

# 如何判定自适应学习系统的有效性

## ——基于因果结构分析框架

□江波 章恒远 魏雨昂

**摘要:** 自适应学习系统是开展大规模个性化教育的重要途径之一,其有效性判定是学界关注的焦点。现有判定自适应学习系统有效性的方法大多采用实验法、问卷调查法等,存在实验周期长、实验变量控制难等问题。基于因果结构的自适应学习系统有效性分析框架可通过分析系统中教育数据间的因果结构关系,来判断系统的有效性及其缘由。该分析框架主要包括多维教育数据与基于因果分析的技术路线两部分。其首先利用描述性检验和显著性检验进行特征关联性分析,判断系统是否有效;再利用相关性检验和马尔可夫毯方法寻找与学习成效具有因果关系的特征子集,并利用机器学习算法进行学习成效预测;最后利用基于搜索的因果结构算法确定特征之间的因果结构及影响强度,来回答系统何以有效。利用该分析框架对“均一教育平台”自适应学习系统进行分析后发现:(1)该系统的个性化推荐功能使用与否与学习收益有强因果关系;(2)学习者的练习数量、解题尝试次数和请求提示次数直接或间接影响学习成效;(3)该系统的个性化推荐可生成符合学生最近发展区的学习路径,在减少学生做题数量和错误次数的同时,提高学习效率。

**关键词:** 自适应学习系统;有效性;因果关系;分析框架;判定方法

**中图分类号:**G434 **文献标识码:**A **文章编号:**1009-5195(2023)02-0095-07 doi10.3969/j.issn.1009-5195.2023.02.011

**基金项目:** 国家自然科学基金面上项目“面向图形化编程的项目式学习的自动化评价研究及应用”(61977058);上海市科技创新行动计划“人工智能”专项项目“教育数据治理与智能教育大脑关键技术研究及典型应用”(20511101600)。

**作者简介:** 江波,博士,副教授,硕士生导师,华东师范大学教育信息技术学系,上海数字化教育装备工程技术研究中心,上海智能教育研究院(上海 200062);章恒远,硕士研究生,华东师范大学教育信息技术学系,上海数字化教育装备工程技术研究中心(上海 200062);魏雨昂,博士研究生,华东师范大学计算机科学与技术学院,上海智能教育研究院(上海 200062)。

随着人工智能技术的不断发展,实现个性化教学的方式和途径变得更加丰富(Xie et al., 2019),其中就包括已在教育领域广泛应用的自适应学习系统(Adaptive Learning System, ALS)。基于学习者的学习行为日志和数据挖掘技术,自适应学习系统能够对不同学习者的学习风格、认知图谱和认知能力等“个体表现”进行动态评估(姜强等, 2011),实时调整学习内容、知识序列及其评估方式(管保霞等, 2017),从而为学习者提供个性化的学习服务与帮助(彭红超等, 2019)。然而,自适应学习系统是否真的有效以及何以有效一直是学界关注的焦点(尚佩瑶等, 2021)。对于自适应学习系统是否有效,学者们运用随机实验法等研究方法开展了大量的实证研究,并借助问卷调查法和访谈法等方法来探究自适应学习系统何以有效。但这些研究方法忽视了

自适应学习系统所生成的大量教育数据,且存在实验周期长、实验变量难以控制等问题。基于此,本研究拟探索能够通过对自适应学习系统中的教育数据进行分析来判断该系统是否有效以及何以有效的方法,以期科学、高效、便利地评估自适应学习系统提供借鉴。

### 一、自适应学习系统的有效性研究现状

关于自适应学习系统的有效性研究一直是学术界关注的焦点。许多研究表明,自适应学习系统能够对学习者产生积极影响,帮助他们取得更好的成绩。Wu等(2017)采用两个等效组进行前后测实验,对比了使用动态呈现学习内容的自适应学习系统和传统电子学习系统的效果,结果显示提供个性化学习材料的自适应学习系统能够促进学习者对学

习内容的理解,进而使其学习结果得到显著提升。Wang等(2012)在一项单向准实验研究中,将144名大学生随机分配到两组。实验组使用具有自适应学习系统推荐功能的英语学习论坛,对照组使用传统论坛。前后测差异性检验、在线调查和博客的数据表明,自适应学习系统的推荐功能可以显著增加学习者的阅读频次,使其在语言测试上更富有成效。在认为自适应学习系统有助于提升学习成效的基础上,学者们进一步对其作用路径进行研究。其中,廖轶(2017)通过随机对照实验和显著性检验发现,应用自适应学习服务参考模型的学习者测试成绩更优、学习时间更短、点击次数更少、学习效率更高。赵艳(2020)通过问卷调查和访谈发现,自适应学习系统能够启发学习者的思维,帮助学生理解学科思想,激发学习兴趣,从而影响学习者的学习成绩。然而,也有研究表明,自适应学习系统对于提高学习者学习成绩的作用是有限的。例如,李建伟等(2020)在考前英语训练中测试了自适应学习APP的效果,结果显示在通过率上,使用自适应学习系统的学习者比未使用的学习者仅高出0.36个百分点,在优秀率上前者甚至比后者低1个百分点。Michlik等(2010)在64名学习者中进行了随机对照实验,根据系统推荐学习内容的适恰情况将学习者分为三组,分别为针对性推荐、随机推荐和不推荐,编程测试的结果显示,针对性推荐有助于提高学习者的学习成绩,但这一结论并未得到统计学意义上的支持。

综上可发现,尽管在自适应学习系统有效性方面,学者们的看法和意见不尽相同,但是他们的研究方法和手段有着一定的趋同性。这些研究主要采

用随机实验和准实验研究方法,旨在揭示自适应学习系统与学习者学习成绩之间的因果关系。此外,学者们还使用问卷调查法、访谈法、数据日志分析等方法综合学习者学习行为的客观数据(如时间、次数、成绩等)和学习者学习体验的主观数据(如适用性、满意度、目的等),来对自适应学习系统进行评价(姜强等,2015)。但在随机对照实验或准实验中,单次实验一般仅针对单一变量,变量控制难度大,实验成本高、周期长,还存在伦理道德限制、霍桑效应(曹浩文等,2015)等;而问卷调查与访谈又易受人的主观影响导致结果失真。针对上述情况,本研究提出可直接基于自适应学习系统中可观测的教育数据,来挖掘数据间的因果关系,从而分析自适应学习系统是否有效以及何以有效的自适应学习系统有效性分析框架。

## 二、基于因果结构的自适应学习系统有效性分析框架

基于观测数据进行因果推测,是有效性评价的一种新兴方法。为此,本文构建了基于因果结构的自适应学习系统有效性分析框架,如图1所示。框架主体可分为两个部分——多维教育数据和基于因果分析的技术路线。

在数据准备环节中,学习者的个人履历数据、认知水平数据和在自适应学习系统中的学习行为数据构成了多维教育数据特征集。个人履历数据包括学习者的编号、年龄、年级、性别、学习风格和学习兴趣;认知水平数据包括学习者的学习状态、知识水平、技能水平、评估记录、测试记录和熟练程度;学习行为数据包括练习时长、尝试次数、练习

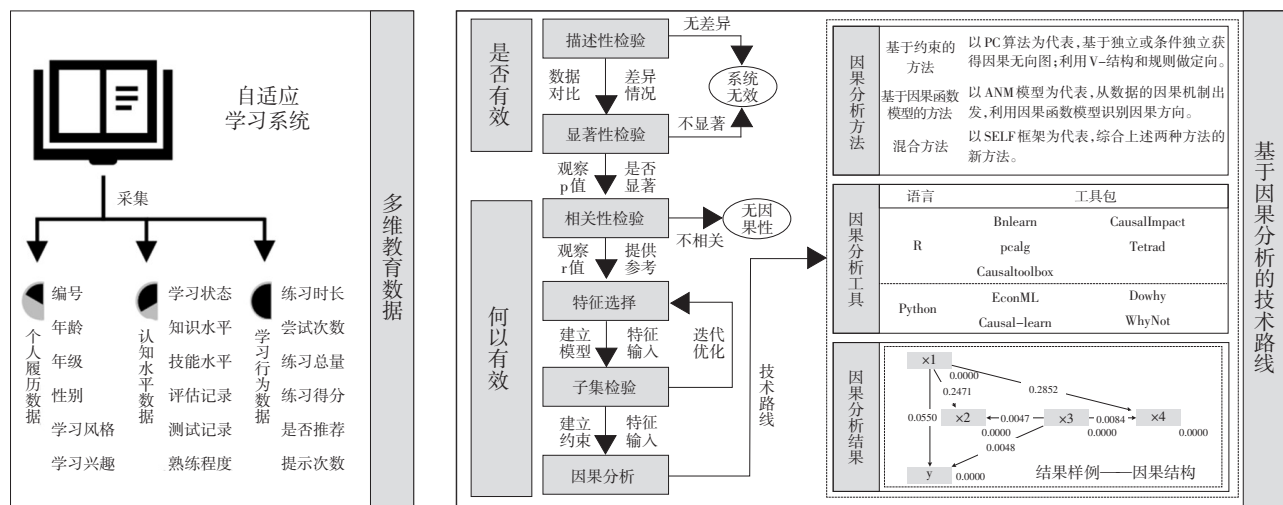


图1 基于因果结构的自适应学习系统有效性分析框架

总量、练习得分、是否使用个性化学习推荐功能和提示次数。

基于因果分析的技术路线主要实现自适应学习系统是否有效与何以有效的判定。在是否有效这一问题上,首先使用描述性检验观察数据间的差异,如果无差异则可得出系统无效,有差异则使用显著性检验从统计学角度分析差异的显著性,如果差异未达到显著性水平,也可得出系统无效。其次,在得出系统有效的基础上探讨何以有效的问题。先对多维教育数据特征与学习结果做相关性检验,若相关程度低,即不相关,可视为无因果性。若相关,则通过特征选择删去不相关特征和冗余特征,得到最佳特征子集,并通过机器学习算法构建学习成效预测模型,检验子集的有效性。如果检验效果不符合预期,则进入迭代优化环节,重新对特征子集进行选取和评估。最后,将最佳子集输入基于搜索的因果结构算法中,以确定各特征间的因果结构及影响强度。

此外,框架还提供了因果分析的常用方法、工具和结果样例。因果分析针对因果关系的方向分析、高维数据中因果关系的误发现、不完全观测数据中的隐藏变量检测等三个关键问题,形成了三类解决方法,分别是基于约束的方法、基于因果函数模型的方法和混合方法,详细介绍如图1所示。同时,也有众多因果分析的开源工具包,其中基于R语言和Python语言的工具有pcalg、CausalImpact、Dowhy、WhyNot、Tetrad等。本研究采用Tetrad分析工具,其可生成一个有向无环的因果结构图,图中不同节点间的因果关系用带箭头的连接线表示,由因指向果,并使用多元线性回归估计影响强度,将其标注于节点间的连接线上,能够清晰、明确地展现学习者应用自适应学习系统时,影响其学习成效的因素、因果关系及其强度。

### 三、自适应学习有效性分析实践

#### 1. 平台介绍

本研究选取了“均一教育平台”作为研究对象,该平台为K-12阶段的学习者提供自适应的学习环境。在2014年至2015年的两年时间里,有近25万名学习者在该平台学习数学课程,留下了超过2500万条学习过程数据。平台以课程为学习单元,课程内容以知识点划分,分为例题、解题、探索 and 基础四个模块。前三个模块以视频学习为主,

基础模块则提供相应的练习题。学习者可以在学习主页的“建议的活动”一栏中查看系统为其推送的个性化推荐内容。如果学习者选择了系统推荐的内容进行练习,则视为学习者使用了系统的个性化推荐功能;反之,则视为没有使用该功能。在习题解答过程中,学习者可以点击“解题说明”按钮查看系统提示,并在完成作答后点击“提交答案”按钮获得正误反馈。如果首次作答正确,将点亮小灯并获得一定的学习收益。但如果该题的作答借助了系统的提示,则该题的学习收益将被系统判定为零。

#### 2. 数据准备

本研究使用了247,606名学习者在“均一教育平台”上进行数学课程练习的25,925,992条学习过程日志数据。从中随机选取1000名学生共118,511条数据进行分析。在特征处理上,删除了与本研究无关的特征,剩余9个特征包括:学生唯一标识符(user\_id)、练习名称(exercise)、练习数量(problem\_number)、是否采纳个性化学习建议(suggested)、解题总时长(time\_taken)、解题尝试次数(count\_attempts)、请求提示次数(count\_hints)、是否达到熟练程度(earned\_proficiency)和学习收益(points\_earned)。其中,采纳个性化学习建议表示学习者的单次学习内容与系统个性化推荐的内容一致,即使用了自适应学习系统;反之,则视为未使用。

#### 3. 数据分析

##### (1) 描述性检验和显著性检验

为了探究“均一教育平台”自适应学习系统的有效性,研究选取学习收益的上四分位数180作为分隔值,将大于等于180的学习收益称为“高收益”,小于180的学习收益称作“低收益”(姚昊等,2022)。该平台中学习者学习收益统计情况如图2所示。横坐标表示是否使用自适应学习系统,纵坐标表示学习收益的占比情况。观察柱状图可以发现,在使用了“均一教育平台”个性化推荐功能的所有学习记录中,约有89.32%的记录取得了高收益,比不使用情况下获得高收益的比例要高得多。此外,独立样本T检验(置信水平为0.05)结果显示,p值等于0.000,效应量等于0.843,即学习者使用“均一教育平台”自适应学习系统与否所获得的学习收益具有显著性差异。因此可以得出,使用“均一教育平台”自适应学习系统能为学习者的学习结果带来显著的正向影响。



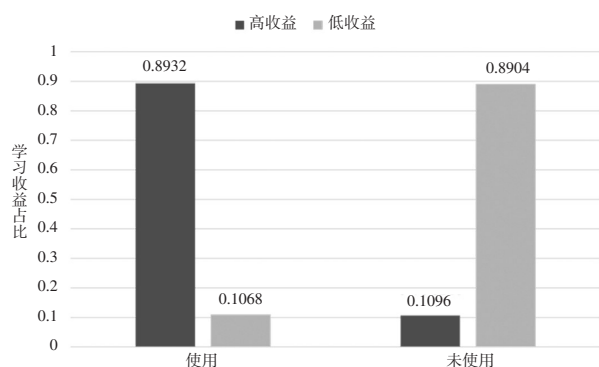


图2 自适应学习系统使用情况-学习收益柱状图

## (2) 相关性检验

采用 Kendall 相关系数、Spearman 相关系数和 Pearson 相关系数测量无序分类变量（是否采纳个性化学习建议）、有序分类变量（练习数量）和连续变量（解题总时长、请求提示次数、解题尝试次数）与学习收益的相关性，检验结果如表1所示。在正负相关关系上，除了是否采纳个性化学习建议这一特征外，其余4个特征与学习收益呈负相关，即指标越高，学习收益越低。在相关关系程度上，是否采纳个性化学习建议与学习收益的相关性最为密切，绝对值达到了0.5087，练习数量次之，绝对值为0.4103，两者均呈中等程度相关。解题尝试次数与学习收益之间呈弱相关关系，绝对值为0.2207。相比之下，请求提示次数和解题总时长与学习收益之间的相关性系数绝对值则小许多，分别为0.1772和0.0404，均呈弱的显著相关。强相关性是因果关系存在的前提（赵森栋等，2014），而解题总时长与学习收益之间的相关性过低，则可认为二者之间不具有因果性。其余四个特征可能与学习收益之间存在潜在的因果关系，可进一步进行因果分析，其中是否采纳个性化学习建议和练习数量两个特征的影响强度较其他变量更大。

表1 各变量与学习收益的相关性分析

变量	X-Grade	
	相关性	Sig 双尾
是否采纳个性化学习建议	0.5087**	0.00
练习数量	-0.4103**	0.00
解题总时长	-0.0404**	0.00
请求提示次数	-0.1772**	0.00
解题尝试次数	-0.2207**	0.00

（注：\*\*代表 $p < 0.01$ 。）

## (3) 基于马尔可夫毯的特征选择

使用 Tetrad 软件，按照特征的先后顺序设置约束条件：是否采纳个性化学习建议节点不能作

为其他节点的子节点；练习数量节点不能作为解题尝试次数和请求提示次数节点的子节点。使用 FGES-MB 特征选择算法（Sanchez-Romero et al., 2017）寻找学习收益特征的马尔可夫毯（Markov Blanket, MB）。该算法的目标是在不造成大量信息丢失的情况下，最大限度地识别并删除与之不相关的冗余特征（Blum et al., 1997）。结果如图3所示，学习收益的马尔可夫毯中包含了是否采纳个性化学习建议、练习数量、解题尝试次数和请求提示次数4个特征，这与前文得出的结论相符。

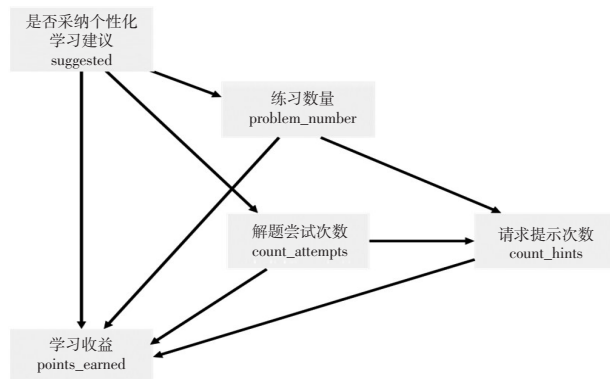


图3 学习者学习收益的马尔可夫毯

## (4) 子集检验

分类器的准确率是评价特征选择方法好坏的重要指标（段洁等，2015）。为了验证基于马尔可夫毯特征选择结果的有效性，本文使用主成分分析（Principal Component Analysis, PCA）将原始特征降至四维，保持与马尔可夫毯特征选择的输出特征数量一致，此时的可解释性方差占比超过99.9%。在支持向量机（Support Vector Machine, SVM）、决策树（Decision Tree, DT）、随机森林（Random Forest, RF）、逻辑回归（Logistic Regression, LR）和人工神经网络（Artificial Neural Network, ANN）等常见的机器学习分类算法中，依次将原始特征、使用主成分分析和马尔可夫毯的特征选择结果作为输入变量，学习收益作为输出变量，比较不同特征选择结果子集在各分类算法中的准确率。结果如表2所示，在随机森林和人工神经网络算法中，使用基于马尔可夫毯特征选择策略建立的模型具有更优异的准确率表现。因此，可以认为马尔可夫毯的特征

表2 基于各特征选择结果的分类算法准确率

模型	SVM	DT	RF	LR	ANN
原始	0.8721	0.8204	0.8932	0.8721	0.8952
PCA	0.8762	0.8265	0.8979	0.8750	0.8978
MB	0.8643	0.8062	0.9037	0.8516	0.9019

选择在此数据集上是有效的,以此为基础进行因果关系建模是合理的。

#### (5) 因果分析

为了评估关键变量(是否采纳个性化学习建议、练习数量、解题尝试次数以及请求提示次数)和学习收益间是否存在潜在的因果关系,并确定它们之间的影响强度,本研究使用因果工具Tetrad软件中的PC算法,得到了图4所示的因果关系模型。分析结果显示,该模型性能良好,卡方检验( $df=1$ )表明其预测结果与实际数据( $p=0.0521$ )之间没有统计学差异。

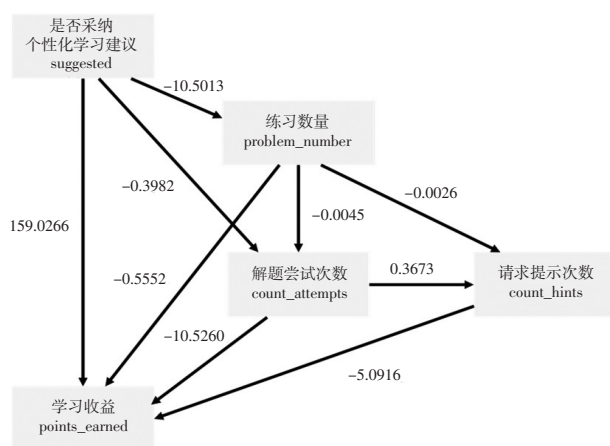


图4 关键变量与学习收益间的因果关系模型

通过观察表1和图4,可以发现因果关系模型中的变量是相关性分析结果的子集,其中补集为解题总时长。这是因为因果关系是特殊的相关关系,相关关系是因果关系的必要条件之一。值得注意的是,在表1中,相关系数仅描述了两个变量之间的相关性大小,并不能说明是哪一个变量的变化导致了另一个变量的变化。相比之下,在图4的因果关系模型中,归一化系数的绝对值表示因果关系强度的大小,归一化系数的正负表示两个变量数值变化的同向性,箭头从自变量出发指向因变量。

根据模型,可以看出练习数量与请求提示次数和解题尝试次数的归一化系数分别为-0.0026和-0.0045,这两个值与其他路径上的归一化系数相比,绝对值相差过大,这表明二者间的因果强度非常低,可以忽略不计。再根据作用路径的不同,将图4中的因果关系模型拆分为图5、图6和图7三个子图,进行逐一分析。

图5显示,使用该自适应学习系统与学习收益间的归一化系数为159.0266,是所有系数中绝对值最高的。这意味着,使用系统推荐功能每增加1个

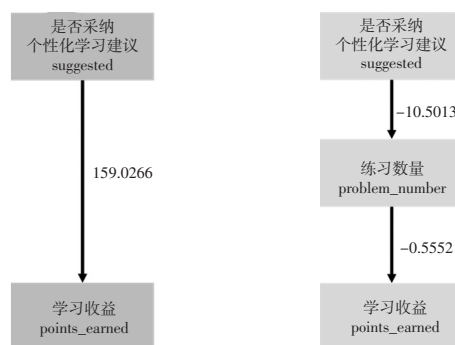


图5 是否采纳个性化学习建议与学习收益之间的因果关系模型

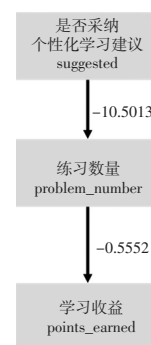


图6 是否采纳个性化学习建议、练习数量与学习收益之间的因果关系模型

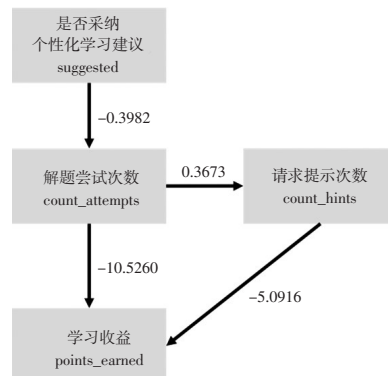


图7 是否采纳个性化学习建议、解题尝试次数、请求提示次数与学习收益之间的因果关系模型

标准差,学习收益就会增加159.0266个标准差。表明该自适应学习系统可以提高学习者的学习成效,且这一提高非常明显。

图6显示,使用自适应学习系统导致练习数量的减少(系数为-10.5013),练习数量的减少导致学习收益的增加(系数为-0.5552)。可以得到,使用自适应学习系统通过减少练习数量,间接导致学习收益的增加,其因果强度为5.8303,计算方法如下:

$$\begin{aligned} \text{suggested} \rightarrow \text{points\_earned} &= \text{suggested} \rightarrow \text{problem\_number} * \\ &\text{problem\_number} \rightarrow \text{points\_earned} \\ &= (-10.5013) * (-0.5552) \\ &= 5.8303 \end{aligned}$$

图7显示,使用自适应学习系统导致解题尝试次数的减少(系数为-0.3982),解题尝试次数的减少导致请求提示次数的减少(系数为0.3673)和学习收益的增加(系数为-10.5260);请求提示次数的减少进一步导致学习收益的增加(系数为-5.0916)。可以得到,使用自适应学习系统通过减少解题尝试次数和请求提示次数,间接导致了学习者学习收益的增加,其因果强度为4.9361,计算方法如下:

$$\begin{aligned} & \text{suggested} \rightarrow \text{points\_earned} = \text{suggested} \rightarrow \text{count\_attempts} * \\ & (\text{count\_attempts} \rightarrow \text{points\_earned} + \text{count\_attempts} \rightarrow \\ & \text{count\_hints} * \text{count\_hints} \rightarrow \text{points\_earned}) \\ & = -0.3982 * [-10.5260 + 0.3673 * (-5.0916)] \\ & = 4.9361 \end{aligned}$$

以上三条路径表明,使用自适应学习系统能够直接导致学习收益的增加,也能通过练习数量、解题尝试次数和请求提示次数这三个关键变量间接导致学习收益的增加,其因果总强度为169.7930。

该自适应学习系统通过精准定位学生的最近发展区,规划知识序列路径,进而减少学习者的练习数量、解题错误次数和使用提示次数,将学生从海量低效甚至无效的重复性学习中抽离出来,使不同层次、不同水平的学生学习效率都得以提升,最终达到减轻学习负担、提高学习效率的目的。这一有效性可以通过教学支架和最近发展区理论来解释。所谓教学支架,是指对学习者的即时支持,通过适当时机和适量支持促进学习者知识与技能的习得(何克抗,2017)。自适应学习系统根据学生学习的特性和需求,动态调整学习内容和学习方式,提供适度挑战的学习资源和学习环境作为支架,促进学生个体的发展。最近发展区理论认为,学习者只有在其认知水平的“边缘”才能获得最有效的学习。自适应学习系统充分考虑学生当前认知状态,判断学生能够承受的挑战程度,避免过度挑战或过度简单化情况,为学生提供个性化、精准的支架,帮助学生在最近发展区内获得有效学习,从而达到提升学习效果的目的。

#### 四、总结与展望

本研究提出基于因果结构的自适应学习系统有效性分析框架,框架主体可分为两部分:多维教育数据和基于因果分析的技术路线。多维教育数据包括了进行有效性分析所需采集的教育数据;基于因果分析的技术路线则提供了分析自适应学习系统是否有效以及何以有效的方法、步骤以及结果样例。本研究以某自适应学习系统为实践案例,利用所提框架分析该自适应学习系统的有效性,分析结果表明:(1)该系统的个性化学习推荐功能使用与否与学习收益有强因果关系,即系统的自适应学习功能有助于提升学习效果;(2)学习者的练习数量、解题尝试次数和请求提示次数直接或间接影响学习收益;(3)系统的个性化推荐功能可生成符

合学生最近发展区的学习路径,在减少学生的做题数量和错误次数的同时提高学习效率。

本研究的创新之处在于,为自适应学习系统的科学评估提供了一种新的视角,即通过采集学习过程数据,建立因果模型来评估系统的有效性。相较于随机对照实验,这种基于观测数据的证据更易获取、更加高效。然而,应用本文所提出的分析框架需要注意以下几点:第一,不同的自适应学习系统所存储的学习者数据种类是不同的,但一些关键特征是能够运用本框架的必要条件,如本文案例中的是否遵循推荐、练习得分、尝试次数等特征。第二,在因果建模后,需对模型进行拟合优度评估,说明得到的因果关系模型与实际情况在统计学上不存在显著差异。第三,在模型解释上,一定要结合系统自身判定规则与条件。例如,本文所使用的“均一教育平台”认为如果学习者在解答问题时使用了系统提示功能,则无论其是否解答正确,该题的学习收益将被判定为零,显然不是所有系统都有类似设定。除此之外,本研究还存在着诸多不足。首先,多维教育数据集的构建上,不同系统对学习者的信息各异,故在实际的应用中需根据系统实际情况进行调整。其次,在分析框架的实践上,本研究仅基于一个公开数据集展开探究活动,还需在更多的数据集上进行应用与实践。最后,在课程方面,本研究所用数据集全部来自数学科目,可能自适应学习系统在不同科目将对学习者的学习成效产生不同的影响,可通过学科的视角展开进一步的探究。

#### 参考文献:

- [1]曹浩文,杜育红(2015).教育研究中的因果推断方法探析——以班级规模与学业成绩的关系研究为例[J].上海教育科研,(6):40-43,31.
- [2]段洁,胡清华,张灵均等(2015).基于邻域粗糙集的多标记分类特征选择算法[J].计算机研究与发展,52(1):56-65.
- [3]何克抗(2017).教学支架的含义、类型、设计及其在教学中的应用——美国《教育传播与技术研究手册(第四版)》让我们深受启发的亮点之一[J].中国电化教育,(4):1-9.
- [4]曾保霞,姜强,赵蔚等(2017).大数据背景下自适应学习个性特征模型研究——基于元分析视角[J].远程教育杂志,35(4):87-96.
- [5]姜强,赵蔚(2011).自适应学习系统述评及其优化机制研究[J].现代远程教育,(6):57-63.
- [6]姜强,赵蔚,王朋娇(2015).自适应学习系统中学习者



中心视角评价学习成效实证研究[J].现代远距离教育,(4):43-49.

[7]李建伟,葛子刚,张爱阳(2020).自适应学习系统在成人本科学士学位英语学习中的应用研究[J].现代教育技术,30(3):59-65.

[8]廖轶(2017).面向基础教育的自适应学习服务系统研究与应用[D].北京:北京交通大学:109-111.

[9]彭红超,祝智庭(2019).人机协同决策支持的个性化适应性学习策略探析[J].电化教育研究,40(2):12-20.

[10]尚佩瑶,汪存友(2021).智适应学习系统的学习效果及其影响因素研究——基于2000—2020年60项实验与准实验研究的元分析[J].数字教育,7(1):45-50.

[11]姚昊,马立超(2022).高学历教师培养的学生成绩更优异吗?——理论争议、实证检验与政策启示[J].开放教育研究,28(2):73-84.

[12]赵森栋,刘挺(2014).因果关系及其在社会媒体上的应用研究综述[J].软件学报,25(12):2733-2752.

[13]赵艳(2020).自适应学习系统中数学习题辅导策略设计及对学生的影响研究[D].锦州:渤海大学:27-43.

[14]Blum, A. L., & Langley, P. (1997). Selection of Relevant Features and Examples in Machine Learning[J].

Artificial Intelligence, 97(1-2):245-271.

[15]Michlík, P., & Bieliková, M. (2010). Exercises Recommending for Limited Time Learning[J]. Procedia Computer Science, 1(2):2821-2828.

[16]Sanchez-Romero, R., Ramsey, J. D., & Liang, J. C. et al. (2017). Identification of Mechanisms of Functional Signaling Between Human Hippocampus Regions[J]. BioRxiv:099820.

[17]Wang, P. Y., & Yang, H. C. (2012). Using Collaborative Filtering to Support College Students' Use of Online Forum for English Learning[J]. Computers Education, 59(2):628-637.

[18]Wu, C. H., Chen, Y. S., & Chen, T. C. (2017). An Adaptive E-Learning System for Enhancing Learning Performance: Based on Dynamic Scaffolding Theory[J]. Eurasia Journal of Mathematics, Science and Technology Education, 14(3):903-913.

[19]Xie, H., Chu, H. C., & Hwang, G. J. et al. (2019). Trends and Development in Technology-Enhanced Adaptive/Personalized Learning: A Systematic Review of Journal Publications from 2007 to 2017[J]. Computers & Education, 140:103599.

收稿日期 2022-08-27 责任编辑 李鑫

## How to Evaluate the Effectiveness of Adaptive Learning Systems

——Based on the Causal Structure Analysis Framework

JIANG Bo, ZHANG Hengyuan, WEI Yuang

**Abstract:** The effectiveness assessment of adaptive learning systems, one of the important approaches for large-scale personalized education, has attracted much attention from literature. Existing methods for evaluating the effectiveness of adaptive learning systems mostly are experimental methods and questionnaire surveys, which suffer from problems such as long experimental duration and difficulty in experimental variables controlling. A causal structure-based framework is proposed to evaluate the effectiveness of an adaptive learning system and analyze how it works by analyzing the causal structure produced. This analysis framework consists of two parts, multidimensional educational data and a causal analysis-based technical framework. Firstly, descriptive and significance tests are used to analyze feature associations and determine whether the system is effective. Secondly, the correlation tests and Markov blanket method are used to identify feature subsets that have causal relationships with learning outcomes, and machine learning algorithms are used for learning outcome prediction. Finally, a search-based causal structure algorithm is used to determine the causal structure and strength of influence between features to explain why the system is effective. The proposed analytical framework is used to evaluate a popular adaptive learning system named “Junyi Education Platform”. The results showed as follows. First, there is a strong causal relationship between the use of personalized recommendation functions and learning outcomes. Second, the number of exercises students practice, the number of attempts to solve problems and the number of hints used directly or indirectly affect learning outcomes. Third, the personalized recommendations of the system improves learning benefit by generating learning paths that are in line with the student's zone of proximal developmental as well as reducing the number of exercises and mistakes.

**Keywords:** Adaptive Learning System; Effectiveness; Causal Relationship; Analysis Framework; Evaluation Method