

数据驱动的学习者画像：认知诊断与自组织映射模型的融合研究

易艳坤¹ 吴宪君^{2*}

1 公诚管理咨询有限公司 2 广东石油化工学院 计算机学院

DOI:10.12238/acair.v3i2.13558

[摘要] 本文聚焦于数据驱动的探索认知诊断与自组织映射模型的融合创新应用,提出了结合认知诊断与自组织映射模型的融合模型。模型尝试了认知诊断技术识别学习者在数据处理、算法设计、模型训练与评估等数字素养方面的能力水平,并借助自组织映射模型的强大聚类分析能力,对学习者的学习兴趣和积极性,提升教学效果与学习效率。实验结果表明,融合模型在学习者能力评估与个性化教学推荐方面表现优良,为教学创新、数字素养培养及评价体系完善提供了有力支持。

[关键词] 认知诊断模型; 自组织映射; 数据驱动学习分类

中图分类号: C37 文献标识码: A

Data driven learner profiling: a fusion study of cognitive diagnosis and self-organizing mapping models

Yankun Yi¹ Xianjun Wu^{2*}

1 Gongcheng Management Consulting Co., Ltd

2 School of Computer Science, Guangdong University of Petrochemical Technology

[Abstract] This article focuses on the innovative application of data-driven exploration in the fusion of cognitive diagnosis and self-organizing mapping models, and proposes a hybrid model that integrates cognitive diagnosis and self-organizing mapping models. The model attempts cognitive diagnostic techniques to identify learners' abilities in digital literacy such as data processing, algorithm design, model training, and evaluation, and utilizes the powerful clustering analysis capabilities of self-organizing mapping models to accurately classify learners. By classifying the results, customized teaching content and learning path recommendations are provided for learners with different abilities, fully stimulating their learning interest and enthusiasm, and improving teaching effectiveness and learning efficiency. The experimental results show that the integrated model performs well in learner ability assessment and personalized teaching recommendation, providing strong support for teaching innovation, digital literacy cultivation, and evaluation system improvement.

[Key words] cognitive diagnostic model; Self-organizing mapping; Data driven learning classification

引言

数字化浪潮席卷全球,人工智能技术正以势不可挡之势重塑各个领域的发展格局,教育进入全方位的变革与升级,以适应未来人才的迫切需求。总书记在中共中央政治局二十次集体学习时强调“要全面推进人工智能科技创新、产业发展和赋能应用”^[1]。在教育领域落实这一重要指示,意味着我们需要以更宏观的视野和更具创新性的思维去探索教育的新形态。而数据驱动的学习分类正是实现这一目标的关键路径之一,它借助先进的技术手段深入挖掘教育数据中蕴含的巨大价值,为精准教学、个性化学习提供有力支撑。

认知诊断模型和自组织映射模型作为新兴的教育技术手段,

在提升教学效果和学习者素养方面发挥着关键作用。认知诊断模型能够精准剖析学习者的学习状态,深入洞察其知识结构中的优势与不足;自组织映射模型则凭借卓越的数据聚类分析能力,清晰勾勒出学习者能力分布的全貌,为教学策略的优化提供有力依据。例如,梁存良等学者(2024)基于自组织映射(SOM)神经网络构建教学认知诊断模型,有效解决了知识点关联松散学科的认知诊断难题,显著提升了教学效率和学习者学习效果^[2]。

为精准评估学习者在课程各方面的能力和素养,实现分类教学与个性化内容推荐。研究主要探索了认知诊断与自组织映射模型在课程教学中的融合应用。通过构建融合模型分析学习

者在数据处理能力、算法设计、模型训练与评估等数字素养方面的能力水平。一方面,利用认知诊断模型识别学习者优势与不足;另一方面,借助自组织映射模型聚类分析,清晰呈现学习者能力分布。依据评估分类结果,为不同层次学习者量身定制教学内容与学习路径,充分激发学习者学习兴趣,提升教学效果与学习效率,推动教育智能化发展。

1 模型相关理论基础

1.1 认知诊断模型

认知诊断模型(Cognitive Diagnosis Model, CDM)是基于心理测量学和认知心理学的教育评估工具,通过分析学习者作答模式,精准识别其知识结构和认知技能。其核心是将学习者潜在认知状态与外显作答行为联系,利用统计模型分类诊断。认知诊断模型始于20世纪80年代,早期有规则空间模型(RSM),后发展出广义DINA模型(G-DINA),可处理复杂的认知属性关系和作答数据^[3]。

在教育领域,认知诊断模型应用广泛,成效显著。例如,马晓梅等学者基于《中国英语能力等级量表》构建英语阅读推理认知诊断测评模型,采用G-DINA模型对1083名学习者进行诊断分析^[4],结果表明其诊断信度高、拟合良好,能生成个性化成绩报告,助力学习者后续学习。认知诊断模型还广泛应用于数学、物理等学科,为教师调整教学方案、实施个性化教学提供支持。

认知诊断模型适用于知识点逻辑关系紧密的学科课程教学,如数学、物理等,但在知识点关联松散的学科中应用存在局限性。如英语、语文等文科类课程,由于知识点关联性弱,传统认知诊断模型难以精准构建属性层级关系和Q矩阵,诊断结果准确性受影响。梁存良等学者指出,认知诊断模型的诊断效率和准确性与知识点依赖关系密切相关,在处理知识点依赖关系松散型学科时,易因生成的可达矩阵稀疏而导致诊断效率和准确度降低^[2]。此外,其构建实施通常需要大量高质量数据支持,这也在一定程度上限制了其应用范围。

1.2 自组织映射模型

自组织映射(Self-Organizing Maps, SOM)模型是一种无监督学习的神经网络模型,由芬兰学者Teuvo Kohonen于20世纪80年代提出^[5]。其原理基于竞争学习机制,通过比较输入向量与神经元权值向量,找到最相似的“获胜神经元”,并调整其及邻域神经元的权值以贴近输入向量,经迭代后形成稳定聚类结果。在数据分类和聚类分析中,SOM模型具有独特优势。它能自动发现数据内在结构和模式,无需预标注数据,还可将高维数据映射到低维空间,便于可视化和解释。但该模型对初始权值敏感,训练易陷入局部最优,聚类结果依赖网络结构参数选择。

在教育领域,SOM模型的应用展现了巨大价值。例如,将学习者学习数据输入SOM模型,可自动分类出具有相似学习特征的学习者群体,助力教师了解学习者个体差异,提供个性化教学支持和学习资源推荐。梁存良等学者(2024)的研究表明,基于SOM神经网络的教学认知诊断模型,能有效提升知识点关联松散型学科的教学效率和学习效果,其在教育领域的应用前景广阔^[2]。

1.3 融合架构

融合模型整合认知诊断模型与自组织映射(SOM)模型优势,构建用于学习者画像的融合模型架构。首先利用认知诊断模型精准识别学习者知识结构和技能水平,并通过该模型确定关键知识属性。再根据SOM模型擅长高维数据聚类分析,用其进行聚类分析。二者结合形成互补。具体考虑模型输入为学习者学习数据,如测验成绩、作业完成情况等。中间处理先用认知诊断模型分析数据生成知识属性掌握情况报告,再将报告输入SOM模型聚类分析。输出为学习者知识状态分类结果及可视化知识图谱。模型训练时,先用最大似然估计法训练认知诊断模型参数,再用无监督学习训练SOM模型。参数优化采用梯度下降法调整认知诊断模型参数,提升诊断准确性;用学习率和邻居函数优化SOM模型,增强聚类效果。

2 学习者画像融合模型构建

为在计算机类基础课中构建学习者画像,本研究融合了认知诊断模型精准识别学习者知识结构和技能水平优势,以及自组织映射(SOM)模型高维数据聚类分析能力,构建了融合模型架构。具体而言,先用认知诊断模型确定关键知识属性,再用SOM模型聚类分析。

输入学习者测验成绩、作业完成情况等;中间处理先用认知诊断模型分析生成知识属性掌握报告,再将报告输入SOM模型聚类分析;输出为学习者知识状态分类结果及可视化知识图谱。直观呈现学习者知识掌握情况。

模型训练过程步骤,先用最大似然估计法训练认知诊断模型参数,再用无监督学习训练SOM模型。参数优化采用梯度下降法调整认知诊断模型参数,提升诊断准确性;用学习率和邻居函数优化SOM模型,增强聚类效果。

在模型评估指标上充分考虑合理性。紧扣课程目标与教学需求,精准锚定评估学习能力和素养的若干关键指标。在编程基础能力板块,考量学习者语法、数据类型等核心要素的熟练度,借助编程作业里的代码规范性、错误率等量化数据,全方位勾勒学习者编程基本功的深浅;算法设计与实现能力方面,结合课程项目实践,围绕学习者编写的算法在解决实际问题场景下的准确性、效率表现,以及面对复杂任务时的逻辑架构合理性,多维度衡量其算法思维的敏锐度与实操水平。针对数据处理与分析这一关键维度,依据学习者在数据预处理环节的精准操作、对复杂数据集的深度分析与可视化呈现成果,精准把握其从原始数据中提炼价值信息的能力层级。模型训练与评估能力聚焦于学习者运用工具搭建、调优模型,以及严谨评估模型性能的专业素养。

在分类标准问题上充分考虑了科学性。认知诊断理论深挖认知诊断模型精准洞察的每个学习者知识属性掌握详情,再凭借自组织映射模型的聚类,精准锚定具有相似能力图谱的学习者群体。依据关键指标的不同权重与组合,构建出细致入微的分类标尺,为每层能力阶梯定义清晰且契合教学实际的门槛。

3 基于融合模型的教学应用

3.1 实验对象与数据处理

为验证融合模型可行性,研究依托项目,具体采集数据和测试模式应用。以“MATLAB基础与应用”课程为场景,选取笔者高校理工科(应化、高分子等)专业参与该课程学习的学习者作为研究对象。通常每类专业中设有两个教学班,每个班级约30名学生。1班作为实验组,采用融合认知诊断与自组织映射(SOM)模型的教学方法;2班作为对照组,采用传统教学方法。两组学生在年龄、性别比例和先修课程成绩上基本一致,确保实验的公平性。

以“MATLAB基础与应用”课程为教学内容,涵盖MATLAB编程基础、数据处理、算法设计、模型训练与评估等知识点。使用通用型在线教学辅助系统作为教学工具记录学生的学习行为数据。平台支持认知诊断测验、诊断报告生成、个性化学习资源推送等功能。应用过程贯穿主要教学周期并同步处理采集的测试数据。数据主要包括课前问卷调查,收集学习者MATLAB基础、学习偏好等背景信息;课程测验与课后作业完成情况记录学习过程数据,了解知识掌握与应用能力;线上学习平台跟踪学习者学习时长、参与度等行为数据,反映学习状态。利用在线问卷平台、教学管理系统以及语言定制脚本采集数据。问卷调查预测经优化,确定问题保持一定的有效性,测验题目和作业任务由教师设计确保与教学内容匹配;线上学习行为数据由专人负责采集后,借助数据清洗等技术手段处理缺失、异常数据,提升质量,保障研究结论可靠性与有效性。

3.2 实验与结果分析

基于认知诊断与自组织映射(SOM)模型的融合模型在处理知识点关联松散型学科时,能够获得较为准确的诊断结果。属于实验组的学生评价分布具体是,67.25%认为诊断准确率较高(80%-100%),16.45%认为准确率在60%-80%之间,16.30%认为准确率低于60%。

在学习效果方面,实验组和对照组学生进行了课程前测试,结果显示两组学生的初始知识水平相当($\text{Sig.}=0.877>0.05$),这为后续的实验提供了公平的起点。课程结束后,两组学生进行了后测。实验组学生的平均成绩为82.48分,而对照组为79.31分,实验组的成绩提升更为显著($\text{Sig.}=0.000<0.05$)。这一显著差异表明,融合模型能够有效支持个性化补救学习,显著提高学生的学习效果。这种提升不仅体现在总体成绩上,还体现在学生对课程各个模块的掌握程度上。例如,在数据处理模块,实验组学生的平均成绩从70分提高到85分,而对照组仅从68分提高到75分。这一结果进一步证明了融合模型在个性化学习路径推荐方面的优势。

通过SOM模型对学习者的聚类分析,生成了5个类别,每个类别对应不同的知识状态和学习需求。例如,类别1主要包含基础薄弱的学生,而类别5主要包含掌握较好的学生。这种分类结果为个性化教学提供了明确的方向,使得教师能够针对不同层次的学生提供定制化的教学内容和学习路径推荐。实验组学生根据诊断报告和推送资源开展个性化补救学习后,其在课程各

个模块的掌握程度显著提高。实验组教学数据显示,分类教学活动不仅提升了学生的学习效率,还激发了他们的学习兴趣和积极性,为教学效果的提升提供了有力支持。

3.3 结果讨论与分析

实验结果显示,融合模型在诊断准确率、学习效果提升和个性化推荐方面表现出色。通过认知诊断模型精准识别学习者知识结构和技能水平,结合SOM模型的聚类分析能力,能够为不同层次的学习者提供定制化的教学内容与学习路径推荐。这种融合模型的应用,不仅充分激发了学习者的学习兴趣与积极性,还显著提升了教学效果与学习效率。但是在实验过程中,我们也发现了一些不足之处。部分学生对认知诊断系统的操作不够熟练,这在一定程度上影响了诊断报告的生成效率。SOM模型的聚类结果对初始权值较为敏感,训练过程中易陷入局部最优,导致部分分类结果不够理想。这些问题提示我们在后续的研究中需要进一步优化模型参数,增强系统的易用性,并拓展模型的应用场景。

4 总结和展望

模型的构建和实施需要较高的技术支持,包括数据采集、处理和分析等环节。此外,模型的优化空间较大,例如在处理大规模数据集时,模型的训练时间较长,影响了教学实践的实时性。后续的研究中需要进一步优化模型参数,例如通过调整模型的初始权值和学习率,减少模型训练过程中的局部最优问题增强系统的易用性,并拓展模型的应用场景。

[基金项目]

基于学习者画像与AI素养的程序语言类课程改革实践(广东省教育科学规划课题高等教育专项2024GXJK560),新质生产力背景下数字素养提升: MATLAB基础与应用教学改革实践(2024年广东石油化工学院校级教改)。

[参考文献]

- [1]新华社.习近平在中共中央政治局第二十次集体学习时强调:坚持自立自强突出应用导向推动人工智能健康有序发展[OL]. <https://www.gov.cn/yaowen/liebiao/202504/content_7021072.htm>.
- [2]梁存良,张玥,黄宏涛,等.基于SOM神经网络的教学认知诊断模型研究[J].现代教育技术,2024,34(09):59-70.
- [3]Jiayuan,Y.(2024).Cognitive Diagnosis.In:The ECPH Encyclopedia of Psychology.Springer,Singapore.
- [4]马晓梅,杜文博.基于《量表》的英语阅读推理能力认知诊断模型构建与成绩报告[J].中国考试,2022,(12):1-9.
- [5]Kohonen,T.(1982).Self-organized formation of topologically correct feature maps. Biological Cybernetics,43(1),59-69.

作者简介:

易艳坤(1994--),女,汉族,广东省茂名市人,本科,从事的研究方向或工作领域:数据共享与治理、大数据挖掘分析及应用,公共数据的开发利用等。

*通讯作者:

吴宪君,硕士,副教授。