arXiv:2304.02205v1 [cs.AI] 5 Apr 2023

MoocRadar：用于提升大规模开放在线课程中认知学生建模的细粒度、多维度知识库

Jifan Yu DCST，清华大学，北京 100084，中国
yujf21@mails.tsinghua.edu.cn

Mengying Lu SIGS，清华大学，深圳 518055，中国
lumy22@mails.tsinghua.edu.cn

Qingyang Zhong DCST，清华大学，北京 100084，中国
zqy20@mails.tsinghua.edu.cn

Zijun Yao DCST，清华大学，北京 100084，中国
yaozj20@mails.tsinghua.edu.cn

Shangqing Tu DCST，清华大学，北京 100084，中国
tsq22@mails.tsinghua.edu.cn

Zhengshan Liao IoE，清华大学，北京 100084，中国
liaozs20@mails.tsinghua.edu.cn

Xiaoya Li IoE，清华大学，北京 100084，中国
lixiaoya21@mails.tsinghua.edu.cn

Manli Li IoE，清华大学，北京 100084，中国
m
arylee@mail.tsinghua.edu.cn

Lei Hou
BNRist，DCST，清华大学，北京 100084，中国
houlei@tsinghua.edu.cn

Hai-Tao Zheng
zheng.haitao@sz.tsinghua.edu.cn
SIGS，清华大学与鹏城实验室，中国深圳 518055

Juanzi Li*BNRist，DCST，清华大学，中国北京 100084
lijuanzi@tsinghua.edu.cn

Jie Tang BNRist，DCST，清华大学，中国北京 100084
jietang@tsinghua.edu.cn

摘要

学生建模是通过学生与课程内容的互动来推断其学习特征的任务，是智能教育中的一个基本问题。尽管近期知识追踪和认知诊断领域的研究提出了若干有望提升现有模型可用性和有效性的方向，但由于缺乏完整的练习情境、细粒度概念和认知标签，现有公开数据集仍不足以满足这些潜在解决方案的需求。本文提出 MoocRadar，这是一个细粒度、多方面的知识库，包含 2，513 道练习题、5，600 个知识概念以及超过 1200 万条行为记录。具体而言，我们提出了一个框架，以确保对细粒度概念和认知标签进行高质量、全面的标注。统计和实验结果表明，我们的数据集为未来改进现有方法奠定了基础。此外，为了方便研究人员使用，我们发布了一套用于数据查询的工具，模型适配，甚至扩展我们的知识库；目前可通过 https://github.com/THU-KEG/MOOC-Radar 获取。

CCS 概念

• **应用计算 → 教育**；• **信息系统 →信息集成**；**专业信息检索**。

关键词

学生建模、数据集、概念挖掘、知识追踪

ACM 参考格式：

Jifan Yu、Mengying Lu、Qingyang Zhong、Zijun Yao、Shangqing Tu、Zheng- shan Liao、Xiaoya Li、Manli Li、Lei Hou、Hai-Tao Zheng、Juanzi Li 和 Jie Tang。2023。MoocRadar：用于改进大规模开放在线课程中认知学生建模的细粒度、多方面知识库。载于第 *46* 届 *ACM* 国际信息检索研究与发展会议（*SIGIR '23*），*2023*年*7*月*23*日至*27*日，混合会议，台北，台湾。ACM，纽约州纽约，美国，11 页。https://doi.org/XXXXXXX.XXXXXXX

*通讯作者。

允许免费为个人或课堂使用而制作本作品全部或部分内容的数字副本或纸质副本，但不得出于营利或商业利益制作或分发副本，且副本首页必须附有本声明及完整引文。对于本作品中由 ACM 以外机构拥有版权的部分，必须尊重其版权。注明出处后可以摘录。以其他方式复制、再版、发布到服务器或向列表重新分发，均须事先获得具体许可和/或支付费用。请向 permissions@acm.org 申请许可。*SIGIR '23，2023*年*7*月*23*日至*27*日，混合会议，台北，台湾© 2023 Association for Computing Machinery。ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 引言

自智能辅导系统 [33],首次提出以来，构建能够推断每位学生特征的学生模型 [39] 一直是支持人工智能驱动教育服务的基础环节，例如学习计划制定 [25, 43], 学习资源推荐 [4, 40] 以及教学助理聊天机器人 [51]。与对学生的显性特征（例如，年龄，年级） [13], 进行建模相比，对隐性认知特征进行建模（例如估计学生对特定知识的掌握程度） [10], 更具前沿性，也更具挑战性，

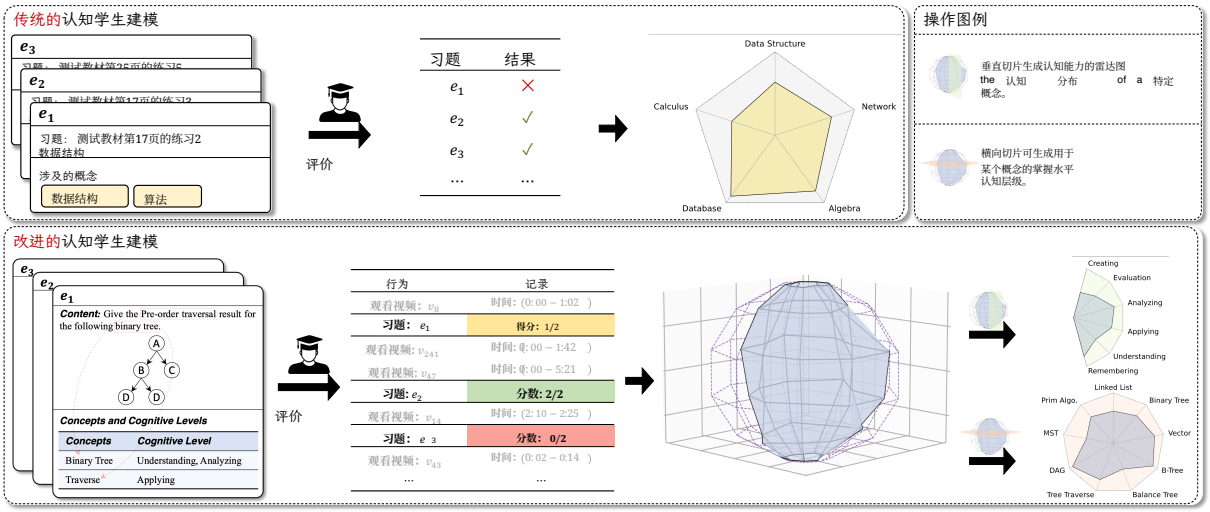


图 1：认知学生建模主题示意图。基于知识追踪或认知诊断方法，MoocRadar 为将此任务扩展到信息更丰富的场景提供了基础，其中包含习题内容、细粒度概念、认知标签及其他学习行为。

因此，知识追踪 [31, 35] 或认知诊断 [41]等相关主题吸引了大量研究关注。

图 1 展示了完成此类任务所需元素的示意图：给定学生在一系列练习题上的行为记录（例如，答对或答错），目标是估计他们在相应概念（例如，数据库）上的实际知识状态。为取得切实可用的效果，除早期仅采用神经方法在序列架构中对上述元素进行建模的尝试外，[31, 48], 研究人员还投入大量精力，寻找有益的辅助信息 [34],挖掘恰当的知识关联 [1, 18]，甚至激发学生的认知过程 [34]，以提升所提方法的准确率和可解释性 [41]。

然而，由于数据资源获取受限且专家标注成本高昂，现有主流数据集仍存在若干固有缺陷 [8, 12]，可能阻碍这些前景广阔的探索，可归纳为以下三大问题：

• **练习行为的上下文不完整**：尽管学生练习行为的多种相关上下文例如教学结构、练习的层级 [26], 练习内容和类型 [20], 以及相关学习行为 [23],已被证明有助于建模，但当前数据集 [8] 通常只考虑其中少数几种，无法支持更多特征组合来利用这些特征。

• **知识概念粒度较粗**：另一个常见问题是，大多数现有数据集仅标注了粗粒度概念例如，ASSISTment 中的 26,688 道练习题仅涉及 123 个概念2009 [12],，而 Junyi 中的 722 道题目仅标注了 41 个概念 [32]。尽管有些数据资源提供细粒度概念 [46],，但它们大多缺少专家标注 [19]，且忽略了一道练习可能对应多个概念的情形。

• **认知过程标注受到忽视**：尽管人们日益关注对学生认知过程的建模，例如学习与遗忘规律 [34], 问题难度等级 [41], 从认知教育学的视角构建监督标签，仍是教育数据集构建中一个研究不足的课题 [30]。如何恰当地运用这类指导来建模认知过程，仍是一个具有挑战性的课题。

本文提出了 MoocRadar，一个用于认知学生建模的大型知识库。MoocRadar 基于开放数据源 [46]和丰富的专家标注构建，收录了 2，513 道练习题及其多方面背景信息、5，600 个细粒度概念、14， 224 名在线学习学生，以及 12，715，126 条行为记录。具体而言，所有学习行为均由教育专家依据广受认可的布鲁姆认知分类法 [17], 进行标注，以构建认知感知型监督，为相关主题的未来发展提供全新视角。为支撑整个构建过程，我们设计了一个框架，包含异构数据集成、自动概念识别，以及一系列便于标注和质量控制的机制。此外，我们还开发并发布了配套工具包，支持多种常见查询和表征训练，可帮助研究人员高效利用这些资源，改进其模型与系统。

我们进行了一系列统计和实验调查，以呈现 MoocRadar 的特征，以及充足的上下文、细粒度概念和认知标签是否确实有利于当前方法。此外，基于我们发布的知识库，可以将当前的知识追踪 或 认知诊断 任务扩展到如图 1 所示的更深入、信息更丰富的设置，从而更详细地展示某位学生的细粒度和多方面认知状态。我们希望我们的努力能够激发研究兴趣，以利用教育学和认知理论，构建更智能的下一代教育应用。

贡献。我们的贡献可概括为 d 为：

（1）一个开放获取、细粒度、多维度且知识完备的知识库，经过高质量标注，可用于支持和推动认知学生建模研究。

（2）一个由一系列标准和工具包组成的框架，便于在教育任务中使用和扩展。

（3）一项研究及若干重要见解，探讨如何运用教育理论来改进学生建模任务，并实现更适用于人工智能驱动教育的服务。

2 背景

2.1 认知学生建模

学生建模（或学生模型）是多种教育技术理论中不可或缺的基础概念 [39],，无论是早期的智能辅导系统 [3]，还是近期的自适应学习 [27]。这项基础任务实际上涵盖对学生各种学习特征的建模，包括显性特征（即性别、年龄等基本信息）和隐性画像（例如知识认知状态和学习风格） [9],，这是开展教育推荐 [4]和教学游戏 [1]等下游应用的前置阶段。

随着大规模开放在线课程（MOOCs）中丰富学习行为数据的不断积累 [45],，推断学生在特定知识上的认知状态逐渐成为一个颇具前景的研究方向 [31]。尽管研究起源各不相同，知识追踪 [20] 和认知诊断 [41]领域的研究者目前都致力于通过挖掘学生的习题记录来预测其未来表现，而这一方法很容易适用于认知学生建模任务。

在序列到序列学习 [1, 42] 和对比式自训练 [18], 等技术进步的推动下，认知学生建模的改进呈现出若干发展方向：（1）丰富习题信息：另一种直观的方向是利用学习资源中的丰富辅助信息来进行建模，例如习题的文本内容和结构 [20, 26],，以及相关学生的其他行为记录 [35]；（2）优化知识概念建模：挖掘习题背后概念的深层含义和内在关系是主要方向之一，例如，Chen 等人 [7] 利用概念之间的先修关系，而 Tong 等人 [37] 设计了一个用于概念建模的独立模块；（3）激发学习的认知过程：另一个新兴趋势是引入教学指导，激发学生的认知学习过程，例如学习与遗忘调节 [6],、习题难度等级 [41],，以解释相应学生表现背后的原因。这些方向也可以结合起来 [34]，并应用于自适应学习 [50],；自适应学习持续引发关于如何利用更多特征来增强现有方法的讨论 [2]。

2.2 开放教育数据集

与学生建模方法的蓬勃发展相比，开放教育数据集的更新和推出仍相对有限 [2, 21, 49]。总体而言，数据集来源大致可分为两类：由人工智能驱动的公开教育挑战赛和开放获取的教育资源库。

由人工智能驱动的公开教育挑战赛是研究人员获取资源的一种自然且易于使用的方式，通常由特定的在线学习平台或团体主办，例如 KDD 杯（代数 2005-2006 、 Bridge 2006-2007 ） [36], ASSISTment [12, 28] 和 Junyi Academy [32]。尽管这类资源方便且广受欢迎，但为了便于任务形式化，它们在数据覆盖范围上有所取舍。对于各类相关行为数据和结构化数据，其中大多数只提供基本标识符（例如，ASSISTment 中的概念仅以概念 ID 表示），而不提供实际且完整的内容。

开放获取的教育资源库更倾向于从多个角度提供有用信息，它们主要通过整合来自异构来源的经处理数据而建立，例如 STATIC2011 [16] EdNet [8], 和 MOOC-CubeX [46]。由于规模庞大，它们通常会缩小标注范围以控制成本，导致缺乏细粒度的知识概念。尽管此前已有一些工作 [46] 尝试通过弱监督自动从视频文本中提取概念，但习题的高质量概念标注仍未受到重视，而且成本相当高。

与此同时，由于缺少具有认知意识的标签，遵循学习规律 [6]的现有方法只能依据其他行为记录构建伪标签。因此，我们认为有必要且迫切需要开展充分的专家标注，并提出 MoocRadar，以支持相关研究人员。

3 MOOC RADAR

本节首先提出构建 MoocRadar 的原则和总体框架，随后详细介绍如何有效开展大规模、高质量的概念与认知标注工作流程。最后，我们介绍如何访问我们的资源库及其配套工具包。

原则。 为确保所提出数据资源的可用性，我们确定在构建该知识库时应遵循两项主要原则。(1) 以习题为中心的数据组织：鉴于做题是当前学生建模方法中最关键的行为 [31],，应优先支持访问学生的做题序列，再以此为索引检索多样且相关的信息；(2) 全面的专家标注：由于知识概念和认知层级属于高度专业化的内容，我们必须安排足够数量的专家进行人工标注，尽管此前的研究因成本高昂而通常以自动化手段取而代之 [46]。

框架。 如图2所示，我们首先大规模收集并处理在线学习数据，然后在适当的智能辅助下，组织一系列专家标注，为习题标注细粒度概念和认知层级。经过质量检查后，将保留的数据进一步存储并打包，以便作为信息丰富的资源供人们轻松访问。该三模块框架可分解如下：

(1) 异构数据集成：鉴于半结构化和弱监督在线学习数据，此阶段主要选择高质量的练习记录并收集异构相关上下文，从而构建整个知识库的数据基础。

(2) 细粒度概念标注：在此阶段，我们对细粒度概念进行标注。具体而言，我们提出一种

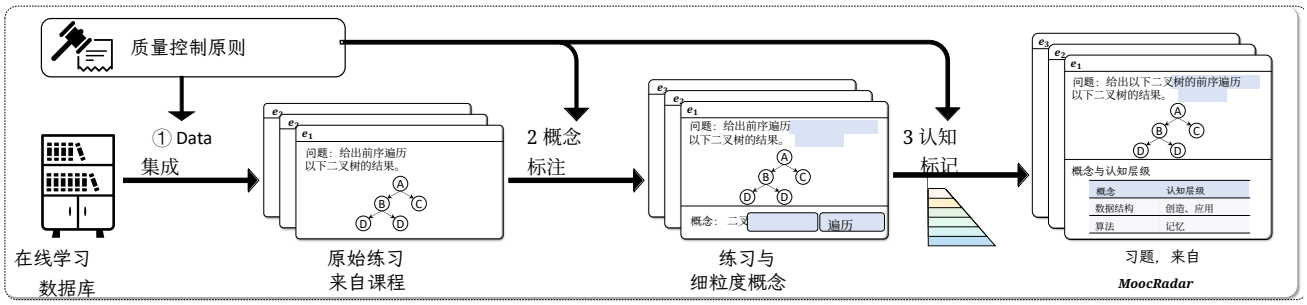


图2：从习题角度展示的 MoocRadar 整体构建框架。随着习题不断细化，相应地收集异构信息。所有处理操作均受到严格的质量控制。

基于预训练模型的概念获取方法，以减轻工作量并协助标注人员。

(3) 认知练习标注：遵循布鲁姆认知分类法 [17],，我们首先进行文本分析，为认知感知标注做准备，然后完成大规模专家标注。最后，所有标注结果均与概念标签对齐，并在严格的质量控制下打包。

3.1 异构数据集成

我们的数据来自 MOOCCubeX [46],，这是一个大规模、以知识为中心的在线教育资源库，包含 358，265 道习题、637，572 个自动抽取的概念（含噪声），以及来自 4，216 门课程的 2.96 亿条学生行为数据。作为依托学堂在线[1]的开放数据资源，学堂在线是中国最大的 MOOC 平台之一；该资源还提供了丰富且维护良好的学习材料和关系丰富的行为记录。因此，我们选择它作为原始数据资源，并启动后续的数据集成流程。

3.1.1 课程选择。通过初步的数据观察，我们发现，要确保练习具有信息量、特征足够丰富且相关学习行为充分，关键前提是恰当选择相应课程。平台标记为"优秀"的课程内容充足、习题设置合理，因此吸引了大量学生。因此，我们根据不同学科的数量分布，从工程、法律、医学和经济学等 12 个领域共选取了 154 门"优秀"课程。这些课程分别由不同的教师授课，平均选课人数为 68，941，足以支持后续的数据收集与标注。

3.1.2 数据收集。基于所选课程，我们需要收集有关练习的多方面数据。

(1) 习题内容与答案：习题内容已被证明有助于学生建模 [20],，也能支持习题检索等任务。我们获取了每道习题的全部文本内容和题型（例如单项选择题、多项选择题）。此外，以往研究常常忽略习题的标准答案，我们也在这一阶段将其收集为补充侧信息。

[1]https://www.xuetangx.com/

(2) 习题图：另一项有益信息是习题之间的结构关系 [26]。由于习题按照一定顺序组织在课程的各个章节中，我们利用这些自然结构构建了数据集中的习题图，可将其概括为三级层次结构和一个偏序图。

（3）练习行为：确定练习后，我们可以抓取学生在这些题目上的行为记录。对于每位学生，我们按时间顺序整理其练习记录，并收集时间戳以构建交互序列。具体而言，得益于标准答案的收集，我们可以自动为每次行为赋予得分。我们还考虑了学生在多项选择题中给出部分正确答案的半对情形，并将得分分为三个等级：0-错误、1-半对、2-正确。研究人员在使用我们的数据集时，可以自行决定是否将其简化为二分类。

(4) 相关资源与行为：除习题信息和行为外，还有其他类型的相关特征，例如位于同一章节的课程视频、观看视频的行为以及课程简介。我们抓取了所有这些信息，以供未来改进学生建模以及研究自适应学习推荐等相关主题时使用 [22]。此外，由于 MOOCCubeX 提供了这些资源的概念集，我们将其保留为远程监督词典，以供后续标注。

3.1.3 噪声过滤。为减轻数据质量问题导致的训练困难，例如过拟合、模型遍历和无法收敛，我们对构建的学生行为序列进行了基本过滤。首先，根据现有数据集的分布，移除过短的行为序列以及分数异常的序列（例如，可能通过作弊手段得到的全对序列）[12, 32]。其次，过滤掉内容不完整的练习。如果某门课程无法保留足够的练习，则剔除整门课程及其所有练习。最后，我们仅保留原始数据的 40% 进行标注。

3.2 细粒度概念标注

概念，习题内容中隐含的知识概念（例如人工智能课程中的图神经网络 ）是知识追踪任务中至关重要且专业性很强的组成部分。即使对于

相关课程的专业教师而言，从头开始标注准确的细粒度概念也并非易事。基于从 MOOCCubeX 自动抽取的概念，我们首先通过提示大型语言模型进行具备噪声容忍能力的概念预识别，随后由一个 40 人专家团队开展广泛标注。

3.2.1 学科感知的概念抽取。以往的概念获取主要依赖传统的短语挖掘或实体链接方法。尽管已有少数研究将此任务视为远程监督命名实体识别（远程监督命名实体识别），但仍有一些问题亟待解决。（1）学科差异：不同学科的课程中，概念在词汇长度、风格和语义方面存在显著差异，现有方法往往难以在不同学科中保持稳定的性能。（2）噪声词典：由于基于给定词典构建的远程监督数据存在噪声，目前大多数方法可能无法区分合适标签与误导性标签，最终导致抽取结果完全不可用。因此，我们尝试利用大型语言模型提示，结合学科信息增强现有的远程监督命名实体识别方法，这一方向可在未来工作中进一步探索。总体而言，我们的自动概念准备分为两个步骤。

学科感知词典清理。在训练远程监督命名实体识别方法之前，我们先进行一个预处理步骤，对自动抽取的概念进行去噪，并将其作为远程监督信号。具体而言，我们通过基于提示的学习 [14]执行学科分类。将 MOOCCubeX 提供的概念词典 $T = \{c_i\}_{i=1,\dots,m}$ 中的每个概念 c_i 作为输入，分类过程会返回一个相关学科的排序列表 $F_{c_i} \subset F$，并为 $f_j \in F = \{f_j\}_{j=1,\dots,k}$ 输出 p_j，以表示其与 f_j 学科相关的可能性，形式化表示如下：

$$p_j(x') = LM(f_{fill}(x', f_j); \theta) \tag{1}$$

其中，$x' = f_{prompt}(c_i)$ 是一个提示，其中填入概念的模板槽位为 $[concept]$；函数 $f_{fill}(x', f_j)$ 则用潜在答案 f_j 填充槽位 $[MASK]$。

之后，我们可以推断某个概念最相关的前 k 个学科，从而排除不匹配的概念。我们将 $\{(X_m, D_m)\}_{m=1}^{M}$ 作为远程监督数据，其中，$X_m = [x_1, x_2, \dots, x_N]$ 由词元组成，并基于 BIO 标注体系构建，$D_m = [d_1, d_2, \dots, d_N]$ 用于训练概念抽取模型。

噪声容忍模型自训练。我们采用 BERT [15]模型作为骨干网络，训练细粒度概念抽取工具。按照常见的表述方式，我们用 $f(\cdot;\theta)$ 表示由 θ 参数化的模型，该模型是一个逐词分类器。$f_{n,c}(\cdot;\cdot)$ 表示 X_m 中第 n 个词元属于 BIO 标注体系中的第 c 类的概率。模型通过最小化 $\{(X_m, D_m)\}_{m=1}^{M}$ 上的交叉熵损失 $\mathrm{L}(\theta)$ 进行训练：

$$\mathcal{L}(\theta) = \frac{1}{M}\frac{1}{N}\sum_{m=1}^{M}\sum_{n=1}^{N} -\log f_{n,d_{m,n}}(X_m;\theta) \tag{2}$$

为了使模型具有抗噪能力，我们采用基于正未标注学习的自训练。为使用该算法，我们应用二元标签分配机制，将“O”映射为 0，将“B”和“I”映射为 1。随后，我们可以得到一个正例集和一个未标注集，用于构建新的 PU 训练损失 $\widehat{\mathrm{L}}(\theta)$，如 Peng et al. [29] 所述。

因此，我们模型 θ^* 的参数通过$\widehat{\ }$交叉熵损失 $\mathrm{L}(\theta)$ 与 PUL损失 $\mathrm{L}(\theta)$ 的组合进行学习：

$$\theta^* = \underset{\theta}{argmin}(\mathcal{L}(\theta) + \beta \cdot \widehat{\mathcal{L}}(\theta)) \tag{3}$$

其中，引入参数 β 以平衡这两个损失函数。采用该模型后，我们得到了性能更好的抽取器，能够识别每道习题中的候选概念，从而显著减轻专家标注的工作量。

3.2.2 概念的专家标注。基于预先抽取的概念，我们组织了一组专家标注员，对每道习题中的概念进行完善、删除和补充。标注员是相关学科的资深教师或助教。在开始标注之前，我们将抽取出的候选概念作为参考，供标注人员确定最细粒度的概念。所有标注人员均须完成以下子任务：（1）给定一道习题及其相关内容，选取或写出该习题考查的细粒度概念；（2）给定同一章节中的练习，选取或写出用于整合这些练习的中级概念；（3）给定同一课程中的习题和视频，完善或写出恰当的高级概念，以概括该课程所培养的技能。这些多层级概念标注可用于构建层次清晰的概念图。

质量控制。每条数据由两名标注人员分别标注，以便交叉核查。合并两名标注人员提供的细粒度概念时，我们采用以下策略确保质量。对于给定候选项中的选择式概念，只有两份结果完全一致时才予以保留。对于书写式概念，我们保留以下两类：a）两份结果完全一致；b）语义相似度超过0.8，且具有较长的公共子序列。实际操作中，按照上述标准完成合并后，我们会补充专家标注，以完善并确认最终结果。我们还通过计算所有课程的细粒度概念标注数据的平均一致性来监控质量，结果为0.673（> 0.6），表明结果质量可靠。

3.3 认知习题标注

学生学习的认知过程是近期方法中的一个新兴特征 [2, 34], , 涉及学习资源的难度、学生的认知模式等。由于目前尚无对此类特征进行标注的实证尝试，我们计划借鉴广受认可的认知理论中的经验，为后续认知标注提供专业起点。在教育科学领域专家（本文共同作者）的参与下，我们决定以布鲁姆认知分类法 [17]，作为后续标注的参考。这是现代教育认知领域最基础的理论之一。

3.3.1 布卢姆认知分类法标注。在布鲁姆认知分类法（该分类法于2001年修订）中，学生的学习可分为六个递进层次，如图3所示：(1) 记忆。保留具体且离散的信息片段，例如事实、定义或方法论。(2) 理解。掌握某个概念或教学材料的含义。(3) 应用。在这一层次，学生能够在新的（但相似的）情境中运用信息。(4) 分析。学生能够拆解已知内容并识别其中的关系。

表 1：MoocRadar 中学生习题行为示例。该示例展示了学生*U*_309 在经济学导论 和机器学习课程中的习题行为。蓝色短语为标注的多层级概念。

用户 ID	课程	章节 ID	习题 ID	内容	选项	认知	答案	分数	Time
U 309_	引言 *to* 经济学	*Ch*_6145	*Ex*_12138	研究对象的 包括：()	A：‘单个消费者’, B：‘单一生产者 ’, C：‘单一市场’, D：	2：理解	参考文献：ABC 用户：AC	1	2020-07-18 15:03:48
		*Ch*_6145	*Ex*_12145	起点的 研究是（）	A:‘ ’, B: ‘欲望’ , C: ‘需求’, D: ‘供给 ’	1: 记忆-bering	参考文献：A 用户：C	0	2020-07-18 15:04:11
	...	...	...	...	...	...	...	...	...
	机 学习	章 1129_	习题 26580_	当 *k* 不同时 的测试结果 *KNN* 可能相同。	A：‘正确’, B：‘错误’	3：应用	参考：A 用户：A	2	2020-09-07 17:46:46

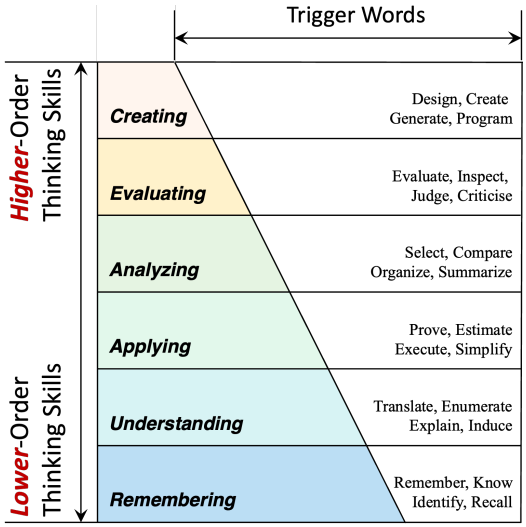


图 3：布鲁姆认知分类法。触发词通过文本挖掘提取，并用作标注的指导。

(5) 评价。 能够审视信息并做出判断。(6) 创造。 能够基于已有知识创造新事物。

尽管当前的评估练习旨在估计学生的认知层级，但在没有指导的情况下标注某个习题问题的特定层级仍然具有挑战性。因此，我们遵循教育学的建议，进行文本挖掘以确定每个认知层级的一系列代表性动词。基于这些触发词，标注人员需要将每个练习标注为最高认知层级。

3.3.2 质量验证。每项认知层级标注任务也会分配给两名标注人员。如果

表2：MoocRadar中多样化数据的统计。过滤率表示因质量控制而从原始数据中过滤掉的相关内容所占比例。

Item	类别	数量	过滤率
习题	单选 多	1,598 915	66.3% 58.2%
学生	-	14,224	91.6%
概念	粗粒度 中粒度 Fine	120 580 5,600	-
认知	-	2,513	-
行为	练习 学习	1,752,319 10,962,807	95.0% 93.5%

在合并过程中出现不一致，我们会分两种情况处理：对于相差不超过 1 个层级的标注，取较大的值，因为这在一定程度上反映了学生的高阶思维。对于相差超过 2 个层级的标注，我们会要求重新标注，并指派第三位专家作出判定。最后，我们还进行了 一致性检验。所有课程的认知层级标注平均一致性为 0。738，这表明认知层级标注的质量甚至优于概念标注。与学生数据匹配后，我们数据中标准学生行为序列的截取片段如表 1 所示。

3.4 可用性

研究人员可通过 https://github.com/ THU-KEG/MOOC-Radar 访问我们的数据仓库。为便于使用数据，我们对多种类型的数据进行了聚合，并提供了一系列带有代码的工具，以简化查询和补充数据。具体而言，已发布的数据仓库包含三个主要组成部分：

- **多个数据集**。由于练习数据可以关联到多层级的知识成分（课程、章节、细粒度的

表3：现有教育数据集的比较。序列长度、密度 和活动 分别对应学生练习序列的平均长度、数据集概念与习题数量的商以及学习行为。

数据集	习题方面信息。			概念方面信息。			认知信息。		手动标注
	序列长度	内容	结构	数量	密度;	图表	级别	活动	
ASSISTments2009	82.2	✗	✗	123	0.004	✗	✗	✗	Yes
ASSISTments2012	131.2	✗	✗	265	0.001	✗	✗	✗	Yes
ASSISTments2015	35.4	✗	✗	✗	✗	✗	✗	✗	No
Junyi Academy	104.6	✓	✓	41	0.056	✗	✗	✗	No
模拟-5	50.0	✗	✗	5	0.100	✗	✗	✗	No
代数 2005-2006	141.5	✗	✗	523	0.103	✗	✗	✗	No
静力学2011	107.8	✓	✗	85	0.069	✗	✓	✗	No
EdNet-KT	121.5	✓	✓	293	0.014	✗	✗	✓	No
MoocRadar	123.2	✓	✓	5,600	2.228	✓	✓	✓	Yes

概念）和其他学习行为，我们提出：（1）带有标注的原始数据；（2）具有不同概念粒度的清洗后练习行为记录；以及（3）练习和视频观看相结合的学生行为数据，这些数据可支持知识追踪、学习推荐及其他相关任务。

• **易用工具包**。与此同时，我们准备了一个工具包，其中包含 12 个常用查询函数和 API，便于信息检索和数据结构构建。通过组合使用多个工具，开发者可以执行更复杂的查询，并根据需求构建更多数据集。

• **复现模型**。我们开展了多个阶段的专家标注，并提出了一些模型（例如自训练概念提取模型）。因此，我们在开放页面上提供了详细的标注指南和模型源代码，以支持对我们资源库的复现和进一步扩展。

我们将继续为更多学科开展专家标注，并计划引入更多认知理论。我们的资源库也将持续更新，以跟进这些工作。此外，我们还将按照第 3.2 节中的讨论，对概念提取工具进行技术改进。

4 数据分析

本节通过分析数据统计信息及其分布，介绍我们知识库的特征。

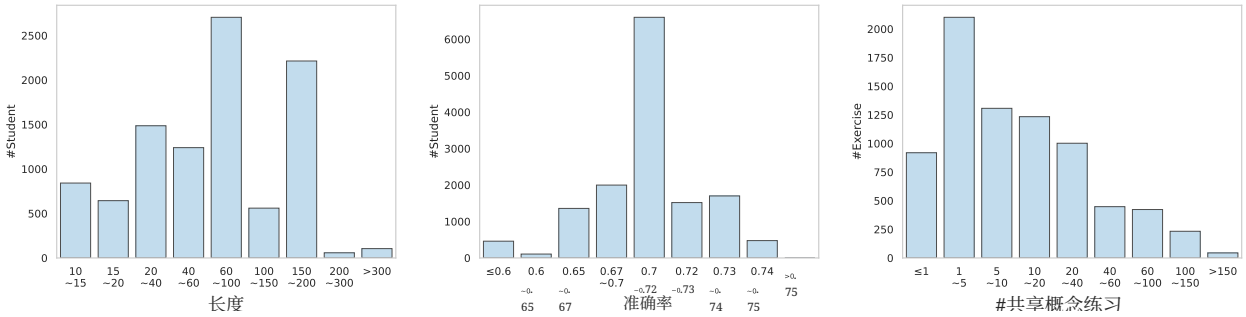
统计信息。 表 2 展示了 MoocRadar 中各类数据的详细统计信息。在严格的数据质量控制下，MoocRadar 删除了大部分不可用的原始数据，并保留了足量的高质量数据（包括多个层级的知识），以支持模型训练与评估。值得注意的是，行为是过滤最多的项目，尤其是其中的练习行为（过滤率为 95.0%）。

与其他数据集的比较。我们选取了几类知名的教育数据集进行比较，包括：1）ASSIST 数据集（2009、2012、2015）[12, 28]；2）公开竞赛数据，即，Junyi Academy [32],模拟-5 [31],KDDcup Algebra 2005-2006 [36]；3）开放数据仓库，即，STATIC2011 [16], EdNET-KT [8]。本次比较未列入数据源 MOOCCubeX [46] 。

我们在表 3 中分别从习题、概念和认知方面比较了各个数据集。总体而言，MoocRadar 对有价值信息的覆盖最为完善，尤其是涵盖的知识概念最广。此外，我们计算了两个统计特征，以展示该仓库的可用性。对于学生练习行为的平均长度（序列长度），MoocRadar 与当前大多数数据集相比具有竞争力。对于习题的概念密度（密度），MoocRadar 相较于其他现有数据集具有显著优势，即，每道习题平均对应 2.228 个新概念。同时，尽管许多现有数据集拥有大量原始数据，但在信息供给、概念粒度等方面仍有待改进，才能更好地支持探索。

练习行为与概念。 我们还在图 4 中展示了练习及其对应行为的分布。从中可以推断出我们知识库的三个特征。首先，我们知识库中的大多数学生在练习行为方面超过 60 次。根据行为序列长度，学生可分为两大群体，即$60-100$ 和 $150-200$；此外，还有一小部分学生完成了 300 多道练习，这再次体现了数据集中学习行为的丰富性。其次，正确率得到了适当控制。经过学生行为过滤后，学生练习的正确率保持在 0.6 至 0.75 之间，且呈正态分布。这表明作弊及其他异常行为得到了有效控制，从而缓解了现有模型的过拟合问题。第三，我们根据概念共享关系计算每道练习的相关练习，如图 4(c) 所示。大多数练习通过概念与其他 $2-20$ 道练习建立关联，从而构建出密度适当的练习图。总体而言，MoocRadar 的练习数据质量高，分布也较为合理。

认知层级分布。我们还在图 5 中展示了学生在六个认知层级上的行为。从统计结果可以看出，目前大多数学生行为仍处于较低层级（例如，记忆和理解），占总数的 70。0%。除记忆层级外，随着认知层级的提高，行为数量呈下降趋势，这符合教育学中的普遍假设 [27]。与此同时，仍有许多练习仅由部分学生完成，



(a) 学生行为长度分布。 (b) 习题准确率分布。 (c) 共享概念练习分布。

图 4：习题的数据分布及其对应的行为。

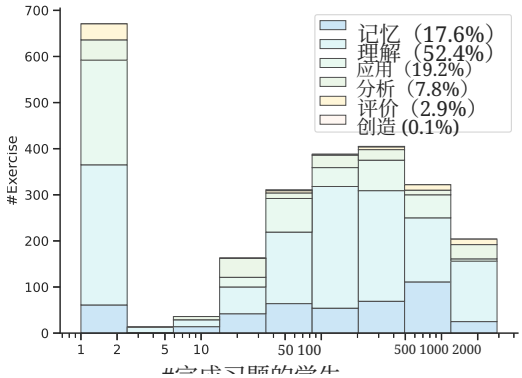


图 5：学生六个认知层级的习题作答行为堆叠分布。

其余练习则由符合正态分布的学生完成。最高阶的创造仅占总数的 0。1%，这可能是因为现有数据主要为选择题。在后续更新中，我们应加入主观题，以丰富数据类型。

5 实验

本节将针对知识追踪和认知诊断开展初步实验，以探究所引入特征在我们的知识库中的潜在应用方向。

5.1 实验设置

5.1.1 数据集。 我们选择具有不同概念粒度的数据集作为核心实验数据基础。同时，我们选择带有认知标签和视频观看行为的粗粒度数据集作为辅助信息，以进一步探索其有效性。总体而言，这些数据集包含多层级概念（粗粒度：120，中粒度：580，细粒度：5600）、2，513 道习题，以及 1，752，319 条行为。在改进方法时，所有 6 种认知标签和 6，567 多个视频都用作辅助信息。

5.1.2 基线。 我们借助两个公开代码库 EduKTM[2] [21] 和 EduCDM [3] [38, 41]，在数据集上构建了若干具有代表性的基线，以观察初步结果。

对于知识追踪 （KT）设置，我们选择（1）**深度知识追踪（DKT）** [31]：一种基于 RNN 的早期方法，将神经网络引入此任务。（2）**深度知识追踪（DKT）+** [44]：一种改进的 KT 方法，考虑了表现的一致性。（3）**DKVMN** [48]：该方法利用动态记忆网络对学生行为进行建模。

对于认知诊断 （CD）设置，我们选择了：**(1) MIRT** [5]：一种多维度 IRT 方法，可应用于更多环境。**(2) GDIRT** [11]：一种采用梯度下降的改进型传统 IRT 方法。需要注意的是，前两种方法均基于传统线性回归。**(3) NCDM** [41]：一种在认知诊断任务中对神经网络的著名改编方法。

5.2 结果分析

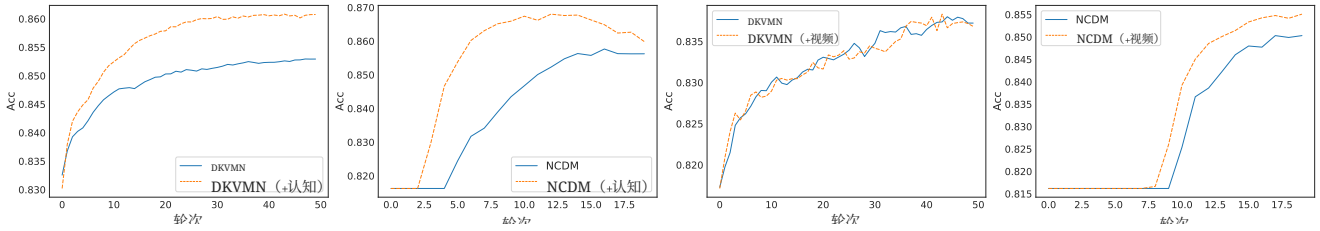
5.2.1 主要结果。 总体性能如表 4 所示。基于这些结果，我们可以推断出该知识库的一些特征，并讨论若干可能指示该主题未来发展方向的重要观察结果。

首先，MoocRadar 支持不同主题和架构的多种模型（RNN、回归、Transformer、记忆网络），体现了该知识库的便捷性和可用性。同时，由于 MoocRadar 提供多层级概念，研究人员可以从更多方面评估其模型性能，并开展更多类型的研究。

其次，使用更细粒度的概念时，所有模型的性能都有所提升，这表明此类标注是有效的。一种可能的原因是，细粒度概念有助于模型更好地理解和解释学生的表现，而粗粒度标注有时会使模型产生混淆（学生在同一粗粒度概念下表现不一致，实际上是因为他们掌握的细粒度知识完全不同）。

第三，一些传统方法（例如，深度知识追踪（DKT）、MIRT）并不适用于细粒度概念场景（因为时间和内存开销过大），这表明近期在人工智能领域的探索至关重要。不过，值得注意的是，传统方法仍能取得颇具竞争力的表现

[2]https://github.com/bigdata-ustc/EduKTM [3] https://github.com/bigdata-ustc/EduCDM



(a) 使用认知特征的 DKVMN。 (b) 使用认知特征的 NCDM。 (c) 使用学习特征的 DKVMN。 (d) 使用学习特征的 NCDM。

图 6：训练轮次中的性能改进曲线

DKVMN 和 NCDM 模型。

表 4：不同概念粒度下知识追踪（KT）和认知诊断（CD）模型的结果。"/"表示该设置下不适用此性能指标。

设置	方法	概念	AUC	ACC
KT	DKT	粗粒度	0.815	0.831
		中粒度	0.848 (+0.033)	0.851 (+0.020)
		Fine	/	/
	DKT+	粗粒度	0.727	0.790
		中粒度	0.741 (+0.014)	0.819 (+0.029)
		Fine	0.842 (+0.075)	0.844 (+0.054)
	DKVMN	粗粒度	0.836	0.835
		中粒度	0.863 (+0.027)	0.850 (+0.006)
		Fine	0.876 (+0.040)	0.862 (+0.027)
CD	MIRT	粗粒度	0.849	0.853
		中粒度	/	/
		Fine	/	/
	GDIRT	粗粒度	0.799	0.789
		中粒度	0.832 (+0.033)	0.836 (+0.047)
		Fine	0.882 (+0.083)	0.867 (+0.078)
	NCDM	粗粒度	0.816	0.837
		中粒度	0.871 (+0.055)	0.853 (+0.016)
		Fine	0.885 (+0.069)	0.868 (+0.029)

（例如，GDIRT），这表明未来研究可以继续考虑将"传统认知理论"与"先进人工智能模型"相结合的范式，以推动该领域的发展。

5.2.2 模型改进。 除了细粒度概念之外，我们还尝试通过认知标签和其他学习行为来改进现有模型。对于认知层级，我们构建独热嵌入，并将其作为新特征加入嵌入层以调整模型。对于学习行为，我们选取最近的五种行为组成一个序列，然后采用学习行为预训练模型 [50] 获取该序列的嵌入，以扩展现有模型。

图 6 展示了 DKVMN 和 NCDM（主实验中表现最佳的模型）改进后的性能（以准确率衡量）。可以观察到，在这些辅助信息的帮助下，这两个模型的性能都有所提升，具体可归纳为：(1) 认知标签能够有效提升模型性能，且在从训练开始到收敛的整个过程中都能发挥作用。除了带来数值上的提升外，添加这类信息还能加快模型收敛，也有助于提高模型的训练稳定性。(2) 经过合理建模后，学习活动能够对部分模型产生积极作用。与认知标签相比，这些含有原始噪声的数据虽不能增强模型的稳定性，但仍能提升具备学习感知能力的模型的表现（例如，NCDM 会考虑学习难度）。(3) 初步实验结果基于自动计算的准确率，这是所有机器学习任务中较为基础且通用的指标。

6 影响

对于人工智能驱动教育领域的研究人员而言，MoocRadar 提供了坚实的实验基础，支持对先进模型和方法的探索。近期，研究人员关注了许多改进方向（例如，概念关系 [7], 练习结构 [26],认知过程 [2]），而 MoocRadar 能有效促进这些想法的实现，尤其是此前受数据限制的研究工作。同时，MoocRadar 通过引入认知层级扩展了现有任务设置，这或许能启发研究人员引入更多认知科学理论，并丰富学生建模任务。

对于智能教育应用的开发者而言，MoocRadar 能发挥更积极的作用，并为打造更具吸引力的教育产品提供分析见解。MoocRadar 中的细粒度概念和认知层级可作为黄金标签，供平台计划对自身数据开展类似标注时使用 [47]。此外，MoocRadar 还可作为公平的基准，用于评价知识追踪、认知诊断和学习推荐预开发模型的性能。

此外，MoocRadar 也是教育学 和教育科学 领域研究者开展在线学习详细分析的数据资源库。在在线教育与线下教育日益融合的主流趋势下 [24], ，我们希望 MoocRadar 能带来构建开放、高质量的人工智能技术与教育研究数据桥梁的经验。

7 结论

本文介绍了 MoocRadar，这是一个细粒度、多方面的知识库，包含 2,513 道习题、5,600 个概念，以及 14,224 名学生的 12,715,126 条行为记录，旨在改进大规模开放在线课程中的认知学生建模。具体而言，我们构建了一个框架和一系列标准，以支持对细粒度概念和认知层级进行大量专家标注。我们对 MoocRadar 开展了统计和实验研究。结果表明，该知识库包含

覆盖面广泛的有益信息，且数据分布合理；所提供的丰富特征确实能够改进采用不同架构的现有模型。该知识库现已公开发布，并附带一个工具包，方便相关研究人员和开发者访问数据。

致谢

本研究项目得到清华大学郭强研究院资助（2019GQB0003）。Jie Tang 得到国家自然科学基金委员会杰出青年科学基金（61825602）资助。本研究还得到国家自然科学基金委员会资助（批准号：62276154）和（批准号：62277034）。

参考文献

[1] Ghodai Abdelrahman 和 Qing Wang。2019。使用序列键值记忆网络的知识追踪。在第*42*届国际*ACM SIGIR*信息检索研究与发展会议论文集。175–184。[2] Ghodai Abdelrahman、Qing Wang 和 Bernardo Pereira Nunes。2022。知识追踪：综述。*Comput. Surveys*（2022）。[3] John R Anderson、C Franklin Boyle 和 Brian J Reiser。1985。智能辅导系统。*Science* 228, 4698（1985），456–462。[4] Sahan Bulathwela、Maria Perez-Ortiz、Emine Yilmaz 和 John Shawe-Taylor。2020。Truelearn：将终身学习者与开放教育资源匹配的贝叶斯算法族。在*AAAI*人工智能会议论文集，第34卷。565–573。[5] R Philip Chalmers。2012。mirt：R环境的多维项目反应理论包。*Journal of statistical Software* 48（2012），1–29。[6] Mingzhi Chen、Quanlong Guan、Yizhou He、Zhenyu He、Liangda Fang 和 Weiqi Luo。2022。具有学习与遗忘行为的知识追踪模型。在第*31*届*ACM*国际信息与知识管理会议论文集。3863–3867。[7] Penghe Chen、Yu Lu、Vincent W Zheng 和 Yang Pian。2018。先修知识驱动的深度知识追踪。在 *2018 IEEE* 国际数据挖掘会议（*ICDM*）。IEEE，39–48。[8] Youngduck Choi、Youngnam Lee、Dongmin Shin、Junghyun Cho、Seoyon Park、Seewoo Lee、Jineon Baek、Chan Bae、Byungsoo Kim 和 Jaewe Heo。2020。Ednet：教育中的大规模层次化数据集。在教育人工智能：第*21*届国际会议，*AIED 2020*，摩洛哥伊夫兰，*2020*年*7*月*6*日至*10*日，会议录，第二部分*21*。Springer，69–73。[9] Konstantina Chrysafiadi 和 Maria Virvou。2013。学生建模方法：过去十年的文献综述。*Expert Systems with Applications* 40, 11（2013），4715–4729。[10] Albert T Corbett 和 John R Anderson。1994。知识追踪：建模程序性知识的获取。用户建模与用户自适应交互 4（1994），253–278。[11] Susan E Embretson 和 Steven P Reise。2013。项目反应理论。Psychology Press。[12] Mingyu Feng、Neil Heffernan 和 Kenneth Koedinger。2009。用边评估边辅导的在线系统应对评估挑战。用户建模与用户自适应交互 19（2009），243–266。[13] Wenzheng Feng、Jie Tang 和 Tracy Xiao Liu。2019。理解大规模开放在线课程中的辍学。在*AAAI*人工智能会议论文集，第33卷。517–524。[14] Yuxian Gu、Xu Han、Zhiyuan Liu 和 Minlie Huang。2022。PPT：少样本学习的预训练提示调优。在计算语言学协会第*60*届年会论文集（第*1*卷：长论文）。8410–8423。[15] Jacob Devlin Ming-Wei Chang Kenton 和 Lee Kristina Toutanova。2019。BERT：用于语言理解的深度双向Transformer预训练。在*NAACL-HLT*论文集。4171–4186。[16] Kenneth R Koedinger、Ryan SJd Baker、Kyle Cunningham、Alida Skogsholm、Brett Leber 和 John Stamper。2010。EDM社区的数据仓库：PSLC DataShop。教育数据挖掘手册 43（2010），43–56。[17] David R Krathwohl。2002。布卢姆分类法的修订：概述。*Theory into practice* 41, 4（2002），212–218。[18] Wonsung Lee、Jaeyoon Chun、Young min Lee、Kyoungsoo Park 和 Sungrae Park。2022。知识追踪的对比学习。在*2022 ACM Web*会议论文集。2330–2338。[19] Irene Li、Alexander R Fabbri、Robert R Tung 和 Dragomir R Radev。2019。我应该先学什么：引入lecturebank用于自然语言处理教育和先修链学习。在*AAAI*人工智能会议论文集，第33卷。6674–6681。[20] Qi Liu、Zhenya Huang、Yu Yin、Enhong Chen、Hui Xiong、Yu Su 和 Guoping Hu。2019。EKT：面向学生成绩预测的练习感知知识追踪。《IEEE Transactions on Knowledge and Data Engineering》33，1（2019），100–115。[21] Qi Liu、Shuanghong Shen、Zhenya Huang、Enhong Chen 和 Yonghe Zheng。2021。知识追踪综述。arXiv 预印本 arXiv:2105.15106（2021）。[22] Qi Liu、Shiwei Tong、Chuanren Liu、Hongke Zhao、Enhong Chen、Haiping Ma 和 Shijin Wang。2019。利用认知结构实现自适应学习。载于《第25届ACM SIGKDD国际知识发现与数据挖掘会议论文集》，627–635。[23] Yiming Mao、Bin Xu、Jifan Yu、Yifan Fang、Jie Yuan、Juanzi Li 和 Lei Hou。2021。为在线教育系统学习行为感知的认知诊断。载于《数据科学：第7届开拓性计算机科学家、工程师与教育工作者国际会议（ICPCSEE 2021）论文集》，第二部分，第7卷，中国太原，2021年9月17–20日。Springer，385–398。[24] Shailendra Palvia、Prageet Aeron、Parul Gupta、Diptiranjan Mahapatra、Ratri Parida、Rebecca Rosner 和 Sumita Sindhi。2018。在线教育：全球现状、挑战、趋势及影响。，233–241页。[25] Liangming Pan、Chengjiang Li、Juanzi Li 和 Jie Tang。2017。学习 MOOC 中概念的先修关系。载于《第55届计算语言学协会年会论文集（第1卷：长篇论文）》，1447–1456。[26] Shalini Pandey 和 Jaideep Srivastava。2020。RKT：面向知识追踪的关系感知自注意力。载于《第29届ACM国际信息与知识管理会议论文集》，1205–1214。[27] Alexandros Paramythis 和 Susanne Loidl-Reisinger。2003。自适应学习环境与电子学习标准。载于《第二届欧洲电子学习会议》，第1卷，369–379。[28] Zachary A Pardos、Ryan SJD Baker、Maria OCZ San Pedro、Sujith M Gowda 和 Supreeth M Gowda。2013。情感状态与状态测试：探究整个学年中的情感如何预测年末学习成果。载于《第三届国际学习分析与知识会议论文集》，117–124。[29] Minlong Peng、Xiaoyu Xing、Qi Zhang、Jinlan Fu 和 Xuan-Jing Huang。2019。利用正例-未标注学习进行远程监督命名实体识别。载于《第57届计算语言学协会年会论文集》，2409–2419。[30] Huy Phuong Phan。2010。学生的学业表现与多种认知学习过程：综合框架与实证分析。《Educational Psychology》30，3（2010），297–322。[31] Chris Piech、Jonathan Bassen、Jonathan Huang、Surya Ganguli、Mehran Sahami、Leonidas J Guibas 和 Jascha Sohl-Dickstein。2015。深度知识追踪。《神经信息处理系统进展》28（2015）。[32] Chen Pojen、Hsieh Mingen 和 Tsai Tzuyang。2020。Junyi Academy 在线学习活动数据集：涵盖小学至高中学生的大规模公开在线学习活动数据集。数据集见
h
t
t
p
s://
w
w
w.kaggle.com/junyiacademy/learning-activity-public-dataset-by-junyi-academy（2020）。[33] Joseph Psotka、Leonard Daniel Massey 和 Sharon A Mutter。1988。智能辅导系统：经验教训。Psychology Press。[34] Shuanghong Shen、Qi Liu、Enhong Chen、Zhenya Huang、Wei Huang、Yu Yin、Yu Su 和 Shijin Wang。2021。学习过程一致的知识追踪。载于《第27届ACM SIGKDD知识发现与数据挖掘会议论文集》，1452–1460。[35] Shuanghong Shen、Qi Liu、Enhong Chen、Han Wu、Zhenya Huang、Weihao Zhao、Yu Su、Haiping Ma 和 Shijin Wang。2020。卷积知识追踪：对学生学习过程中的个体差异进行建模。载于《第43届ACM国际信息检索研究与发展会议论文集》，1857–1860。[36] J Stamper、A Niculescu-Mizil、S Ritter、GJ Gordon 和 KR Koedinger。2010。KDD Cup 2010教育数据挖掘挑战赛数据集。（2010）。[37] Shiwei Tong、Qi Liu、Wei Huang、Zhenya Hunag、Enhong Chen、Chuanren Liu、Haiping Ma 和 Shijin Wang。2020。基于结构的知识追踪：一种影响传播视角。载于《2020年IEEE国际数据挖掘会议（ICDM）》，IEEE，541–550。[38] Shiwei Tong、Qi Liu、Runlong Yu、Wei Huang、Zhenya Huang、Zachary A Pardos 和 Weijie Jiang。2021。面向认知诊断的项目反应排序。载于IJCAI，1750–1756。[39] Kurt VanLehn。1988。学生建模。《智能辅导系统基础》55（1988），78。[40] Shanshan Wan 和 Zhendong Niu。2019。一种基于学习者影响传播的混合式电子学习推荐方法。《IEEE Transactions on Knowledge and Data Engineering》32，5（2019），827–840。[41] Fei Wang、Qi Liu、Enhong Chen、Zhenya Huang、Yuying Chen、Yu Yin、Zai Huang 和 Shijin Wang。2020。面向智能教育系统的神经认知诊断。载于《AAAI人工智能会议论文集》第34卷，6153–6161。[42] Pengfei Wang、Yu Fan、Long Xia、Wayne Xin Zhao、ShaoZhang Niu 和 Jimmy Huang。2020。KERL：一种用于序列推荐的知识引导强化学习模型。载于《第43届ACM国际信息检索研究与发展会议论文集》，209–218。

[43] Meng Xia、Mingfei Sun、Huan Wei、Qing Chen、Yong Wang、Lei Shi、Huamin Qu 和 Xiaojuan Ma。2019。Peerlens：受同伴启发的在线题库交互式学习路径规划。载于*2019*年*CHI*人因计算系统会议论文集，1–12。[44] Chun Kit Yeung 和 Dit Yan Yeung。2018。通过预测一致性正则化解决深度知识追踪中的两个问题。载于第*5*届*ACM Learning @ Scale*会议论文集。ACM，5:1–5:10。[45] Jifan Yu、Gan Luo、Tong Xiao、Qingyang Zhong、Yuquan Wang、Wenzheng Feng、Junyi Luo、Chenyu Wang、Lei Hou、Juanzi Li 等。2020。MOOCCube：面向大规模开放在线课程自然语言处理应用的大规模数据仓库。载于第*58*届计算语言学协会年会论文集，3135–3142。[46] Jifan Yu、Yuquan Wang、Qingyang Zhong、Gan Luo、Yiming Mao、Kai Sun、Wenzheng Feng、Wei Xu、Shulin Cao、Kaisheng Zeng 等。2021。MOOCCubeX：面向大规模开放在线课程自适应学习的大型以知识为中心的仓库。载于第*30*届*ACM*国际信息与知识管理会议论文集，4643–4652。[47] Jifan Yu、Xiaohan Zhang、Yifan Xu、Xuanyu Lei、Xinyu Guan、Jing Zhang、Lei Hou、Juanzi Li 和 Jie Tang。2022。XDAI：一种无需调优的框架，用于利用知识引导的对话生成中的预训练语言模型。载于第*28*届*ACM SIGKDD*知识发现与数据挖掘会议论文集。4422–4432。[48] Jiani Zhang、Xingjian Shi、Irwin King和Dit-Yan Yeung。2017。用于知识追踪的动态键值记忆网络。载于第*26*届万维网国际会议论文集。765–774。[49] Bowen Zhao、Jiuding Sun、Bin Xu、Xingyu Lu、Yuchen Li、Jifan Yu、Minghui Liu、TingjianZhang、Qiuyang Chen、Hanming Li等。2022。EDUKG：一个异构、可持续的K-12教育知识图谱。*arXiv*预印本 *arXiv:2210.12228*（2022）。[50] Qingyang Zhong、Jifan Yu、Zheyuan Zhang、Yiming Mao、Yuquan Wang、Yankai Lin、Lei Hou、Juanzi Li和Jie Tang。2022。迈向大规模开放在线课程自适应学习的通用预训练框架。*arXiv*预印本 *arXiv:2208.04708*（2022）。[51] Yutao Zhu、Jian-Yun Nie、Kun Zhou、Pan Du、Hao Jiang和Zhicheng Dou。2021。基于相关知识和目标的主动式检索聊天机器人。载于第*44*届国际*ACM SIGIR*信息检索研究与发展会议论文集。2000–2004。