

# 因果驱动的自适应去噪认知诊断框架

张桂衍<sup>1),2)</sup> 袁冠<sup>1),2)</sup> 张艳梅<sup>1)</sup> 闫秋艳<sup>1)</sup> 刘上<sup>1)</sup>

<sup>1)</sup>(中国矿业大学计算机科学与技术学院/人工智能学院 江苏 徐州 221116)

<sup>2)</sup>(中国矿业大学矿山数字化教育部工程研究中心 江苏 徐州 221116)

**摘要** 在教育领域,认知诊断旨在通过学生的作答记录来了解他们对知识的掌握水平,对习题推荐、个性化学习路径生成等下游应用有着重大影响,在智慧教育系统中扮演着重要的角色。尽管现有的认知诊断模型利用图神经网络等方法在准确性上取得了显著的进展,但这些方法往往忽略了数据中噪声引发的错误引导,使诊断结果可能严重偏离学生真实的知识掌握状态。在本文中,我们从因果角度分析了教育数据的生成过程,通过因果图揭示了噪声的影响机制。为此,本文提出了一种因果驱动的自适应去噪认知诊断框架,通过双阶段去噪策略提高了认知诊断模型的鲁棒性。首先,将学生、题目和知识构建为异构图并利用图神经网络进行表示学习。其次,设计了一种基于因果关系的学生表示去噪方法以获得更可靠的学生表示,然后基于学生表示和题目表示计算边的可靠性,自适应地去除不可靠的作答记录。最后,我们使用基于去噪结构的表示和基于原始结构的表示进行自监督对齐,在保证模型准确性的同时提升鲁棒性,从而获得了准确且鲁棒的可信认知诊断结果。在三个真实数据集上的大量实验表明,该框架有效地提升了认知诊断模型的效果,同时在不同数量级噪声的情况下证明了本框架可以一直保持最佳的准确性,特别是在增加了15%的噪声情况下,相比于最先进的方法平均提高了32.82%的准确率。

**关键词** 认知诊断; 图神经网络; 因果关系; 噪声; 鲁棒性

中图法分类号 TP18 DOI号 10.11897/SP.J.1016.2026.00557

## Causality-Driven Adaptive Denoising Cognitive Diagnostic Framework

ZHANG Gui-Xian<sup>1),2)</sup> YUAN Guan<sup>1),2)</sup> ZHANG Yan-Mei<sup>1)</sup> YAN Qiu-Yan<sup>1)</sup> LIU Shang<sup>1)</sup>

<sup>1)</sup>(School of Computer Science and Technology/School of Artificial Intelligence, China University of Mining and Technology, Xuzhou, Jiangsu 221116)

<sup>2)</sup>(Mine Digitization Engineering Research Center of the Ministry of Education, China University of Mining and Technology, Xuzhou, Jiangsu 221116)

**Abstract** In the field of education, cognitive diagnosis, which aims to understand students' knowledge mastery level through their response logs, plays an important role in intelligent education systems and has a significant impact on downstream applications such as exercise recommendation and personalized learning path generation. With the rapid development of deep learning, cognitive representation modeling has become an important paradigm in the field of cognitive diagnosis. Although existing cognitive diagnostic models have made significant progress in accuracy using methods such as graph neural networks, existing cognitive diagnostic methods often assume that the data is trustworthy and ignore the impact of noise, which may lead to the model being misled by noise and reduce the credibility of cognitive

收稿日期: 2025-06-16; 在线发布日期: 2025-11-06。本课题得到国家自然科学基金(6250071514)、徐州市重点研发计划项目(KC23296)、徐州市科技基金(KC22047)、中国矿业大学研究生创新计划项目(2024WLKXJ183)、中央高校基本科研业务费专项资金(2024-10949)、江苏省研究生科研与实践创新计划(KYCX24\_2781)资助。张桂衍, 博士研究生, 中国计算机学会(CCF)学生会员, 主要研究领域为可信任人工智能、图学习。E-mail: guixian@cumt.edu.cn。袁冠(通信作者), 博士, 教授, 中国计算机学会(CCF)高级会员, 主要研究领域为智能信息数据处理、大规模图数据计算。E-mail: yuanguan@cumt.edu.cn。张艳梅(通信作者), 博士, 副教授, 中国计算机学会(CCF)专业会员, 主要研究领域为软件分析与测试、软件缺陷预测。E-mail: ymzhang@cumt.edu.cn。闫秋艳, 博士, 教授, 中国计算机学会(CCF)专业会员, 主要研究领域为教育大数据挖掘、时序数据挖掘。刘上, 博士, 副教授, 中国计算机学会(CCF)专业会员, 主要研究领域为图数据分析、隐私保护。

diagnostic results in real-world scenarios. In particular, with the popularization of computing devices and cloud computing technology, more and more students are learning and answering questions through online education systems, and the errors caused by erroneous clicks have exacerbated the impact of noise. Existing cognitive diagnostic methods tend to ignore noise-induced misdirection in the data, which makes the diagnostic results may seriously deviate from the students' real knowledge mastery status. In this paper, we analyze the process of generating educational data from a causal perspective and analyze the impact of noise in the data modeling process. To this end, this paper proposes a causality-driven adaptive denoising cognitive diagnosis framework named CADCD, which improves the robustness of the cognitive diagnosis model through a two-stage denoising strategy. First, students, exercises and knowledge concepts are constructed as heterogeneous graphs and represented for learning using graph neural networks. Faced with the complex noise present in the real world, we categorize noise into two types: institutional noise and occasional noise. We use a causal graph to illustrate the differences and impacts of these two types of noise, and design methods to remove each type of noise. Specifically, a causality-based denoising method for student representations is designed to obtain more reliable student representations, and then unreliable response logs are adaptively removed using Bernoulli distribution based on the reliability of the computed edges of student and exercise representations. Finally, we perform self-supervised alignment between the denoised structure-based representation with the original structure-based representation to improve robustness while ensuring model accuracy, resulting in accurate and robust trusted representations of student, exercise, and knowledge concepts. We leverage existing cognitive diagnostic models to perform response prediction based on node representations to evaluate the effectiveness of cognitive diagnostics. In this paper, we conduct experiments based on three publicly available educational datasets. We compare state-of-the-art cognitive diagnostic models and frameworks, and the experimental results show that the CADCD framework effectively improves the performance of existing cognitive diagnostic models. Extensive ablation experiments validate the necessity of each module. We demonstrate the rationality of the denoising model through a case study. Furthermore, to verify the robustness of each framework, we artificially added different orders of magnitude of noise to the three datasets. The final experimental results prove that CADCD has the best robustness, especially at a noise level of 15%, where it achieves an average accuracy improvement of 32.82% compared to the best existing method.

**Key words** cognitive diagnosis; graph neural network; causality; noise; robustness

## 1 引 言

智慧教育是以学生全周期行为数据为核心,依托大规模数据处理技术构建的新型教育范式<sup>[1]</sup>。随着计算设备与云计算技术的普及,教育场景产生的数据呈现指数级增长,智慧教育逐渐成为社会各界关注和研究的热点<sup>[2]</sup>。认知诊断(Cognitive Diagnosis, CD)作为教育数据挖掘的关键技术,其目的是通过对海量学习数据进行特征提取与知识状态建模<sup>[3]</sup>以评估学生知识掌握情况。如图 1 所示,认知诊断是智慧教育系统的重要基础组成部分<sup>[4]</sup>,为多项下游任务提供了决策基础。

近年来,认知诊断在教育测量领域取得了显著

进展,其中最具代表性的模型包括经典的项目反应理论(Item Response Theory, IRT)<sup>[5]</sup>和新兴的神经认知诊断模型(Neural Cognitive Diagnosis Models, NCDM)<sup>[6]</sup>。认知诊断的核心框架由两个相互关联的组件构成:(1)状态诊断模块,负责推断学生的知识掌握水平;(2)作答预测模块,基于交互函数实现对学生在题目上的作答行为的准确拟合。IRT 框架采用单维潜在变量来表征学生的整体能力水平,其基于 Logistic 回归方程构建的交互函数虽然具有参数解释性强的优势,但难以捕捉多维知识结构中的复杂交互关系。NCDM 通过引入多层感知机(Multi-Layer Perceptron, MLP)作为非线性交互函数,配合面向知识概念的向量化表征,显著提升了

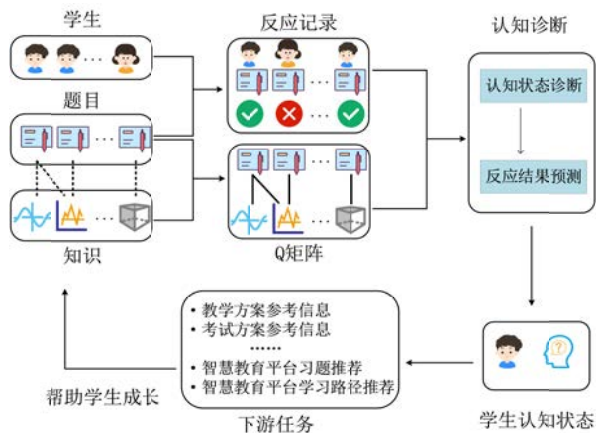


图1 认知诊断的应用

模型对高阶认知交互的建模能力。

随着深度学习的飞速发展, 认知表示建模已成为认知诊断领域的重要范式。近期研究普遍将学生作答记录(Response logs)和题目知识关联矩阵( $Q$ 矩阵)作为输入<sup>[6-8]</sup>, 其中 $Q$ 矩阵 $Q \in \{0,1\}^{M \times Z}$ 是题目和知识之间关系的结构化表示, 其元素 $Q_{ij} = 1$ 表示题目 $x_i$ 与知识 $k_j$ 相关, 否则为0。该矩阵能够精确建立题目与知识体系的映射关系, 是构建认知诊断模型的重要基础。认知诊断模型将学生认知状态建模为可更新的认知状态表示, 题目则建模为包含知识点关联权重的题目表示。配合图神经网络(Graph Neural Network, GNN)<sup>[9]</sup>等技术, 能够有效捕捉知识组件间的潜在关联<sup>[10-11]</sup>。认知诊断方法的精确性依赖于两个重要因素: 其一是表示学习的精确性, 要求模型既能准确量化学生对特定知识点的掌握程度, 又能识别概念间的迁移关系; 其二是响应预测的可靠性, 要求模型确保能够基于历史作答数据, 准确推断学生在新型题目组合上的表现。

然而, 现实场景中的数据往往存在噪声。在认知诊断的实际应用中, 数据质量隐患已成为制约模型可信性的关键瓶颈, 例如学生答题时的失误可能会导致模型错误评估学生的知识状态、题目表述模糊或知识点关联错误可能导致大量学生的诊断结果出现偏差。本研究首次通过结构因果方程<sup>[12]</sup>从噪声角度构建了认知诊断场景的因果图, 从因果角度将噪声分为两类并探究其影响, 进而揭示了两种噪声的表现形式及作用机制: (1)持续性系统噪声(Institutional noise, I)反映数据内部固有偏差对 $Q$ 矩阵标注质量和作答记录的深层干扰<sup>[13-14]</sup>, 其影响 $Q$ 矩阵和学生作答记录 $L$ 的可靠性。例如, 一些复杂题目可能作为经典例题被老师讲解或者进行大

量重复训练(如题海战术), 虽然学生们并没有掌握这个知识点, 但被重复训练内化为程序性记忆, 从而也能准确回答相关题目。持续性系统噪声会导致学生作答记录 $L$ 与真实知识状态间产生虚假相关性(Spurious correlation)<sup>[15]</sup>, 导致这种内在系统性噪声更加隐蔽, 很容易被模型错误地学习。(2)瞬时性偶然噪声(Occasional noise, O)表征学生在特定时刻通过猜测获得正确响应或通过失误出现错误响应的偶然性偏差<sup>[14,16]</sup>, 其指向过去的学生作答记录 $L$ , 影响了学生作答记录的可靠性。图2展示了从因果角度构建的认知诊断数据生成过程。其中, $Q$ 代表题目与知识的关联矩阵, $L$ 代表学生对题目的作答记录, $I$ 代表持续性系统噪声, $O$ 代表瞬时性偶然噪声。需要说明的是, 本文构建的因果图旨在厘清噪声如何影响观测数据。主要目标是分析噪声影响的主干因果结构, 而非穷尽所有变量。为清晰展示噪声在因果角度的影响, 教育场景中的部分复杂变量被简化, 如学生个体能力差异、题目语言表述偏差、交互时序等。 $I$ 指向 $Q$ 代表着 $I$ 会影响 $Q$ 矩阵的可靠性, 做对某个题目不代表学会了其背后的知识。 $I$ 指向 $L$ 代表 $I$ 会影响作答记录的可靠性, 部分同学可能通过老师讲解或者题海战术做对高难度的题目。 $O$ 指向 $L$ 代表 $O$ 会影响作答记录的可靠性, 部分同学可能在回答某个题目时猜对或失误, 导致单次作答无法反映其真实水平。现有的认知诊断方法往往假设数据是可信任的, 忽略了这两类噪声的影响, 导致模型可能被噪声误导, 降低了认知诊断结果在实际场景中的可信性。因此, 如何设计一种自适应去噪机制, 从而在保留有效认知信息的同时消除噪声对表示学习的影响, 成为了一个关键问题。

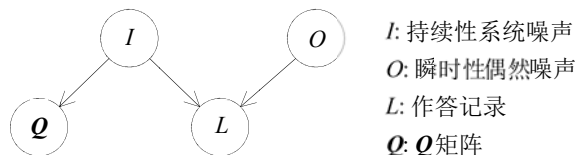


图2 噪声角度的认知诊断因果图

为了解决这个问题, 本文提出了一种因果驱动的去噪自适应去噪认知诊断框架(Causality-driven Adaptive Denoising Cognitive Diagnostic framework, CADCD), 通过两阶段去噪分别移除瞬时性随机噪声和持续性系统噪声以提高认知诊断方法的鲁棒性。首先, 构建基于 $Q$ 矩阵和作答记录的学生-题目-知识交互图, 通过图神经网络捕获交互图中的

信息并更新交互图中的节点表示,并可以将现有认知诊断算法作为作答预测模块以进行验证和约束。其次,针对同时影响作答记录和 $Q$ 矩阵的持续性系统噪声,采用基于题目表示对学生认知表示进行去噪,从而获得了更可信的学生认知表示。然后,利用节点表示来评估节点之间交互的可靠性,并通过自适应采样获得去噪后的邻接矩阵以消除交互中存在的噪声。接下来,通过去噪后的邻接矩阵更新节点表示,获得更加鲁棒的学生认知状态表示,从而提高了认知诊断结果的可信性。最后,在三个真实世界数据集上,本文提出的因果驱动自适应去噪认知诊断框架取得了最优的效果,并且在不同程度的噪声干扰下实现了最佳的鲁棒性。本文的贡献可以总结如下:

(1)首次构建了噪声角度的认知诊断因果图,根据不同的出现原因将数据中的噪声划分为瞬时性偶然噪声和持续性系统噪声,并从因果角度分析了不同类型噪声的影响。

(2)提出了一种因果驱动的去噪认知诊断框架,针对同时影响作答记录和 $Q$ 矩阵的持续性系统噪声提出了自适应学生认知表示去噪方法,然后基于去噪后的学生表示实现自适应交互结构去噪,最终通过自监督对齐方法实现更加鲁棒的诊断效果。

(3)三个真实数据集上的大量实验证明了CADCD框架有效地提升了认知诊断模型的准确性,同时在人为增加的5%、10%和15%三个不同数量级噪声的情况下都表现出了最佳的鲁棒性。

## 2 认知诊断相关工作

认知诊断作为心理测量领域的重要方法,历经数十年发展已形成成熟的评估体系。该方法通过建立数学模型,基于被试者在标准化测验中的外显反应,对其内在隐藏的认知特征进行量化推断<sup>[3]</sup>。相较于可直接观测的物理属性,人类认知能力属于潜在心理特质。与物理测量对象不同,一个人的能力水平是一种不能直接观察到的心理特征,隐藏在反应中的能力水平。因此,个人能力测量的基本思想是通过测试,根据个体的表现来推断被测者的能力。认知诊断已被广泛应用于教育<sup>[16]</sup>、医疗<sup>[17]</sup>等领域,从而为下游服务提供更好的认知状态感知。一个完整的认知诊断过程包含多个步骤:1)制作严谨的问卷或测试题目以构建用于教育和心理治疗等领域的反应收集。题目和相关属性或知识之间的关系通

常由专家提供;2)收集学生做题的结果;3)认知诊断模型设计;4)将诊断出的能力水平反馈给下游任务。根据认知诊断模型的不同,反馈可能会有所不同。在本节中,我们将认知诊断模型分为三种类型来进行介绍,分别为基于浅层模型的认知诊断方法、基于神经网络的认知诊断方法和基于图神经网络的认知诊断方法。

### 2.1 基于浅层模型的认知诊断方法

基于浅层模型的认知诊断方法可以分为基于统计模型和基于传统机器学习两类方法。作为心理测量学的理论演进,项目反应理论(Item Response Theory, IRT)<sup>[5]</sup>的提出开创了潜变量建模的先河。IRT通过构建项目特征曲线(Item Characteristic Curve, ICC),采用数理函数刻画被试者潜在特质与项目反应概率间的非线性关系,奠定了计算机自适应测试的理论基础。早期研究聚焦于单维IRT及多维IRT(Multidimensional IRT, MIRT)<sup>[18]</sup>的拓展,这些模型被统称为能力水平范式<sup>[19]</sup>,因其采用一维或多维潜变量表征被试者的整体能力水平。然而,传统测量理论仅能评估个体的宏观能力水平,难以揭示知识结构的微观特征<sup>[19]</sup>。 $Q$ 矩阵( $Q$ -matrix)的引入对解决这一问题具有里程碑意义。该矩阵以二元组形式编码测验项目与目标认知属性间的映射关系,为诊断模型提供了认知验证的理论锚点。基于 $Q$ 矩阵的约束条件,研究者相继提出AHM<sup>[20]</sup>,DINA<sup>[21]</sup>等代表性认知诊断模型。这些模型通过将被试者分类至特定的掌握模式,实现对多维度认知属性的联合诊断,标志着认知诊断从宏观能力评估向微观认知结构诊断的范式转变。然而,这类方法高度依赖于人工先验知识,难以刻画认知属性的连续发展过程,非常容易受主观构建误差的影响。

近年来,认知诊断研究呈现显著的跨学科融合趋势,机器学习技术的介入推动了该领域的范式革新。Chiu等人<sup>[22]</sup>采用K-means聚类结合层次凝聚聚类分析。Liu等人<sup>[23]</sup>将诊断模型问题转换为超维空间中的二次优化问题,利用支持向量机(Support Vector Machine, SVM)取得了优秀的结果。一些研究将认知诊断视为用户建模问题,采用协同过滤和矩阵分解方法来模拟学习者的能力并预测学习者的测试表现<sup>[24-25]</sup>。这些数据驱动的方法可以自动挖掘属性间的潜在关联,然而,这类方法往往基于线性假设或低维空间来构建认知诊断逻辑,难以捕捉学生认知能力与多维知识点之间复杂的非线性关

联。同时，这类方法往往依赖人工设计的显式特征，无法从原始交互数据中自主挖掘潜在的行为模式。当面对海量的真实教育数据时，浅层模型易陷入维度灾难(Curse of dimensionality)与过拟合陷阱，其诊断结果的解释性和泛化性显著受限。

## 2.2 基于神经网络的认知诊断方法

随着计算设备的进步，深度神经网络(Deep Neural Network, DNN)被引入认知诊断方法之中。Gierl 等人<sup>[26]</sup>提出了一种基于神经网络的能力分类器，该分类器使用预先训练的属性分层模型生成的数据进行训练。Wang 等人<sup>[6]</sup>认识到专家设计的交互功能的局限性，并提出了一种新的数据驱动的认知诊断框架，将学生因素、题目因素和交互函数融入基于神经网络的认知诊断方法中。Wang 等人<sup>[27]</sup>将学生情绪融入认知诊断当中，实现了认知状态和情绪状态的联合建模。然而这些方法虽然通过神经网络可以很好地拟合学生和习题的交互关系，但神经网络模型的大量参数难以将诊断结果映射到教育心理学理论维度，缺乏可解释性。朱天宇等人<sup>[28]</sup>基于认知诊断模型实现了精准的习题推荐，展示了认知诊断在智慧教育领域的必要性。张所娟等人<sup>[29]</sup>提出利用模糊测度的深度神经网络来进行认知诊断。Shen 等人<sup>[30]</sup>提出了一种符号认知诊断框架来混合优化表示和参数，增强了认知诊断的泛化能力和可解释性。基于深度学习的模型与心理测量学的理论相结合，具有更好的复杂认知过程拟合能力以及更好的可解释性的优势<sup>[31]</sup>。

另一方面，针对现实场景中的不同需求，Dong 等人<sup>[32]</sup>提出了一种用于认知诊断的增量学习框架以应对这个问题。Zhang 等人<sup>[33]</sup>提出了一种可靠认知诊断框架，通过量化诊断反馈的置信度使得其针对不同的应用场景拥有更高的灵活性。Zhang 等人<sup>[34]</sup>分析了现有认知诊断模型的不公平现象，通过确保敏感信息独立来消除敏感属性对学生熟练度的影响。Xu 等人<sup>[35]</sup>提出了一种兼顾公平性和准确性的认知诊断模型，保证了不同群体之间的预测公平。Zhang 等人<sup>[36]</sup>整合了单调性假设，以建立实现公平和准确的数据增强约束。Zhao 等人<sup>[37]</sup>通过多视角下的条件扩散模型实现了更精准的认知诊断。Li 等人<sup>[38]</sup>针对教育数据的有序多元素类别的特点，提出了一个可解释认知诊断模型。面向跨学科场景，Liu 等人<sup>[39]</sup>提出了一种基于软提示的认知诊断框架。虽然现有方法考虑了特定场景下认知诊断模型的改

进需求，但是这些方法高度依赖于优秀的的数据质量，很容易受到现实数据中噪声的干扰。同时，传统深度学习方法通常将学生、题目和知识点视为独立实体，难以学习三者间错综复杂的交互关系，导致认知状态推理过程中交互结构信息的损失。

## 2.3 基于图神经网络的认知诊断方法

尽管认知诊断模型在交互建模方面取得显著进展，但传统方法中诊断因子仅依赖标识符进行初始嵌入(ID-based embedding)，导致潜在语义信息缺失<sup>[3]</sup>。这一瓶颈促使研究者着力提升诊断因子的表征能力，通过融合多源异构信息构建具有认知可解释性的嵌入空间。随着图神经网络的广泛研究与应用<sup>[40-42]</sup>，一些研究者已经意识到学生、题目甚至知识概念之间的相关性可以形成图结构，因此将图神经网络纳入他们的认知诊断模型中。

在这些工作中，图神经网络主要用于改进学生、题目和知识概念的嵌入或对不同知识概念掌握之间的影响传播进行建模。许多研究引入基于教育先验的关系图，例如知识概念图和项目概念关联图，以增强学习者和项目的表示<sup>[43-44]</sup>。Gao 等人<sup>[45]</sup>对学习-项目-知识的异构图结构进行建模，以充分探索概念图中节点之间的高阶交互关系和知识概念之间的依赖关系，从而增强学习者认知状态和项目特征表示。Li 等人<sup>[7]</sup>提出了 HierCDF 框架来模拟分层知识结构对认知诊断的影响。这些方法虽然构建了知识依赖图，但未区分依赖关系之间的强度差异，将项目难度、区分度等特征与知识静态绑定，忽视了学生之间的潜在关联。Song 等人<sup>[46]</sup>专注于知识概念图与知识概念依赖关系和项目特征的有效融合。金天成等人<sup>[47]</sup>在认知诊断的基础上引入知识图谱来作为外部知识，实现了更为准确的习题推荐。Li 等人<sup>[31]</sup>揭示了知识概念和项目之间的关系以及知识概念图中的概念依赖关系，增强了项目和学生特征的表示。这样虽然提高了认知诊断效果的准确性，但在实际应用时往往导致学生认知状态建模的过拟合，使得模型容易被噪声误导。

然而，现有方法往往忽略了噪声的影响，噪声会通过图神经网络的迭代传播机制逐层扩散，模型可能将系统噪声误认为真实认知规律，导致诊断结果产生结构性偏移。现有的一些图神经网络方法已经注意到噪声的影响，例如 Qian 等人<sup>[16]</sup>通过数据增强和自监督对齐来获得了相对鲁棒的表示，Yao 等人<sup>[48]</sup>通过损失函数来约束表示的鲁棒性。但这些方

法并没有从因果视角分析不同类型噪声对认知诊断任务的影响,从而导致认知诊断模型难以自适应地应对噪声数据,使得模型容易受到噪声干扰,影响了认知诊断模型在现实应用的可信性。

### 3 因果驱动的自适应去噪认知诊断框架

本节首先定义了认知诊断问题,然后展开了对

CADCD框架的介绍。在本框架中,我们首先设计了交互图构建策略和节点表示学习方法,然后分别给出了自适应学生表示去噪和自适应交互结构去噪方法的详细说明,最后给出优化方法,实现了对学生认知状态进行鲁棒诊断,形成状态诊断模块。最后将现有的认知诊断方法引入作为作答预测模块,从而实现准确且鲁棒的学生作答预测结果。CADCD的研究框架如图3所示。

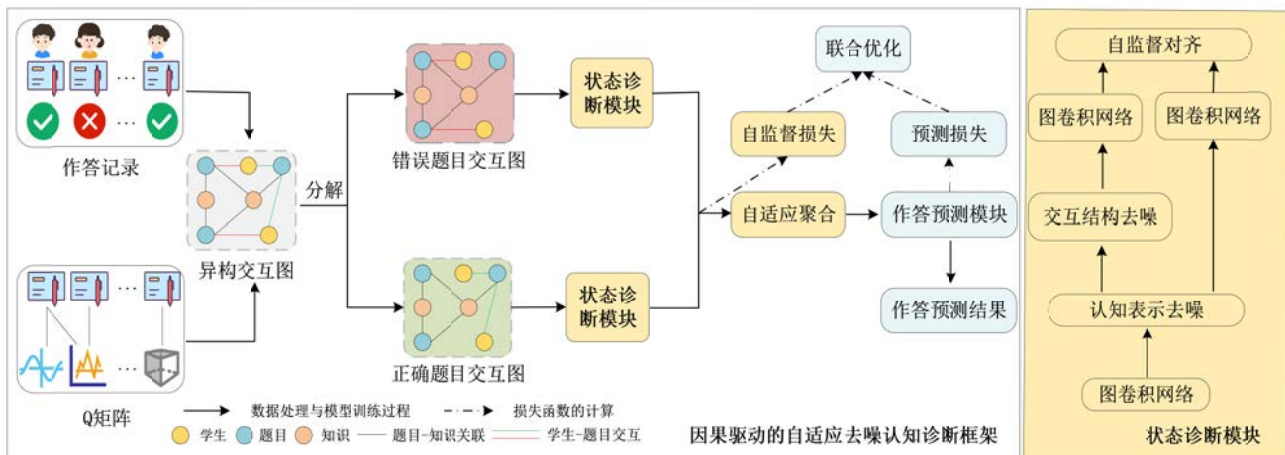


图3 CADCD框架图

#### 3.1 问题定义

对于给定的教育认知诊断场景,设学生数量为 $N$ 、题目数为 $M$ ,知识数量为 $Z$ ,则教育数据可以表示为三个集合:学生集合 $S = \{s_1, s_2, \dots, s_N\}$ ,题目集合 $X = \{x_1, x_2, \dots, x_M\}$ 和知识概念集合 $K = \{k_1, k_2, \dots, k_Z\}$ 。

**定义 1(Q矩阵).** 题目和知识的关联矩阵,在认知诊断过程中称为 $Q$ 矩阵 $Q \in \{0, 1\}^{M \times Z}$ ,其中 $Q_{ij}$ 表示题目 $x_i$ 与知识 $k_j$ 是否相关。

**定义 2(作答记录).** 学生根据自身兴趣和学习需求选择题目进行作答,进而形成回答题目是否正确的作答记录 $L$ 。

**定义 3(认知诊断任务).** 给定作答记录 $L$ 和题目-知识关系矩阵 $Q$ ,认知诊断的任务是推断学生对知识概念的潜在掌握水平向量,以表示学生对每个知识概念的掌握程度。

#### 3.2 交互图构建与节点表示学习

在认知诊断中,主要输入数据是题目-知识二元关系矩阵 $Q$ 和代表学生-题目关系的作答记录 $L$ 。为了更好地利用交互信息并进行建模,本文将这些复杂信息解构为学生、题目和知识三类节点并将其统一到异构图之中。为了清晰地描述本文提出的因

果驱动的自适应去噪认知诊断(CADCD)框架,本节只描述交互图的构建以及节点表示学习过程,将去噪过程放在第3.3节和第3.4节之中。需要说明的是,学生-题目-知识异构图的节点表示学习和去噪过程共同组成了CADCD框架中的状态诊断模块。

首先使用可训练嵌入 $H_S \in \mathbb{R}^{N \times d}$ ,  $H_X \in \mathbb{R}^{M \times d}$ ,  $H_K \in \mathbb{R}^{Z \times d}$ 对学生、题目和知识进行编码,从而获得了每个节点的初始表示。跟随之前的工作<sup>[16]</sup>,本文根据学生做题是否正确将学生、题目和知识之间的关联划分成正确交互图和错误交互图两个异构图。如式(1)所示,  $A_R$ 为答题正确的异构图,  $A_W$ 为答题错误的异构图。

$$\tilde{A}_R = \begin{bmatrix} \mathbf{O} & I_R & \mathbf{O} \\ I_R^\top & \mathbf{O} & Q \\ \mathbf{O} & Q^\top & \mathbf{O} \end{bmatrix}, \quad \tilde{A}_W = \begin{bmatrix} \mathbf{O} & I_W & \mathbf{O} \\ I_W^\top & \mathbf{O} & Q \\ \mathbf{O} & Q^\top & \mathbf{O} \end{bmatrix} \quad (1)$$

其中 $I_R$ 代表正确的做题记录交互矩阵,  $I_W$ 代表错误的做题记录交互矩阵,  $\mathbf{O}$ 代表零矩阵,  $T$ 代表将矩阵进行转置。

在认知诊断场景中,学生、题目、知识仅为ID向量的特性和稀疏的交互结构,本文采用LightGCN<sup>[49]</sup>作为节点表示方法,其舍弃了传统图卷积网络中的特征变换矩阵和非线性激活函数,有效

地减轻了过拟合风险<sup>[16,49]</sup>，仅通过邻域聚合传播学生、题目和知识的嵌入，从而保留了三者之间更纯粹的协同信号。同时，通过对交互图进行对称归一化以避免度分布不平衡的影响，实现了更稳定的梯度传播：

$$\tilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \quad (2)$$

其中， $\mathbf{A} \in \{\mathbf{A}_R, \mathbf{A}_W\}$  为原始邻接矩阵，度矩阵  $\mathbf{D} \in \mathbb{R}^{(N+M+Z) \times (N+M+Z)}$ ，其对角元素满足：

$$D_{ii} = \sum_j \mathbb{I}(A_{ij} \neq 0) \quad (3)$$

其中， $\mathbb{I}(\cdot)$  为指示函数， $D_{ii}$  表示邻接矩阵第  $i$  行的非零元素数量。

然后，节点表示基于归一化的交互图进行分层信息传播与融合：

$$\mathbf{H}^{(l)} = \tilde{\mathbf{A}} \mathbf{H}^{(l-1)} \quad (4)$$

其中， $\mathbf{H}^{(l)} \in \mathbb{R}^{(N+M+Z) \times d}$  代表包含学生、题目和知识点的第  $l$  层节点表示， $d$  为节点表示的维度。

接下来，分别在正确交互图和错误交互图上学习得到节点表示。为防止直接叠加反应信号导致的信息混淆，使用双通道自适应聚合器来融合两个图上的节点表示：

$$\mathbf{H}_{WR}^{(l)} = \phi(\mathbf{W}_R^{(l)} \mathbf{H}_R^{(l)} + \mathbf{W}_W^{(l)} \mathbf{H}_W^{(l)}), \quad (5)$$

其中， $\mathbf{H}_{WR}$  代表融合了正确交互图表示  $\mathbf{H}_W$  和错误交互图表示  $\mathbf{H}_R$  的融合表示， $\mathbf{W}_R^{(l)}, \mathbf{W}_W^{(l)} \in \mathbb{R}^{d \times d}$  是可训练参数矩阵， $\phi(\cdot)$  代表双通道自适应聚合器，在本文中我们使用多层感知机。最终的节点表示  $\mathbf{H}$  由对每层的节点表示进行平均池化得到：

$$\mathbf{H} = \frac{1}{L} \sum_{l=0}^L \gamma^{(l)} \mathbf{H}_{WR}^{(l)}, \quad (6)$$

其中， $\gamma^{(l)}$  为相同的重要性权重，且满足  $\sum_{l=0}^L \gamma^{(l)} = 1$ 。

### 3.3 自适应学生认知表示去噪

学生的认知状态和题目所代表的知识具有高度相关性，然而，题目本身可能因教学过程中答案泄露或题海战术等情况，导致学生作答数据存在噪声干扰。为了确保教育评估的客观性，经典的项目反应理论将学生能力与题目参数分离为两个相互的独立变量进行建模<sup>[5]</sup>。这种分离保证了学生能力评估不受题目特性干扰，同时题目参数的校准可以独立于特定学生群体。

然而，在基于神经网络的方法中，特别是基于信息传递机制的神经网络方法中，节点之间的相

互作用使得不同类型的节点不可避免地相互影响，这可能导致在节点表示更新过程中，学生的真实能力信息被题目的特性信息所污染，带来持续性系统噪声。如果发现某些题目与学生之间存在某种相关性，而这种相关性并非由学生真实的认知引起，而是由题目设计中的某种“缺陷”或日常学习中的“题海战术”引起，那么我们就可以尝试将这种“缺陷”带来的影响从学生认知状态中剥离以实现去噪。因此，当我们假设所有学生都可以准确捕捉题目信息，不会因为个人能力而错误理解题目时，即真实认知水平与题目独立，我们可以基于半同胞回归<sup>[50]</sup>的思想来实现学生认知表示的去噪：

$$\widehat{\mathbf{H}}_S = \mathbf{H}_S - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X] \quad (7)$$

其中， $\widehat{\mathbf{H}}_S$  为重构后的学生认知表示。

通过这种去噪方式，可以实现更准确的学生认知建模。本文通过定理 1 在理论上证明了这一点。对于存在系统噪声的场景，研究人员往往对数据进行中心化处理以消除整体均值偏移<sup>[51]</sup>，即  $\mathbf{H}_S - \mathbb{E}[\mathbf{H}_S]$ 。在理论分析中，我们引入一个潜在变量  $T$  来表示学生的真实认知状态，这是一个不受任何噪声干扰的理想表示。而现实中我们建模得到的学生表示  $\mathbf{H}_S$  是对  $T$  的一个有偏观测估计，不可避免的受到噪声影响，即  $Z = T - \mathbf{H}_S$ ，其中  $Z$  为噪声引起的偏差。本节的目标正是要获得一个更接近  $T$  的重构表示  $\widehat{\mathbf{H}}_S$ 。过往研究已经证明<sup>[50,52]</sup>，基于条件期望的去噪思想在解决混淆变量问题上具有理论上的优越性。定理 1 表明，我们的重构表示  $\widehat{\mathbf{H}}_S$  在期望上更接近真实认知状态  $T$ ，即实现了去噪。

**定理 1.** 假设存在满足  $T$  和  $\mathbf{H}_X$  相互独立的任意随机变量  $T$ ， $\mathbf{H}_S$ ， $\mathbf{H}_X$ ，以下期望不等式成立：

$$\mathbb{E}[(T - \mathbb{E}[T] - (\mathbf{H}_S - \mathbb{E}[\mathbf{H}_S]))^2] \geq \mathbb{E}[(T - \mathbb{E}[T] - \widehat{\mathbf{H}}_S)^2] \quad (8)$$

其中， $\widehat{\mathbf{H}}_S = \mathbf{H}_S - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X]$ 。

证明. 首先，对于任意随机变量  $\mathbf{H}_S$ ，根据条件方差的定义，可以得到：

$$\text{Var}[\mathbf{H}_S | \mathbf{H}_X] = \mathbb{E}[(\mathbf{H}_S - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X])^2 | \mathbf{H}_X] \quad (9)$$

在此处使用全期望定律 (Law of Total Expectation)，对两边关于  $\mathbf{H}_X$  取期望可得：

$$\mathbb{E}[\text{Var}[\mathbf{H}_S | \mathbf{H}_X]] = \mathbb{E}[(\mathbf{H}_S - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X])^2] \quad (10)$$

令  $Z = T - \mathbf{H}_S$ ，即认知诊断场景中，学生表示  $\mathbf{H}_S$  相对于其真实认知状态  $T$  的偏差。式(8)中的右边可重写为  $\mathbb{E}[(Z - \mathbb{E}[Z | \mathbf{H}_X])^2]$ 。那么根据式(10)，我们可以将其重写为

$$\mathbb{E}[\text{Var}(Z | \mathbf{H}_X)] \quad (11)$$

式(8)中的目标表达式的左边可重写为

$$\mathbb{E}[(Z - \mathbb{E}[Z])^2] = \text{Var}(Z) \quad (12)$$

由全方差定律(Law of total variance)可知,  $\text{Var}(Z) = \mathbb{E}[\text{Var}(Z | \mathbf{H}_X)] + \text{Var}(\mathbb{E}[Z | \mathbf{H}_X])$ 。由此, 式(12)的右侧可以视为

$$\mathbb{E}[\text{Var}(Z | \mathbf{H}_X)] + \text{Var}(\mathbb{E}[Z | \mathbf{H}_X]) \quad (13)$$

由于  $T$  和  $\mathbf{H}_X$  相互独立, 对  $Z = T - \mathbf{H}_S$  的条件期望为  $\mathbb{E}[Z | \mathbf{H}_X] = \mathbb{E}[T | \mathbf{H}_X] - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X]$ , 可以将其转化为  $\mathbb{E}[T] - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X]$ , 其方差为

$$\text{Var}(\mathbb{E}[Z | \mathbf{H}_X]) = \text{Var}(\mathbb{E}[T] - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X]) = \text{Var}(\mathbb{E}[\mathbf{H}_S | \mathbf{H}_X]) \quad (14)$$

因此, 公式(13)可以重写为

$$\mathbb{E}[\text{Var}(Z | \mathbf{H}_X)] + \text{Var}(\mathbb{E}[\mathbf{H}_S | \mathbf{H}_X]) \quad (15)$$

由于  $\text{Var}(\mathbb{E}[\mathbf{H}_S | \mathbf{H}_X])$  恒为正数, 因此可得

$$\mathbb{E}[(Z - \mathbb{E}[Z])^2] \geq \mathbb{E}[\text{Var}(Z | \mathbf{H}_X)] \quad (16)$$

综上所述, 我们可以得到

$$\mathbb{E}[(T - \mathbb{E}[T] - (\mathbf{H}_S - \mathbb{E}[\mathbf{H}_S]))^2] \geq \mathbb{E}[(T - \mathbb{E}[T] - \widehat{\mathbf{H}}_S)^2] \quad (17)$$

证毕。

由此, 我们得到了重构的学生认知状态表示  $\widehat{\mathbf{H}}_S = \mathbf{H}_S - \mathbb{E}[\mathbf{H}_S | \mathbf{H}_X] = T - \mathbb{E}[T] + n$ , 其中  $n$  代表难以通过回归消除的噪声。定理 1 证明了重构的学生认知表示至少比直接使用学生认知表示更接近于真实认知状态  $T$ , 这证明了本研究提出的表示去噪模块实现了针对持续性系统噪声的去噪。对于瞬时性猜测噪声, 本文通过下一节的自适应交互结构去噪方法进行进一步处理。

### 3.4 自适应交互结构去噪

邻域聚合机制的同质性假设会将噪声信号等同有效信息进行扩散, 因此这种噪声会在消息传递期间被放大并影响节点表示, 从而产生认知表征的偏移。在历史学习数据的交互图的表示学习过程中, 由于持续性系统噪声和瞬时性猜测噪声的影响, 噪声边不仅可能造成学生表示和题目表示的错误学习, 还可能扭曲题目与知识概念的映射。

对于认知诊断任务来说, 噪声边绝大部分处于学生与题目的交互边之中, 例如做题过程中常见的猜测行为、失误行为。但直接基于学生表示和题目表示计算交互的可信度是不可靠的, 因为存在着持续性系统噪声同样影响着学生表示。因此, 只有在

经历自适应学生表示去噪后, 我们才可以基于学生表示来更准确地计算学生和题目的交互可靠性。具体来说, 为了减轻噪声对交互结构的影响, 本文提出了一种自适应交互结构去噪方法。该方法旨在对学生和题目的交互边进行可靠性建模, 可靠性代表着对应边被保留的可能性。首先, 本文将不同类型的节点表示连接起来, 然后应用不同的变换矩阵分别获取均值和方差参数:

$$\mu_{ui} = \mathbf{W}_1^{\text{SX}} \cdot [\mathbf{h}_S^u \oplus \mathbf{h}_X^i] \quad (18)$$

$$\log \sigma_{ui}^2 = \mathbf{W}_2^{\text{SX}} \cdot [\mathbf{h}_S^u \oplus \mathbf{h}_X^i] \quad (19)$$

其中,  $\mathbf{W}_1^{\text{SX}}, \mathbf{W}_2^{\text{SX}}$  是学生一题目边的独立参数矩阵。

接下来, 分别将学生  $s_u$  和作答题目  $x_i$  的关联性建模为高斯分布, 并进行边权重采样, 从而获得每个边的可靠性表示:

$$w_{ui} = \mu_{ui} + \epsilon \cdot \sigma_{ui}, \quad \epsilon \sim \mathcal{N}(0, 1) \quad (20)$$

较大的可靠性代表着这条边更重要并且应当被保留, 如稳定知识掌握或典型错误的模式。而较小的可靠性表明这条边更可能是因为噪声而出现的, 如随机猜测、操作失误或答案泄露等因素。因此, 本文应用伯努利分布(Bernoulli distribution)来自适应的进行边去噪。在本文中, 经过伯努利分布采样得到的输出值是 0 或 1, 其通过式(21)进行计算:

$$p_{ui} = \sigma \left( \log t - \log(1-t) + \log \frac{\sigma(w_{ui})}{1 - \sigma(w_{ui})} \right) \quad (21)$$

其中,  $\sigma$  为 sigmoid 函数,  $t$  为分布(0,1)的采样。

由于正确作答和错误作答具有不同的模式, 本文提出基于作答正确性的二分图解耦去噪策略。正确交互图保留了学生正确作答题目的边来反映对知识掌握的能力, 需过滤因偶然猜测或临时记忆产生的噪声。正确交互图上的可靠边可以表达为  $a_{ui}^R = p_{ui}^R$ 。错误交互图保留了学生错误作答题目的边来反映认知缺陷或系统性误解, 需抑制由操作失误、题目歧义或错误标注导致的噪声。同理, 错误交互图上的可靠边可以表达为  $a_{ui}^W = p_{ui}^W$ 。

为防止去噪过程中过度剔除有效边导致信息损失, 本文使用一个自监督约束来强制去噪邻接矩阵与原始邻接矩阵在学生表示空间保持一致性:

$$L_{\text{ssl}} = - \sum_{s_u \in S} \log(\exp(\cos(\mathbf{h}_S^{du} \mathbf{h}_S^u)) / \tau) \quad (22)$$

其中,  $\cos$  代表余弦相似度,  $\mathbf{h}_S^{du}$  代表基于去噪邻接矩阵的学生  $s_u$  的认知表示,  $\mathbf{h}_S^u$  代表基于原始邻接矩阵的学生  $s_u$  的认知表示。

### 3.5 认知诊断输出与优化

为了实现认知诊断的鲁棒性增强,本文提出的CADCD框架采用模块化结构设计,可与任意现有的认知诊断模型实现即插即用式组合,即现有的认知诊断模型都可以作为框架中的作答预测模块。首先,对于部分需要固定维度的认知诊断模型,本文设计了特征转换层来解决模型的维度匹配问题:

$$\mathbf{H}^t = \text{ReLU}(\mathbf{H}\mathbf{W}_t + \mathbf{b}_t) \quad (23)$$

其中,  $\mathbf{W}_t \in \mathbb{R}^{d \times Z}$  为权重矩阵,  $\mathbf{b}_t \in \mathbb{R}^{(N+M+Z) \times 1}$  为偏置项。

在确保维度对齐之后,节点表征将作为下游诊断模型的输入,即进行学生与题目的交互预测:

$$\hat{Y}_{SX} = \text{CDM}(\widehat{\mathbf{H}}_S^{(t)}, \mathbf{H}_X^{(t)}, \mathbf{H}_K^{(t)}) \quad (24)$$

其中,CDM(Cognitive Diagnosis Model)表示任意认知诊断模型,  $(t)$  代表可能存在的特征转换。

在CADCD框架训练中,我们通过联合训练机制实现端到端优化。具体来说,本文首先采用负对数似然函数  $\mathcal{L}_{\text{BCE}}$  来衡量作答预测能力,也就是利用学生表示、题目表示和知识表示来预测学生是否可以做对题目:

$$-\sum_{(s,x,y_{sx}) \in L} [y_{sx} \log \sigma(\hat{y}_{sx}) + (1 - y_{sx}) \log(1 - \sigma(\hat{y}_{sx}))] \quad (25)$$

其中,  $L$  代表作答记录,  $y_{sx}$  代表记录中当前学生是否做对该题目。

然后,本文将预测损失与自监督对齐损失相结合以得到最终损失函数:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}} + \lambda_{\text{ssl}} \mathcal{L}_{\text{ssl}} \quad (26)$$

其中,  $\lambda_{\text{ssl}}$  为调整去噪强度的超参数。

## 4 实验分析

为了验证方法的有效性,本文在三个真实世界的教育数据集上进行了大量的实验验证。本节首先介绍了实验数据集和训练设置的具体情况,然后基于现有最先进的算法进行了大量的对比实验。同时,针对每个模块进行了消融实验。为了进一步解释去噪方法的原理,我们通过案例分析从统计特征上分析了噪声交互的判断依据。为了验证本文所提出方法的鲁棒性,实验中人为地增加不同程度的噪声以进行了鲁棒性测试。大量的实验数据表明,本文提出的CADCD框架增强了现有认知诊断模型的准确性和鲁棒性,并取得了最佳的效果。

### 4.1 实验数据

本文在三个真实公开的教育数据集上进行了大量实验,包括 Assist17、Neurips2020 和 Junyi。表1给出了数据集所对应的详细信息。

为了使数据集的情况更加清晰,本文引入了稀疏度、平均正确率和  $\mathbf{Q}$  矩阵密度三个指标来展示每个数据集的特点。稀疏性代表学生-题目交互矩阵的观测值比例,本质是数据缺失程度的量化指标,计算方式为|作答记录数量|/(|学生数量|×|题目数量|)。平均正确率反映了题目的平均难度。 $\mathbf{Q}$  矩阵密度可以理解为主题密度,用于衡量题目与知识点关联的复杂程度。

表1 数据集信息

数据集	学生数量	题目数量	知识数量	作答记录	稀疏度	平均正确率	$\mathbf{Q}$ 矩阵密度
Assist2017	1709	3162	102	390 311	0.072	0.815	1.22
Neurips2020	2840	6000	268	214 328	0.012	0.631	4.14
Junyi	10 000	734	734	408 057	0.055	0.687	1.00

Assist17数据集<sup>[53]</sup>是由ASSISTment智能导学系统构建的认知诊断基准数据集,具有三个数据集中最高的平均正确率和稀疏度。Neurips2020数据集<sup>[54]</sup>来源于来自一项名为“The NeurIPS 2020 Education Challenge”的竞赛,具有三个数据集中最高的题目数量和  $\mathbf{Q}$  矩阵密度。它包含Eedi教育平台上的学生在两个学年(2018~2020年)内对数学问题的回答记录。Junyi数据集<sup>[55]</sup>来自均一教育平台,其具有三个数据集中最多的学生数量、知识数量和答题记录数量。

### 4.2 评价指标

机器学习模型的验证往往依赖于标签,但在智慧教育场景中,由于学生的认知状态(如对特定知识点的理解深度)无法直接观测,本研究遵循认知诊断领域的标准验证范式<sup>[43-44]</sup>,通过评估模型对学生作答表现(即能否正确解答题目)的预测准确性来验证认知诊断的效果。预测准确率越高,表明模型对学生潜在认知状态的建模越精确。本文将原始学生反应日志数据集按7:1:2的比例分割为训练集、验证集和测试集。其中,训练集专门用于模型参数估计与学生认知状态的学习,验证集用于调整超参数,最终在测试集上通过二元分类指标进行性能验证。本文使用四个准确性指标和一个可解释性指标来全方面评价实验结果。

对于认知诊断的准确性,本文用四个指标分别从不同维度衡量分类模型的性能特征:(1)准确率

(Accuracy, ACC): 预测正确的样本占比, 直观地反映模型的全局预测精度; (2) 平均精度 (Average Precision, AP): 通过计算精度-召回率曲线 (Precision-Recall Curve) 的加权面积, 该指标重点关注正例样本的预测质量。与 AUC 不同, AP 对假正例赋予更高惩罚权重, 在认知诊断这种误报成本较高的应用中具有特殊意义。AP 值越接近 1, 表明模型在保持高召回率的同时具备更高的预测置信度; (3) 受试者工作特征曲线下面积 (Area Under receiver operating characteristic Curve, AUC): 通过积分受试者工作特征曲线下的面积来量化模型对正负样本的区分能力, 适用于类别不平衡场景下的性能评估; (4) F1: 作为精确率与召回率的调和平均数, 可以平衡查全率与查准率的价值。

在认知诊断中, 诊断结果的可解释性不仅是模型效度的核心验证标准, 更是实现个性化学习干预的关键前提。本研究采用在过往研究中被广泛使用且具有理论完备性的一致性度 (Degree of Agreement, DOA) 作为可解释性验证指标, 其构建逻辑基于对知识掌握状态的显性表征假设。假设存在两个学生  $s_i$  和  $s_j$ , 针对关联目标知识概念  $k_a$  的题目  $x_a$ , 若满足  $\text{Acc}(s_i, x_a) > \text{Acc}(s_j, x_a)$ , 则理论上应保证学生  $s_i$  在  $k_a$  上的掌握概率估计值优于学生  $s_j$ , 即对于作答记录中包含概念  $k_a$  且两人都做过但回答不同的练习, 应该有  $s_i$  答对而  $s_j$  答错的练习数量等于  $s_j$  和  $s_i$  回答不同的练习数量。参考之前的研究<sup>[16,30]</sup>, 本文将每个数据集中作答记录最多的十个知识点来作为 DOA 的验证依据。具体来说, 知识  $k_a$  的 DOA 的计算公式为

$$\text{DOA}_a = \frac{\sum_{s_i, s_j \in S} \delta(\text{Mas}_{s_i, k_a}, \text{Mas}_{s_j, k_a}) \cdot \alpha}{\sum_{s_i, s_j \in S} \delta(\text{Mas}_{s_i, k_a}, \text{Mas}_{s_j, k_a})} \quad (27)$$

$$\alpha = \frac{\sum_{e=1}^M Q_{ea} \wedge \varphi(x_e, s_i, s_j) \wedge \delta(r_{ei}, r_{ej})}{\sum_{e=1}^M Q_{ea} \wedge \varphi(x_e, s_i, s_j) \wedge I(r_{ei} \neq r_{ej})} \quad (28)$$

其中,  $\delta(\text{Mas}_{s_i, k_a}, \text{Mas}_{s_j, k_a})$  代表学生  $s_i$  在  $k_a$  上的认知水平是否高于  $s_j$ 。  $Q_{ea}$  代表题目  $x_e$  和知识点  $k_a$  是否有关。如果学生  $s_i$  答对了题目  $x_e$  但  $s_j$  没有, 则  $\delta(r_{ei}, r_{ej}) = 1$ , 否则为 0。  $\varphi(x_e, s_i, s_j)$  检查  $s_i$  和  $s_j$  是否都回答了题目  $x_e$ 。  $I(r_{ei} \neq r_{ej})$  代表  $s_i$  和  $s_j$  的回答是否相同。如果  $s_i$  答对了但是  $s_j$  答错了, 则  $\delta(r_{ei}, r_{ej}) = 1$ , 否则为 0。

### 4.3 对比算法

为系统评估 CADCD 的算法优势, 本文将框架与多个当前最先进的认知诊断模型进行融合, 将现有认知诊断模型作为作答预测模块:

(1) IRT<sup>[5]</sup>: 作为潜在特质模型的奠基性方法, IRT 通过单维能力参数构建学生认知状态表征, 其采用 Logistic 反应函数建立题目特征与学生表现的映射关系。

(2) MIRT<sup>[18]</sup>: 针对传统 IRT 的单维表征局限, MIRT 创新性地引入多维潜在空间以实现复杂认知状态的分布式表征, 从而有效捕捉知识结构中的隐式关联。

(3) NCDM<sup>[6]</sup>: 该模型通过多层感知机实现响应函数的自动化学习, 突破传统人工设计函数的结构限制, 通过端到端训练自动捕获题目与学生特征的深层交互模式。

(4) CDMFKC<sup>[43]</sup>: 该模型通过设计知识概念的难易程度和判别性, 进一步考虑了知识概念在认知诊断中的影响。

(5) KaNCD<sup>[44]</sup>: 该模型通过隐式建模知识点关联, 利用已覆盖知识点推断未覆盖知识点的熟练度, 提升未覆盖知识点的诊断可靠性。

为了更全面地评估 CADCD 框架的优势, 我们分别采用 LightGCN 和 ORCDF 两种基于图的学生状态建模框架进行表示学习, 并融合上述认知诊断模型来进行对比实验结果:

(1) LightGCN<sup>[49]</sup>: 该方法是一个经典的协同过滤算法。本文在构建学生、题目和知识交互图之后使用 LightGCN 进行学生认知状态诊断, 然后使用现有认知诊断模型作为作答预测模块。

(2) ORCDF<sup>[16]</sup>: 该框架通过构建异构交互图来增强现有的认知诊断模型, 同时通过对交互边的随机反转实现了更加鲁棒的认知诊断效果。

### 4.4 实施细节

在本文中, 我们采用 Xavier 参数初始化方法进行权重分配, 同时使用 Adam 优化器进行模型优化。学习率通过网格搜索来调整, 范围为  $\{0.004, 0.005\}$ 。自监督损失权重设置为 0.003。在进行作答预测时, 认知诊断模型的维度跟随各个方法原文的要求, NCDM 和 CDMFKC 的维度被设置为知识点的数量  $Z$  维, 其他方法的维度被统一设置为 32 维。所有数据集的批量大小 (Batch Size) 均设为 4096。本文所有的实验都在一张 NVIDIA Tesla V100 32GB 的

GPU上运行。为确保模型性能对比的严谨性和公平性，本文严格遵循各对比模型在其原始文献中公布的超参数配置。IRT和MIRT作为潜在特质模型的代表，它们输出的学生掌握度本质上是全局能力估计值，与具体知识点的掌握状态无因果关联，因此不适用于计算DOA。在后续的对比如实验结果中，本文使用“-”来表示这种不适用性。

#### 4.5 对比实验

本文分别在表2、表3和表4中展示了三个数据集的对比实验结果。从实验数据上可以看出，本文提出的CADCD框架在所有数据集上一致地改善了所有基础认知诊断模型的准确性和可解释性。同时超过了现有的认知诊断框架，在每一个基础认知诊断模型上都对准确性和可解释性实现了最大的提升。这证实了CADCD有效去除了数据中的噪声并实现了更准确的认知诊断。

表2 Assist17数据集上的对比实验结果

Method	ACC	AP	AUC	F1	DOA
IRT	85.56	96.46	88.27	91.66	-
LightGCN-IRT	86.75	97.08	89.43	92.11	-
ORCDF-IRT	86.83	97.08	89.43	92.13	-
CADCD-IRT	<b>86.93</b>	<b>97.15</b>	<b>89.57</b>	<b>92.25</b>	-
提高	0.12%	0.07%	0.16%	0.13%	-
MIRT	86.38	96.99	89.32	91.98	-
LightGCN-MIRT	86.07	96.52	88.58	91.89	-
ORCDF-MIRT	86.76	97.14	89.52	92.11	-
CADCD-MIRT	<b>88.38</b>	<b>97.67</b>	<b>91.58</b>	<b>92.99</b>	-
提高	1.87%	0.55%	2.30%	0.96%	-
NCDM	82.22	93.93	79.56	89.39	55.54
LightGCN-NCDM	83.29	94.96	82.94	89.87	56.94
ORCDF-NCDM	85.63	96.47	87.59	91.27	60.27
CADCD-NCDM	<b>86.57</b>	<b>97.05</b>	<b>89.49</b>	<b>91.79</b>	<b>63.90</b>
提高	1.10%	0.60%	2.17%	0.57%	6.02%
CDMFKC	83.01	95.76	84.34	89.93	55.98
LightGCN-CDMFKC	83.09	95.16	83.46	89.70	57.20
ORCDF-CDMFKC	86.94	97.05	89.58	92.08	62.70
CADCD-CDMFKC	<b>87.32</b>	<b>97.15</b>	<b>89.95</b>	<b>92.36</b>	<b>67.76</b>
提高	0.44%	0.10%	0.41%	0.30%	8.07%
KaNCD	85.03	96.12	86.45	90.98	57.9
LightGCN-KaNCD	85.49	96.31	87.05	91.25	58.38
ORCDF-KaNCD	87.14	97.10	89.73	92.27	62.42
CADCD-KaNCD	<b>87.59</b>	<b>97.23</b>	<b>90.23</b>	<b>92.50</b>	<b>67.53</b>
提高	0.52%	0.13%	0.56%	0.25%	8.19%

从实验结果中可以看出，MIRT通过关联学生的多个潜在维度获得了更强大的认知诊断能力，突破传统IRT单维能力假设的局限性，在每个数据集

上都取得了高于IRT的实验结果。CDMFKC在NCDM基础上，通过知识概念动态参数化实现诊断能力的跃升。通过引入动态知识概念影响参数量化不同知识点对得分的差异化影响，并结合习题区分度与知识难度构建多维嵌入空间。其采用基于知识影响强度的自适应判定边界机制，使预测更贴合实际教学场景。KaNCNCD通过潜在向量隐式学习知识点间关系，利用已覆盖知识点推断未覆盖知识点的熟练度，从而获得了更为准确的认知诊断效果。

LightGCN、ORCDF和CADCD利用图神经网络对学生、题目和知识的异构交互图进行学生认知状态诊断，并将现有模型作为作答预测模块的做法，显著提升了认知诊断模型的效果。这是因为教育场景中的核心实体本质上构成异构拓扑关系：学生通过答题行为与题目交互，题目通过知识标记与知识点关联。传统方法通常将学生和题目的交互视为独立事件，基于信息传递机制的图神经网络可以有效建模此类多跳关系交互的图结构数据。这种高阶关系的学习能力使模型能够从稀疏的交互数据中提取更深层的认知模式，从而帮助模型理解了学生、题目和知识之间的关系。

在认知诊断框架的架构设计中，学生-题目交互数据的特性与协同过滤算法展现出显著的范式适配性。学生作答记录本质上构成了一个隐式反馈系统，其中正确/错误二值响应可视为非显式评分行为的代理指标。这一特性十分契合协同过滤的假设：一方面，具有相似知识状态的学习者倾向于呈现相近的作答模式；另一方面，相同知识点的题目在解答过程中会引发类似的认知路径。因此，LightGCN作为优秀的协同过滤算法，可以很好地建模学生认知状态，从而取得了更好的认知诊断效果。ORCDF和CADCD中的图卷积部分也均采用了LightGCN而非传统的图卷积神经网络。

ORCDF通过将交互图分为正确题目交互图和错误题目交互图，并人为地翻转正确错误交互作为数据增强来增强模型的泛化性，进一步提高了认知诊断的效果。这是因为正确作答和错误作答反映了不同的行为模式，正确回答通常体现学生对知识点间逻辑关系(如数学中的定理依赖性)的稳定掌握，错误作答可能因为学生错误地理解了知识点或者无法关联相关知识点。随机反转本质上是扩充了训练数据集的多样性，使得模型能够更好地捕捉真实信号。然而，这些方法并没有从因果角度分析并处

理不同类型的噪声, 本文提出的 CADCD 有效地去除了噪声的影响并获得了最佳的认知诊断效果。

表 3 Neurips2020 数据集上的对比实验结果

Method	ACC	AP	AUC	F1	DOA
IRT	69.95	83.34	75.22	76.85	-
LightGCN-IRT	71.43	84.70	76.56	78.63	-
ORCDF-IRT	71.45	84.78	76.68	78.24	-
CADCD-IRT	<b>71.56</b>	<b>84.85</b>	<b>76.78</b>	<b>78.51</b>	-
提高	0.15%	0.08%	0.13%	0.35%	-
MIRT	70.09	83.06	74.51	77.84	-
LightGCN-MIRT	70.49	83.89	75.47	78.25	-
ORCDF-MIRT	71.54	84.68	76.58	78.35	-
CADCD-MIRT	<b>71.74</b>	<b>84.90</b>	<b>76.91</b>	<b>78.55</b>	-
提高	0.28%	0.26%	0.43%	0.26%	-
NCDM	69.73	82.77	74.19	76.80	66.55
LightGCN-NCDM	70.55	83.48	75.58	77.00	67.57
ORCDF-NCDM	71.36	84.80	76.68	77.72	69.25
CADCD-NCDM	<b>72.12</b>	<b>85.40</b>	<b>77.5</b>	<b>79.01</b>	<b>70.86</b>
提高	1.07%	0.71%	1.07%	1.66%	2.32%
CDMFKC	70.86	84.22	76.05	77.21	68.35
LightGCN-CDMFKC	71.62	84.69	76.82	78.42	69.90
ORCDF-CDMFKC	71.66	84.68	76.72	78.55	72.15
CADCD-CDMFKC	<b>71.97</b>	<b>85.15</b>	<b>77.23</b>	<b>78.86</b>	<b>72.42</b>
提高	0.43%	0.56%	0.66%	0.39%	0.37%
KaNCD	71.01	84.17	75.97	77.91	68.67
LightGCN-KaNCD	71.08	84.78	76.59	76.99	69.14
ORCDF-KaNCD	71.59	84.92	76.92	78.18	70.09
CADCD-KaNCD	<b>71.78</b>	<b>85.11</b>	<b>77.02</b>	<b>78.41</b>	<b>72.66</b>
提高	0.27%	0.22%	0.13%	0.29%	3.67%

值得注意的是, IRT 采用单维能力参数建模和 Logistic 响应函数, 这种线性决策边界难以捕捉教育数据中的复杂结构信息, 所以当 CADCD 等基于图神经网络的框架引入高维表示时, IRT 会因为维度过低而导致信息损失, 使得 CADCD 等框架对其增益效果相对不如对其他模型明显。另一方面, 在三个数据集上, 经过 LightGCN 和 ORCDF 增强的 MIRT 和相同条件增强下的 IRT 算法取得了相近的结果, 并没有充分展现 MIRT 的优势。这是因为单维 IRT 通过最大似然估计分离能力与题目参数, 在数据存在噪声时, 这种参数解耦设计避免了噪声在参数间的交叉污染。而 MIRT 依赖的多维参数空间对数据噪声敏感, 噪声通过 LightGCN 的邻域聚合被扩散到相邻节点, 引入了过拟合风险, 反而阻碍了认知诊断效果的提升。也正是因为这个原因, 当本研究使用 CADCD 框架来分别增强 IRT 和 MIRT 时, 由于 CADCD 框架实现了有效的去噪, 此时经

过 CADCD 框架增强后的 MIRT 依然取得了优于 IRT 的结果。

表 4 Junyi 数据集上的对比实验结果

Method	ACC	AP	AUC	F1	DOA
IRT	76.27	89.07	80.37	83.49	-
LightGCN-IRT	77.46	89.71	81.74	84.31	-
ORCDF-IRT	77.57	89.38	81.68	84.42	-
CADCD-IRT	<b>77.62</b>	<b>89.73</b>	<b>81.84</b>	<b>84.56</b>	-
提高	0.10%	0.39%	0.20%	0.17%	-
MIRT	77.09	89.18	80.95	84.25	-
LightGCN-MIRT	77.29	89.25	81.17	84.28	-
ORCDF-MIRT	77.51	89.15	81.33	84.25	-
CADCD-MIRT	<b>77.59</b>	<b>89.31</b>	<b>81.42</b>	<b>84.53</b>	-
提高	0.10%	0.18%	0.11%	0.33%	-
NCDM	74.27	87.92	78.07	82.59	50.41
LightGCN-NCDM	75.00	82.45	74.8	82.43	55.44
ORCDF-NCDM	76.86	89.33	80.89	83.78	59.43
CADCD-NCDM	<b>77.35</b>	<b>89.58</b>	<b>81.46</b>	<b>84.21</b>	<b>60.94</b>
提高	0.34%	0.27%	0.42%	0.13%	0.49%
CDMFKC	74.87	88.03	78.43	82.51	49.24
LightGCN-CDMFKC	76.19	88.73	80.83	82.47	59.33
ORCDF-CDMFKC	77.26	89.05	81.05	84.37	60.64
CADCD-CDMFKC	<b>77.52</b>	<b>89.29</b>	<b>81.39</b>	<b>84.48</b>	<b>61.24</b>
提高	0.64%	0.28%	0.70%	0.51%	3.05%
KaNCD	75.27	83.48	75.21	83.32	53.86
LightGCN-KaNCD	76.61	88.25	79.89	83.6	58.88
ORCDF-KaNCD	77.54	89.43	81.43	84.35	60.58
CADCD-KaNCD	<b>77.79</b>	<b>89.71</b>	<b>81.88</b>	<b>84.62</b>	<b>61.30</b>
提高	0.32%	0.31%	0.55%	0.32%	1.19%

#### 4.6 案例分析

为了清晰地展示交互去噪的合理性, 本节基于 Assist17 数据集中 ID 为 2 的学生进行了案例分析, 该学生共有 207 条作答记录。部分被视为噪声的交互记录如表 5 所示, 其中题目难度为所有学生的作答记录中的错误率, 值越高表示难度越大。掌握水平代表该学生在该知识点的相关题目上的正确率, 值越高代表该学生对该知识点的掌握越好。

认知诊断的目标是通过学生可观测的题目响应(作答结果), 去推断其不可观测的内在知识状态。其内涵的假设是一个作答行为的结果应当取决于学生的知识状态和题目属性。只有当学生的能力水平足以充分覆盖题目所要求的综合知识需求时, 其答对该题才符合认知规律; 反之, 若学生的能力无法满足题目需求, 则其答错属于预期内的结果。在表 5 的作答记录中, 题目#36 和#522 的难度极高, 而学生掌握水平不足。因此, 学生在此类题目上的正确应答, 极有可能是源于猜测所导致的假阳性结果。对于题目#128、#244 和#491 来说, 题目难度低且学

生掌握水平高。在此条件下出现的答错情形，与学生的实际能力明显不符，应属于因疏忽或错误点击等因素引起的假阴性结果。

表5 Assist17 数据集集中学生#2的部分作答记录

题目编号	是否答对	题目难度	知识编号	掌握水平
36	是	0.80	5	0.69
128	否	0.13	33	0.86
244	否	0.05	58	0.85
491	否	0.11	35	0.86
522	是	0.83	1	0.66

为本文基于学生表示和题目表示计算交互的可靠性以进行交互去噪，其本质是将学生知识状态和题目属性转化为可计算的匹配度指标。学生节点表示是包含所有知识点的掌握水平的综合编码。题目节点表示是其知识需求、难度、区分度等属性的综合编码。当学生与题目的匹配度较高，意味着学生的知识状态与题目的需求高度匹配，理论上应答对。若实际答错，则被判为假阴性噪声，如题目#128、#244和#491。移除这类假阴性噪声，防止模型对学生能力的过低估计。当学生与题目的匹配度较低，意味着学生的知识水平无法满足题目要求，理论上应答错。若实际答对，则会被判为假阳性噪声，如

题目#36和#522。移除这类假阳性噪声，防止了模型对学生能力的虚高估计。去噪后的作答记录不仅可以更加准确地评估学生知识水平，同时也更加准确地评估了题目属性中的难度和区分度。因此，我们的框架不仅提高了准确性指标，更提高了可解释性指标 DOA，这说明我们对学生作答的预测与学生的真实表现更为一致，更加值得信任。

#### 4.7 鲁棒性测试

为了验证 CADCD 对噪声的抵抗能力，本节在三个数据集上针对三个框架进行了鲁棒性测试。我们人为地在交互图中分别随机增加 5%，10%和 15%的交互并进行测试，即在学生节点和题目节点中间增加随机边。所有的鲁棒性实验都基于 KaNCD 来进行。为了使结果更加清晰，本文选用准确度作为评价标准。实验结果如图 4 所示，其中的(a)(b)(c)分别代表 Assist17、Neurips2020 和 Junyi 数据集上的鲁棒性测试实验结果。从图 4 中可以看出，本文提出的 CADCD 框架具有最佳的抗噪能力，无论在何种程度的噪声上，本文的 CADCD 都能展现出最佳的认知诊断能力。

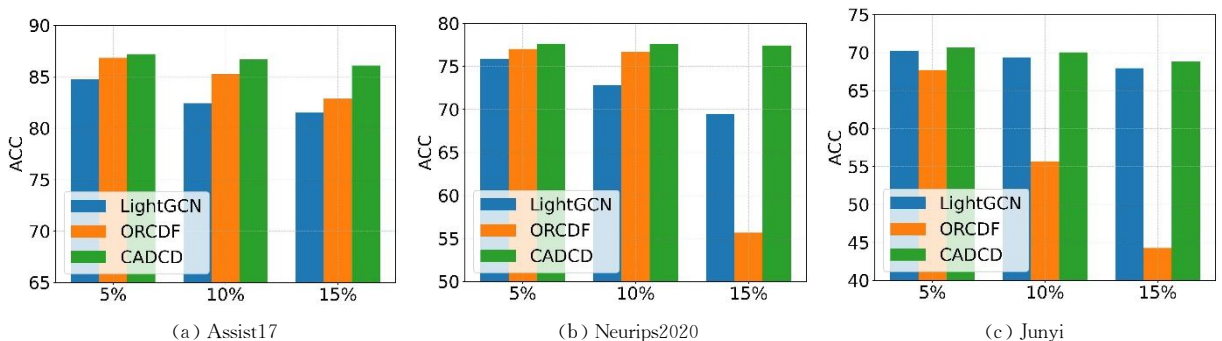


图4 在三个数据集上的鲁棒性测试，横坐标代表增加了不同比例的噪声

当增加 15%的噪声时，与最先进的 ORCDF 相比，CADCD 在 Assist17 数据集上提高了 3.87%，在 Neurips2020 数据集上提高了 39.11%，在 Junyi 数据集上提高了 55.48%。在三个数据集上平均提高了 32.82%的实验结果充分证明了 CADCD 的鲁棒性。在平均正确率最高的 Assist17 数据集上，ORCDF 展现出了比 LightGCN 更为优秀的抗噪能力。但在平均正确率较低的 Neurips2020 数据集和 Junyi 数据集上，ORCDF 反而受到了噪声较为强烈的干扰。尤其是在题目数量最多的 Neurips2020 数据集上，其结果相对受到了最大的影响。

这表明，虽然在正确作答较多的数据集上

ORCDF 可以通过随机反转交互类型来实现数据增强，从而获得更鲁棒的效果。这种增益主要源于当正确模式存在显著分布偏差时，随机反转策略能创造有意义的虚拟负样本，有效增加负样本的多样性，缓解了模型对正确模式的过拟合。但当面对正确错误作答都比较均匀且交互较多的情况时，随机反转交互类型并不能实现有效的数据增强，反而进一步扰乱了模型对于学生认知状态的学习，混淆了模型对学生知识状态演变规律的捕捉。与之形成鲜明对比的是，CADCD 通过双阶段自适应去噪实现了对噪声信号的精准处理，从而获得了最佳的鲁棒性。这种性能优势印证了自适应去噪策略有效消除了

基于随机反转的增强策略存在的干扰问题，证明了其在增强认知诊断鲁棒性上的重要作用。

#### 4.8 消融分析

为了清晰每个模块的重要性和必要性，本文在三个数据集上使用 KaNCD 作为反应预测模块进行了消融实验，实验结果如表 6、表 7 和表 8 所示。为了验证去噪模块的性能，本文基于控制变量的思想分别移除了表示去噪模块和结构去噪模块以观察其必要性和贡献。

表 6 Assist17 数据集上的消融实验结果

变体	ACC	AP	AUC	F1	DOA
CADCD	87.59	97.23	90.23	92.50	67.53
去除表示去噪	87.51	97.21	90.14	92.42	66.23
5%噪声	85.70	96.72	88.38	91.68	57.59
10%噪声	84.62	96.79	88.78	91.29	56.94
15%噪声	84.24	96.61	88.16	91.11	56.31
去除结构去噪	87.18	97.12	89.83	92.29	61.30
5%噪声	86.77	97.05	89.41	92.09	68.53
10%噪声	86.35	96.94	89.08	91.93	68.63
15%噪声	85.58	96.58	88.04	91.44	54.10

表 7 Neurips2020 数据集上的消融实验结果

变体	ACC	AP	AUC	F1	DOA
CADCD	71.78	85.01	77.02	78.41	72.66
去除表示去噪	77.76	89.52	81.64	84.57	60.73
5%噪声	77.08	89.50	81.29	84.38	54.94
10%噪声	76.69	89.41	81.11	84.45	52.61
15%噪声	76.86	88.25	80.44	84.09	54.35
去除结构去噪	77.65	89.80	81.93	84.39	53.03
5%噪声	77.39	89.39	81.50	84.50	54.25
10%噪声	77.31	89.33	81.33	84.40	54.11
15%噪声	77.27	89.54	81.40	84.43	53.86

表 8 Junyi 数据集上的消融实验结果

变体	ACC	AP	AUC	F1	DOA
CADCD	77.79	89.71	81.88	84.62	61.30
去除表示去噪	71.61	84.92	76.85	78.18	73.40
5%噪声	70.27	84.33	76.10	75.56	55.07
10%噪声	69.40	84.16	75.79	74.27	55.34
15%噪声	63.48	82.69	74.08	77.61	71.93
去除结构去噪	71.64	84.83	76.82	78.38	70.17
5%噪声	69.78	84.49	76.28	79.40	71.94
10%噪声	65.44	84.01	75.62	78.28	72.79
15%噪声	62.12	84.20	75.90	61.49	57.58

从实验结果中可以看出，每一个模块的去除都会引发准确性、鲁棒性和可解释性的下降，这证明了每个模块都可以有效地提升认知诊断的效果。为了评价每个模块对噪声的影响，本文在每个变体下进行了鲁棒性测试。与第 4.7 节类似，我们分别在异构图中增加 5%、10% 和 15% 的噪声交互边并进行

认知诊断。从消融实验结果中可以发现，不同数据情况下结构去噪模块与表示去噪模块的贡献度存在显著差异。

在 Assist17 数据集和 Neurips2020 数据集上，我们发现去除表示去噪模块后的鲁棒性普遍低于去除结构去噪模块的变体。这种差异源于其高题目密度与稀疏交互的数据特性，即该平台学生较少、题目较多，同时作答记录较少。在这种情况下，表示去噪所针对的持续性系统噪声虽然具有重要意义，但交互结构上的去噪显得更为关键。当单个学生仅提供有限信号时，知识拓扑的准确性成为关键锚点，少量的噪声就可能带来较大的影响。结构去噪通过修正交互异构图的错误关联边，确保了在数据稀疏条件下的推理可靠性。因此对于 Assist17 数据集和 Neurips2020 数据集来说，结构去噪模块更为必要。

然而，在学生数量和作答记录最多、题目最少的 Junyi 数据集上，表示去噪模块展现了更为关键的作用。该现象与 Junyi 数据集的高作答密度和低题目密度有关，即学生较多、作答记录较多但题目数量较少。海量学生数据和作答数据中的持续性系统噪声会展现广泛而强烈的影响，在这种情况下首先进行表示去噪显得更为必要。只有进行了表示去噪，才能获得可信任的学生认知表示，才能实现准确的结构去噪。直接进行结构去噪反而可能会被持续性系统噪声误导而去除了原本正确的交互边。

综上所述，消融实验证明了本文提出的 CADCD 通过双阶段自适应去噪可以适应不同特点的数据，以实现准确而鲁棒的可信认知诊断。

## 5 总结与未来工作

为了实现在教育领域的可信认知诊断，本文提出了一个因果驱动的自适应去噪认知诊断框架。通过在学生认知状态诊断阶段进行自适应去噪，集成并增强现有的认知诊断模型进行作答预测，从而提高认知诊断方法的准确性和鲁棒性。本文首次构建了去噪角度的认知诊断因果图，从因果角度出发设计了双阶段自适应去噪方法来去除持续性系统噪声和瞬时性偶然噪声。具体包括，从因果角度来去除表示中的噪声，并基于去噪后的表示实现了更为精准的结构去噪，然后通过自监督对齐方法来学习去噪后的交互结构，兼顾了认知诊断的鲁棒性和准确性。在三个真实数据集上的大量实验证明了本文

方法的准确性, 同时通过人为地增加了不同程度的噪声验证了本文方法的鲁棒性。在未来的工作中, 将针对认知诊断过程中的可解释性进行进一步研究。我们将考虑把学生个体能力差异、题目语言表述偏差和交互时序等变量纳入因果图中, 从而更加精细化地识别和去除噪声。

## 参 考 文 献

- [1] Bhutoria A. Personalized education and artificial intelligence in the united states, china, and india: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence*, 2022, 3: 100068
- [2] Zhang K, Aslan A B. Ai technologies for education: Recent research & future directions. *Computers and Education: Artificial intelligence*, 2021, 2: 100025
- [3] Liu Y, Zhang T, Wang X, et al. New development of cognitive diagnosis models. *Frontiers of Computer Science*, 2023, 17(1): 171604
- [4] Holmes W, Tuomi I. State of the art and practice in AI in education. *European Journal of Education*, 2022, 57(4): 542-570
- [5] Hambleton R K, Swaminathan H. *Item response theory: Principles and applications*. Springer Science & Business Media, 2013
- [6] Wang F, Liu Q, Chen E, et al. Neural cognitive diagnosis for intelligent education systems//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA. 2020, 34: 6153-6161
- [7] Li J, Wang F, Liu Q, et al. Hiercdf: A Bayesian network-based hierarchical cognitive diagnosis framework//*Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Washington, USA. 2022: 904-913
- [8] Yang S, Wei H, Ma H, et al. Cognitive diagnosis-based personalized exercise group assembly via a multi-objective evolutionary algorithm. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023, 7(3): 829-844
- [9] Zhou Y, Zheng H, Huang X, et al. Graph neural networks: Taxonomy, advances, and trends. *ACM Transactions on Intelligent Systems and Technology*, 2022, 13(1): 1-54
- [10] Ma C, Ouyang J, Xu G. Learning latent and hierarchical structures in cognitive diagnosis models. *Psychometrika*, 2023, 88(1): 175-207
- [11] Xie P, Li G, Li T. Knowledge tracing model based on exercise-knowledge point heterogeneous graph and multi-feature fusion. *Computer Science*, 2025, 52(03): 197-205(in Chinese)  
(解培中, 李冠进, 李汀. 基于试题-知识点异构图和多特征融合的知识追踪模型. *计算机科学*, 2025, 52(03): 197-205)
- [12] Pearl J. *Causality*. Cambridge: Cambridge University Press, 2009
- [13] Chew S L, Cerbin W J. The cognitive challenges of effective teaching. *The Journal of Economic Education*, 2021, 52(1): 17-40
- [14] Noorbehbahani F, Mohammadi A, Aminazadeh M. A systematic review of research on cheating in online exams from 2010 to 2021. *Education and Information Technologies*, 2022, 27(6): 8413-8460
- [15] Bao J, Zhang K, Wu L, et al. Conformity-aware debiased neural news recommendation with causal reasoning. *Chinese Journal of Computers*, 2024, 47(10): 2333-2351(in Chinese)  
(鲍纪敏, 张琨, 吴乐等. 从众性感知的因果去偏新闻推荐方法. *计算机学报*, 2024, 47(10): 2333-2351)
- [16] Qian H, Liu S, Li M, et al. Orcdf: An oversmoothing-resistant cognitive diagnosis framework for student learning in online education systems//*Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Barcelona, Spain. 2024: 2455-2466
- [17] Hallock H, Mantwill M, Vajkoczy P, et al. Sport-related concussion: A cognitive perspective. *Neurology: Clinical Practice*, 2023, 13(2): e200123
- [18] Chalmers R P. Mirt: A multidimensional item response theory package for the r environment. *Journal of Statistical Software*, 2012, 48: 1-29
- [19] Wang F, Huang Z, Liu Q, et al. Dynamic cognitive diagnosis: An educational priors-enhanced deep knowledge tracing perspective. *IEEE Transactions on Learning Technologies*, 2023, 16(3): 306-323
- [20] Leighton J P, Gierl M J, Hunka S M. The attribute hierarchy method for cognitive assessment: A variation on tatsuoaka's rule-space approach. *Journal of Educational Measurement*, 2004, 41(3): 205-237
- [21] De La Torre J. Dina model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 2009, 34(1): 115-130
- [22] Chiu C Y, Douglas J A, Li X. Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 2009, 74: 633-665
- [23] Liu C, Cheng Y. An application of the support vector machine for attribute-by-attribute classification in cognitive diagnosis. *Applied Psychological Measurement*, 2018, 42(1): 58-72
- [24] Pang Y, Jin Y, Zhang Y, et al. Collaborative filtering recommendation for mooc application. *Computer Applications in Engineering Education*, 2017, 25(1): 120-128
- [25] Yu S, Zeng Y, Pan Y, et al. Snmcf: A scalable non-negative matrix co-factorization for student cognitive modeling. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 36(7): 3486-3500
- [26] Gierl M J, Cui Y, Hunka S. Using connectionist models to evaluate examinees' response patterns to achievement tests. *Journal of Modern Applied Statistical Methods*, 2008, 7(1): 19
- [27] Wang S, Zeng Z, Yang X, et al. Boosting neural cognitive diagnosis with student's affective state modeling//*Proceedings of the AAAI Conference on Artificial Intelligence*. Vancouver, Canada. 2024, 38: 620-627
- [28] Zhu T, Huang Z, Chen E, et al. Cognitive diagnosis based personalized question recommendation. *Chinese Journal of Computers*, 2017, 40(1): 103-124(in Chinese)

- (朱天宇, 黄振亚, 陈恩红等. 基于认知诊断的个性化试题推荐方法. 计算机学报, 2017, 40(1): 176-191)
- [29] Zhang S, Yu X h, Chen E, et al. A concept interaction-based cognitive diagnosis deep model. *Pattern Recognition and Artificial Intelligence*, 2023, 36(1): 22-33(in Chinese)  
(张所娟, 余晓晗, 陈恩红等. 融合知识交互关系的认知诊断深度模型. 模式识别与人工智能, 2023, 36(1): 22-33)
- [30] Shen J, Qian H, Zhang W, et al. Symbolic cognitive diagnosis via hybrid optimization for intelligent education systems//*Proceedings of the AAAI Conference on Artificial Intelligence*. Vancouver, Canada. 2024, 38: 14928-14936
- [31] Li J, Liu Q, Wang F, et al. Towards the identifiability and explainability for personalized learner modeling: An inductive paradigm//*Proceedings of the ACM Web Conference 2024*. Singapore. 2024: 3420-3431
- [32] Tong S, Liu J, Hong Y, et al. Incremental cognitive diagnosis for intelligent education//*Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Washington, USA. 2022: 1760-1770
- [33] Zhang Y, Qin C, Shen D, et al. Relicd: A reliable cognitive diagnosis framework with confidence awareness//*Proceedings of the 2023 IEEE International Conference on Data Mining*. Shanghai, China. 2023: 858-867
- [34] Zhang Z, Wu L, Liu Q, et al. Understanding and improving fairness in cognitive diagnosis. *Science China Information Sciences*, 2024, 67(5): 152106
- [35] Xu H, Hou M, Wu L, et al. Fair personalized learner modeling without sensitive attributes//*Proceedings of the ACM on Web Conference 2025*. Sydney, Australia. 2025: 4612-4624
- [36] Zhang Z, Song W, Liu Q, et al. Towards accurate and fair cognitive diagnosis via monotonic data augmentation//*Proceedings of the 38th International Conference on Neural Information Processing Systems*. Vancouver, Canada. 2024: 47767-47789
- [37] Zhao G, Huang Z, Cheng C, et al. Multi-perspective consolidation enhanced cognitive diagnosis via conditional diffusion model//*Proceedings of the AAAI Conference on Artificial Intelligence*. Philadelphia, USA. 2025, 39: 1174-1182
- [38] Li X, Guo S, Wu J, et al. An interpretable polytomous cognitive diagnosis framework for predicting examinee performance. *Information Processing & Management*, 2025, 62(1): 103913
- [39] Liu F, Zhang Y, Liu S, et al. Prompt transfer for dual-aspect cross-domain cognitive diagnosis. *IEEE Transactions on Computational Social Systems*, 2025: 1-14
- [40] Zhang G, Zhang S, Yuan G. Bayesian graph local extrema convolution with long-tail strategy for misinformation detection. *ACM Transactions on Knowledge Discovery from Data*, 2024, 18(4): 1-21
- [41] Xu B, Cen T, Huang J, et al. A survey on graph convolutional neural networks. *Chinese Journal of Computers*, 2020, 43(5): 755-780(in Chinese)
- (徐冰冰, 岑科廷, 黄俊杰等. 图卷积神经网络综述. 计算机学报, 2020, 43(5): 755-780)
- [42] Zhang G, Yuan G, Cheng D, et al. Disentangled contrastive learning for fair graph representations. *Neural Networks*, 2025, 181: 106781
- [43] Li S, Guan Q, Fang L, et al. Cognitive diagnosis focusing on knowledge concepts//*Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. Atlanta, USA. 2022: 3272-3281
- [44] Wang F, Liu Q, Chen E, et al. Neuralcd: A general framework for cognitive diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(8): 8312-8327
- [45] Gao W, Liu Q, Huang Z, et al. Rcd: Relation map driven cognitive diagnosis for intelligent education systems//*Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Virtual, Canada. 2021: 501-510
- [46] Ma H, Song S, Qin C, et al. Dged: An adaptive denoising gnn for group-level cognitive diagnosis//*Proceedings of the 33rd International Joint Conference on Artificial Intelligence*. Jeju, Republic of Korea. 2024: 2261-2269
- [47] Jin T, Dou L, Xiao C, et al. Personalized OJ exercise recommendation method with memory and cognition merging. *Chinese Journal of Computers*, 2023, 46(1): 103-124(in Chinese)  
(金天成, 窦亮, 肖春芸等. 记忆与认知融合的个性化 OJ 习题推荐方法. 计算机学报, 2023, 46(1): 103-124)
- [48] Yao F, Liu Q, Yue L, et al. Adard: An adaptive response denoising framework for robust learner modeling//*Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024: 3886-3895
- [49] He X, Deng K, Wang X, et al. Lightgcn: Simplifying and powering graph convolution network for recommendation//*Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020: 639-648
- [50] Schölkopf B, Hogg D W, Wang D, et al. Modeling confounding by half-sibling regression. *Proceedings of the National Academy of Sciences*, 2016, 113(27): 7391-7398
- [51] Robinson C, Schumacker R E. Interaction effects:centering, variance inflation factor, and interpretation issues. *General Linear Model Journal*, 2009, 35(1): 6-11.
- [52] Yang Z, Feng J. A causal inference method for reducing gender bias in word embedding relations//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA. 2020, 34: 9434-9441.
- [53] Feng M, Heffernan N, Koedinger K. Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, 2009, 19: 243-266.
- [54] Wang Z, Lamb A, Saveliev E, et al. Results and insights from diagnostic questions: The neurips 2020 education challenge//*Proceedings of the NeurIPS 2020 Competition and Demonstration Track*. Vancouver, Canada. 2021: 191-205.

[55] Chang H S, Hsu H J, Chen K T. Modeling exercise relationships in e-learning: A unified approach//Proceedings of the 8th International

Conference on Educational Data Mining. Madrid, Spain. 2015: 532-535.



**ZHANG Gui-Xian**, Ph.D. candidate.

His research interests include trustworthy artificial intelligence, graph learning.

**YUAN Guan**, Ph.D., professor. His research interests include intelligent information and data processing, large-scale graph data computation.

### Background

Cognitive diagnosis, as a key technology of educational data mining, has the core task of assessing students' knowledge mastery through feature extraction and knowledge state modeling of massive learning data. Compared with traditional methods, graph neural network-based cognitive diagnosis methods have realized significant breakthroughs in data processing paradigms and model performance. While traditional methods mainly rely on linear assumptions and manual feature engineering, which are difficult to capture the higher-order associations between students, exercises, and knowledge concepts. Graph neural networks can naturally model the topology and interactive dependencies by constructing a student-exercise relational graph. These properties make graph neural networks a cutting-edge technological path in the field of cognitive diagnosis.

Existing cognitive diagnostic methods often assume that the data is trustworthy and ignore the impact of noise, which may lead to the model being misled by noise and reduce the credibility of cognitive diagnostic results in real-world scenarios. In particular, with the popularization of computing devices and cloud computing technology, more and more

**ZHANG Yan-Mei**, Ph.D., associate professor. Her research interests include software analysis and testing, software defect prediction.

**YAN Qiu-Yan**, Ph.D., professor. Her research interests include educational big data mining, time series data mining.

**LIU Shang**, Ph.D., associate professor. His research interests include graph data analysis, privacy protection.

students are learning and answering questions through online education systems, and errors caused by erroneous clicks exacerbate the impact of noise. However, existing methods often ignore the existence of different types of noise in student-exercise interaction data, which is amplified through the aggregation propagation mechanism of graph neural networks. In this paper, we categorize noise into two types based on causality and design different methods for denoising. The experimental results show that the framework proposed in this paper can effectively improve the accuracy and robustness of the existing cognitive diagnostic models with the best results.

This work was supported in part by the National Natural Science Foundation of China under Grant 6250071514, Xuzhou K&D Program under Grant KC23296, the Science and Technology Program of Xuzhou under Grant No. KC22047, the Graduate Innovation Program of China University of Mining and Technology 2024WLKXJ183, the Fundamental Research Funds for the Central Universities 2024-10949, and the Postgraduate Research & Practice Innovation Program of Jiangsu Province KYCX24\_2781.