

大模型驱动认知诊断的成绩预测与教学实践^{*}

刘凡** 崔金凤

吴铭

沈荣

淮海大学计算机与软件学院
南京 211100

淮海大学人工智能与自动化学院
常州 213200

淮海大学计算机与软件学院
南京 211100

摘要 随着在线教育体系的快速发展,教育者对学生学习效果以及认知水平掌握不准确的问题逐渐成为教育者面临着一个关键挑战。传统评估方法往往依赖于有限的答题数据,难以全面反映学生的知识掌握情况。针对这一问题,本研究创新性地提出了一种基于大模型驱动的认知诊断成绩预测方法。该方法的核心在于将先进的大模型技术与神经认知诊断模型(NeuralCDM)相结合,通过构建一个智能化的诊断预测系统来提升评估准确性。具体而言,该方法基于大模型技术,将微调后最优权重的Qwen大模型引入NeuralCDM神经认知诊断过程中,对学生未作答试题进行答题预测,筛选预测结果得到高置信度伪标签以扩充学生答题数据集,并重新训练NeuralCDM得到更优的成绩预测效果。实验结果显示,此方法在数据集上的准确率、AUC值及F1值分别提升了约1.5%、2%及1.3%,表明其在成绩预测准确度上优于传统成绩预测方法,有助于教育者准确预估学生的能力值,对教学策略和教学重点的改进有着重要的指导意义。

关键词 成绩预测, 认知诊断, 大语言模型, NeuralCDM

Performance Prediction and Teaching Practice of Cognitive Diagnosis Driven by Large Model

Fan Liu** Jinfeng Cui

Ming Wu

Rong Shen

School of Computer and Software
Hohai University
Nanjing 211100, China

School of Artificial Intelligence and
Automation Hohai University
Changzhou 213200, China

School of Computer and Software
Hohai University Nanjing 211100,
China

Abstract—With the rapid development of online education systems, the challenge of educators accurately assessing students' learning outcomes and cognitive levels has become a critical issue. Traditional evaluation methods often rely on limited answer data, making it difficult to comprehensively reflect students' knowledge mastery. To address this, this study innovatively proposes a cognitive diagnostic performance prediction method driven by large models. The core of this approach lies in integrating advanced large model technology with the Neural Cognitive Diagnostic Model (NeuralCDM), building an intelligent diagnostic prediction system to enhance evaluation accuracy. Specifically, this method leverages fine-tuned Qwen large model weights in the NeuralCDM's cognitive diagnostic process to predict students' responses to unanswered questions. By filtering high-confidence pseudo-labels from prediction results to expand the answer dataset, the NeuralCDM is retrained to achieve improved performance prediction. Experimental results show that this method improves accuracy, AUC, and F1 values by approximately 1.5%, 2%, and 1.3% respectively, demonstrating superior performance in academic prediction compared to traditional methods. This advancement helps educators accurately estimate students' competency levels, providing crucial guidance for refining teaching strategies and instructional priorities.

Keywords—Performance prediction, cognitive diagnosis, large language models, NeuralCDM

1 前言

党的二十大报告指出:“教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑。

必须坚持科技是第一生产力、人才是第一资源、创新是第一动力,深入实施科教兴国战略、人才强国战略、创新驱动发展战略,开辟发展新领域新赛道,不断塑造发展新动能新优势。”^[1]构建以“人工智能+”国家治理为代表的数字化、智能化治理体系,是深入推进国家治理体系和治理能力现代化的必由之路和实现人民群众美好生活的关键一招^[2]。2024年两会政府工作报告将加快发展新质生产力作为首要任务,进一步提出要实施“人工智能+”行动,以期加快人工智能创新和赋能现代化产业体系建设^[3]。在“人工智能+教育”方面,当前人工智能处于教育变革的最前沿,与教育的结合体现出融合、创新、跨界、终身

***基金资助:** 本文得到{江苏省高等教育“电子信息类专业建设、课程建设、教学研究”重点课题“大模型与认知诊断联合驱动的成绩预测及自动组卷研究”(2024JSDZJG12);教育部产学研合作协同育人项目“基于华为云AI开发平台的融合课程建设与交流”(231100007112225);中国软件行业协会国产软件进课堂教学改革项目“基于华为AI的人工智能课程教学改革”(JGLX-CSIA-HW-20241088)}资助。

** 通讯作者: 刘凡 fanliu@hhu.edu.cn。

化的新特征^[4]。传统的教学实验往往受限于资源、时间和空间,如实验设备、课程时间安排、教学场所等的限制,而人工智能的赋能则有望打破这些限制,为学习者提供更加个性化、高效的学习体验。

人工智能技术可以有效地解决传统教学实验中的问题,而数据挖掘作为人工智能技术的一个重要技术分支,逐渐成为教育研究的重要手段^[5]。在教学信息中挖掘学生的能力值对于教育者来说是十分重要的,具体可以表现在对知识点的掌握程度。认知诊断是一种评估个体认知过程和理解水平的方法,能够有效反映学生的知识掌握程度,老师可以根据认知诊断结果改变教学重点,调整教学策略。然而部分学生在在线教育系统中的答题记录非常少,使得认知诊断在这些学生中运行效果较差。

大语言模型利用大规模文本数据进行预训练,具有强大的语言理解和生成能力^[6],可以很好地解决学生答题数据匮乏导致的认知诊断结果不佳的问题,进而为教育领域提供更深入、更全面的数据分析和智能化支持。因此,论文提出了一种大模型驱动认知诊断的成绩预测方法。本方法采用在线教育系统中产生的学生真实答题记录进行模型训练、验证和测试,并引入大语言模型,基于大语言模型的微调技术和问答机制进行成绩预测,最终验证了将大模型用于成绩预测任务的可行性和高效性。

2 相关工作

认知诊断理论^[7]可以应用于心理学^[8]、医学^[9]、人机交互^[10]和教育学^[11]等领域。它在智慧教育领域应用最为广泛^[12],如学业成绩预测、课程设计优化、教学质量评估等,其中,最常见的应用场景即为通过分析学生的历史数据进行学业成绩的预测。在这个方面,它能够超越传统的成绩测量方法,深入剖析学生在特定知识领域和认知技能上的掌握程度及存在的不足。通过精细的诊断,教师可以了解每个学生的具体认知强项和弱点,为个性化教学提供依据。

在线教育的兴起及教育数据的大量产生促进了认知诊断理论的发展,在现代课程不断发展、教育方法不断创新的背景下,认知诊断作为教育数据挖掘中的一项重要研究,在学生评价中发挥着举足轻重的作用^[13-15]。对于教育政策的制定,大规模的认知诊断数据能够为教育部门提供有关学生群体整体认知水平的信息,有助于制定合理的教育政策和资源分配方案。

3 本文方法

认知诊断模型 NeuralCDM^[16]通过神经网络学习学生与练习之间的复杂交互作用,实现智能教育系统中准确和可解释的认知诊断。Qwen^[17]是阿里云推出的一个超大规模的语言模型,具备强大的语言处理能力,可以辅助认知诊断模型完成成绩预测。基于以上两种模型,论文提出了一种大模型驱动认知诊断的成绩预测方法。

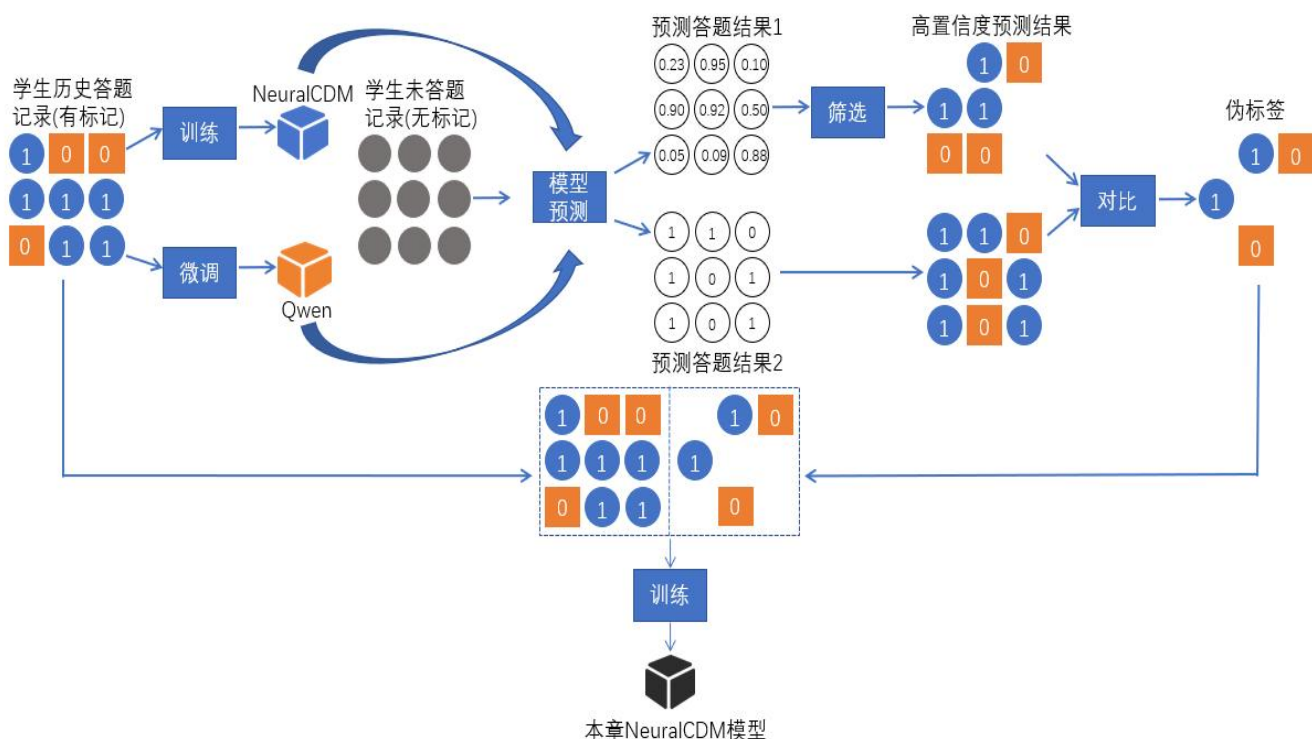


图 1 大模型驱动认知诊断的成绩预测方法

该方法的具体设计与实现过程如图 1 所示，主要涉及两个模块：一是研究基于 Qwen 大模型的成绩预测方法，利用大模型的优势并融合认知诊断 Q 矩阵进行问答预测，具体来说，是通过微调技术训练出不同权重的大模型，找到最优权重以适应成绩预测任务；二是采用伪标签半监督学习进行成绩预测，对无标签数据集（也就是学生未答题记录）设置伪标签，达到扩充训练数据集的效果。具体来说，采用微调后最优权重的 Qwen 大模型和初步训练好的 NeuralCDM 对学生未作答过的试题进行预测，对比选取高置信度的伪标签数据，与历史答题记录合并扩充学生答题记录，在合并后的数据集上重新训练 NeuralCDM，得到预测效果更优的 NeuralCDM，最终形成对班级全体学生更全面的认知诊断结果。

3.1 Prompt 设计

在大语言模型中，“Prompt”通常指的是作为输入提供给模型的一段文本或问题，这段文本或问题被用作模型生成输出的起点或参考。通过给模型提供 Prompt，用户可以引导模型生成特定主题或内容的文本，从而更好地控制模型的输出。Prompt 的选择和质量直接影响了模型生成文本的准确性和相关性，一个好的 Prompt 可以有效引导模型输出期望的答案，而一个不恰当的 Prompt 则可能导致模型产生误导甚至有害的回应。

Background: You are now a smart model for predicting student answers. The array[[1, 1, '0.68', '4', 1], [1, 2, '0.94', '2', 1], [1, 3, '0.95', '3', 1], [1, 4, '1.00', '1', 1], [1, 5, '0.86', '5', 1], [1, 6, '1.00', '2', 1], [1, 24, '0.81', '7', 1], [1, 25, '0.95', '8', 1], [1, 26, '0.40', '8', 1], [1, 27, '0.91', '9', 1], [1, 28, '0.90', '10', 1], [1, 29, '0.82', '10', 1], [1, 50, '0.80', '11,12', 1], [1, 51, '0.88', '11', 1], [1, 52, '0.89', '11,13', 1], [1, 49, '0.87', '14', 1], [1, 53, '0.96', '14', 1], [1, 54, '0.86', '12', 1], [1, 69, '0.83', '16,17', 1], [1, 66, '0.82', '17', 1], [1, 72, '0.96', '16', 1], [1, 73, '0.92', '17', 1], [1, 68, '0.95', '17', 1], [1, 70, '0.76', '16', 0], [1, 87, '0.88', '19', 1], [1, 88, '0.94', '18', 1]]represents a set of student answer records, forming a two-dimensional array. Each element is a one-dimensional array representing a single answer record. In the one-dimensional array, the first element denotes the student ID, the second element represents the question ID, the third element indicates the overall class accuracy for this question, the fourth element indicates the knowledge point ID associated with the question. If a question involves multiple knowledge points, they are separated by commas, for example, '1,2' represents a question with knowledge points 1 and 2. The fifth element represents the student's answer, where 1 indicates a correct answer and 0 indicates an incorrect answer.\n**Requirement:** Now, please predict the answer for the following question: [1, 251, '0.93', '43', '?].Please fill in the question mark with a predicted value, and return the array.

图 2 Prompt 示例

如图 2 所示，论文将 Prompt 划分成了结构化的两个部分，第一部分介绍背景，首先指定了大模型的角色是一个优秀的成绩预测模型，然后给出某个学生的全部做题记录，做题记录用二维数组表示（如下表 1 所示）：二维数组中的一维数组代表学生的一条条做题记录，一维数组中的元素分别代表学生 ID，试题 ID，该题目在整个班上的正确率，该题目所包含的知识点 ID（有些题目可能包含多个知识点，用逗号分割表示），以及学生对该题的作答结果（1 表示学生答对该题，0 表示答错）；第二部分提出要求，

要求大模型基于以上背景去预测学生在一道新题目上的答题结果，并引导大模型生成期望的答案格式。

表 1 学生做题记录示例

记录序号	学生 ID	试题 ID	试题班级正确率	试题知识点 ID	答题结果
记录 1	21101	1	0.65	2	1
记录 2	21105	2	0.70	5	1
.....
记录 n	21199	55	0.75	3	0

3.2 Qwen 微调与推理

Qwen 大模型的微调训练数据需要满足一定的格式，一个字典对应一条训练数据，包含 id 和 conversations 两个属性，conversations 是一个列表，包含两个元素，第一个元素代表给大模型的输入，第二个元素代表微调时期望大模型给到的输出。

为了严谨对比使用大模型进行成绩预测的效果与其他成绩预测模型（例如 DINA^[18]、NeuralCDM）预测的效果，微调使用的训练数据需要完全从训练集中进行抽取。具体来说，对每个学生的答题记录，随机抽取答题记录总数的十分之一条记录作为微调训练数据中 assistant 的回答，未被抽取到的答题记录作为背景信息输入到大模型中，期望大模型能从学生作答数据中习得学生属性与试题属性的关联，最终正确预测新试题上的答题结果。

在微调时，可以采用全参数微调、LoRA 微调与 Q-LoRA 微调三种方式，选择预训练 Qwen 模型，设置好各种参数信息，即可进行微调，通过对比 4 种训练权重，最终选择 checkpoint-32 的训练权重。在大模型推理阶段，基于微调时选择的预训练 Qwen 模型加载训练权重，并给定推理数据集的路径，即可进行推理预测。

3.3 伪标签生成与筛选

针对部分学生历史答题记录数量偏少导致的认知诊断出现偏差问题，使用伪标签半监督学习来解决，对无标签数据集，也就是学生未答题记录设置伪标签，以增加学生答题记录，扩充训练集数据量。需要注意的是，由于模型预测存在误差，伪标签不一定是可靠的，所以需要确定一个置信度阈值，只有当模型对预测的标签非常有信心时，才将其视为伪标签，这有助于过滤掉模型不确定或不可靠的预测。为了进一步得到可靠的伪标签，论文对比筛选通过 Qwen 大模型以及 NeuralCDM 模型得到的预测结果。

参考图 1 所示，具体来说，通过初次训练好的 NeuralCDM 模型和微调好的 Qwen 大模型可以预测学

生未作答试题的答题结果，得到 $result_{NCDM}$ 和 $result_{Qwen}$ ， $result_{NCDM}$ 是一个 0 到 1 之间的概率值，论文将概率值大于等于 0.5 的预测结果表示为学生答对该题，用 1 表示，与之相反将概率值小于 0.5 的预测结果表示为学生答错该题，用 0 表示，其计算过程如公式 1 所示：

$$value = \begin{cases} 1, & p \geq 0.5 \\ 0, & p < 0.5 \end{cases} \quad (1)$$

为了得到高置信度的伪标签，论文将预测学生答对试题的置信度阈值设置为大于等于 0.9，预测学生答错试题的置信度阈值设置为小于等于 0.1，也就是当预测的概率值大于等于 0.9 或者小于等于 0.1 才保留该伪标签，否则舍弃，其计算过程如公式 2-2 所示：

$$result = \begin{cases} \text{retain}, & p \leq 0.1 \parallel p \geq 0.9 \\ \text{abandon}, & 0.1 < p < 0.9 \end{cases} \quad (2)$$

对 $result_{NCDM}$ 设置置信度阈值得到高置信度预测结果后，为了进一步得到高置信度伪标签，需要将 $result_{Qwen}$ 与 $result_{NCDM}$ 的高置信度预测结果对比，对于同一学生的同一试题，若两者的预测结果相同，则保留该条答题记录，若不同则舍去该条答题记录，得到最终的伪标签数据集。将原有的标记数据集与伪标签数据集合并，在新的数据集上重新训练 NCDM 模型。

4 实验设计及结果分析

4.1 实验准备

(1) 数据准备

论文所提出的方法将在真实教学环境中进行验证，使用的数据集来源于在线教育系统中《软件开发环境》课程的课堂测验内容，选择某班级共 126 名学生对 271 道试题的 17521 条答题记录（每章的测验是随机抽取一部分题目让学生作答，故每个学生都存在一些未作答试题）。其中试题来源于《软件开发环

境》课程的 13 章测验内容，试题类型都是选择题，答题结果 1 表示学生答对该题，答题结果 0 表示学生答错该题。

(2) 数据预处理

数据集的处理主要有以下四块内容：试题知识点标注、数据集划分、微调推理数据集生成、学生未作答试题数据集生成。

试题知识点标注：认知诊断模型的关键是需要构建一个试题与知识点间关联的 Q 矩阵，描述试题与知识点之间的包含关系。论文所使用的 271 道试题数据由领域内专家进行知识点标注，标注后共产生了 48 个知识点，试题与知识点的包含关系映射于 item.csv 文件，试题知识点映射关系标注实例如下表 2 所示。

数据集划分：为了模型训练、验证模型效果及测试模型泛化能力，需要将全部答题记录划分成训练集、验证集与测试集三个数据集。在数据集划分时，按章节进行随机且不放回的采样，确保在训练集、验证集与测试集上试题的知识点分布均匀，且按照 3:1:1 的比例划分成训练集、验证集与测试集。此外，由于原始的学生 ID 与试题 ID 被设置为唯一且随机的长字符串，不利于进行模型训练，需要将学生 ID 与试题 ID 映射到从 1 开始的整形数值。

微调推理数据集生成：微调数据集用于 Qwen 大模型进行 LoRA 或 Q-LoRA 微调训练，推理数据集用于 Qwen 大模型预测测试集中试题的答题结果。微调使用的训练数据需要完全从训练集中进行抽取，对每个学生的全部答题记录，随机抽取记录总数的十分之一条记录作为微调训练数据中 assistant 的回答，其他作为背景信息输入到大模型中。期望大模型推理预测的试题即测试集中的试题记录，并以学生为单位将训练集中学生的答题记录作为背景信息输入到大模型中。此外，输入给大模型的背景信息还包括试题难度属性，需要对训练集中的答题记录按试题计算答对的正确率。

表 2 试题知识点映射关系标注实例

试题 ID	题干	对应知识点	试题知识点 ID
1	第二代电子计算机使用的电子器件是 ()	计算机硬件发展史	[2]
2	Web 三要素中超文本技术 (HTML) 实现的功能是 ()	web 的三要素	[5]
3	标准的 URL 由 ()、主机名、路径及文件名组成	统一资源定位 URL	[7]
4	下面对 HTML 描述准确的一项是 ()	HTML 概述	[8]
5	align 属性用来标记 ()	HTML 的标记和元素	[10]
6	Tomcat 服务器的默认端口号是: ()	安装配置 JSP 运行环境	[6]
...

学生未作答试题数据集生成：虽然每章测试是随机抽取部分题目，但需要注意，若学生缺席了某次测验，则学生在该测验上的全部未作答试题不能包括在学生未作答试题数据集中，因为这代表着学生在该测验上缺少数据进行认知诊断。

4.2 实验设计及结果分析

（1）不同 Prompt、模型规模及微调策略进行大模型成绩预测对比实验

在本次的对比实验中，首先分别在三种不同的 Prompt 下使用四种不同的权重（Checkpoint-16、Checkpoint-32、Checkpoint-48、Checkpoint-64）进行成绩的预测，确定效果最好的 Prompt；其次在不同参数规模及不同微调策略下分别使用四种权重再次进行成绩的预测；最后选择预测效果最佳的 Prompt 和参数规模以及微调策略。

具体来说，首先进行 Prompt 的确定，分别对三种不同的 Prompt 进行验证对比成绩预测效果，统一使用模型大小为 7b 的 Qwen 大模型进行 Q-LoRA 微

调，划分得到 4 个阶段的训练权重，在这 4 个阶段的训练权重下去推理测试。

如表 3 所示，在同一权重下，添加试题难度属性并引入失误和猜测属性的 Prompt 在微调训练后的预测效果普遍比不添加试题难度属性和失误猜测属性的 Prompt 的预测效果好，只添加试题难度属性的 Prompt 又比添加试题难度属性并引入失误和猜测属性的 Prompt 的预测效果好。说明 Prompt 中试题难度属性的引入对大模型的成绩预测效果有显著提升，通过试题难度属性，大模型可以有效地获取试题属性信息。

确定好 Prompt 后，对不同参数规模的大模型进行不同微调策略，通过对比可以得知使用 14b 的大模型进行 Q-LoRA 微调得到的实验结果比使用 7b 或 1.8b 的大模型的实验结果都要好，并且在 Checkpoint-32 的权重下取到最好的预测效果。根据实验结果推断出大模型的参数规模越大，成绩预测的效果越好，这可能是因为更多的参数提供了更大的模型容量，使得模型能够更好地拟合训练数据，并学习到更复杂的模式和结构。

表 3 不同 Prompt 的成绩预测结果

不同 Prompt	Accuracy	AUC	Precision	Recall	F1
不添加试题难度属性和失误猜测属性	79.03%	53.61%	81.25%	92.48%	86.50%
添加试题难度属性	81.12%	59.11%	84.48%	94.10%	89.03%
添加试题难度属性并引入失误和猜测属性	80.31%	59.11%	84.52%	92.82%	88.47%

（2）大模型驱动认知诊断的成绩预测算法与已有认知诊断算法对比实验

对比实验主要将基于 Qwen 大模型的成绩预测算法、大模型驱动认知诊断的成绩预测算法与传统的 DINA 和 NeuralCDM 认知诊断算法进行对比，对比实验结果如下表 4 所示。

对比基于 Qwen 大模型的成绩预测算法与 IRT、

DINA 和 NeuralCDM 认知诊断算法，虽然 Accuracy 值有少量提高，但 AUC 值明显下降。通过对比 Recall 值的提高及 Precision 值的下降，可知是由于 Qwen 大模型更倾向于将样本预测为正例，从而导致召回率提高，同时由于预测为正例的样本出现很多误报，因此精确率降低。所以相比 NeuralCDM 模型，Qwen 大模型在正例上的预测准确率提高了，但在负例上的预测表现不够理想。

表 4 成绩预测算法对比实验结果

	Accuracy	AUC	Precision	Recall	F1
IRT	81.73%	75.48%	83.46%	89.98%	86.59%
DINA	70.56%	66.72%	83.03%	80.26%	81.62%
NeuralCDM	82.39%	75.07%	85.78%	93.11%	89.29%
基于 Qwen 成绩预测	83.03%	60.62%	83.84%	98.05%	90.38%
本文方法	83.95%	77.03%	85.84%	96.15%	90.70%

基于以上问题，本章结合了 Qwen 大模型与 NeuralCDM 模型进行算法改进，提出大模型驱动认知诊断的成绩预测算法，可以有效提高预测的准确率和 AUC 值，算法在预测正例和负例上的表现结果都有明显提高。实验结果显示 Accuracy 值达到了 83.95%，相比普通 NeuralCDM 算法提高了 1.5 个点，AUC 值达

到了 77.03%，提高了近 2 个点，且在 Precision、Recall 和 F1 上都有不同程度的提高。

5 教学改革与效果分析

论文使用的数据集来源于在线教育系统中“软件

开发环境”课程的课堂测验内容，选择某班级共 126 名学生对 271 道试题的 17521 条答题记录，图 3 所示为系统答题界面。

实验结果表明，该方法优于传统成绩预测方法，在论文数据集上的准确率、AUC 值及 F1 值分别提升了约 1.5%、2%及 1.3%。

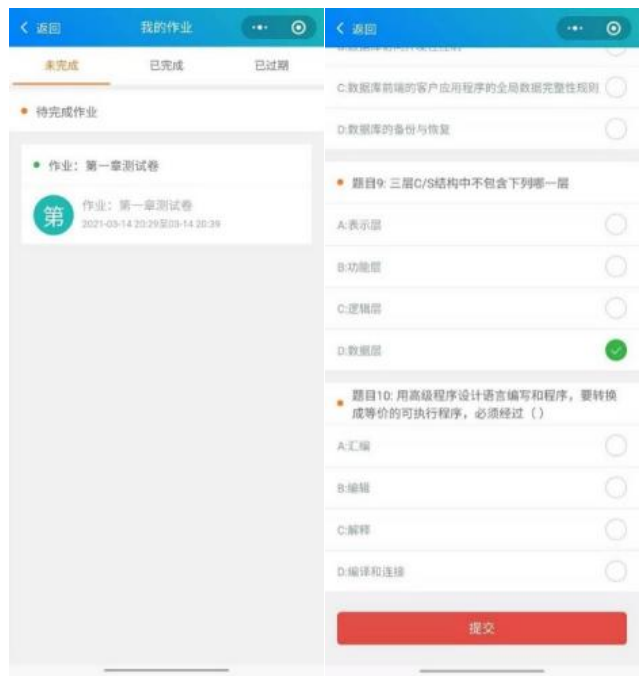


图 3 在线教育系统课堂测验界面

结合“软件开发环境”课程的实际教学应用场景，认知诊断功能模块分为知识点诊断和章节诊断两部分。知识点诊断是学生对全部知识点的诊断结果，章节诊断是学生对课程所有章节的诊断结果。在这个过程中，不仅综合评估学生认知能力的强弱，而且以可视化的形式展现出所有选课学生对于知识点和章节掌握情况（如图 4 所示）。



图 4 认知诊断可视化实践案例

根据认知诊断结果以及学生的历史答题数据，可以获得每位学生对于新试题的成绩预测结果表，从而获得学生易错题等统计数据，如图 5 所示。这有助于教育者充分掌握学生的学习和认知情况，在此基础上对试题难度进行适当的调整，筛选出难度适中又能反映学生学习效果的试题。



图 5 成绩预测与统计分析界面

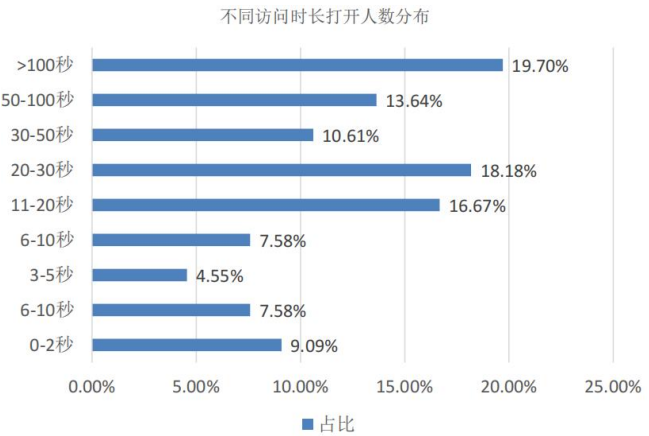


图 6 不同访问时长打开人数分布

在高校教学改革中，通过使用大模型驱动认知诊断的方法进行成绩预测，可以提高教育者对学生成绩的全面掌握，及时优化教学策略与资源配置，如利用大模型驱动的认知诊断技术，为学生单独组卷，自动筛选出检测其薄弱点的定制试题，精准评估学习效果。根据图 6 不同访问时长打开人数分布来看，有 62.13% 的打开人数的访问时长大于 20 秒，且每次使用中，访问时长集中分布于 20 到 100 秒之间，说明本系统的访客留存能力较强，个性化教育的提升效果较好。根据认知诊断和成绩预测的结果，在教学过

程中针对性地巩固强化学生薄弱的知识点,使学生专注于需要加强的内容,提高学习效率。这些教育改革应用方向可以有效促进个性化教育,确保每个学生都能在适合自己的节奏和方式下获得最佳的学习成果,从而推动整体教学质量的提升。

6 结束语

大模型作为新兴的智能技术,具备强大的数据处理和分析能力,可以为教育决策提供数据支撑和科学依据。论文提出了大模型驱动认知诊断的成绩预测算法,使用大语言模型的微调技术和问答机制进行成绩预测,从而推动个性化教学并完善教学评价体系。这种评价体系不仅关注学生的学习结果,还关注学生的学习过程和学习态度,有助于促进学生的全面发展。

参考文献

- [1] 习近平.高举中国特色社会主义伟大旗帜为全面建设社会主义现代化国家而团结奋斗[N].人民日报 2022-10-26(001).
- [2] 余海宁.人工智能赋能国家治理现代化的前景、挑战与优化[J].领导科学论坛,2024,(06):75-79.
- [3] 尹西明,苏雅欣,陈泰伦,等.场景驱动型人工智能创新生态系统:逻辑与进路[J].中国科技论坛,2024,(06):35-45.
- [4] 王兆瑞,武雯宇.人工智能时代教育工作的定位探究[J].郑州师范教育,2024,13(04):60-64.
- [5] 徐悦,黄子文,宋雨轩,等.从AI大模型看高校计算机教育面临的机遇与挑战[J].计算机技术与教育学报,2024,12(03):99-106.
- [6] 孙争艳,陈磊,陈宝国.生成式人工智能在计算机通识教育中的价值定位、挑战与实施路径[J].计算机技术与教育学报,2025,13(02):115-119.
- [7] DiBello L V, Roussos L A, Stout W. 31a review of cognitively diagnostic assessment and a summary of psychometric models[J]. Handbook of statistics, 2006, 26: 979-1030.
- [8] Templin J L, Henson R A. Measurement of psychological disorders using cognitive diagnosis models[J]. Psychological methods, 2006, 11(3): 287.
- [9] Gilhooly K J. Cognitive psychology and medical diagnosis[J]. Applied cognitive psychology, 1990, 4(4): 261-272.
- [10] Duric Z, Gray W D, Heishman R, et al. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction[J]. Proceedings of the IEEE, 2002, 90(7): 1272-1289.
- [11] Leighton, Jacqueline, and Mark Gierl, eds. Cognitive diagnostic assessment for education: Theory and applications[M]. Cambridge University Press, 2007.
- [12] Wang F, Liu Q, Chen E, et al. Neural cognitive diagnosis for intelligent education systems[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(04): 6153-6161.
- [13] 胡心颖,何钰,孙广中.基于概率图模型的计算机课程教学认知诊断框架[J].中国科学技术大学学报, 2021,51(1):12-21
- [14] 马华,黄卓轩,唐文胜.面向个性化学习的认知诊断模型及其应用综述[J].计算机技术与发展, 2021,31(11):35-40,45
- [15] 王炼红,帅智康,许加柱,等.基于模糊认知诊断模型的慕课学习成效评估[J].电气电子教学学报, 2022,44(5):108-113
- [16] Wang F, Liu Q, Chen E, et al. Neural cognitive diagnosis for intelligent education systems[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(04): 6153-6161.
- [17] Bai J, Bai S, Chu Y, et al. Qwen technical report[J]. arXiv preprint arXiv:2309.16609, 2023.
- [18] Junker B W, Sijtsma K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory[J]. Applied Psychological Measurement, 2001, 25(3): 258-272.