Big), \tag{11}$$

where h_j^s denotes the final student state representation used for cognitive diagnosis.

To preserve concept-level ability information, we reconstruct $\hat{\mathbf{a}}_j = \mathrm{MLP}_r(h_j^s) \in \mathbb{R}^K$ and minimize

$$\mathcal{L}_{rec} = \mathbb{E}_{(H,e_j,a)\sim\mathcal{D}}\left[\frac{1}{K}\left\|\widehat{a}_j - a_j\right\|_2^2\right]. \tag{12}$$

This regularizes concept-level information without requiring identical latent representations.

4.3 Content-Enhanced Target Exercise Representation

Given the enhanced student state, DBSCD incorporates both exercise and knowledge concept information for diagnosis.

For the target exercise e_j, its ID embedding and textual semantic embedding are denoted as $e_j^{id}, e_j^{text} \in \mathbb{R}^d$, respectively. Meanwhile, $e_j^0 \in \mathbb{R}^d$ denotes the fixed initial ID embedding of exercise e_j saved before training. Appendix C provides the detailed construction. Based on the variation between the current ID embedding and the initial ID embedding, an adaptive fusion gate is calculated

$$g_j^q = \max\left(0.2,\, \sigma\left(\mathrm{MLP}_q\left(\left[\left|e_j^{id} - e_j^0\right| \parallel \left(e_j^{id} \odot e_j^0\right)\right]\right)\right)\right), \tag{13}$$

where $\odot$ denotes element-wise multiplication, and $g_j^q \in [0.2, 1), h_j^q \in \mathbb{R}^d$ represent the exercise computation. The gate adaptively balances the learned ID representation and textual semantics, while the lower bound 0.2 ensures that exercise-specific ID information always retains a minimum contribution.

The fused exercise representation is then obtained as $h_j^q = g_j^q e_j^{id} + (1 - g_j^q) e_j^{text} \in \mathbb{R}^d$.

For knowledge concept c_k, the concept text and graph representations, $c_k^{\text{text}} \in \mathbb{R}^{d_t}$ and $c_k^{\text{graph}} \in \mathbb{R}^{d_g}$, are fused as $h_k^c = \text{MLP}_c([c_k^{\text{text}} \| c_k^{\text{graph}}]) \in \mathbb{R}^{d_c}$. Graph construction is detailed in the Appendix D.

For notational simplicity, let $\mathcal{K}_j = \{k \mid q_{j,k} = 1\}$ denote the index set of knowledge concepts associated with target exercise e_j, and define $K_j^+ = \max(|\mathcal{K}_j|, 1)$.

The target exercise's aggregated concept representation is $h_j^k = (K_j^+)^{-1} \sum_{k \in \mathcal{K}_j} h_k^c \in \mathbb{R}^{d_c}$.

We concatenate the exercise representation and the aggregated concept representation, then project the result to obtain the target exercise representation

$$h_j = \text{MLP}_e\left([h_j^q \| h_j^k]\right) \in \mathbb{R}^d. \tag{14}$$

4.4 Cognitive Diagnosis Prediction

Based on the enhanced student state h_j^s and the target exercise representation h_j, the model estimates the student's knowledge mastery level, exercise difficulty, and exercise discrimination

$$\widehat{\mathbf{m}}_j = \sigma\left(\text{MLP}_\theta\left(h_j^s\right)\right) \in (0,1)^K, \tag{15}$$

$$\beta_j = \sigma\left(\text{MLP}_\beta\left(h_j\right)\right) \in (0,1)^K, \tag{16}$$

$$\alpha_j = \text{Softplus}\left(\text{MLP}_\alpha\left(h_j\right)\right) \in \mathbb{R}_{>0}. \tag{17}$$

Let $b_j \in \mathbb{R}$ denote the learnable bias term of the target exercise e_j. For the knowledge concepts involved in the target exercise, the model compares the student's knowledge mastery level with the exercise difficulty and calculates the final response probability

$$\widehat{y}_j = \sigma\left(\alpha_j \frac{1}{K_j^+} \sum_{k \in \mathcal{K}_j} (\widehat{m}_{j,k} - \beta_{j,k}) + b_j\right). \tag{18}$$

We train response prediction using binary cross-entropy

$$\mathcal{L}_{bce} = -\frac{1}{|\mathcal{D}|} \sum_{(H, e_j, y_j) \in \mathcal{D}} \left[y_j \log \widehat{y}_j + (1 - y_j) \log\left(1 - \widehat{y}_j\right)\right], \tag{19}$$

where $y_j \in \{0, 1\}$ denotes the observed response outcome on target exercise e_j.

Model Optimization. DBSCD is trained by jointly optimizing the response prediction loss, diffusion denoising loss, and ability reconstruction loss. The overall objective is defined as

$$\mathcal{L} = \lambda_{bce}\mathcal{L}_{bce} + \lambda_{diff}\mathcal{L}_{diff} + \lambda_{rec}\mathcal{L}_{rec}. \tag{20}$$

The coefficients λ_{bce}, λ_{diff}, and λ_{rec} control the three objectives.

5 Experiments

To evaluate the effectiveness of DBSCD, we conduct extensive experiments on three real-world educational datasets. Specifically, we aim to answer the following research questions:

RQ1: How does DBSCD perform compared with existing baseline methods?

RQ2: How do different components contribute to the performance of DBSCD?

RQ3: How do different hyperparameter settings affect the performance of DBSCD?

RQ4: How does DBSCD perform in real-world offline examination scenarios?

Table 1: Dataset statistics.

Dataset	#Students	#Exercises	#Concepts	#Interactions
XES3G5M	5000	7313	839	1542391
MoocRadar	5000	2151	2490	334344
Computer	3847	253	198	87614

5.1 Datasets and Experimental Protocol

To evaluate the effectiveness of our method, we conduct experiments on three real-world educational datasets. Table 1 summarizes detailed statistics for the three datasets.

Datasets. We evaluate DBSCD on two online educational datasets, XES3G5M (Liu et al., 2023) and MoocRadar (Yu et al., 2023), and one offline examination dataset, Computer. XES3G5M and MoocRadar contain timestamped interactions, question texts, and knowledge concept annotations. We randomly sample 5,000 students from each dataset and construct chronological interaction sequences for next-response prediction. The data are split into training, validation, and test sets at a ratio of 7:1:2, using only interactions preceding the target response as historical input. The Computer dataset is collected from an offline introductory computer science examination at a university, with preprocessing detailed in Appendix E.

Baselines. We compare DBSCD with nine representative baseline methods, including traditional psychometric methods (IRT (Lord, 1980)), neural cognitive diagnosis methods (NCD (Wang et al., 2020), ACD (Wang et al., 2024), ICDM (Liu et al., 2024), DFCD (Liu et al., 2025), HCD (Sun et al., 2025), DMC-CDM (Zhao et al., 2025)), debiased diagnosis methods (DBCD (Yang et al., 2026)), and inductive cognitive diagnosis methods (ID-CDM (Li et al., 2024)).Detailed descriptions of all baseline methods are provided in Appendix F.

Implementation Details. DBSCD is implemented in PyTorch 2.0.0 and optimized using AdamW on an RTX 4090 GPU (Loshchilov & Hutter, 2019). The hidden dimension is set to 128, with a two-layer Transformer containing four attention heads. The dimension of time-interval embeddings is set to 32. Exercise and concept texts are encoded offline using BERT to obtain 768-dimensional semantic representations (Devlin et al., 2019). Following previous cognitive diagnosis studies, we use AUC, ACC, and RMSE as evaluation metrics. Higher AUC and ACC values indicate better performance, while lower RMSE values indicate smaller prediction errors.

5.2 Analysis of Experimental Results (RQ1)

Table 2: Performance comparison on XES3G5M and MoocRadar.

Method	XES3G5M			MoocRadar		
	AUC ↑	ACC ↑	RMSE ↓	AUC ↑	ACC ↑	RMSE ↓
IRT	0.6560±0.0022	0.7020±0.0025	0.4493±0.0007	0.6774±0.0018	0.7186±0.0026	0.4456±0.0013
NCD	0.6801±0.0013	0.6945±0.0497	0.4542±0.0175	0.7068±0.0020	0.6128±0.0296	0.5131±0.1569
ID-CDM	0.7007±0.0004	0.5678±0.0177	0.5058±0.0086	0.7195±0.0020	0.7169±0.0109	0.4289±0.0066
ACD	0.5797±0.1211	0.7939±0.0046	0.4027±0.0062	0.5032±0.0062	0.8116±0.0000	0.3910±0.0000
ICDM	0.7063±0.0006	0.7902±0.0028	0.3969±0.0045	0.7389±0.0021	0.8196±0.0019	0.3785±0.0005
DFCD	0.7167±0.0022	0.7894±0.0030	0.3897±0.0023	0.7764±0.0028	0.8181±0.0060	0.3619±0.0034
HCD	0.6710±0.0016	0.7975±0.0011	0.3933±0.0015	0.7383±0.0049	0.8264±0.0016	0.3646±0.0015
DMC-CDM	0.7809±0.0006	0.8090±0.0010	0.3694±0.0002	0.8534±0.0018	0.8453±0.0036	0.3325±0.0031
DBCD	0.7158±0.0023	0.7933±0.0010	0.3919±0.0017	0.7646±0.0022	0.8184±0.0012	0.3654±0.0029
DBSCD	**0.8029±0.0007**	**0.8160±0.0018**	**0.3615±0.0016**	**0.8716±0.0022**	**0.8543±0.0012**	**0.3208±0.0012**

Table 2 compares DBSCD with nine baseline methods on XES3G5M and MoocRadar. All results are reported as the mean and standard deviation over three random seeds. DBSCD achieves better performance across all six evaluation settings. Compared with DMC-CDM, the strongest baseline on both datasets, DBSCD further improves AUC and ACC while reducing RMSE, showing a consistent advantage across different evaluation metrics. Notably, DMC-CDM also employs conditional diffusion, suggesting that diffusion-based cognitive modeling is a competitive direction; the additional

gains of DBSCD indicate that refining explicit ability with historical behavioral and target-content representations provides a more effective formulation.

5.3 Ablation Study (RQ2)

Table 3: Ablation study on XES3G5M and MoocRadar.

Method	XES3G5M			MoocRadar		
	AUC ↑	ACC ↑	RMSE ↓	AUC ↑	ACC ↑	RMSE ↓
DBSCD	**0.8029**	**0.8160**	**0.3615**	**0.8716**	**0.8543**	**0.3208**
w/o diffusion	0.8020	0.8150	0.3628	0.8694	0.8522	0.3233
w/o time	0.8006	0.8138	0.3640	0.8593	0.8505	0.3266
w/o concept	0.8028	0.8158	0.3618	0.8692	0.8529	0.3213
w/o graph	0.8020	0.8150	0.3623	0.8690	0.8506	0.3246
w/o e&c text	0.8023	0.8154	0.3619	0.8687	0.8523	0.3325

Table 4: Performance on Computer.

Method	Computer		
	AUC ↑	ACC ↑	RMSE ↓
IRT	0.8047	0.711	0.4274
ACD	<u>0.8217</u>	<u>0.7802</u>	0.4302
DMC-CDM	0.8201	0.7543	0.4110
DFCD	0.7939	0.7791	0.4093
DBCD	0.8150	0.7784	<u>0.3999</u>
DBSCD	**0.8373**	**0.7953**	**0.3791**

To investigate the contribution of different components in DBSCD, we construct five variants by removing the conditional diffusion module (w/o diffusion), time-interval information (w/o time), knowledge concept text information (w/o concept), knowledge structure information (w/o graph), and exercise-concept text information (w/o e&c text), respectively. As shown in Table 3, the complete DBSCD model consistently achieves the best performance on both datasets compared with all variants. The performance degradation of different variants demonstrates the effectiveness and necessity of each component in DBSCD, including temporal behavior modeling, conditional diffusion-based ability refinement, and content-aware exercise representation learning.

5.4 Hyperparameter Sensitivity Analysis (RQ3)

To investigate the sensitivity of DBSCD to different hyperparameter settings, we evaluate the performance variations under different values of the diffusion loss weight λ_{diff}, DDIM sampling steps, dropout ratio, and weight decay coefficient on the MoocRadar dataset, as shown in Figure 3. The results show that DBSCD achieves relatively stable performance across different hyperparameter configurations, suggesting that the model is not overly sensitive to moderate parameter changes. Among the investigated hyperparameters, λ_{diff} and DDIM sampling steps have a noticeable influence on performance, while dropout and weight decay mainly affect the regularization during model training. Overall, the results verify the robustness of DBSCD under different hyperparameter settings.

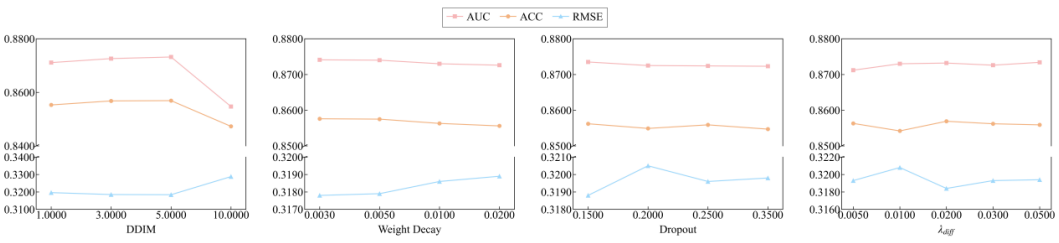

Figure 3: Hyperparameter sensitivity analysis on MoocRadar.

5.5 Evaluation in Real-World Examination Scenarios(RQ4)

Overall Performance on Real-World Examination Data. We evaluate the overall diagnostic performance of DBSCD on the Computer dataset in a real-world course examination setting. As shown in Table 4, DBSCD achieves strong performance among the compared methods.These results indicate that DBSCD maintains effective cognitive diagnosis performance on real-world examination data.

Class-Level Knowledge Mastery Analysis. Figure 4 presents the relative mastery of knowledge concepts across different classes. The classes exhibit distinct relative strengths and weaknesses

across knowledge concepts, indicating that DBSCD can identify differences in knowledge mastery across classes.

Group-Level Diagnostic Reliability Analysis. Figure 5 evaluates the reliability of DBSCD from the perspectives of predictive accuracy and probability calibration across different class groups. Figure 5(a) reports the prediction accuracy of five representative classes, while Figure 5(b) compares the mean predicted probabilities with the observed correct rates for different class–concept combinations. The predicted probabilities are generally consistent with the observed response outcomes. Taken together, these results show that DBSCD can not only identify class-specific strengths and weaknesses across knowledge concepts, but also provide accurate and well-calibrated response probability estimates.

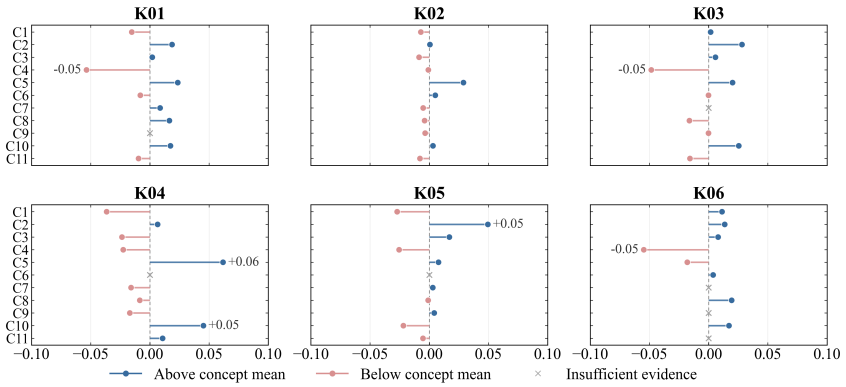

Figure 4: Relative mastery estimates across representative classes and knowledge concepts.

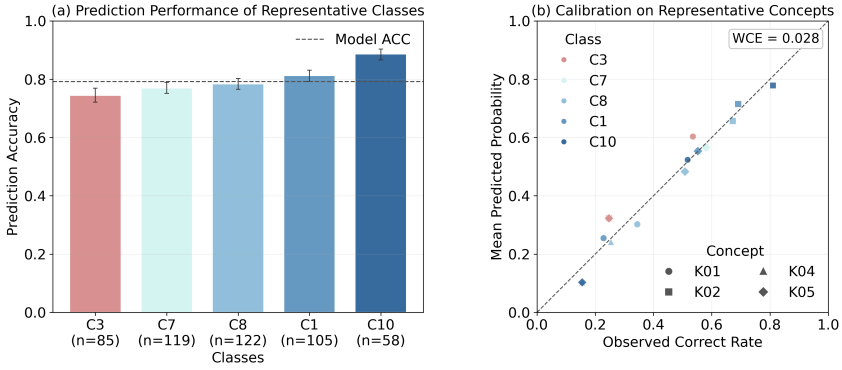

Figure 5: Prediction performance and probability calibration on representative classes. (a) Prediction accuracy across classes. (b) Predicted probabilities versus observed correct rates on representative knowledge concepts.

6 Conclusion

This paper proposes a Dual Behavior Sequence Enhancement and Conditional Diffusion for Cognitive Diagnostic method (DBSCD). The proposed method characterizes students' historical behaviors by jointly modeling response outcomes and time intervals, refines explicit ability representations through conditional diffusion guided by historical behavioral states, and integrates item and knowledge concept information for cognitive diagnosis. Experimental results on multiple real-world educational datasets demonstrate the effectiveness of DBSCD and further verify the contributions of dual-behavior-sequence modeling and conditional diffusion-based ability refinement to student cognitive state modeling.

AI USE STATEMENT

To enhance the manuscript's clarity and style, a generative language model (GPT-5.6) was used to assist with grammar correction, academic expression, and sentence fluency, based solely on the authors' original draft. All core ideas, including the main technical contributions, method design, experimental setup, data analysis, and interpretation of results, were entirely conceived and developed by the authors without AI assistance. All figures, including method diagrams and experimental result visualizations, were created solely by the authors using standard scientific plotting tools, without any assistance from generative models. The authors bear full responsibility for the scientific content, accuracy, and integrity of this work.

REFERENCES

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

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*, pp. 4171–4186. Association for Computational Linguistics, 2019.

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*, pp. 501–510. ACM, 2021.

Hengnian Gu, Guoqian Luo, Xiaoxiao Dong, Shulin Li, and Dongdai Zhou. Revisiting cognition in neural cognitive diagnosis. In *KDD (1)*, pp. 402–412. ACM, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

Jiatong Li, Qi Liu, Fei Wang, Jiayu Liu, Zhenya Huang, Fangzhou Yao, Linbo Zhu, and Yu Su. Towards the identifiability and explainability for personalized learner modeling: An inductive paradigm. In *WWW*, pp. 3420–3431. ACM, 2024.

Ruidi Liu, Yong Niu, and Hu Li. Diffkt: a diffusion model for fine-grained knowledge tracing. *Applied Intelligence*, 56(14):398, 2026.

Shuo Liu, Junhao Shen, Hong Qian, and Aimin Zhou. Inductive cognitive diagnosis for fast student learning in web-based intelligent education systems. In *WWW*, pp. 4260–4271. ACM, 2024.

Yuanhao Liu, Shuo Liu, Yimeng Liu, Chanjin Zheng, Wei Zhang, and Hong Qian. A dual-fusion cognitive diagnosis framework for open student learning environments. In *KDD (2)*, pp. 1915–1926. ACM, 2025.

Zitao Liu, Qiongqiong Liu, Teng Guo, Jiahao Chen, Shuyan Huang, Xiangyu Zhao, Jiliang Tang, Weiqi Luo, and Jian Weng. XES3G5M: A knowledge tracing benchmark dataset with auxiliary information. In *NeurIPS*, 2023.

Frederic M Lord. *Applications of item response theory to practical testing problems*. Routledge, 1980.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR (Poster)*. OpenReview.net, 2019.

Mark D Reckase. 18 multidimensional item response theory. *Handbook of statistics*, 26:607–642, 2006.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*. OpenReview.net, 2021.

Xinjie Sun, Qi Liu, Kai Zhang, Shuanghong Shen, Fei Wang, Yan Zhuang, Zheng Zhang, Weiyin Gong, Shijin Wang, Lina Yang, and Xingying Huo. HCD: A hierarchy constraint-aware neural cognitive diagnosis framework. *Neural Networks*, 190:107668, 2025.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, pp. 5998–6008, 2017.

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*, pp. 6153–6161. AAAI Press, 2020.

Fei Wang, Zhenya Huang, Qi Liu, Enhong Chen, Yu Yin, Jianhui Ma, and Shijin Wang. Dynamic cognitive diagnosis: An educational priors-enhanced deep knowledge tracing perspective. *IEEE Trans. Learn. Technol.*, 16(3):306–323, 2023a.

Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yu Yin, Shijin Wang, and Yu Su. Neuralcd: A general framework for cognitive diagnosis. *IEEE Trans. Knowl. Data Eng.*, 35(8):8312–8327, 2023b.

Shanshan Wang, Zhen Zeng, Xun Yang, Ke Xu, and Xingyi Zhang. Boosting neural cognitive diagnosis with student's affective state modeling. In *AAAI*, pp. 620–627. AAAI Press, 2024.

Shangshang Yang, Xuewen Duan, Xiaoshan Yu, Ziwen Wang, Haiping Ma, and Xingyi Zhang. Debiased cognitive diagnosis: A contrastive counterfactual modeling method via variational autoencoder. In *AAAI*, pp. 27611–27620. AAAI Press, 2026.

Jifan Yu, Mengying Lu, Qingyang Zhong, Zijun Yao, Shangqing Tu, Zhengshan Liao, Xiaoya Li, Manli Li, Lei Hou, Hai-Tao Zheng, Juanzi Li, and Jie Tang. Moocradar: A fine-grained and multi-aspect knowledge repository for improving cognitive student modeling in moocs. In *SIGIR*, pp. 2924–2934. ACM, 2023.

Kai Zhang, Tao Ji, and Huiling Zhang. Knowledge tracing via multiple-state diffusion representation. *Expert Syst. Appl.*, 255:124797, 2024.

Guanhao Zhao, Zhenya Huang, Yan Zhuang, Haoyang Bi, Yiyan Wang, Fei Wang, Zhiyuan Ma, and Yixia Zhao. A diffusion-based cognitive diagnosis framework for robust learner assessment. *IEEE Trans. Learn. Technol.*, 17:2281–2295, 2024.

Guanhao Zhao, Zhenya Huang, Cheng Cheng, Yan Zhuang, Qingyang Mao, Xin Li, Shijin Wang, and Enhong Chen. Multi-perspective consolidation enhanced cognitive diagnosis via conditional diffusion model. In *AAAI*, pp. 1174–1182. AAAI Press, 2025.

A KNOWLEDGE CONCEPT EMBEDDING AGGREGATION

For the ℓ-th historical interaction, the knowledge concept embedding is obtained by averaging the embeddings of all concepts associated with exercise e_ℓ:

$$e_\ell^c = \frac{\sum_{k=1}^{K} q_{\ell,k} v_k^c}{\max\left(\sum_{k=1}^{K} q_{\ell,k},\, 1\right)}.$$

Here, $v_k^c \in \mathbb{R}^d$ denotes the embedding of knowledge concept c_k, and $q_{\ell,k} = 1$ if exercise e_ℓ involves concept c_k, and 0 otherwise.

B TARGET-RELEVANT LOCAL FEATURE CONSTRUCTION

For target exercise e_j, the target-relevant local state is constructed from historical interactions associated with its knowledge concepts. In addition to exact concept overlap, structural concept similarity is used to softly incorporate historical interactions involving structurally related concepts.

For the ℓ-th historical interaction, exact knowledge-concept relevance is defined as $o_{\ell j}^{exact} = \mathbb{I}\left(q_\ell^\top q_j > 0\right)$, where $o_{\ell j}^{exact} \in \{0, 1\}$. Let the normalized structural representation of knowledge concept c_k be $\bar{c}_k^{graph} = \frac{c_k^{graph}}{\left\|c_k^{graph}\right\|_2} \in \mathbb{R}^{d_g}$.

The structural representation of target exercise e_j is obtained by averaging the structural representations of its associated concepts and normalizing the result

$$u_j = \mathrm{Norm}\left(\frac{\sum_{k=1}^{K} q_{j,k} \bar{c}_k^{graph}}{\max\left(\sum_{k=1}^{K} q_{j,k}, 1\right)}\right) \in \mathbb{R}^{d_g}.$$

The structural similarity between historical interaction ℓ and target exercise e_j is $s_{\ell j} = \max_{k:q_{\ell,k}=1} \left(\bar{c}_k^{graph}\right)^\top u_j$.

Given a structural similarity threshold τ and soft-relation strength γ, the soft structural relevance is calculated as

$$o_{\ell j}^{soft} = \gamma \left[\mathrm{clip}\left(\frac{s_{\ell j} - \tau}{1 - \tau}, 0, 1\right)\right]^2 \in [0, \gamma].$$

The final relevance weight of historical interaction ℓ is

$$o_{\ell j} = \max\left(o_{\ell j}^{exact}, o_{\ell j}^{soft}\right) \in [0, 1].$$

Let $d_\ell = L - \ell$ denote the relative distance between historical interaction ℓ and the current target interaction. A recency-aware relevance weight is defined as

$$w_{\ell j} = o_{\ell j} \exp\left(-\frac{d_\ell}{10}\right).$$

The total relevance mass is $C_j = \sum_{\ell=1}^{L} o_{\ell j}$, and the indicator of whether relevant historical evidence exists is $I_j^{rel} = \mathbb{I}(C_j > 0)$.

The student's global historical correct-response ratio is $p^{glob} = \frac{1}{L}\sum_{\ell=1}^{L} r_\ell$. The recency-weighted target-relevant correct-response ratio is $\widetilde{p_j^{loc}} = \frac{\sum_{\ell=1}^{L} w_{\ell j} r_\ell}{\sum_{\ell=1}^{L} w_{\ell j} + \varepsilon}$. When no relevant historical interaction exists, the local response ratio falls back to the global response ratio $p_j^{loc} = I_j^{rel}\widetilde{p_j^{loc}} + \left(1 - I_j^{rel}\right) p^{glob}$. The relative amount of target-relevant historical evidence is defined as

$$\rho_j^{cnt} = \frac{\log(1 + C_j)}{\log(1 + L)}.$$

Let $d_j^{\min} = \min_{\ell:o_{\ell j}>0} d_\ell$ denote the distance to the most recent relevant historical interaction. The recency feature is defined as

$$\rho_j^{rec} = \begin{cases} \dfrac{1}{1+d_j^{\min}}, & C_j > 0, \\ 0, & C_j = 0. \end{cases}$$

For temporal information, the logarithmic time interval is $\widetilde{t_\ell} = \log(1+\Delta t_\ell)$. The global average temporal feature is $t^{\bar{glob}} = \frac{1}{L}\sum_{\ell=1}^{L}\widetilde{t_\ell}$, and the target-relevant temporal feature is $t_j^{\bar{loc}} = \frac{\sum_{\ell=1}^{L} w_{\ell j}\widetilde{t_\ell}}{\sum_{\ell=1}^{L} w_{\ell j}+\varepsilon}$. Their normalized difference is defined as $\delta_j^{time} = I_j^{rel}\tanh\left(t_j^{\bar{loc}} - t^{\bar{glob}}\right)$. Finally, the target-relevant local statistical feature vector is constructed as $r_j^{loc} = \left[p_j^{loc}, p^{glob}, p_j^{loc}-p^{glob}, \rho_j^{cnt}, \rho_j^{rec}, \delta_j^{time}, I_j^{rel}\right]^\top \in \mathbb{R}^7$ and is mapped into the hidden space according to

$$\Delta h_j^{local} = \mathrm{MLP}_{loc}\left(r_j^{loc}\right) \in \mathbb{R}^d.$$

C Initial Exercise ID Representation

To characterize the evolution of exercise-specific ID representations during training, we retain a fixed copy of the initial ID embedding for each exercise.

Let $E^{id} \in \mathbb{R}^{N\times d}$ denote the learnable exercise ID embedding matrix, where N is the number of exercises and d is the hidden dimension. After model parameter initialization and before training, a fixed copy of the embedding matrix is retained as

$$E^0 = \mathrm{StopGrad}\left(E^{id}\right) \in \mathbb{R}^{N\times d},$$

where $\mathrm{StopGrad}(\cdot)$ indicates that the copied representation is excluded from gradient updates. For target exercise e_j, its initial ID embedding is therefore defined as $e_j^0 = E_{j,:}^0 \in \mathbb{R}^d$, while the current learnable ID embedding is $e_j^{id} = E_{j,:}^{id} \in \mathbb{R}^d$.

During training, e_j^{id} is updated through gradient optimization, whereas e_j^0 remains fixed. Therefore, $\left|e_j^{id} - e_j^0\right|$ captures the change of the learned ID representation relative to its initialization, while $e_j^{id} \odot e_j^0$ characterizes their element-wise interaction. These two representations are jointly used as the input to the adaptive fusion gate that balances exercise ID information and textual semantics.

D Knowledge-Concept Structural Representation

To capture structural relationships among knowledge concepts induced by their exercise associations, we first construct a concept-relation graph from the Q-matrix and then derive fixed structural representations through spectral decomposition.

Concept Co-occurrence. Given the exercise–concept association matrix $Q \in \{0,1\}^{N\times K}$, the concept co-occurrence matrix is computed as

$$M = Q^\top Q \in \mathbb{R}^{K\times K},$$

where $M_{k,k'}$ denotes the number of exercises involving both concepts c_k and $c_{k'}$, and $M_{k,k}$ corresponds to the occurrence frequency of concept c_k. The co-occurrence-based cosine similarity is defined as

$$S_{k,k'} = \frac{M_{k,k'}}{\sqrt{M_{k,k}M_{k',k'}}},$$

and a positive pointwise mutual information term is further computed as

$$P_{k,k'} = \max\left[\log\frac{M_{k,k'}+\varepsilon}{M_{k,k}M_{k',k'}/N}, 0\right].$$

The preprocessing procedure combines cosine similarity and PPMI to characterize structural relations between concepts.

For concept pairs satisfying both the similarity threshold and minimum co-occurrence requirement, we define

$$A^{sim}_{k,k'} = \mathbb{I}\left(S_{k,k'} \geq \tau\right) \mathbb{I}\left(M_{k,k'} \geq m_{\min}\right) S_{k,k'}.$$

The normalized PPMI score is $P^{-}_{k,k'} = \frac{P_{k,k'}}{\max_{u,v} P_{u,v} + \varepsilon}$, and the final concept-relation matrix is constructed as $A_{k,k'} = \max\left(A^{sim}_{k,k'}, P^{-}_{k,k'}\right)$.

Relations with insufficient co-occurrence are removed. In the current implementation, $m_{\min} = 2$, while the similarity threshold is automatically determined from the positive cosine similarities when no explicit threshold is specified.

Spectral Structural Embedding. After obtaining the concept-relation matrix A, self-loops are added and symmetric normalization is applied

$$\widetilde{A} = A + I, A_{norm} = D^{-\frac{1}{2}} \widetilde{A} D^{-\frac{1}{2}},$$

where $D_{k,k} = \sum_{k'} \widetilde{A_{k,k'}}$.

We then perform eigendecomposition $A_{norm} = U \Lambda U^{\top}$. The d_g leading eigencomponents are retained to construct the structural representation matrix

$$C^{graph} = U_{d_g} \sqrt{\max\left(\Lambda_{d_g}, 0\right)} \in \mathbb{R}^{K \times d_g}.$$

The structural representation of knowledge concept c_k is given by

$$c^{graph}_{k} = C^{graph}_{k,:} \in \mathbb{R}^{d_g}.$$

These representations are precomputed offline and kept fixed during DBSCD training. The implementation performs eigendecomposition of the normalized relation matrix and constructs the embedding from the leading eigenvectors weighted by the square roots of the corresponding nonnegative eigenvalues.

E Preprocessing of the Computer Dataset

The Computer dataset used in this study was collected from course examination materials provided by university instructors. The original data contain examination papers completed by students from different classes and their corresponding response records. Since the raw data were stored as course examination files, we perform a unified preprocessing procedure to construct data suitable for cognitive diagnosis.

E.1 Construction of Student–Exercise Interaction Data

We first parse the original examination files and extract exercise information and student responses, including question contents, question types, options, correct answers, and students' answers.

Since examinations from different classes may contain duplicated exercises, we match and remove duplicate exercises according to exercise text, question type, options, and correct answers. Each unique exercise is assigned a unified identifier $\mathrm{exer_id}$.

The student response is then converted into a binary response variable according to the comparison between the student's answer and the correct answer $y_i \in \{0, 1\}$, where $y_i = 1$ indicates that the student correctly answers exercise, otherwise $y_i = 0$.

Finally, each interaction record contains the student identifier, exercise identifier, and response outcome, forming the student–exercise interaction sequences.

E.2 Knowledge Concept Extraction and Q-matrix Construction

Since the original Computer dataset does not contain predefined knowledge concept annotations, we further construct exercise–concept associations based on course textbooks and examination exercises.

First, the computer fundamentals textbook is divided according to its chapter and section structure to obtain corresponding learning materials. Then, each examination exercise, its correct answer, and the related textbook contents are provided as input to a large language model for assisting the extraction of core knowledge concepts involved in the exercise.

For each candidate concept, the large language model generates the concept name, description, and supporting textbook evidence to ensure that the extracted concepts are grounded in the course materials. To reduce potential errors introduced by automatic generation, the extracted concepts and exercise–concept associations are further manually reviewed by annotators with relevant course knowledge. The review process examines the validity of concepts, consistency with textbook evidence, and correctness of exercise–concept associations. After human verification, only validated concepts and associations are retained.

Based on the verified knowledge concepts, we construct the exercise–concept association matrix $Q \in \{0,1\}^{N\times K}$, where N denotes the number of exercises and K denotes the number of knowledge concepts. Specifically, $Q_{j,k} = 1$ indicates that exercise j involves knowledge concept k, while $Q_{j,k} = 0$ indicates no association.

The resulting Q-matrix describes the relationships between exercises and knowledge concepts and serves as the knowledge structure input for DBSCD.

E.3 Summary of Data Processing

Through the above preprocessing procedure, the original examination materials are transformed into standardized cognitive diagnosis data, including student–exercise interaction records; exercise information; knowledge concept set; exercise–concept association matrix. This procedure enables the offline examination data to be utilized for training and evaluating cognitive diagnosis models.

F Baseline Descriptions

We provide brief descriptions of the nine baseline methods used in our experiments.

1. **IRT.** IRT models students' response probabilities by estimating latent abilities and item parameters based on psychometric theory.
2. **NCD.** NCD employs neural networks to capture nonlinear interactions among students, exercises, and knowledge concepts while maintaining cognitive diagnosis interpretability.
3. **ID-CDM.** ID-CDM adopts a response–proficiency–response paradigm, using an inductive diagnostic module to infer students' cognitive states from response data and a predictive module to estimate response probabilities.
4. **ACD.** ACD incorporates students' affective states into cognitive diagnosis to improve the estimation of students' knowledge mastery.
5. **ICDM.** ICDM introduces item and concept information interaction modeling to enhance student cognitive state representations.
6. **DFCD.** DFCD integrates multi-source information with a dual-fusion framework to improve cognitive diagnosis performance.
7. **HCD.** HCD utilizes hierarchical structures and attention mechanisms to capture complex relationships among students, exercises, and knowledge concepts.
8. **DMC-CDM.** DMC-CDM employs conditional diffusion with multi-perspective observations to refine latent cognitive state representations.
9. **DBCD.** DBCD adopts counterfactual learning and contrastive learning strategies to alleviate bias in student response modeling.