

基于引导式精通的协同设计支持机器学习赋能中学科学探究

A Guided Mastery Co-Design Approach Supporting Secondary School Teachers in Designing ML-Supported Scientific Inquiries

周晓菲¹

¹ 上海交通大学教育学院

* zhouxf626@sjtu.edu.cn

【摘要】 在大数据与人工智能时代，培养数据与计算素养对 K-12 教育至关重要。本研究旨在通过基于引导式精通理论的协同设计，赋能教师创建机器学习支持的科学探究学习活动，并提升其教学动机。我们开展了两项研究。首先，14 名中学教师与机器学习专家合作创建了 26 个涵盖生物、历史等学科的探究活动；10 名高中生对这些活动进行了学习评估。实证显示教师能够成功将聚类和分类等 ML 实践与科学探究行为（如提出假设、模式解释）深度整合；通过协同设计，教师 AI 教学自信与动机和学生的机器学习知识均显著提升，并支持了科学探究学习。本研究表明，基于引导式精通的协同设计为非专业教师提供递进式挑战与结构化支持，是提升教师 AI 教学素养及学生数据驱动探究能力的有效路径。

【关键词】 人工智能；机器学习；科学探究；数据分析；协同设计；引导式精通

1. 前言

在大数据与人工智能（artificial intelligence, AI）时代，培养数据素养与计算素养已成为使下一代在以数据为中心的世界中蓬勃发展的关键 [1]。诸如聚类和分类等机器学习（machine learning, ML）技术能够揭示大型数据集中以往难以处理的模式，从而加速了数据驱动的科学探究（scientific inquiry, SI）进程 [2]。尽管 AI4K12、CSTA 和 AI4ALL 等倡议已针对 K-12 阶段的 ML 制定了国家级指南（涵盖从高阶训练流水线到现实模型评估的各个环节），但受限于在职教师有限的时间和技术专长，目前仍缺乏将 ML 整合进 K-12 课程的跨学科资源；同时教师在教授 AI 相关内容时缺乏信心 [3]。在利用 ML 进行数据处理、模式发现和意义识别之前，统计学领域的高知识门槛往往令 K-12 教师对 AI/ML 产生畏难情绪 [4]。研究表明，即使学习者掌握了 AI 知识，其 AI 相关动机也未必能得到改善，这表明提升 AI 相关的学习动机或自我效能感仍是当前面临的主要挑战 [5]。

因此，本研究旨在提出一种协同设计（Co-design）方法，赋能教师为其课堂创建 ML 支持的科学探究（ML-SI）学习活动，并提升教师在 AI 相关教学中的动力。科学探究学习（或称探究式学习）[6] 是一种常见的 K-12 教学法，鼓励学生采取类似于专业科学家的实践来进行知识构建。本设计方案受到引导式精通（Guided Mastery）理论的启发，并基于一项针对 18 名在职 K-12 教师的形成性设计研究，通过让他们为其课堂创建 ML-SI 活动的概念设计来实施。引导式精通通过为学习者提供具有挑战性的递进任务，辅以结构化支持和反馈，使其能够在复杂技能上逐步建立能力和信心 [7]。

我们开展了两项研究。研究 1 中，我们邀请了 14 名中学教师，通过与 AI 专家协同设计，创建了符合其教学兴趣的 ML-SI 活动，专家负责确保 ML 应用的有效性。研究 2 中，我们邀请了 10 名高中生，对教师设计的 ML-SI 活动进行了评估。本研究探讨了三个研究问题：RQ1.

教师如何将 ML 整合到符合课程大纲的科学探究活动中？（研究 1）；RQ2. ML-SI 活动能否增强学生的 ML 知识并支持有意义的科学探究学习？（研究 2）

2. 研究 1：与中学教师的协同设计

2.1 实验设计

通过教育相关的邮件列表和教师教育项目协调员，我们招募了 14 名中学 STEM 教师。本研究已获得机构审查委员会（IRB）的批准。中学教师与 AI 专家协同设计了 ML-SI 学习活动。该研究由三次一小时的环节组成，通过 Zoom 在线进行。

环节 1：教师作为学习者（Teacher-as-Learner）在本环节开始前，教师填写了包含三个部分的预研究调查问卷：(1) 人口统计信息和教学背景；(2) 在编程、ML、数据可视化和分析方面的过往经验；(3) 在课堂教学中引入 ML 概念和方法的动机。我们采用了科学动机问卷 [8] 来评估不同的动机成分，包括自我效能感、内在价值和职业动机。首先，教师与研究者讨论了教师希望在设计环节中使用哪些数据集。随后，研究者通过剔除缺失值和非数值特征对新数据集进行预处理，以确保数据集与 ML 算法的兼容性。

环节 2 & 3：教师作为设计者（Teacher-as-Designer）在两次设计环节中，教师首先对他们选择的两个数据集进行探索性分析。研究者和 AI 专家在需要时提供协助。教师在每个步骤中提出他们感兴趣的问题，根据数据可视化结果回答询问，并规划后续行动。在每一步之后，研究者都会提示教师考虑是否需要 ML 和科学探究组件的配对进行调整。完成后，教师被要求回顾其设计过程，并思考对其创建的学习活动进行修改。我们收集了设计环节的 Zoom 屏幕录像，以及教师与 Web 应用程序交互行为和文本输入的日志。最后一次设计环节结束后，教师完成了评估整合 ML 教学动机的后测问卷。

2.2 研究结果 (RQ1. 教师如何将 ML 整合到符合课程要求的科学探究活动中？)

在教师设计的 26 个 ML-SI 学习活动中，15 个使用了 K-means 聚类后接 KNN 的模式，9 个仅使用 K-means 聚类，1 个在聚类前应用了 KNN，还有 1 个先后使用了 K-means 聚类、KNN，最后再次进行聚类。在教师设计的活动中，按使用频率从高到低排列：ML 算法调试元素依次为特征选择 (Mean = 2.35, SD = 2.29)、参数调整 (Mean = 2.15, SD = 2.71)、更新输入数据 (Mean = 0.81, SD = 1.07)；ML 评估与分析元素依次为自动聚类 (Mean = 2.12, SD = 1.45)、KNN 预测 (Mean = 2.12, SD = 3.29)；ML 探索元素依次为读取属性 (Mean = 0.46, SD = 0.50)、手动聚类 (Mean = 0.35, SD = 0.83)；科学探究学习行为依次为假设迭代 (Mean = 2.69, SD = 2.07)、提问 (Mean = 1.46, SD = 1.45)、模式解释 (Mean = 1.27, SD = 1.26)、结论 (Mean = 1.08, SD = 0.47)。

3. 研究 2：学生评估研究

3.1 实验设计

我们根据机构审查委员会（IRB）批准的方案，对高中生进行了评估研究。在一个共有 14 名学生的夏季项目中，研究人员亲自向学生说明了研究程序，并获得了 10 名学生的知情同意及其家长的许可。未报名参加研究的学生完成了相同的学习活动，但未采集其数据。研究在校内计算机实验室进行，连续两天，每天约 2 小时。研究人员为计算机标记了 P1-P10 的编号。所有学生在研究环节中均独立工作。

第一天：ML-SI 活动简介学生完成了 15 分钟的预调查，以提供人口统计信息并评估其 ML 知识。随后，他们体验了 ML-SI 活动中所有的 ML 支持数据实践，这遵循了逐步揭示 ML-SI 学习复杂性的设计考量 [9]。此过程侧重于培养 ML 知识：包括 ML 概念和相关数据实践，同时尽量减少因教师设计的活动差异而产生的潜在偏差。最后，学生完成了包含相同问题的后测，以评估其 ML 知识。

第二天:教师设计的 ML-SI 学习学生体验了由研究 1 中的两名教师设计的两个 ML-SI 活动。第一个活动使用了一个关于蓝莓产量及其相关生态特征的环境科学数据集;学习目标是确定影响美加地区蓝莓总产量的因素。另一个活动使用了一个收入数据集,邀请学生调查不同的个人因素如何影响收入水平。系统后台记录了学生与 ML 及科学探究组件的交互行为,以及他们在科学探究组件中的文本响应。

3.2 测量指标与数据分析

研究 2 的核心研究问题是 RQ3: ML-SI 活动能否增强学生的 ML 知识,并且支持学生进行有意义的科学探究学习?针对 ML 的学习增益,我们开展了预测试和后测试,以评估学生的 ML 知识(如多维相似度、聚类、质心、KNN)和相关数据分析技能(如聚类分析、KNN 解释)。测试题目基于 K-means 和 KNN 概念 [10,11],每匹配到一个关键词给 1 分,满分为 3 分,最终得分由两名评分员的平均值决定。由于数据不符合正态分布(Shapiro-Wilk 检验),我们应用了 Wilcoxon 符号秩检验进行配对比较,并在统计量为零时使用 Cliff's Delta 确定效应量。

学生应用 ML 进行科学探究学习的表现通过以下两个方面衡量:(1) 学生是否正确分析了 ML 的输出;(2) 学生是否基于 ML 实践的发现生成并迭代了有意义的假设。首先,为了评估学生对 ML 输出的分析,我们将学生在“数据分析”中的输入与教师预期的标准模式列表进行了比较。评分规则如下:(1) 错误答案计 0 分;(2) 每个正确识别的模式计 1 分,最高 3 分。其次,为了评估学生的假设迭代,我们分析了科学探究组件“假设生成”中的响应,以评估学生是否生成了与数据属性和关系相符的有意义假设。通过“结论”组件,我们评估了学生是否通过与发现结果进行对比,正确评价了他们的原假设。

3.3 研究结果 (RQ2. ML-SI 活动能否增强学生的 ML 知识并支持有意义的科学探究学习?)

3.3.1. ML 的学习增益

预测试和后测试结果(表 1)表明,ML-SI 实践有效地增强了学生的 ML 知识,包括他们对多维相似度计算和质心等概念的理解,以及聚类分析和 KNN 解释等相关数据实践。除了关于多维相似度计算的第一题外,所有测试项均显示出较大的效应量。

表 1 机器学习知识的前后测结果(显著性标记为 $p < 0.05^*$ 和 $p < 0.01^{**}$)

(1) Target concept and assessment question (scores range 0-3)	Mean _{Pre} (SD)	Mean _{Post} (SD)	Statistic	P value	Effect Size
Multidimensional similarity computation: What makes two multidimensional data points similar or dissimilar?	0.20 (0.60)	1.40 (1.20)	0.000	0.038*	0.120
Clustering: What does it mean to cluster a dataset?	0.60 (1.02)	2.70 (0.46)	0.000	0.007**	0.740
Centroid: What is the centroid of a cluster of data points?	0 (0)	2.30 (1.00)	0.000	0.006**	0.800
Cluster analysis: What patterns can we pay attention to when we look at the clustering result?	0.20 (0.60)	2.60 (0.92)	0.000	0.005**	0.780
KNN algorithm: Briefly describe how the KNN algorithm predicts an unknown data point.	0 (0)	2.40 (0.66)	0.000	0.002**	1.000
KNN interpretation: What does it mean the KNN algorithm makes a very different prediction from the actual value (the truth)?	0 (0)	2.20 (0.87)	0.000	0.002**	1.000

3.3.1. 学生应用机器学习进行科学探究学习的表现

该评估通过学生使用生物学和社会科学数据集开展的两个 ML-SI 活动进行。在关于蓝莓产量影响因素的第一个活动中,学生的表现分为三步评估:(1) K-means 聚类的数据分析;(2) KNN 预测结果的解释;(3) 假设更新。首先,10 名学生中有 9 名在聚类分析中获得了满分,他们识别出高产量组具有更高的果实质量和更小的克隆尺寸;P1 因遗漏一个模式被扣除 1 分。其次,8 名学生充分解释了 KNN 结果,指出果实质量和克隆尺寸的预测能力比大黄蜂密度更强;P4 和 P7 仅提到准确率较低表示相关性较弱,缺乏具体语境细节。最后,8 名学生正确利用聚类和 KNN 的发现更新了他们的假设;P1 和 P5 未能充分综合这些模式。在关于收入水平影响因素的第二个活动中,评估方式类似:(1) 基于 KNN 结果生成假设;(2) 聚类结果的数据分析;(3) 假设更新。首先,8 名学生根据 KNN 预测生成了全面

的假设；P3 和 P5 仅注意到高学历与高收入相关，忽略了工作时长。其次，所有 10 名学生均成功识别了类内和类间模式。最后，8 名学生基于综合见解迭代了假设；然而，P1 未能综合模式得出明确结论，P5 忽略了年龄和工作时长的影响。

4. 结论

本研究探讨了 K-12 教师在将 ML 整合到课程中所面临的挑战，这些挑战主要源于他们在统计学和 ML 领域的专业知识有限。我们的研究表明，ML 支持的科学探究活动能够有效支持学生的学习与意义建构。研究结果阐明了教师如何利用 ML 实践来增强科学探究学习，并指出未来的工具和协作社区可以进一步支持初学者的 ML 实践，从而为学生应对人工智能驱动的未来做好准备。

参考文献

- [1] Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of science education and technology*, 25(1), 127-147.
- [2] Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, 346(6206), 171-172.
- [3] Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2023). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*, 28(5), 5967-5997.
- [4] Marques, L. S., Gresse von Wangenheim, C., & Hauck, J. C. (2020). Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education*, 19(2), 283-321.
- [5] Kaspersen, M. H., Musaeus, L. H., Bilstrup, K. E. K., Petersen, M. G., Iversen, O. S., Dindler, C., & Dalsgaard, P. (2024, May). From primary education to premium workforce: Drawing on K-12 approaches for developing AI literacy. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).
- [6] Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., ... & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational research review*, 14, 47-61.
- [7] Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.
- [8] Glynn, S. M., Brickman, P., Armstrong, N., & Taasobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of research in science teaching*, 48(10), 1159-1176.
- [9] Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).
- [10] Mining, W. I. D. (2006). *Introduction to data mining*. Mining Multimedia Databases, Mining Time Series and.
- [11] Steinbach, M., & Tan, P. N. (2009). kNN: k-nearest neighbors. In *The top ten algorithms in data mining* (pp. 165-176). Chapman and Hall/CRC.