**DOI:** 10.16725/j.1673-808X.2025309
**引用本文:** 郭美美, 肖刚. 基于多维项目反应理论的可解释动态学习评估模型[J]. 桂林电子科技大学学报, 2026, 46(3): 221-229.
Guo Meimei, Xiao Gang. Interpretable dynamic learning assessment model based on multidimensional item response theory[J]. Journal of Guilin University of Electronic Technology, 2026, 46(3): 221-229.

# 基于多维项目反应理论的可解释动态学习评估模型

郭美美，肖　刚

（韩山师范学院 数学与统计学院, 广东 潮州　521041）

**摘　要：** 精确地刻画学生的动态学习行为并输出具有认知解释性的诊断反馈, 是学习分析与智能教育评估领域的核心问题。为此, 构建了一种基于多维项目反应理论(MIRT)的可解释动态学习评估模型 DK-MIRT。该模型由知识状态表征层、认知建模层以及预测与诊断层三部分构成: 前者利用长短期记忆网络(LSTM)学习学生作答序列中的时序依赖, 中间层在 MIRT 的结构化框架下显示建模题目难度与区分度等认知参数, 后者将学生能力向量与题目属性进行多维交互以生成预测结果与诊断依据。通过这种结构化的协同机制, DK-MIRT 在统一框架内同时实现了较高精度预测与可解释认知诊断, 使模型不仅具备数据驱动的预测能力, 也具备心理测量意义上的透明度与可解释性。实验结果表明, DK-MIRT 在多个真实数据集上具有优于传统深度知识追踪方法的性能, 并能在教育实践中支持学习瓶颈识别、个性化练习路径优化以及知识掌握演化呈现等应用, 为自适应教育系统提供了一种兼具性能与解释性的动态学习评估工具。

**关键词：** 动态知识追踪; 多维项目反应理论; 可解释性学习评估; 深度学习; 认知诊断
**中图分类号：** TP391.7　　**文献标志码：** A　　**文章编号：** 1673-808X(2026)03-0221-09

# Interpretable dynamic learning assessment model based on multidimensional item response theory

GUO Meimei，XIAO Gang

（School of Mathematics and Statistics, Hanshan Normal University, Chaozhou 521041, China）

**Abstract:** Accurately characterizing students' dynamic learning behaviors and providing cognitively interpretable diagnostic feedback are two central challenges in learning analytics and intelligent educational assessment. To address these challenges, an interpretable dynamic learning assessment model, termed DK-MIRT (Dynamic Knowledge-aware Multidimensional Item Response Theory), is developed within the framework of multidimensional item response theory (MIRT). The proposed model consists of three functional modules: a knowledge-state representation layer, a cognitive modeling layer, and a prediction-and-diagnosis layer. The first module employs a long short-term memory (LSTM) network to capture temporal dependencies in response sequences; the second explicitly models item difficulty, discrimination, and other cognitive parameters within a structured MIRT framework; and the third integrates student ability vectors with item attributes through multidimensional interactions to produce both performance predictions and diagnostic evidence. Through this coordinated architecture, DK-MIRT simultaneously achieves high predictive accuracy and cognitive interpretability, providing not only strong predictive capability but also psychometrically grounded transparency and traceability. Experimental results on multiple real-world datasets demonstrate that DK-MIRT outperforms conventional deep knowledge tracing methods while supporting practical educational applications, including learning bottleneck identification, personalized practice recommendation, and knowledge mastery visualization. These findings demonstrate that DK-MIRT provides an effective assessment framework that combines high predictive performance with cognitive interpretability for adaptive educational systems.

**收稿日期:** 2025-12-09
**基金项目:** 韩山师范学院教育教学改革(HSJG-GK231133)
**通信作者:** 肖刚(1968—), 男, 教授, 博士, 研究方向为应用数学、医学影像学。E-mail: xiaogang@hstc.edu.cn

**Key words**: dynamic knowledge tracing; multidimensional item response theory; interpretable learning assessment; deep learning; cognitive diagnosis

随着教育技术的迅猛发展, 智能辅导系统(intelligent tutoring systems, 简称 ITS)[1] 和在线学习平台在全球范围内得到广泛应用, 尤其在 K-12 和高等教育场景中, 已成为个性化学习与实时反馈的重要基础设施[2]。近期研究表明, ITS 能够通过分析学习轨迹生成定制化教学内容, 从而提升学习动机与理解深度[3]。与此同时, AI 驱动的学习分析技术使教育者能够实时监控学习状态, 识别薄弱知识点, 并生成精确干预策略, 从而在大规模环境下支持个性化教育决策。近年来, 深度学习方法在复杂模式识别与高维特征建模任务中展现出显著优势, 并已在缺陷检测、目标识别及工程智能分析等多个领域得到应用[4-6]。

在智能教育与学习分析领域中, 如何从时序交互数据中刻画学习过程, 并形成具备认知解释性的诊断结果, 是教育数据挖掘中的关键问题。在线学习平台与 ITS 的普及使系统能够采集包含答题行为、练习顺序与时间间隔等结构化事件数据, 这为基于行为序列的知识建模提供了数据基础。然而, 从此类时序结构中捕获知识状态随时间演化的规律, 并同时实现预测可靠性与解释能力, 仍然具有挑战性。

现有知识追踪(knowledge tracing, 简称 KT)途径可分为隐马尔可夫建模 (hidden Markov model,简称 HMM)[7] 与基于深度神经网络的序列学习范式。深度知识追踪 (deep knowledge tracing,简称 DKT) [8] 利用循环神经网络实现答题序列表示, 在预测准确性方面表现突出, 但模型内部结构不透明, 难以提供参数层面的解释。这类模型通常作为“黑箱”系统, 缺乏对学生能力和题目特征的明确解释, 难以为教育实践提供可解释的认知诊断结果[9-11]。相对而言, 认知诊断模型 (cognitive diagnosis model, 简称 CDM)[12-19] 与多维项目反应理论(multidimensional item response theory, 简称 MIRT)[18] 通过显式参数化学生能力向量与题目属性矩阵, 能生成可解释评估结果, 然而通常忽略学习过程中的时间依赖性与知识演化特征。

为解决上述问题, 构建了一种基于多维项目反应理论的可解释动态学习评估(dynamic knowledge-aware multidimensional item response theory, 简称 DK-MIRT)模型。该模型融合了长短期记忆网络 (long short-term memory, 简称 LSTM)[19] 的时间序列建模能力与 MIRT 的认知测评结构, 实现对知识掌握水平的动态估计与可解释预测。在 DK-MIRT 中, LSTM 编码学习序列以捕捉知识状态演化趋势; MIRT 结构用于参数化题目难度与区分度, 从而在神经网络结构下实现认知测评的可解释性。与传统 CDM 相比, DK-MIRT 能捕捉时间依赖; 与纯 LSTM 模型相比, 可将预测结果显式映射至知识点维度。DK-MIRT 在预测性能与可解释性之间实现一致性, 为自适应教育系统提供知识点级的可操作评估依据。

## 1 相关研究

学习评估旨在通过对学生交互数据进行建模, 刻画知识掌握水平与学习能力的动态变化。现有方法大体可分为两类: 一类是基于心理测量与认知诊断的静态建模方法; 另一类是基于序列建模与深度学习的动态预测方法。近年来, 融合式可解释混合模型通过结合认知测量结构与神经网络特征抽取机制, 在预测性能与解释能力之间取得新平衡, 为相关研究提供了理论依据与方法学基础。

### 1.1 基于认知诊断与项目反应理论的模型

项目反应理论 (item response theory, 简称 IRT)[20] 及其扩展模型, 如 MIRT, 通过参数化学生能力与题目属性之间的交互关系, 为教育测评提供了严谨的理论基础。典型的 IRT 模型假设学生的作答概率取决于能力参数、题目难度和区分度, 从而在解释性方面具有显著优势。与此同时, 认知诊断模型如 DINA[21]、DINO[22]、 NCDM[23] 等, 通过显式利用 Q-matrix[24] 连接题目与知识点, 实现了更细粒度的知识掌握评估。然而, 这类模型通常依赖人工构建的知识结构, 难以处理大规模数据与复杂的时间依赖性, 因此在动态学习情境下性能受限。

### 1.2 基于深度学习的动态知识追踪模型

随着深度学习的发展, 循环神经网络 (RNN)[25] 及其变体, 如 LSTM、门控循环单元 (GRU)[26], 被应用于知识追踪任务, 以建模答题序列的时序模式。DKT 首次利用 LSTM 表征学生行为序列并取得显著性能增益。后续方法如动态键值记忆网络(DKVMN)[27] 与注意力机制[28-29] 进一步增强了模型在长序列依赖与知识点关联建模方面的能力。近年来, 研究趋向于在保持预测性能的同时提升可解释性与过程感知能力[30]。例如, 孔超等[31] 提出一种融合学习者遗忘行为的多特征深度知识追踪模型, 通过引入遗忘机制增强了模型对真实学习过程的刻画能力, 进一步验证了时序特征

与认知行为建模在知识追踪任务中的重要性。另外, 基于 Transformer [32]、SAINT[33] 系列及其扩展模型 SAINT+[34], 以及分层/会话感知结构 (如 HiTSKT)[35] 在大规模在线学习数据上表现出良好的预测性能, 逐渐成为知识追踪任务的重要研究方向。同时, 也出现了一类基于图神经网络增强知识结构建模的研究, 如 DyGKT[36]、GKT[37] 和 HGKT[38] 等。这些方法通过引入图结构信息, 在一定程度上提升了模型对知识点关联关系的刻画能力。尽管上述深度模型在预测性能方面取得了一定进展, 但大多仍属于黑箱结构, 缺乏对学生能力变化与题目属性的显式解释机制, 因此难以支持可追溯的教育诊断与细粒度反馈。

### 1.3　融合可解释建模与序列建模的最新研究

为同时兼顾预测性能与可解释性, 近年来出现了一类融合心理测量理论与深度学习的混合模型。如 NeuralCDM [39]、Deep-IRT[40]、ReliCD[41] 和 TC-MIRT[42] 等模型将 IRT 或 CDM 的参数化机制嵌入神经网络结构中[43], 实现端到端的可解释建模。类似地, 在其他时序感知与记忆引导任务中, 引入显式时空记忆结构也被证明能够有效提升模型对动态行为模式的建模能力[44]。然而, 多数现有方法仍难以充分刻画学生知识状态随学习时间演化的动态规律。

在此背景下, 提出的 DK-MIRT 模型将多维项目反应理论的参数化机制与 LSTM 的序列数据处理能力结合, 构建统一框架以同时估计学生能力演化特征。该结构支持对学习轨迹的可解释分析, 并为个性化学习干预提供可操作的认知诊断依据。

## 2　DK-MIRT 模型的构建与实现

DK-MIRT 模型可在时序学习场景下同时实现预测性能与认知可解释性的统一。

### 2.1　任务描述

设系统中有$N$个学生、$M$道题目以及$K$个知识点, 分别记作$U=\{u_1,u_2,\cdots,u_N\}$, $E=\{e_1,e_2,\cdots,e_M\}$, $P=\{p_1,p_2,\cdots,p_K\}$。学生在学习过程中会与多个题目交互, 形成作答日志$X=\{x_1,x_2,\cdots,x_t\}$, 其中$x_t=\{(u_t,e_t,r_t)\}$, $r_t\in\{0,1\}$表示学生$u_t$在题目$e_t$上的作答结果。知识点关联由 Q-matrix $\boldsymbol{Q}\in\{0,1\}^{M\times K}$, 若 $Q_{m,k}=1$, 则题目$e_m$关联知识点$p_k$。研究目标是利用作答日志$X$和 Q-matrix 建模学生在多个时间步知识熟练度的动态演化, 联合学习序列交互驱动的潜在知识状态、题目参数(难度与区分度)以及学生−题目交互机制, 从而在预测未来作答表现的同时提供知识掌握水平的可解释诊断。

### 2.2　方法

DK-MIRT 在多维项目反应理论(MIRT)框架下, 引入长短期记忆网络(LSTM)对学生的作答行为序列进行动态建模, 以刻画学生的知识状态在时间维度上的演化。如图 1 所示, 模型结构包含 3 层功能模块: 知识状态序列表征层、认知建模层、预测与诊断层。

#### 2.2.1　知识状态表征层

学生的作答过程不仅受当前题目属性的影响, 也受历史交互轨迹的影响。在序列表征层中, 为表征学生作答序列中的时序依赖关系, 将交互序列$X_{t=1}^{T}$输入至嵌入层(Embedding Layer)。首先对于时刻$t$的题目$e_t$, 通过嵌入映射得到向量:

$$\boldsymbol{E}_t=\text{Embedding}(e_t),$$

同时, 通过知识点嵌入矩阵$\boldsymbol{W}_Q\in\mathbf{R}^{k\times d}$得到知识点表示$\boldsymbol{Q}_t=\boldsymbol{q}_t\boldsymbol{W}_Q$,其中$\boldsymbol{q}_t\in\{0,1\}^K$表示题目$e_t$在 Q-matrix 中的行向量, 用于指示该题目所涉及的知识点。为捕获学习过程中知识状态的时间依赖性, 将题目嵌入与知识点嵌入进行向量拼接后输入 LSTM 网络, 得到输出隐藏状态:

$$\boldsymbol{H}_t=\text{LSTM}([\boldsymbol{E}_t\oplus\boldsymbol{Q}_t],\boldsymbol{H}_{t-1}),$$

其中: $\oplus$表示向量拼接操作; $\boldsymbol{H}_t\in\mathbf{R}^d$表示学生在时刻$t$的作答历史所编码的潜在特征; $d$为潜在维度。$\boldsymbol{H}_t$通过全连接层映射至知识点维度, 得到学生在各知识点上的潜在能力向量:

$$\boldsymbol{\theta}_t=\text{ReLU}(\boldsymbol{W}_1\boldsymbol{H}_t+\boldsymbol{b}_1),$$

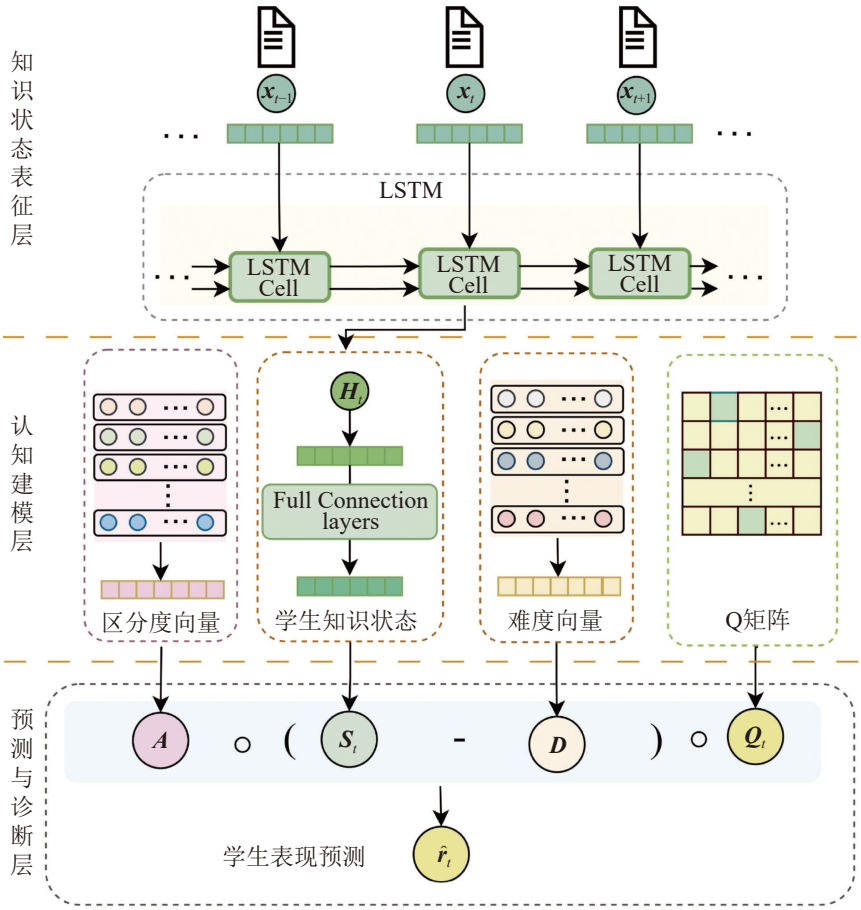
其中: $\boldsymbol{W}_1\in\mathbf{R}^{K\times H}$、$\boldsymbol{b}_1\in\mathbf{R}^K$为可学习参数; ReLU(·) 表示修正线性单元激活函数。进一步通过 Sigmoid 函数进行归一化, 得到学生在各知识点上的掌握水平$\boldsymbol{S}_t=\sigma(\boldsymbol{\theta}_t)$, 其中: $\sigma(\cdot)$ 为 Sigmoid 函数; $\boldsymbol{S}_t\in[0,1]^K$为学生在时刻$t$的知识点的掌握水平。上述过程实现从序列交互信息到时变潜在能力空间的映射, 最终生成学生能力嵌入向量$\boldsymbol{S}_t$。

#### 2.2.2　认知建模层

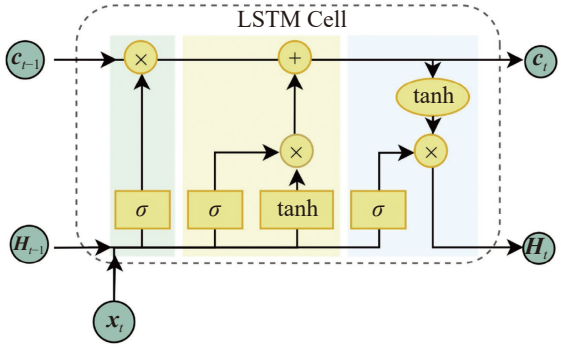
在教育测量理论中, 不同题目在区分学生能力与检验知识掌握方面具有显著差异。为刻画题目特征并保持模型的可解释性, DK-MIRT 在认知建模层中引入了题目难度与区分度参数的向量化表示。对于题目集合$\boldsymbol{E}$, 其难度向量参数通过嵌入函数映射得到: $\boldsymbol{D}=\sigma(\text{Embedding}_D(\boldsymbol{E}))$,其中, $\boldsymbol{D}\in\mathbf{R}^{M\times k}$为$M$道题目在$K$个知识维度上的难度参数。同理, 题目区分度参数向量定义为

$$\boldsymbol{A}=\alpha\times\sigma(\text{Embedding}_a(\boldsymbol{E})),$$

其中: $\boldsymbol{A}\in\mathbf{R}^{M\times K}$ 表示$M$道题目对不同能力水平学生的




(a) 动态知识感知多维项目反应理论模型




(b) 长短期记忆网络单元

**图 1　动态知识感知多维项目反应理论框架**

敏感程度；$\alpha$为缩放系数(通常设为 10)，用于增强模型对能力差异的刻画能力。

通过上述参数化方式，题目特征被映射到与学生知识状态一致的多维空间，使得后续学生能力向量$\boldsymbol{S}_t$与题目参数能够在统一的知识维度上进行交互建模。该设计不仅继承了多维项目反应理论中参数可解释性的优势，同时保留认知测量维度的结构透明性，并使答题表现具备可解释的测量学意义。

### 2.2.3　预测与诊断层

在预测阶段，模型引入多维项目反应理论(MIRT)的参数化思想，通过学生能力向量、题目难度与区分度的交互计算，生成答题正确概率。

对于第$t$时刻的题目$e_i$，其区分度向量记为$a_i \in \boldsymbol{A}$，难度偏置记为$d_i \in \boldsymbol{D}$，则正确回答$e_i$的概率可展开为

$$P_t = \sum_{k=1}^{K} a_i(\boldsymbol{S}_t - d_i)Q_i(k)。$$

由于$Q_i(k) \in \{0,1\}$，仅当题目涉及知识点时，该维度才参与计算，因此模型仅在 Q-matrix 指示的相关知识维度上计算能力与题目属性的差异与加权作用。于是，学生能力与题目参数的多维交互定义为：基于学生知识状态$\boldsymbol{S}_t$，题目参数$\boldsymbol{D}$、$\boldsymbol{A}$结合 Q-matrix 的稀

疏选择机制, 经过 sigmoid 激活得到答题正确概率, 其预测概率定义为

$$\hat{r}_t = \sigma(\boldsymbol{A} \circ (\boldsymbol{S}_t - \boldsymbol{D}) \circ \boldsymbol{Q}_t),$$

其中, ∘表示哈达玛积。该结构实现了学生能力维度与题目特征维度之间的非线性交互映射, 从而得到学生在时间步$t$时对当前题目的回答正确概率。

模型训练目标为最小化预测概率$\hat{r}_t$与真实作答结果$r_t$之间的差异。采用二元交叉熵损失:

$$\mathcal{L} = -(r_t \log(\hat{r}_t) + (1 - r_t)\log(1 - \hat{r}_t))。$$

通过最优化上述损失函数, 学生能力向量与题目参数在联合更新过程中实现自适应收敛, 使预测偏差逐步降低。

总之, DK-MIRT 同时具备时序建模与认知参数化能力: LSTM 捕获学生知识状态的动态演化, MIRT 参数保持了题目特性的可解释建模结构。与传统深层 KT 模型相比, 该模型不仅提升了预测精度, 同时明确刻画了学生能力变化与题目特征之间的对应映射关系, 为学习行为分析与智能教育评估提供了可量化与可溯源的认知层面描述。

## 2.3　实现细节

### 2.3.1　数据描述

为了验证 DK-MIRT 的有效性与通用性, 模型在 3 个具有代表性的知识追踪数据集 Synthetic-5、 ASSISTments2009 与 ASSISTments2012 上进行实验。如表 1 所示, 这三类数据覆盖从可控模拟环境到真实在线学习场景的不同层级, 为模型在多样化数据特征下的表现提供了系统性评估基础。

**表 1　数据集信息**

| 特征 | Synthetic-5 | ASSISTments2009 | ASSISTments2012 |
|---|---|---|---|
| 学生数量 $N$ | 4 000 | 4 163 | 29 018 |
| 题目数量 $M$ | 50 | 17 746 | 53 091 |
| 知识点数量 | 5 | 123 | 265 |
| 交互记录 | 200 000 | 324 572 | 29 748 |

1) Synthetic-5

该数据集是知识追踪研究中常用的模拟实验数据, 旨在提供一个可控、无噪声的实验环境, 用于验证模型的建模能力。学生的知识掌握状态在作答序列中会动态演化, 其状态转移由预设学习率与遗忘率控制。题目正确率由能力状态与 [“粗心”/“猜测”] 参数共同生成; 最终学生每步的作答结果取值均为{0,1}。该数据集包含 4 000 名学生、50 道题目与 5 个知识点, 总交互数为 200 000。

2) ASSISTments2009

该数据集来自 ASSISTments 在线辅导平台真实行为日志, 是知识追踪与认知诊断任务中的经典 benchmark, 包含 4 163 名学生、17 746 道题目、123 个知识点, 总交互数超过 300 000。

每条记录提供学生 ID、题目 ID、知识点 ID 与答题结果。由于来源真实, 其包含显著噪声、能力分布差异与真实的学习行为复杂性, 是评估模型泛化能力、序列建模能力及稀疏鲁棒性的重要测试资源。按照领域通用预处理协议: 1) 删除序列过短的学生; 2) 保证每题至少映射一个专家标注知识点 Q-matrix; 3) 将答题结果二值化为正确 (1) 或错误 (0)。

3) ASSISTments2012

该数据集来源于 ASSISTments 平台, 但规模更大, 包含 29 018 名学生、53 091 道题目与 265 个知识点, 总交互记录为 29 748。相比 2009 版本, 该数据在学生与题目规模上显著提升, 知识点划分更细, 但交互次数更稀疏, Q-matrix 结构更复杂, 更适合检验模型在大规模、多知识维场景下的建模能力。

综上, 三类数据在学生规模、题目数量、知识维度覆盖以及稀疏性方面差异显著, 为模型能力验证提供了多维实验环境。实验首先以 Synthetic-5 作为初始验证与调优, 再分别在 ASSISTments2009 与 ASSISTments2012 上开展进一步评估, 以联合模拟数据与真实数据, 验证 DKMIRT 在保持可解释性的同时, 是否仍具备在真实场景下的适应性与稳健性。

### 2.3.2　实验设置

对模型结构与训练过程中的关键超参数进行统一设定。

在序列建模部分, 学生交互序列由单层单向 LSTM 网络进行建模, 其隐藏层维度设置分别为 50 (Synthetic-5) 和 200(ASSISTments2009/ASSISTments-2012), 以在模型表达能力与计算效率之间取得平衡。LSTM 输出通过一层全连接映射至知识点维度, 用于生成学生在各知识点上的潜在能力表示。为防止过拟合, 在 LSTM 输出层后引入 dropout 正则化, 其概率设为 0.2。

在认知参数建模部分, 题目难度参数与区分度参数均采用可学习嵌入向量进行表示。其中, 区分度参数在经过 sigmoid 激活后额外乘以常数因子 10, 以增强不同能力水平学生之间的区分性, 并提升训练初期的梯度响应稳定性。题目难度参数与学生能力嵌入均采用 Xavier 初始化, 以保持参数分布的数值稳定。

模型训练采用 Adam 优化器, 基础学习率设为 0.001, 权重衰减设为$1 \times 10^{-5}$, 以对模型参数施加$\ell_2$正

则约束。训练过程中未采用学习率衰减或预热策略。批次大小设为 64，最大训练轮数（epoch）为 200。为避免梯度爆炸，在训练过程中对梯度范数进行裁剪，其最大阈值设为 5.0。

在数据划分方面，训练集、验证集与测试集均按学生维度进行划分，以避免同一学生的作答记录同时出现在不同数据子集中，从而确保评估结果的公平性。模型性能在验证集上进行监控，并引入 Early Stopping 训练策略：若验证集 AUC 在连续 5 个 epoch 内无提升，则提前终止训练，以进一步抑制过拟合风险。

实验基于 PyTorch 框架搭建，在相同随机种子条件下重复运行 5 次，并输出平均结果，以保证实验结论的稳定性与可复现性。

### 2.4 实验结果与分析

为系统性验证 DK-MIRT 模型的有效性，在 3 个数据集上开展实验，并与多种具有代表性的基线方法进行比较。对比方法包含基于隐马尔可夫模型对知识掌握演化建模 BKT[45]、引入循环神经网络以建模学生作答行为的时间依赖 DKT、通过记忆增强结构追踪知识状态 DKVMN、在认知诊断框架下结合神经网络优化预测性能的 NeuralCDM。

#### 2.4.1 性能对比

实验对 BKT、DKT、DKVMN、NeuralCDM 与 DK-MIRT 在 3 个数据集上的 AUC 指标进行对比。AUC 衡量模型预测下一步答题表现能力，其值越高，表明预测区分能力越强。5 种模型的 AUC 如图 2 所示，DK-MIRT 在所有数据集上均获得较优性能，并呈现稳定的泛化能力。DK-MIRT 将长短期记忆网络（LSTM）与多维项目反应理论（MIRT）进行结构级融合。LSTM 捕获时间序列上的知识获取与遗忘规律，MIRT 则量化题目参数与学生能力之间的多维交互结构。二者协同不仅提升了预测性能，也使模型具备参数可解释性。因此，无论是在相对规整的模拟环境（Synthetic-5），还是在噪声更高、稀疏性更强的真实环境（ASSIST2009 / ASSIST2012）中，DK-MIRT 均表现出显著的性能稳健性。

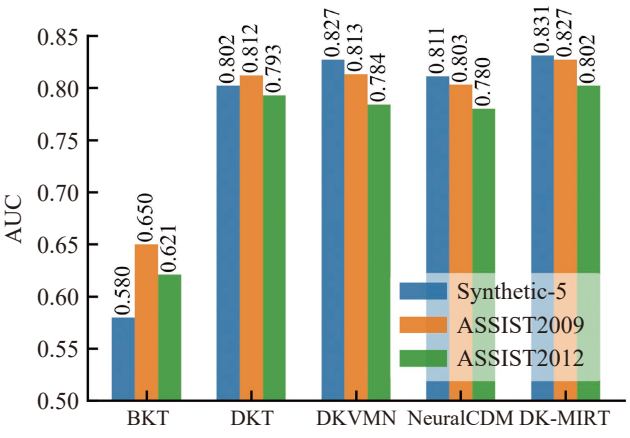



图 2 预测性能 (AUC)

在 Synthetic-5 数据集，传统 BKT 模型性能较弱，而合成数据分布通常更加规整，有利于模型学习，因此深度学习模型整体优于基于隐马尔可夫的传统模型。其中，DK-MIRT 在该数据集上的 AUC 均超过 DKVMN、NeuralCDM。

在 ASSIST2009 数据集上，深度学习方法同样显著优于 BKT。DK-MIRT 获得最佳 AUC，相较于 DKT、DKVMN 均有小幅提升，并优于 NeuralCDM。这表明，DK-MIRT 在真实教育环境中能够有效刻画知识状态的动态演化。

对于 ASSIST2012 数据集，其数据噪声更高，分布更复杂。DK-MIRT 在该数据集上的 AUC 高于 DKVMN、NeuralCDM、DKT。这说明即使在数据特性差异显著的情况下，DK-MIRT 仍然保持稳定的预测性能。

综合可知，BKT 的隐状态空间通常为二值，难以描述知识掌握的多维连续变化，且无法显式估计题目参数，使其在涉及复杂认知结构的场景中性能受限。基于序列建模的 DKT 与 DKVMN 虽能够刻画时间依赖，但缺乏心理测量层面的可解释参数，因而解释性不足。NeuralCDM 在解释性与预测性之间取得一定平衡，但整体性能仍不及 DK-MIRT。相比之下，DK-MIRT 通过 LSTM 层刻画学生知识状态的时序演化，同时在模型结构中显式引入 MIRT 的难度与区分度参数，使题目属性与能力估计之间建立可解释交互结构。此外，区分度参数放大因子的引入显著增强了不同能力群体之间的区分能力，并有助于提升训练收敛速度。

#### 2.4.2 消融实验

为了进一步分析 DK-MIRT 模型中各结构组件的贡献，设计了一组消融实验，通过逐步移除关键模块来评估其影响。这些模块包括时间序列建模层（DMLSTM-）、可解释的 MIRT 参数化结构（DMIRT-）、题目区分度参数（DMdis-）、题目难度参数（DMdif-）以及 Q-matrix 掩码机制（DMKC-）。实验结果如图 3 所示。

首先，在移除 LSTM 模块后，模型的 ACC 和 AUC 均出现明显下降，表明时间依赖对于捕捉学生知识状态随时间演化的动态过程至关重要。缺少显式的序列建模后，模型无法有效表达学习轨迹，从而导

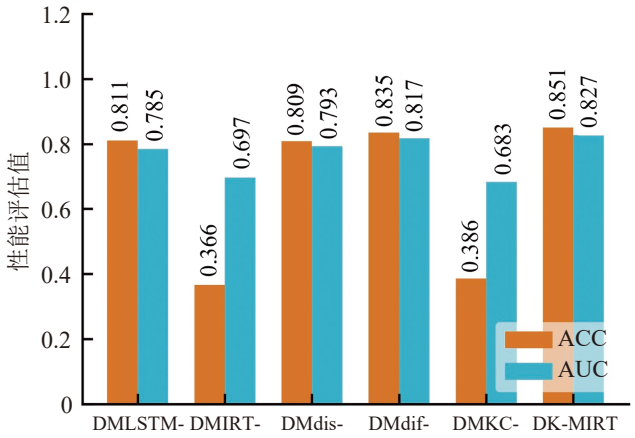



**图 3　消融实验性能**

致预测性能明显退化。进一步地, 当移除 MIRT 参数化结构后(DMIRT-), 模型性能下降最严重。此时 ACC、AUC 分别降至 0.366、0.697, 表现出所有消融设置中最显著的崩溃趋势。这表明项目难度与区分度等心理测量参数在建模学生–题目交互行为中发挥着核心作用, 也进一步强调了在深度模型中融合心理测量结构的重要性。在更细粒度的参数层面, 题目区分度与难度参数的移除均会造成一定程度的性能下降, 其中移除难度参数带来的负面影响略大于移除区分度, 说明难度在将潜在能力表示与题目特征对齐过程中起着更为基础的作用。知识点掩码机制(DMKC-)同样对模型性能具有显著影响。当移除该结构后, ACC 和 AUC 出现明显下降, 其劣化程度接近 DMIRT-的情况。这表明在 MIRT 计算式中仅激活与题目相关的知识维度, 对于抑制噪声、保持能力向量与题目特征的一致性具有关键作用。

综上所述, DK-MIRT 的性能优势并非来源于单一模块, 而是源于动态能力建模、可解释项目参数化结构与知识点结构化表示之间的协同作用。其中, MIRT 参数模块与知识点掩码机制对性能的影响最显著, 这验证了模型整体结构设计的有效性与必要性。

### 2.4.3　可解释性分析

与传统 DKT 对所有知识概念进行统一更新不同, DK-MIRT 能够根据学生的实际作答, 仅对关联概念进行有针对性的动态增益或衰减, 从而更贴近真实的学习认知变化。为验证这一点, 从测试集中随机抽取一名学生, 并在不同时间步 ($t$) 展示其知识状态变化的可视化结果, 如图 4 所示。

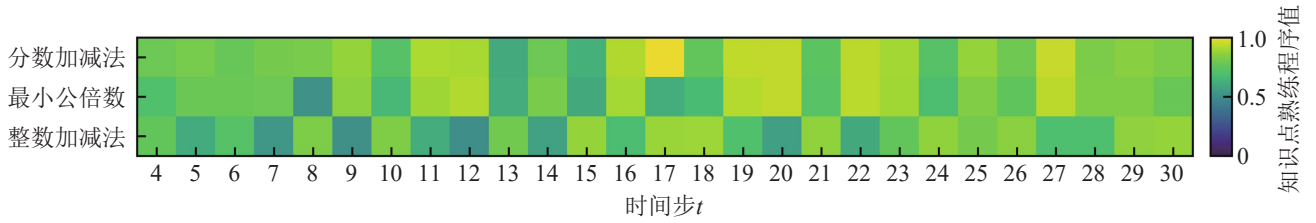



**图 4　概念子集预测知识状态的可视化**

在 $t=5$ 时, 该学生正确回答了关于“最小公倍数”的题目, DK-MIRT 既上调了该概念的掌握程度, 也同步增强了与之存在依赖关系的“分数加减法”。这一趋势与数学学习路径一致: 对最小公倍数的掌握有助于后续的分母统一处理, 从而提高分数加减的成功率。

在 $t=18$ 时, 该学生在“最小公倍数”上回答错误, 模型降低了该概念的能力估计, 同时“分数加减法”的掌握也随之下降; 而“整数加减法”并未受到明显影响, 表明 DK-MIRT 避免了对与当前概念无依赖关系的知识产生不必要扰动。

在 $t=27$ 时, 该学生再次正确回答“最小公倍数”, 模型提升该知识点的状态, 并保持“分数加减法”稳定, 无额外波动。

在 $t=30$ 时, 该学生正确回答了“整数加减法”, 此基础概念得到强化, 并间接提升了“分数加减法”的掌握水平。这符合层级依赖的教育逻辑: 整数运算是分数运算的底层基础技能。

综上, “整数加减法”作为算术的基础技能, 为“分数加减法”提供了运算支撑, 而“最小公倍数”又是“分数加减法”的前置知识, 形成了一条具有层次性的认知依赖链。 DK-MIRT 能沿该链条进行方向性更新, 既能保持单一概念更新的合理性, 又能反映概念间的交互与跨概念的增益效应, 从而产生符合教育学规律的知识演化轨迹。

## 3　结束语

围绕“如何在时间序列上刻画学生知识状态的动态演化并同时实现具有认知解释性的诊断”这一核心问题, 提出了一种基于多维项目反应理论的可解释动态学习评估模型 DK-MIRT。模型在统一框架中融合了 LSTM 的序列建模优势与 MIRT 的结构化测量机制, 实现了对学生潜在知识状态的连续时序建模, 并克服了传统 MIRT 只能提供静态能力估计的局限。通过这一动态能力建模机制, 学生的知识演化过程得

以被细粒度捕获，为理解学习进程提供了更精准的刻画。

在参数可解释性方面，DK-MIRT 将题目难度与区分度以可学习向量形式嵌入深度网络，同时保持其在心理测量学中的原有含义，使深度学习模型具备题目参数的透明性与可溯性。模型显式建模学生能力与题目属性之间的多维交互，使预测结果由能力差异和题目特性共同驱动，从而实现“可预测 + 可解释”的双重目标，兼顾性能提升与认知理论一致性。

在计算效率方面，DK-MIRT 在时间复杂度上与主流基于 LSTM 的知识追踪模型（如 DKT）保持同阶，其主要计算开销来自序列建模阶段的 LSTM 前向传播。相较之下，引入的 MIRT 参数化模块仅涉及与知识点维度相关的向量运算，其计算复杂度与知识点数量呈线性关系，在实际数据集中通常远小于序列长度带来的时序计算开销。因此，DK-MIRT 在保持可解释性的同时，未显著增加推理成本，具备良好的在线预测效率与可扩展性，适合在大规模在线教育与智能辅导系统中进行部署。

实验结果表明，DK-MIRT 在多个数据集上的预测性能均优于主流深度知识追踪模型（DKT、DKVMN、NeuralCDM），同时在认知解释性方面保持 MIRT 框架的解析优势。更重要的是，模型输出的多维能力向量随时间的演化轨迹、题目参数的显式表示以及学生题目交互的结构化解释，为教育实践提供了切实可用的诊断价值：能够支持学习瓶颈识别、个性化练习路径优化、难度梯度调控以及动态教学策略设计，为自适应学习系统提供可实施的智能反馈机制。

未来的研究可从多个方向进一步拓展：一是将 DK-MIRT 延伸至跨学科与多模态学习场景，使模型能够融合过程数据（如点击流）、文本或视频中的认知线索；二是将 MIRT 参数化与现代认知诊断理论深度结合，构建兼具预测性能与心理测量一致性的解释型人工智能（XAI）框架[46]，以增强模型输出的理论可验证性与教育可信性。

**参考文献：**


[1] Ma W, Adesope O O, Nesbit J C, et al. Intelligent tutoring systems and learning outcomes: a meta-analysis[J]. Journal of Educational Psychology, 2014, 106(4): 901-918.

[2] Lin Chienchang, Huang Anna Y Q, Lu Owen H T. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review[J]. Smart Learning Environments, 2023, 10(1): 41.

[3] Li Qingyao, Xia Wei, Du Kounianhua, et al. Learning structure and knowledge aware representation with large language models for concept recommendation[PP/OL]. arXiv (2024-05-21)[2025-05-28]. https://arxiv.org/abs/2405.12442.

[4] 万海斌, 孙洪民, 农丝静, 等. 基于深度学习的馆藏图书索书号识别算法[J]. 桂林电子科技大学学报, 2025, 45(3): 221-228.

[5] 彭智勇, 吴磊, 肖博. 基于深度学习的高精度相机相对位姿估计网络[J]. 桂林电子科技大学学报, 2024, 44(6): 568-578.

[6] 林乐平, 朱静, 欧阳宁. 基于协同对抗优化网络的图像压缩感知重建[J]. 桂林电子科技大学学报, 2024, 44(6): 592-598.

[7] Eddy S R. Hidden Markov models[J]. Current Opinion in Structural Biology, 1996, 6(3): 361-365.

[8] Piech C, Bassen J, Huang J, et al. Deep knowledge tracing[J]. Advances in Neural Information Processing Systems, 2015, 28: 505-513.

[9] Şahin E, Arslan N N, Özdemir D. Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning[J]. Neural Computing and Applications, 2025, 37(2): 859-965.

[10] Hulsen T. Explainable artificial intelligence (XAI): concepts and challenges in healthcare[J]. AI, 2023, 4(3): 652-666.

[11] Ibrahim R, Shafiq M O. Explainable convolutional neural networks: a taxonomy, review, and future directions[J]. ACM Computing Surveys, 2023, 55(10): 1-37.

[12] Templin J, Henson R A. Diagnostic measurement: Theory, methods, and applications[M]. New York: Guilford Press, 2010: 32-40.

[13] Templin J L, Henson R A. Measurement of psychological disorders using cognitive diagnosis Models[J]. Psychological Methods, 2006, 11(3): 287.

[14] Leighton J, Gierl M. Cognitive diagnostic assessment for education: theory and applications[M]. Cambridge: Cambridge University Press, 2007: 205-241.

[15] Wang Fei, Gao Weibo, Liu Qi, et al. A survey of models for cognitive diagnosis: new developments and future directions[PP/OL]. arXiv (2024-07-07)[2025-05-28]. https://doi.org/10.48550/arXiv.2407.05458.

[16] Liu Yingjie, Zhang Tiancheng, Wang Xuecen, et al. New development of cognitive diagnosis models[J]. Frontiers of Computer Science, 2023, 17(1): 171604.

[17] De La Torre J, Douglas J A. Higher-order latent trait models for cognitive diagnosis[J]. Psychometrika, 2004, 69(3): 333-353.

[18] Reckase M D. Multidimensional item response theory[J]. Handbook of Statistics, 2006, 26: 607-642.

[19] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[20] Embretson S E, Reise S P. Item response theory for psychologists[M]. New York, NY: Psychology Press, 2000.

[21] De La Torre J. Dina model and parameter estimation: a didactic[J]. Journal of Educational and Behavioral

Statistics, 2009, 34(1): 115-130.
[22] Zhang Hao, Li Feng, Liu Shilong, et al. Dino: DETR with improved denoising anchor boxes for end-to-end object detection[PP/OL]. arXiv (2022-07-11)[2025-05-28]. https://doi.org/10.48550/ arXiv.2203.03605.
[23] Li Jiatong, Liu Qi, Zhu Mengxiao. Generative cognitive diagnosis[PP/OL]. arXiv (2025-07-13)[2025-05-28]. https://doi.org/10.48550/arXiv.2507.09831.
[24] Barnes T. The q-matrix method: Mining student response data for knowledge[C]//American Association for Artificial Intelligence 2005 Educational Data Mining Workshop. Pittsburgh, PA: AAAI Press, 2005: 1-8.
[25] Pearlmutter. Learning state space trajectories in recurrent neural networks[C]//International 1989 Joint Conference on Neural Networks. Piscataway, NJ: IEEE Press, 1989: 365-372.
[26] Chung J, Gulcehre C, Cho K, et al. Empirical evaluation of gated recurrent neural etworks on sequence modeling[PP/OL]. arXiv (2014-12-11)[2025-05-28]. https://doi.org/10.48550/arXiv.1412.3555.
[27] Zhang Jiani, Shi Xingjian, King I, et al. Dynamic key-value memory networks for knowledge tracing[C]// Proceedings of the 26th International Conference on World Wide Web. New York, NY: ACM Press, 2017: 765-774.
[28] Pandey S, Karypis G. A self-attentive model for knowledge tracing[PP/OL]. arXiv (2019-07-16)[2025-05-28]. https://doi.org/10.48550/arXiv.1907.06837.
[29] Ghosh A, Heffernan N, Lan A S. Context-aware attentive knowledge tracing[C]//Proceedings of the 26th ACM SIGKDD International ,Conference on Knowledge Discovery & Data Mining. New York, NY: ACM Press, 2020: 2330-2339.
[30] Abdelrahman G E A. Knowledge tracing: a survey[J]. ACM Computing Surveys (CSUR), 2022, 55(6): 1-37.
[31] 孔超, 高祥云, 孙先兰, 等. 融合学习者遗忘行为的多特征深度知识追踪[J]. 桂林电子科技大学学报, 2025, 45(3): 229-236.
[32] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008
[33] Choi Y, Lee Y, Cho J, et al. Towards an appropriate query, key, and value computation for knowledge tracing[C]//Proceedings of the seventh ACM conference on learning at scale. New York, NY: ACM Press, 2020: 341-344.
[34] Shin D, Shim Y, Yu H, et al. Saint+: integrating temporal features for ednet correctness prediction[C]//LAK21: 11th International Learning Analytics and Knowledge Conference. Irvine, New York, NY: ACM Press, 2021: 490-496.
[35] Ke Fucai, Wang Weiqing, Tan Weicong, et al. HiTSKT: a hierarchical transformer model for session-aware knowledge tracing[J]. Knowledge-Based Systems, 2024, 284: 111300.
[36] Cheng Ke, Peng Linzhi, Wang Pengyang et al. DyGKT: dynamic graph learning for knowledge tracing[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York, NY: ACM Press, 2024: 409-420.
[37] Nakagawa H, Iwasawa Y, Matsuo Y. Graph-based knowledge tracing: modeling student proficiency using graph neural network[C]//IEEE/WIC/ACM International Conference on Web Intelligence. Piscataway, NJ : IEEE Press, 2019: 156-163.
[38] Tong Hanshuang, Wang Zhen, Zhou Yun, et al. Introducing problem schema with hierarchical exercise graph for knowledge tracing[C]//Proceedings of the 45th Iinternational ACM SIGIR Conference on Research and Development in Information Rretrieval. New York, NY: ACM Press, 2022: 405-415.
[39] Wang Fei, Liu Qi, Chen Enhong, et al. Neural cognitive diagnosis for intelligent education systems[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020: 6153-6161.
[40] Yeung C K. Deep-IRT: make deep learning based knowledge tracing explainable using item response theory [PP/OL]. arXiv (2019-04-26)[2025-05-28]. https://doi.org/10.48550/arXiv.1904.11738.
[41] Zhang Yunfei, Qin Chuan, Shen Dazhong, et al. Relicd: A reliable cognitive diagnosis framework with confidence awareness[C]//2023 IEEE International Conference on Data Mining (ICDM). Piscataway, NJ: IEEE Press, 2023: 858-867.
[42] Su Yu, Cheng Zeyu, Luo Pengfei, et al. Time-and-concept enhanced deep multidimensional item response theory for interpretable knowledge tracing[J]. Knowledge-Based Systems, 2021, 218: 106819.
[43] Tsutsumi E, Kinoshita R, Ueno M. Deep item response theory as a novel test theory based on deep learning[J]. Electronics, 2021, 10(9): 1020.
[44] 王健聪, 杨海燕, 刘书宁. 时空记忆引导边缘学习的视频异常检测算法[J]. 桂林电子科技大学学报, 2025, 45(1): 33-40.
[45] Corbett A T, Anderson J R. Knowledge tracing: Modeling the acquisition of procedural knowledge[J]. User Modeling and User-adapted Interaction, 1994, 4(4): 253-278.
[46] Chamola V, Hassija V, Sulthana A R, et al. A review of trustworthy and explainable artificial intelligence (XAI)[J]. IEEE Access, 2023, 11: 78994-79015.

编辑：张所滨