

C7: 學習分析與學習評估

跨越隐性鸿沟：基于可解释机器学习的数字资源对学业表现的预测关联及算法

公平性研究 ——以 PISA 2022 新加坡与台湾数据为例

Bridging the Hidden Divide: Explainable ML Predictive Evidence on Digital Resources, Achievement, and Algorithmic Fairness —— Evidence from PISA 2022 in Singapore and Taiwan

朱雯^{1*}, 刘炫妤^{1†}

¹ 浙江大学教育学院

* 1246656240@qq.com

【摘要】 随着智能教育系统的普及，数字鸿沟正由“资源接入差异”转向“算法适应性差异”。本研究提出“算法鸿沟”概念，指因特征表征不足与数据异质性导致算法对不同群体产生系统性预测差异。基于 PISA 2022 新加坡与台湾 11,591 名学生数据，采用梯度提升树与 SHAP 可解释机器学习框架，检验 ICT 因素与数学素养的非线性预测关联及其公平性表现。结果显示，ICT“使用”维度的预测力显著强于“接入”，低 ICT 群体预测误差更高且极端误差更大，而重加权干预改善有限。研究为学习分析系统的公平设计提供实证依据。

【关键词】 PISA 2022；数字鸿沟；可解释机器学习；算法公平性；教育数据挖掘

Abstract: As intelligent education systems proliferate, the digital divide is shifting from resource access to algorithmic adaptability. This study introduces the Algorithmic Divide, defined as systematic predictive disparities arising from insufficient feature representation and data heterogeneity. Using PISA 2022 data from 11,591 students in Singapore and Taiwan, we employ Gradient Boosting with SHAP to examine nonlinear predictive associations between ICT factors and mathematics literacy and evaluate algorithmic fairness. Results show that ICT use outperforms mere access in predictive power, and low-ICT students exhibit greater error variability and more extreme errors, while reweighting interventions yield limited improvement. These findings inform equitable learning analytics design.

Keywords: PISA 2022; Digital Divide; Explainable Machine Learning; Algorithmic Fairness; Educational Data Mining

1. 前言

在教育数字化转型的背景下，尽管物理接入层面的“第一道数字鸿沟”逐步缩小，学生在数字素养与技术赋能效果上的“第二道数字鸿沟”依然显著（OECD, 2023; van Dijk, 2020; Scheerder et al., 2017），数字技术究竟是学业“均衡器”还是差距“放大器”尚无定论。当前学业预警系统、智能推荐与自适应学习平台日益依赖预测模型进行资源分配与干预触发，若模型对弱势群体的预测质量系统性更低，则技术部署本身可能成为不平等的新载体。然而，现有研究多依赖传统统计方法，难以捕捉非线性关系与交互效应（Vigdor et al., 2014），且鲜少关注智能算法对不同群体的公平性问题（Baker & Hawn, 2022）。

近年来，算法公平性研究已揭示预测模型在不同群体间可能产生系统性性能差异（Corbett-Davies et al., 2023），教育领域也开始关注智能系统对弱势群体的潜在偏见（Baker & Hawn, 2022）。然而，已有研究多聚焦分类任务中的准确率差异，鲜少从连续型

学业指标的预测误差分布角度审视公平性。基于此，本研究提出“算法鸿沟”概念——在智能教育系统中，由于弱势群体的数据异质性或学习路径分化，导致算法在其上的预测误差或极端误判显著高于优势群体。与传统鸿沟关注“资源接入”不同，算法鸿沟聚焦技术对既有不平等的隐性放大。本研究选取了 PISA 2022 新加坡与台湾的数据(N=11,591)，采用可解释机器学习框架，

基于"接入—使用—能力"三维框架，探讨：(1)数字资源如何非线性地预测学生数学素养？(2)模型在不同数字资源群体上的预测效能是否存在系统性差异？本研究的创新在于提出并界定"算法鸿沟"概念，采用 XAI 框架通过 SHAP 分析揭示非线性机制并评估算法公平性；针对典型教育技术场景提出公平性设计建议。

2. 研究方法

本研究使用 PISA 2022 公开数据库（学生问卷，STU_QQQ），选取新加坡与台湾作为研究样本。选取 PISA 2022 是因其为目前唯一同时包含 ICT 接入、使用与自评能力三维度测量的大规模跨国教育评估，且数据公开可复现。新加坡与台湾均属 GCCCE 关注的华人/亚洲教育情境，且两者在 PISA 数学表现位居前列、数字化程度较高，适合检验"在数字基础设施相对完善的条件下算法鸿沟是否仍然存在"的命题。采用完整案例分析后有效样本量为 $N=11,591$ （新加坡 6,147 人，台湾 5,444 人），需要说明的是，所有统计分析均基于该样本且未引入 PISA 复杂抽样权重，推断性结果应理解为样本内预测性关联证据。因变量为数学素养（PV1MATH），ICT 变量基于"接入—使用—能力"框架（van Dijk, 2020）涵盖三个维度，接入（ICTRES 家庭 ICT 资源量、ICTHOME 家用设备数量）、使用（ICTSCH 校内使用频率、ICTOUT 校外使用频率、ICTWKDY 工作日使用时长）、能力（ICTEFFIC 数字自我效能）。

控制变量包括 ESCS、性别与移民背景。分析方法采用了可解释机器学习框架，首先以梯度提升回归树(Gradient Boosting Regressor)构建预测模型，其次引入 SHAP 方法（Lundberg & Lee, 2017）进行可解释性分析，最后按 ICTRES 三分位点将样本划分为低资源组与高资源组，以预测误差差异（error parity）作为公平性度量的操作化指标，并通过 Welch's t 检验评估组间差异。

3. 实验结果与分析

3.1. 描述性统计与模型性能

有效样本（ $N=11,591$ ）的数学平均成绩为 559.6 分（ $SD=107.5$ ），高于 OECD 均值，ESCS 均值为 0.03（ $SD=0.91$ ），女生与本地生占比分别约 50.0%和 85.0%。测试集性能对比显示，梯度提升树（ $R^2=0.388$ ）较 OLS 基线（ $R^2=0.284$ ）提升 36.6%，表明非线性建模更适合同捕捉 ICT 与学业表现的复杂关联结构。

3.2. 关键特征的可解释性分析

图 1 展示 SHAP 全局重要性排序。定量结果显示，ESCS 仍为最强预测因子（Mean $|SHAP|=31.39$ ），但 ICT 使用频率（ICTWKDY, Mean $|SHAP|=24.55$ ）与数字自我效能（ICTEFFIC, Mean $|SHAP|=14.26$ ）分别位列第二、第三，均显著高于 ICT 接入变量（ICTRES, Mean $|SHAP|=7.30$; ICTHOME, Mean $|SHAP|=5.21$ ），支持"使用维度预测力强于接入维度"的判断。

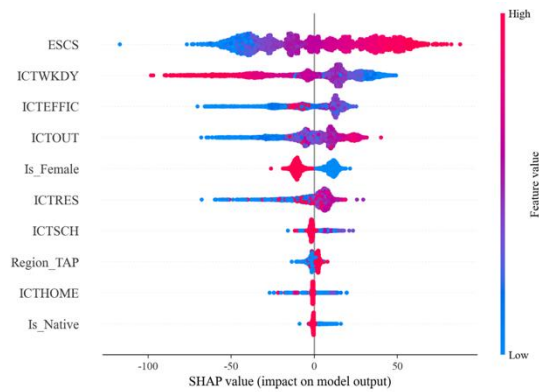


图 1 SHAP 特征重要性排序

图 2 (a) 揭示了非线性关联: ICT 资源对成绩的预测贡献存在显著饱和阈值($x=-0.56$)。在阈值左侧, 资源增加与预测值提升呈陡峭正相关 ($k=20.6$); 跨过阈值后, 关联强度转负或趋于平缓 ($k=-8.7$), 覆盖约 73% 样本。教育学上, 图 2 (a) 显示资源变量的预测贡献在阈值后趋于平缓, 提示仅增加设备数量未必伴随更优学业表现, 使用质量可能更为关键。图 2 (b) 展示 ICTRES 与 ESCS 的交互效应贡献 (interaction value)。颜色代表 ESCS 数值: 红色=高 ESCS, 蓝色=低 ESCS; 纵轴为 ICTRES $\times$ ESCS 交互项对模型预测的贡献, 用于刻画“在不同 SES 水平下, ICT 资源的边际预测贡献如何变化”。

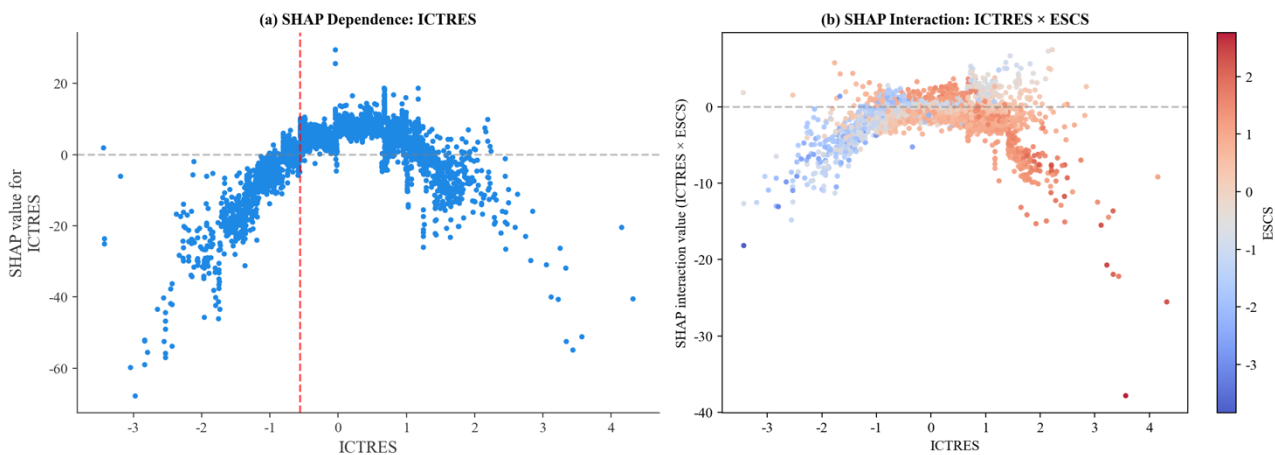


图 2 ICTRES 的 SHAP 结果 (a) 对学业表现的预测 (b) 与 ESCS 的交互贡献

3.3. 算法鸿沟的量化验证与干预测试

测试集评估验证了算法鸿沟的存在, 结果如表 1 所示, 低 ICT 组的预测误差 (MAE=70.59) 显著高于高 ICT 组 (MAE=62.89), 存在 7.70 分的显著差距 (相对差距 12.2%, $p<0.001$)。进一步的误差结构分析显示, 两组的均值偏差 (Mean Signed Error) 均接近 0, 但低 ICT 组的残差方差高出 21.0% ($p=0.006$), 且 P95 极端误差 (167.2 分) 高出优势组 12.1 分。这表明算法鸿沟主要源于预测的不确定性 (Variance) 而非方向性偏差 (Bias)。在应用场景模拟中, 若设预警系统阈值为 450 分, 低 ICT 组的真实学困生召回率仅为 36.1%, 存在极高的漏判风险。本研究进一步测试样本重加权策略 (权重 1.0–3.0 倍), MAE 差距稳定在 7.39–7.93 之间, 表明简单的模型微调难以消除鸿沟。

表 1 算法鸿沟 (组间预测误差差异)

Group	N	MAE	Variance	P95 Error
-------	---	-----	----------	-----------

Low ICT (Bottom 33%)	1,160	70.59	2724.3	167.2
High ICT (Top 33%)	1,160	62.89	2250.7	155.1
Δ (Gap)	—	+7.70 (p<0.001)	+21.0%(p<0.01)	+12.1

这些证据表明，算法鸿沟的根源并非模型没调好，更可能反映了低资源群体学习生态的更高异质性。同样的观测特征背后，弱势学生的学习路径更分化、受环境约束更多，因而更难被少量问卷变量稳定刻画。结合 SHAP 分析，低 ICT 组与高 ICT 组在使用频率、质量、目的上的差异可能比问卷能捕捉的更深层，导致误差方差上升。因此，低 ICT 组更高的误差波动性可能源于弱势群体更为多样化的成长生态与学业路径（Lareau, 2011）、数据信噪比的群体差异、以及测量误差的群体差异。

4. 讨论与结论

4.1. 理论贡献与发现

本研究拓展了数字鸿沟的理论内涵。实证表明，“使用维度”比“设备接入”在成绩预测方面更为重要。此外，研究揭示了“算法鸿沟”的隐性特征，弱势群体因学习生态的非典型性，现有常规特征难以稳定刻画其学习状态，导致算法在其上的误差方差和极端误判率显著上升。

4.2. 面向智能教育系统的公平设计启示

本研究针对典型智能教育应用场景，提出以下建议。第一，从系统设计方面，对于学业预警与资源分配系统，应将“低 ICT 特征”设为保守触发条件。当识别到该群体时，系统应自动降低机器预测的决策权重，引入人工复核机制（人机协同），避免高方差预测导致资源错配。第二，从数据处理方面，智能推荐平台不应局限于调整算法超参数，而应开发基于多模态或过程性日志（如学习节律、交互策略）的底层数据采集框架，从源头填补弱势群体的特征表征空白。

4.3. 局限性与未来展望

本文存在以下局限。第一，本研究采用完全案例剔除(Listwise Deletion)可能导致幸存者偏差，最弱势群体可能未进入样本，真实的算法鸿沟可能被低估；第二，对 PISA 数据的处理进行了简化，分析仅采用 PV1MATH 且未引入 PISA 抽样权重，可能低估标准误，后续应采用全部似真值与 BRR 权重，结论更适理解解为样本内的预测性关联；第三，基于截面问卷数据的预测本质上为相关性而非严格的因果效应。未来研究需结合纵向行为数据，以更准确地识别算法鸿沟的成因并提出有效的干预策略。

参考文献

- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R., & Goel, S. (2023). The measure and mismeasure of fairness. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1801–1816). Association for Computing Machinery. <https://doi.org/10.1145/3593013.3594011>
- Lareau, A. (2011). *Unequal childhoods: Class, race, and family life* (2nd ed.). University of California Press.

- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. <https://doi.org/10.1787/53f23881-en>
- Scheerder, A., van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2017). Determinants of Internet skills, uses and outcomes: A systematic review of the second- and third-level digital divide. *Telematics and Informatics*, 34(8), 1607–1624. <https://doi.org/10.1016/j.tele.2017.07.007>
- van Dijk, J. A. G. M. (2020). *The digital divide*. Polity Press.
- Vigdor, J. L., Ladd, H. F., & Martinez, E. (2014). Scaling the digital divide: Home computer technology and student achievement. *Economic Inquiry*, 52(3), 1103–1119. <https://doi.org/10.1111/ecin.12089>