

基于计算教育学的人工智能课堂分析 框架和技术实现

张 治¹ 程杼一² 王天蓉³

(1. 上海市宝山区教育局 上海市未来学习研究与发展中心, 上海 201999; 2. 上海理工大学
光电信息与计算机工程学院 上海 200093; 3. 上海市宝山区教育事务服务中心
上海市未来学习研究与发展中心, 上海 201999)

【摘要】 人工智能技术为基于计算的教育教学研究提供了全新场景, 驱动课堂教学分析进入计算时代。本研究基于“素养导向, 学为中心”的课堂价值观, 形成了“主动学、深度学、互动学、高效学、多样学”五维观测视角和高质量课堂分析框架模型, 建立了计算教育学视域下的人工智能课堂分析技术框架, 开发了包含采集层、数据层、计算层、分析层、支架层、资源层和循证层的七大功能模块, 共154项学习分析指标的人工智能课堂分析系统。在此基础上, 本研究以上海市宝山区161所学校10109节课为样本开展课堂分析, 构建了“微观—中观—宏观”三层次课堂改进和“诊断—反思—研究”三梯度赋能教师发展的实践路径, 并形成了“理论驱动”“数据驱动”“问题驱动”三类应用模式。

【关键词】 计算教育学; 课堂分析; 人工智能循证实践; 教学评价

【中图分类号】 G434

【文献标识码】 A

【文章编号】 1007-2179(2024)05-0087-14

人工智能与大数据为基于计算的教育教学研究提供了全新场景, 驱动课堂教学分析进入计算时代。然而, 目前智能技术支持的课堂分析, 仍面临以下困难: 一是惯性地关注教的行为, 尚未形成“以学习为中心”的分析逻辑和框架; 二是难以实现常态化的数据采集与分析; 三是停留在浅层行为分析, 缺乏深度认知和思维分析, 未能让学生思维发展和素养提升实现可计算。由于课堂是由复杂变量构成的教育场景, “课堂的价值导向”“计算的底层逻辑”“技术的架构搭建”和“研究的范式突破”等问题, 仍是人工智能课堂分析亟需面对

与解决的关键挑战。因此, 本研究基于计算教育学视角, 重塑课堂分析的价值导向和计算的底层逻辑, 建立人工智能课堂分析技术框架, 并基于区域真实课堂的全过程数据, 探索人工智能课堂分析的应用路径与模式, 以期丰富计算教育学的实践应用场景、创新人工智能课堂分析范式提供借鉴。

一、研究现状与范式突破

(一) 研究现状

自1970年弗兰德斯互动分析系统(Flanders Interaction Analysis System)形成以来, 经过近五十

【收稿日期】 2024-09-01

【修回日期】 2024-09-05

【DOI编码】 10.13966/j.cnki.kfjyjj.2024.05.009

【作者简介】 张治, 上海市宝山区教育局局长, 上海市未来学习研究与发展中心常务副主任, 华东师范大学特聘教授(87476088@qq.com); 程杼一, 助理研究员, 上海理工大学光电信息与计算机工程学院; 王天蓉, 正高级教师, 上海市宝山区教育事务服务中心, 上海市未来学习研究与发展中心办公室主任。

【引用信息】 张治, 程杼一, 王天蓉(2024). 基于计算教育学的人工智能课堂分析框架和技术实现[J]. 开放教育研究, 30(5): 87-100.

年的发展,课堂分析技术已逐步走向专业化、合作化与标准化(肖思汉等,2016)。传统的课堂分析存在标准化不足、专业性缺失等问题,片面运用技术手段展开课堂分析,容易陷入指标单一化和标签化的风险(杨晓哲,2021)。随着人工智能与大数据技术的发展,课堂分析逐步走向大规模采集、多模态数据和智能化分析。基于人工智能技术的课堂教学分析,正经历从全人工、弱人工智能、强人工智能向人机协同进化(孙众等,2020)。无论是分析指标体系的建构与侧重,还是数据采集分析的途径与技术,已从基于单一数据走向运用多模态数据,通过对课堂教学过程的语音、表情、行为和空间等多模态数据进行自动化采集、识别与分析,客观解读与评价教师教学行为和学生状态。

然而,现有研究在落实核心素养、关注思维发展、体现以学习为中心的价值导向等方面存在明显不足。同时,基于人工智能技术的课堂分析研究仍处于调查测试阶段,尚难以做到伴随式采集、常态化分析与普及化应用,系统采集的课堂大数据还未形成。此外,目前课堂分析大部分是自上而下的“标准驱动”或“理论驱动”模式,分析在很大程度上是为了验证专家提出的假设,而不是探寻课堂学习行为数据间的内在关联,挖掘现象背后隐藏的问题。因此,人工智能课堂分析需要新的理论指导,并在研究范式上有所突破。计算教育学的发展为此带来了新的机遇。

(二) 范式突破

1. 计算教育学何以赋能课堂分析

21世纪以来,数据科学和大数据技术蓬勃发展。在此背景下,格雷(Gray)提出了科学研究的第

四种范式——数据密集型科学范式(Hey, 2011)。2009年,拉泽尔(Lazer, 2009)提出了“计算社会科学”这一崭新概念。随着慕课的兴起和学习行为数据研究的深入,计算教育学逐步发展起来。李政涛等(2019)在总结国内学者关于计算教育学内涵的基础上,提出计算教育学是以大数据为基础、以计算和模型为手段、以信息科学为支撑的定量精确学科。郑永和等(2020)立足于学科建设,将计算教育学定义为运用量化的教育数据和计算方法,揭示日常难以发现的教育教学规律,融合多学科知识的跨领域新兴学科。计算教育学带来了教育科学研究范式的转变,体现为从经验判断走向数据驱动、从抽样调查走向全样本分析、从现象的局部观察走向全景式深度解剖,不仅成为教育理论发展与实践创新的助推器,也被期待能够破解教育的规模化覆盖和个性化发展的难题(刘三女牙等,2020)。

课堂是一个由复杂变量构成的教育场景,课堂分析属于教育场域中的数据密集型研究。传统研究方法难以实现对复杂过程和多维素养的评价,计算教育学视域下的研究范式则为探索课堂奥秘、打开课堂“黑箱”提供了可能。与传统课堂定量研究范式相比,基于计算教育学的课堂研究的突破之处表现在不再始于假设,而是基于数据驱动;不但采用自上而下的专家知识指导,而且依靠自下而上的数据归纳(见表1)。

2. 基于计算教育学的课堂分析新内涵

真实场景的课堂研究,本质上是一种循证实践,需要基于理论假设的数据分析诊断课堂,基于数据挖掘的预测功能发现规律,还要基于现实问题,运用数据验证既有的行动假设或者补充原先的认知,

表1 传统听评课、课堂定量研究与基于计算教育学研究范式比较

课堂分析范式	传统听评课	传统定量研究	基于计算教育学的课堂研究
数据特征	定性为主	主要通过问卷调查获得数据 以有限的样本数据为主	通过数据采集挖掘获得海量的自然数据、 过程数据、结果数据等多维度数据
获取效率	即时获得	延迟获得	即时获得
研究路径	基于经验	始于假设:现象—假设—抽样—验证	基于数据、数据汇聚、数据清洗、数据挖掘与 分析——数据预测和服务
研究目标	改进实践	理论修正	数据解释与理论构建
认识过程	经验判断	自上而下	自下而上(数据分析)+自上而下(专家知识)
关系认知	不确定	因果关系	关联关系+因果关系,基于关联关系的因果关系, 为干预提供依据

三者缺一不可。因此,在对课堂研究本质认识的基础上,本研究构建了“理论驱动”“数据驱动”“问题驱动”的多元驱动研究范式。具体而言,理论驱动即“标准驱动”,指根据现有教育理论制定标准、形成算法和模型后对数据进行分析,是一种对教育理论在不同情境下的验证,可用于教育诊断与评价。数据驱动指通过数据挖掘,发现教育要素之间潜在的相关性,具有预测功能,可基于已有数据进行实时教学决策,或通过数据挖掘进行教育治理,也可以对潜在的相关问题展开研究,解释、完善甚至建构理论。问题驱动指以解决教育实际问题为导向,综合运用假设检验和发现归纳两种方式对教育大数据进行影响因素分析,形成研究结论。

立足于上述认识,本研究认为基于计算教育学的课堂分析,是以真实课堂中学与教的伴随性数据为研究对象,以智能化课堂诊断、大规模教学治理与个性化课堂改进为目的,以科学价值观引领的课堂教学数据定量分析为研究方法,在数据驱动与理论驱动双向互动中形成智能算法,通过循证实践改进教学并优化算法,在数据挖掘中发现数据的相关关系和因果关系,从而发现课堂教学规律,为课堂改进提供依据,实现课堂育人价值的一种新型课堂研究范式。在这种课堂研究范式下,本研究构建了人工智能课堂分析的实践框架(见表2)。

二、技术框架

在基于计算教育学研究范式与实践框架下,本研究提出人工智能课堂分析技术框架,包括系统架构(计算原理)、数据要求(计算什么)、数据采集(如何采集)、数据分析(如何计算)、数据循证(计算后如何循证改进教学)和功能模块(系统如何运行)六部分。

(一)基于教育大脑的人工智能课堂分析系统架构

计算教育学以教育过程可计算为基础,其技术手段涉及硬件、平台、算法、数据、应用方式等(郑永和等,2020)。面向智能教育的三个基本计算问题是认知计算、行为计算和环境计算(黄荣怀等,2019)。基于计算教育学的思想,本研究参考教育大脑仿生学架构(张治等,2022),以课堂教学为应用场景,建立以数据为驱动要素的课堂分析系统架构,包括采集、分析、计算、管理、循证五大模块(见图1)。

(二)素养导向的多维数据要求

在素养导向的教育数字化改革中,数据应全面、真实、可靠、共享、智能。其一,数据内容需涵盖学生的情感态度、知识结构、能力特征、合作互动、认知过程和思维水平,以刻画课堂学习的总体面貌;

表2 计算教育学视域下的人工智能课堂分析实践框架

研究视域	基于计算教育学的课堂研究范式	实践框架	人工智能课堂分析实践框架(各章内容)
研究路径	基于数据:数据汇聚—数据清洗—数据挖掘与分析—数据预测和服务	系统架构	教育大脑仿生学系统架构(第二章第一节) 支持多应用生态的功能模块(第二章第六节)
数据特征	通过数据采集挖掘获得海量的自然数据、过程数据、结果数据等多维度数据	数据要求	素养导向的多维数据要求(第二章第二节)
获取效率	即时获得	数据采集	真实课堂中学与教的大规模伴随式数据采集(第二章第三节)
计算方法	密集型课堂教学数据的定量分析为主要方法,体现计算思维	数据分析	运用微积分思想解构与建构教学过程(第二章第四节,第三章)
研究目标	科学价值观引领的数据解释与理论构建	实践目标	建立学习为中心的课堂循证模型(第二章第五节) 以智能化课堂诊断、大规模教学治理与个性化课堂改进为实践目标(第四章)
认识过程	自下而上(数据分析)+自上而下(专家知识)+循证实践迭代改进	驱动方式	理论驱动、数据驱动、问题驱动(第四章)
关系认知	关联关系+因果关系,基于关联关系的因果关系,为干预提供依据	认知路径	数据挖掘—发现关联关系—探索因果关系—发现课堂教学规律,如课堂质量体系、教师能力预测、课堂特征观察等(第四章)

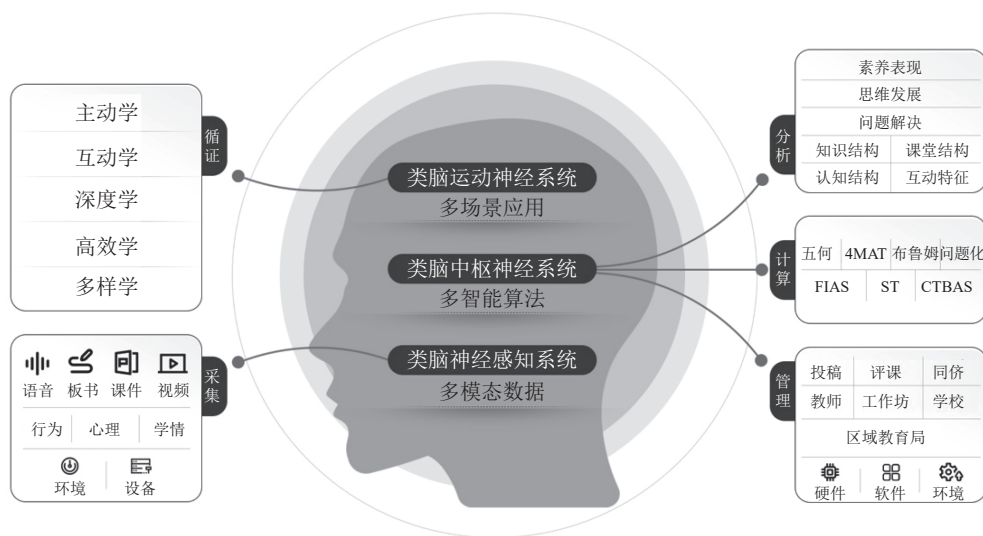


图1 基于教育大脑的人工智能课堂分析系统架构

其二,数据应来源于真实课堂环境的师生课堂行为,确保数据的真实性和过程性;其三,数据形态需满足可描述、可保持、可复现的要求(Bandura, 1977),确保数据的可靠性和有效性,这意味着数据应清晰描述课堂事件,并能在不同时间和情境下保持一致,便于重复观察和分析;其四,数据机制能支持教师、学校和区域教研部门之间的数据共享和协作,形成“分析—诊断—改进”的工作闭环;其五,数据计算能通过人工智能减轻教师的工作负担,即自动计算和分析关键数据,使课堂智能分析成为日常教学改革的一部分。只有满足上述数据要求才能有效支持智能化课堂诊断、大规模教学治理与个性化课堂改进。

(三)大规模伴随式数据采集

计算教育学的数据来自真实教育场景的各类常态化、伴随式、多维度、多模态数据,课堂作为教育教学的核心场景,每天产生大量的认知、心理、行为、环境等学与教的过程性数据,是计算教育学的“富矿”。同时,互联网背景下全媒体学习生态环境(祝智庭等, 2020),很大程度上从实时通讯、认知交互、学习资源等方面改变了课堂结构,但也极大地增加了课堂数据采集的干扰度、复杂度、耦合度。这就需要构建各类传感器融合系统,采集师生语音、板书、课件、视频、行为、心理、学情、环境、设备等数据,实现真实课堂的数字孪生,为后续智能分析提供高质量数据基础。

(四)运用微积分思想的数据分析

长期以来,课堂被视为一个难以琢磨的“黑箱”,被类比为数学领域复杂的高阶非线性系统,即一种高阶函数。研究者可以借助数学微分方法,将高阶函数分解为低阶函数,使课堂动态过程变得可感知和可理解。本研究构建了课堂分析的微积分V字模型(见图2),包括微分解构和积分重构两个关键步骤,即通过采集课堂的多模态异构数据进行微分和积分处理,灵活解构与重构课堂,满足不同维度的分析需求。这种方法将课堂量化分析与质性评价有机结合,可为课堂研究与改进提供便捷的实现路径。

微分解构位于V字模型的左侧,它通过将课堂活动细分为时间序列捕捉课堂的细微变化,即通过每3到30秒(即微分时间 dt)对课堂切片分析,以便细致观察课堂动态。这种细分方法有助于深入分析课堂的每个片段,从多个维度揭示课堂特征,为后续智能课堂分析和教学改革奠定基础。具体而言,基于多媒体设备,研究者对学与教过程中采集到的视频、音频等数据进行微分处理,将课堂行为、师生语言、表情、问题、互动、关键词等要素解构成一个个一维的函数,这样数据计算就融合了国内外常见的课堂量化分析方法,包括弗兰德斯FIAS(Flanders, 1963)、ST(Student-Teacher)分析法(张敏霞等, 2004)、五何分类法(祝智庭, 2001)、4MAT(McCarthy, 2005)、布鲁姆教学目标分类

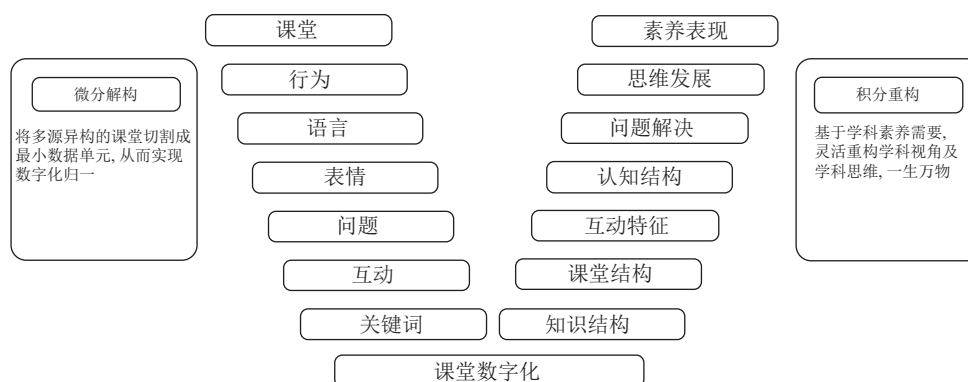


图2 基于微积分思想的数据分析方法

(Anderson, 2018)、问题化学习 15 问(王天蓉等, 2023)等。

积分重构位于 V 字模型的右侧,是基于课堂分析多种目的的积分重构数据。本系统构建了包含教与学行为各维度,指向师生的知识结构、课堂结构、互动特征、认知结构、问题解决、思维发展和核心素养七大类 154 项指标,其中包括学生表现(A 类 51 项数据指标)、教师表现(B 类 38 项数据指标)、课堂互动(C 类 16 项数据指标)、课堂结构(D 类 5 项数据指标)、课堂认知(E 类 23 项数据指标)、问题化学习 PSL(Problem Systematized Learning)课堂互动分析模型(F 类 21 项数据指标)(见表 3)。

(五)以学习为中心的课堂循证指标体系

从系统论角度看,任何可持续优化的系统都需要形成 PDCA(Plan-Do-Check-Act)质量闭环。运用计算教育学进行课堂分析,通过解构与建构学与教的过程,实现全流程多维度数据计算,是实现教学质量提升的第一步。而基于数据的课堂循证,则是落实课堂价值质量闭环的核心。

从认知建构角度看,课堂的本质是一种问题逻辑,即师生以问题为纽带开展互动。因此,本研究基于上海市宝山区问题化学习 20 年研究成果(中华人民共和国教育部, 2023)和上海市未来学习研究与发展中心建立的学习评价的五个维度,以“问题”为线索,将课堂计算、分析的数据应用到以学习为中心的课堂循证过程,从学生的“主动意愿、互动对话、认知深度、目标达成、学习方式”五方面入手,建立包括主动学、深度学、互动学、高效学和多样学五个维度的“五学”课堂循证模型,以

及相对应的行为指标(见表 4),并基于数据统计建立各指标的初始常模值。

在上述模型中,主动学指学生课堂学习的主动意愿和维持主动学习的时间与动态分布;互动学指学生在课堂中的社会参与,即师生、生生对话状态;深度学指学生课堂认知加工的深度;高效学指课堂学科学习目标的达成效率、重难点与学科核心素养落实;多样学指学生课堂学习路径、学习方式、思维视角、活动类型多元化。在此模型的基础上,本研究从学生表现、教师表现、课堂互动、课堂结构、课堂认知五方面采集与统计数据,建立指标体系(见表 5)。

(六)支持多应用生态的功能模块

本研究开发了人工智能课堂分析系统,通过真实课堂的多模态数据采集,运用微积分分析方法,构建以学习为中心的教学质量闭环。该系统由七个层次构成:采集层、数据层、计算层、分析层、支架层、资源层和循证层。这些层级与区域数字基座平台整合,支持区域教育数据治理和师生发展。其中,采集层负责采集课堂中音视频、课件、板书、作业等多模态数据,支持多种数据采集设备,实现跨设备的数据采集;数据层对采集的数据进行清洗、分类、标签,以备计算层使用;计算层应用人工智能技术,如机器学习、深度学习等,自动计算国内外课堂学习分析模型,实现教学过程的微分化;分析层利用计算层的数据,对教学质量进行多维度积分,形成数字化评价结果;支架层提供在线平台,支持教师教学分析、反思和设计,促进教师专业发展;资源层将有价值的数据转化为教学资源,发挥数据

表3 人工智能课堂分析系统课堂核心154项指标

编号	学生表现 A-51	教师表现 B-38	课堂互动 C-16	课堂结构 D-5	课堂认知 E-23	问题化学习 F-21
1	学生课堂高频词	教师课堂高频词	教师言语比例 Rt	师生言语转换率 Ch	师生记忆性言语	表达情感
2	学生独立发言总次数	教师发言总次数	学生言语比例	混合型	师理解性言语	鼓励表扬
3	学生独立发言总时长	教师发言总时长	教师语言动态图	对话型	师生运用性言语	回应
4	学生提问总次数	教师提问总次数	学生语言动态图	练习型	师生分析性言语	提问
5	学生提问总时长	教师提问平均等待时间	沉寂混乱语言动态图	讲授型	师生评价性言语	讲授
6	男、女生独立发言总次数	问题价值度	对话深度一		师生创造性言语	指令
7	男、女生独立发言总时长	教师追问15问	对话深度二		课堂问题 SOLO 分类	评价(批评)
8	男、女生提问总次数	教师五何问题分类	对话深度三		目标素养	应答
9	男、女生提问总时长	教师布鲁姆问题六级分类	对话深度四		情景	主动
10	男女生提问比例	教师4MAT问题分类(8)	对话深度五		问题化五力	沉寂混乱
11	学生回答次数		师生提问时间分布图			沉寂混乱比率
12	生生追问次数		师生追问总次数			教师提问比率
13	学生追问次数		师生发言交替次			内容十字区比率
14	学生补充次数		师生言语次数比			学生主动与学生应答比率
15	学生反思次数		师生言语时长比			教师间接与直接影响比率
16	学生追问15问		师生提问次数比			教师积极影响与消极影响比率
17	学生五何问题分类					
18	学生布鲁姆问题六级分类					
19	学生4MAT问题分类(8)					

注:学生4MAT问题分类(8)代表该指标包含8个下级指标。

表4 以学习为中心的课堂循证模型

一级维度	二级指标	三级行为数据
主动学(主动意愿)	学生主动回应(评价一),学生主动提问(评价二)	回答、补充、提问、追问等行为数据
互动学(对话状态)	体现师生互动(评价一),关注生生互动(评价二)	师生互动、RtCh课堂风格PSLias问题解决互动分析模型
深度学(认知深度)	指向问题解决(评价一),迈向高阶思维(评价二)	布鲁姆、SOLO、五何问题、15问、对话深度等
高效学(目标达成)	聚焦核心问题(评价一),落实学科素养(评价二)	学生提问、追问等与学业成绩显著相关的行为数据
多样学(学习方式)	学习方式多样态(评价一),思维类型多元化(评价二)	学习行为类型、互动方式、问题解决五力模型、问题系统等

表5 以学习为中心的课堂循证指标体系

目标	一级指标	学生表现 A	教师表现 B	课堂互动 C	课堂结构 D	课堂认知 E	PSLias(F)
主动学	高参与	A2~A11	B2~B5	C3、C4、C11、C16			F2、F7、F9、F14、F16
互动学	高互动			C1、C2、C13~C15	D1		F1、F4~F6、F10~F13、F15
深度学	高思维	A12~A16	B6、B8	C6~C10、C12		E1~E6、E10	
高效学	高效能	A1、A18	B1、B9			E7~E9	F8
多样学	高样态	A16、A17、A19	B7、B10		D2~D5	E10	F3

价值;循证层以“五学”课堂循证模型为核心,关联课前、课中、课后的数据质量,实现课堂质量评价的循证闭环。系统功能模块结构及关系见图3。

从数据采集到数据分析,系统的计算均需根据业务场景的数据特点和需求,融合多种人工智能技术解构多源异构高结构化数据(见表6)。

7	循证层	教学设计循证				作业系统循证				
		互动学	有效学	主动学	深度学	多样学				
6	资源层	学校教学策略库		名师课堂		问题系统		知识结构		
5	支架层	课堂对比		数据看板		EduGPT		云同侪		
		课堂分析		名师课堂		教学反思		数据治理		
4	分析层	问题解决		思维发展		素养表现		…		
		知识结构		课堂结构		互动特征		认知结构		
3	计算层	五何	4MAT	布鲁姆	15问	RtCh	ST	PSLias	FIAS	IRE
2	数据层	数据清洗			数据分类			数据标签		
1	采集层	视频、音频			常态化课堂采集			录播教室		
区域数字基座										

图3 人工智能课堂分析系统

三、研究方法过程

本研究以指标体系中的“五何问题分类”为例,阐述课堂分析的具体方法与过程。“五何问题分类”(即表1中的学生表现A17和教师表现B8)由国内著名学者祝智庭(2001)提出,旨在通过问题的分析分类,观察课堂学与教过程中师生的思维发展过程和状态。这些指标是“五学”循证模型中体现深度学(迈向高阶思维)和多样学(思维类型多元化)的重要指标,也是落实高阶思维发展、促进核心素养落地的核心计算模块。

(一)数据来源

研究样本是上海市宝山区项目实验学校的常态化课堂实录素材,课堂数据集由2124节课组成,其中,1439节小学课,685节初中课。本研究通过人工智能技术、NLP技术对上述课堂素材进行语音识别、文字转录、问题提炼,呈现师生对话生成的问题,共得到110252条问题话语。

(二)编码框架

本研究基于问题化学习理论对五何问题的定义(王天蓉,2023)编码数据。各种问题类型的编码、定义和示例见表7。

数据分析双向编码器BERT(Bidirectional Encoder Representations from Transformers)是一种预训练的语言表征模型,它基于Transformer架构实现对文本的深度理解。该表征模型不再采用传统的单向语言模型或简单拼接两个单向模型的方法进行预训练,而是通过新的掩码语言模型MLM

(Masked Language Model)对双向的Transformers进行预训练,生成深度双向语言表征,以更好地理解词语在语境中的含义,全面、准确地捕捉文本的语义关系。

本研究选用chinese-bert-wwm-ext预训练模型并基于已经标记的问题语料对语言模型进行参数调整,从而获得问题分类模型(见表8)。

在分类问题中,num_classes指数据集中不同类别的总数;epoch指模型完整地经历一次训练的数据集,包括前向传播和反向传播过程;batch_size是指在单次模型参数更新中使用的样本数量;padding的设置旨在以矩阵形式批量处理数据,提高计算效率;max_length为设定输入序列的最大长度,超过此长度的文本将被截断,而短于此长度的文本将被填充,以确保输入的一致性。学习率是控制模型参数更新步长的关键参数,将影响模型训练的效率 and 稳定性;优化器是一种算法,用于调整模型参数,使函数损失最小化,提高模型性能。这些参数相互作用,确保模型有效地从数据中学习并做出准确预测。

基于上述编码框架,12名经过培训的研究人员独立完成问题话语编码。在编码过程中,任何不一致的结果都经过了讨论和修正,确保数据的一致性和准确性。基于这一过程,本研究构建了用于机器学习的初始数据集,并按照60%训练集、20%验证集、20%测试集的比例进行划分,以便于训练和评估模型(见表9)。

表6 融合多种人工智能技术解构数据

数据处理及教学分析模型	应用人工智能模型
常态化语音采集(边缘端)	多通道处理技术、小波变换等
语音角色识别	机器学习:支持向量机 SVM
语音转文字	深度学习:循环神经网络(RNN)
关键词提取	NLP:分词
问题提取	基于规则的方法
IRF	
PPT 课件提取	机器视觉:卷积神经网络 CNN
行为提取	
情感提取	深度学习:BERT 编解码+Transformer 架构
五何模型	
布鲁姆教学目标分类	
问题化 15 问、追问	
弗兰德斯 FIAS	
……	……

表7 “五何”问题分类编码

分类	编码	知识类型	定义	例句
是何	Q1	关注概念	是何(what)问题通常指向事实或概念,探讨具体的事实内容和事物的本质等。	树叶一般是什么颜色的?——事实 什么是正义?——本质
为何	Q2	关注意义	为何(why)问题通常指向动机与价值,探讨目的与理由。	树叶为什么一般都是绿色的?——原由 你为什么这样做——目的和意义
如何	Q3	关注应用	如何(how)问题通常指向方式与途径,具体探讨方法、程序、手段与路径。	你是如何做这件事的?——方式
若何	Q4	关注创造	若何(if)问题通常指向应对改变,即具体探讨情境条件发生变化时如何解决问题。	如果农村没有人了,会是什么样的结果?
由何	Q5	关注情境	由何问题(context question)通常指向历史根源,溯源问题发生的条件、来历、起因,通过分析问题产生的情境,认识历史沿革。	黄金分割点 0.618 由何而来?——根源
其他	Q6	其他问题	上述五类问题外的问题,如教师确认类问题	是不是,对不对等

表8 调整参数表

参数	值	说明
num_classes	6	五何类型数
epochs	4	训练轮次
batch_size	64	批大小
padding	max_length	每个问题长度对齐为 max_length
max_length	128	每个问题处理成的长度
learning_rate	1.00E-05	学习率
optimizer	AdamW	优化器

(三)分析结果

历经模型训练、参数微调和指标评价等一系

列过程,本研究获取了“五何”问题分类水平结果(见表10)。其中,准确率(Precision)指在所有被预测为正的样本中,实际为正的样本所占比例。召回率(Recall)指在所有实际为正的样本中,被正确预测为正的样本的比例。F1值作为准确率和召回率的调和平均数,在机器学习是个至关重要的性能指标,它能够在准确率与召回率之间达成平衡,对分类模型的性能进行综合评估。

经过严格验证,本研究证实了课堂“五何”问题分类模型具有良好的信度和效度。模型与人工编码的结果高度一致,显著提升了课堂问题分类和

表9 “五何”问题数据集划分

类别	训练集	验证集	测试集
为何	13084	2617	2617
如何	6103	1221	1221
若何	4022	805	805
是何	25082	5017	5017
由何	1132	227	227
其他	29325	5865	5865
总计	78748	15752	15752

表10 “五何”问题分类模型评估结果

类别	准确率	召回率	F1分数
为何	0.8846	0.8789	0.8817
如何	0.8909	0.8026	0.8445
若何	0.7632	0.7205	0.7412
是何	0.9388	0.9169	0.9277
由何	0.6100	0.5374	0.5714
其他	0.8333	0.8525	0.8428
平均值	0.8201	0.7848	0.8016

循证分析的效率。这一模型有助于提升课堂观察和评价的精准性、广泛性和快速性。

四、实践路径与应用模式

2022年12月至2024年6月,上海市宝山区在全区范围内启动了人工智能课堂分析项目。该项目深入161所学校,涵盖从幼儿园到高中的全学段全学科的1708位教师,共分析了10109节课。在上述大规模实践的基础上,本研究构建了两类人工智能课堂分析路径,并提炼了三种应用模式。

(一)纵横交错,构建人工智能课堂分析应用路径

本研究一方面针对实践对象,从微观—中观—宏观三层次,分别聚焦课堂、班级、学科、学校、区域等应用场景,纵深改进课堂;另一方面,基于实践功能,从诊断—反思—研究三梯度,借助分析/诊断—反思/改进—研究/治理的应用赋能教师发展(见图4)。

基于人工智能技术的课堂分析指向实现数据驱动的教师教学成效分析与精准化的教师专业发展,并最终促进个性化的课堂改进、智能化的课堂诊断和大规模教学数据治理(蔡慧英等,2023)。从

理论指导和技术干预的视角看,人工智能课堂分析一是要充分挖掘技术优势,实现多元化的课堂数据收集,包括课例数据、师生对话数据、班级学业表现数据等,更好地生成教师画像,促进教师进行课例反思和专题研究,建立“分析/诊断”“反思/改进”“研究/治理”为主要功能的应用;二是要充分释放数据治理的潜能,分别从微观(教师个体)—中观(校本教研)—宏观(区域治理)层面推动教师发展,从点到面提升人工智能时代教师的课堂教学能力。

(二)多元驱动,形成人工智能课堂分析应用模式

本研究提炼出了计算教育学视域下人工智能课堂分析的三类应用模式。

1.“双向驱动”模式:以大规模课堂质量评估为例

为推动基于数据的大规模课堂质量评估,本研究先通过理论驱动,构建评估课堂质量的基本维度与指标体系,然后通过数据驱动,建立各学科、学段、学校、区域的课堂质量常模值,形成课堂改进的数据起点。在此基础上,教师开展总结性反思、对比性反思、批判性反思,学校和区域逐步建立精准化、个性化、系统化的课堂质量提升方案,助力课堂循证实践。某初中针对11门学科进行常态化课堂数据采集与分析,涉及85名老师477节课,包括教师有效提问数量、教师高阶提问数量、学生有效提问数量、学生高阶言语数等11个维度(见表11)。基于上述课堂分析数据统计结果(见表12,图5),教学人员可以基于课堂分析统计数据进行教学质量的反思和研究,找到明确、适切的课堂改进方向,实现理论与数据的双向驱动。

2.“数据驱动”模式:以教师课堂提问能力预测与评估为例

成熟教师与新手教师的课堂教学技能和水平存在差异,这主要体现在教师课堂教学话语中,如提问的类型(“五何”问题的分布)、提问的深度(布鲁姆六级认知问题)、对学生的反馈等(如追问、补充、复述、聚焦、转化、搁置、整理)。本案例对74名成熟教师(区教学能手)与150名新手教师课堂录像的行为数据进行挖掘,构建了混合效应Logit模型,预测与评估教师课堂提问能力,并在此

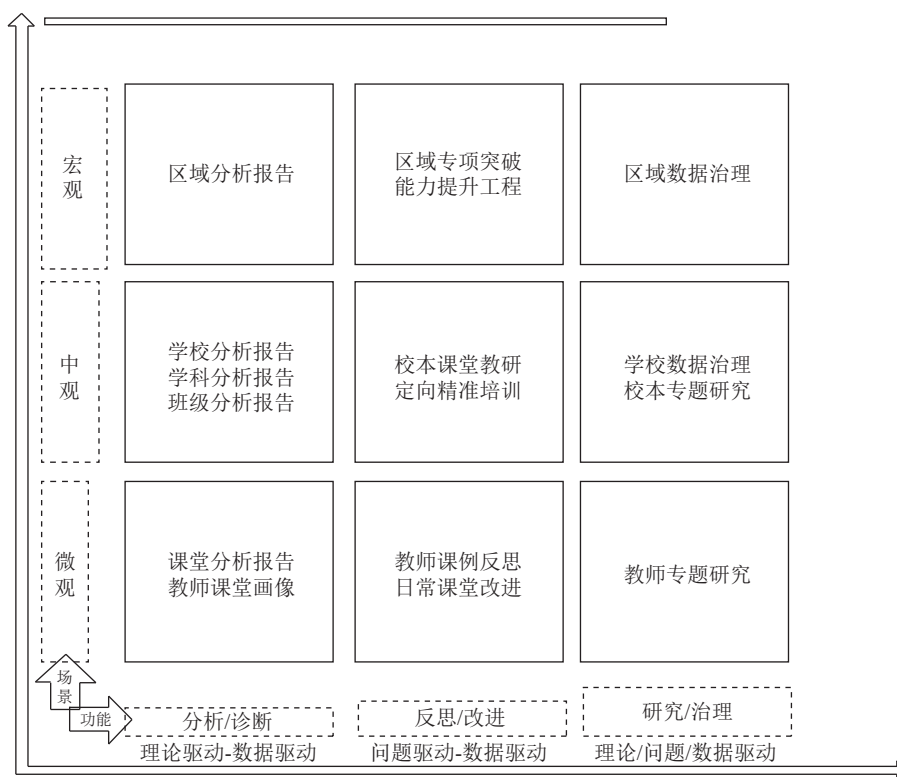


图 4 人工智能课堂分析应用路径

表 11 学科常模值对比

学段学科	教师有效提问数量	教师高阶提问数量	教师高阶言语数量	课堂互动 CH 值	学生有效提问数量	学生高阶言语数
校级均值	48.88	7.46	5.01	0.32	3.19	0.91
小学语文	↓ 37.86	↓ 6.30	↓ 4.88	↑ 0.34	↑ 5.64	↑ 1.31
小学数学	↑ 61.30	↑ 10.28	↓ 4.67	↑ 0.36	↓ 3.02	↓ 0.60
小学英语	↓ 28.49	↓ 1.24	↓ 1.02	↓ 0.30	↓ 0.78	↓ 0.09
小学科学	↓ 36.07	↑ 9.07	↑ 5.14	↓ 0.29	↓ 2.14	↓ 0.79
小学道德与法治	↑ 52.70	↓ 6.30	↑ 6.40	↑ 0.35	↑ 7.60	↓ 0.90

表 12 学科教师常模值对比

学段学科	学生有效提问数量	ST 学生占比 (%)	学生言语深度	学生“五何”种类数	学生高阶言语数
校级均值	3.19	45.46	4.22	1.16	0.91
一(3)班	↓ 1.40	↑ 48.50	↓ 2.20	↓ 0.80	↓ 0.00
一(2)班	↑ 4.07	↑ 50.67	↓ 3.80	↑ 1.53	↓ 0.13
一(1)班	↑ 3.83	↑ 50.33	↓ 3.83	↑ 1.83	↓ 0.33

基础上提出针对性课堂改进建议。考虑到不同学段和学科背景对教师教学方式的影响,该模型还设置了学段和学科作为随机效应。

1)模型公式与说明

模型设计旨在实现两个关键功能:评估教师教学能力和提供教师能力提升建议。因而模型选择

不仅要考虑预测和评估教师教学能力的准确性,还要考虑其可解释性,以便根据模型参数诊断教师课堂提问行为和层次的不足。因此,本研究放弃使用常见的机器学习分类模型,如 SVM 和神经网络,采用 Logit 线性模型捕捉行为变量对教师教学成熟度的普遍影响。本项目利用 Python 的 statsmodels 包

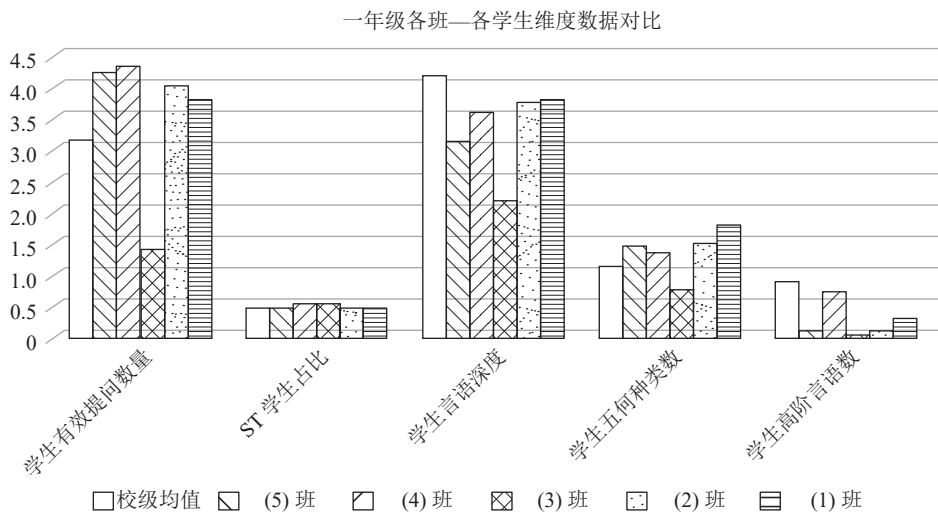


图5 班级常模值对比

进行拟合,并对参数的估计值、标准误差、显著性水平等进行评估,获得以下公式:

$$\begin{aligned} \text{Logit}(P(\text{成熟教师})) = & \beta_0 + \\ & \beta_1 \times \text{学生提问数量(是何)} + \\ & \beta_2 \times \text{老师提问数量(是何)} + \\ & \beta_3 \times \text{学生提问数量(为何)} + \\ & \beta_4 \times \text{老师提问数量(为何)} + \\ & \beta_5 \times \text{学生高阶提问数量(如何、若何、由何)} + \\ & \beta_6 \times \text{老师高阶提问数量(如何、若何、由何)} + \\ & \beta_7 \times \text{布鲁姆(记忆层)的数量} + \\ & \beta_8 \times \text{布鲁姆(理解层)的数量} + \\ & \beta_9 \times \text{布鲁姆(应用层)的数量} + \\ & \beta_{10} \times \text{布鲁姆(分析层)的数量} + \\ & \beta_{11} \times \text{布鲁姆(评价层)的数量} + \\ & \beta_{12} \times \text{布鲁姆(创造层)的数量} + \mu_{\text{学段}} + \nu_{\text{学科}} \end{aligned}$$

其中, $P(\text{成熟教师})$ 是成熟教师的概率, β_0 为截距项, $\beta_1-\beta_{12}$ 为固定效应系数, $\mu_{\text{学段}}$ 和 $\nu_{\text{学科}}$ 分别代表学段和学科随机效应的成分。固定效应系数 β 表示每个自变量对于因变量的平均影响,随机效应 μ 和 ν 则考虑了不同学段和学科对平均影响的潜在变异。

2) 分析结果与建议

本研究的混合效应Logit模型的准确率达90.18%,精确率为96.43%,具体结果见表13。

学生提问数量无论是基础层次的“是何”问题,还是较高层次的“为何”问题、高阶层次的“如何”“若何”“由何”问题,其系数均为正,且都

通过了显著性检验($P < 0.05$)。这表明在课堂互动中,学生提问的频率越高,教师被识别为成熟型的可能性越大。教师提问中仅有为何类型提问的数量系数通过显著性检验($P = 0.05$),这说明教师提问的效果可能取决于多种情境因素,如提问的时机、质量、课堂环境和学生的反应。因此,单纯的数量可能不足以反映其对教学效果的影响。评价层指向布鲁姆认知层次中的高阶思维,学生不仅需要回忆和理解信息,还需要在更深层次上分析、整合和评判信息。仅有评价层系数通过显著性检验且为正,记忆层、理解层、应用层、分析层和创造层均未通过显著性检验。这表明教师在课堂上提出的评价性问题越多,越可能被识别为成熟教师,评价性问题引导的活动能够促进学生深入思考和讨论,同时也是区分教师教学能力的重要参考。

由混合Logit模型的研究结果可见,教师可以从多个维度优化课堂:一是培养学生提问的技能,尤其是关注“为何”“若何”类问题,引导学生深入思考和讨论;二是提升自身提问的技能,特别是深层次和思考性问题,这类问题更能培养学生批判性思维;三是重视评价层次的教学活动,激发学生对知识和学习过程的深层次反思和评价,促进学生的认知与元认知发展。

3. “问题驱动”模式:以教师课堂成长规律探究为例

教师在实际教学中,经常会出现理论与实践不一现象。计算教育学可在更多教学环节、更长教

表 13 Logit 模型分析结果

term	coef	std	z	P> z	[0.025	0.975]
学生提问数量(是何)	0.08	0.03	2.73	0.01***	0.02	0.14
老师提问数量(是何)	0.00	0.03	-0.14	0.89	-0.05	0.05
学生提问数量(为何)	0.14	0.03	5.37	0.00***	0.09	0.18
老师提问数量(为何)	0.05	0.03	1.96	0.05**	0.00	0.11
学生高阶提问数量(如何、若何、由何)	0.11	0.03	3.46	0.00***	0.05	0.17
老师高阶提问数量(如何、若何、由何)	-0.02	0.03	-0.83	0.41	-0.08	0.03
布鲁姆(记忆层)的数量	-0.03	0.02	-1.27	0.21	-0.08	0.02
布鲁姆(理解层)的数量	0.04	0.03	1.21	0.23	-0.02	0.09
布鲁姆(应用层)的数量	0.02	0.03	0.87	0.39	-0.03	0.07
布鲁姆(分析层)的数量	0.03	0.02	1.31	0.19	-0.02	0.08
布鲁姆(评价层)的数量	0.07	0.03	2.70	0.01***	0.02	0.12
布鲁姆(创造层)的数量	0.03	0.02	1.13	0.26	-0.02	0.07
1 学段	0.005	0.05	0.10	0.92	-0.10	0.11
1 学科	0.062	0.02	2.90	0.00***	0.02	0.10
Group Var	0.018	0.07				

注: * $P<0.1$, ** $P<0.05$, *** $P<0.01$.

学周期、更多教学维度、更客观的过程性数据支持下,以解决实际问题为驱动,进行数据挖掘与影响因素分析,从而发现潜在规律,促进理论与实践的互动发展。

1) 问题显现

Rt-Ch 教学模式分析是一种常见的定量分析方法, Rt 代表一节课中的教师行为占有率, Ch 代表师生行为转换率。通过分析这两个值,本研究将课堂教学类型分为讲授型、混合型、练习型、对话型四种。在以学习为中心的教育理念引导下,教学管理者通常要求教师课堂实践的 Rt 值必须小于 0.7 甚至 0.5, 凸显学生主体地位,但这种刻板方式可能脱离教学实际并违背教师成长规律。

2) 数据收集与分析

针对上述悖论,本研究分析了长周期(一学期)的教师日常课堂教学,包括 853 个新授课样本,涉及初中教师 36 人,小学教师 63 人,其中区级骨干教师 10 人,校级骨干教师 10 人,普通教师 79 人。

由数据(见图 6)可知,混合型($0.3<Rt<0.7$; $Ch<0.4$)占 50.29%,为课堂主体类型;讲授型($Rt\geq 0.7$; $Ch<0.4$)占 23.33%,为第二大类型;对话型课堂和练习型课堂依次为第三、四种类型。后续研究需要进一步挖掘三类教师群体的课堂教学类型特征,分

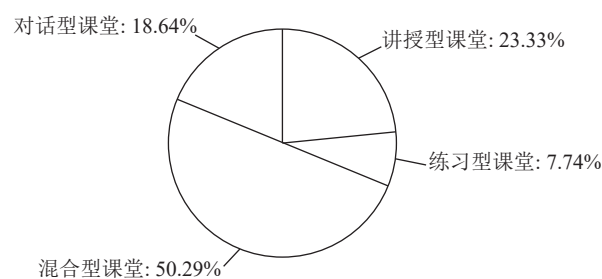


图 6 课堂教学类型占比

析教师教学能力与课堂类型之间的关系。

3) 教师成长规律探究

本研究利用单因素方差(见表 14)分析上述三类教师的课堂教学类型差异发现,不同教师的课堂教学类型呈 0.05 水平显著性($F=3.026$, $P=0.049$)。

本研究对数据进行小提琴图分析发现,图中弧线为核密度分布对称展示,中间部分展示 2.5%、25%、50%、75% 和 97.5% 分位数(见图 7)。由此可见,普通教师的课堂结构更多呈现混合型特征;校级骨干教师课堂类型以混合型为中心,两端是讲授型和对话型课堂。这表明该群体能驾驭更多类型的教学手段;区级骨干教师总体呈向对话型课堂发展趋势,这说明他们更加关注课堂互动中学生的学习建构(见图 7)。这些规律的发现,不仅可以客观地帮

表 14 三类教师课堂教学类型差异分析

	教师群体(平均值±标准差)			F	P
	区级骨干教师(n=120)	校级骨干教师(n=100)	普通教师(n=633)		
课堂结构	2.84±1.08	2.52±1.13	2.62±1.01	3.026	0.049*

注: * $P<0.05$, ** $P<0.01$.

助教师分析自己的教学风格,也可以为教师的专业发展提供指引。

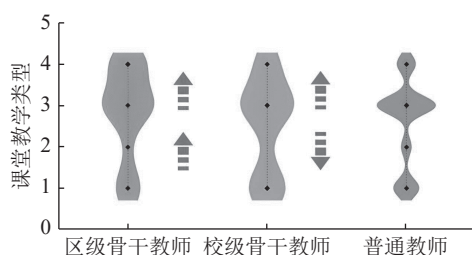


图 7 三类教师群体的课堂教学类型占比分析

注: 课堂教学类型中的数字 1 代表讲授型, 2 代表练习型, 3 代表混合型, 4 代表对话型。

以上三个案例,是运用基于计算教育学的人工智能课堂分析的典型模式。常模值数据的采集、积累与分析,有助于建立课堂质量的基础特征,为计算教育学提供数据标准的计算来源,而进一步的数据挖掘,有助于开展学与教的相关预测与评估,进而发现潜在的规律,揭示“已知的未知”。

五、余论

课堂的过程性评价一直是教育研究的难点,传统的定量与定性研究很难实现课堂整体性和动态性分析,难以刻画课堂教学的总体面貌与复杂过程。人工智能技术通过采集课堂学与教过程的大规模、伴随式、多模态教学数据,有助于全面客观地记录教学过程。基于计算教育学的课堂研究,除了自下而上的数据驱动,也需要结合自上而下的理论解释,在发现“关联”的同时探索“因果”。而以课堂改进为目的的循证实践研究范式更多的是基于数据驱动与数据的价值判断,这是计算教育学研究范式的重要特征,也是创新人工智能课堂分析范式的重要立场。

人工智能技术为基于计算的课堂分析提供了全新场景。然而,在具体实践中,教师还需基于解释学、现象学的原理进行解释与经验概括。因为

课堂仅仅以“数据说话”还远远不够,需要其他研究范式“接着说”,充分关注数据挖掘的工具性与教育学科的人文性相融合,才能触及教育领域的“真实”和“本然”。

[参考文献]

- [1] Anderson, L. W. (2018). 布鲁姆教育目标分类学(修订版)学习、教学和评估的分类学[M]. 皮连生,主译. 上海:华东师范大学出版社: 1-5.
- [2] Bandura, A. (1977). Social learning theory [M]. New York: General Learning Press: 5-12.
- [3] 蔡慧英,卢琳萌,顾小清(2023). 人机协同教研会促进教师教学反思能力的发展吗?——基于课堂视频智能分析技术的实证研究[J]. 现代远距离教育, (1): 40-49.
- [4] Flanders, N.(1963). Action and feedback: A preparation for teaching[J]. Journal of Teacher Education, (14): 25-26.
- [5] 黄荣怀,周伟,杜静等(2019). 面向智能教育的三个基本计算问题[J]. 开放教育研究, (5): 11-12.
- [6] Hey, T.(2011). The fourth paradigm-data-intensive scientific discovery[J]. Proceedings of the IEEE, (8): 12-14.
- [7] Lazer, D.(2009). Life in the network: The coming age of computational social[J]. Science, 323(5915): 33-35.
- [8] 李政涛,文娟(2019). 计算教育学:是否可能,如何可能?[J]. 远程教育杂志, (11): 12-18.
- [9] 刘三女牙,杨宗凯,李卿(2020). 计算教育学:内涵与进路[J]. 教育研究(3): 152-159.
- [10] McCarthy, B. (2005). Teaching around the 4MAT (R) Cycle: Designing instruction for diverse learners with diverse learning styles[M]. Thousand Oaks: Corwin Publishers: 2-5.
- [11] 王天蓉,徐谊(2023). 问题化学习[M]. 北京:教育科学出版社: 44-55.
- [12] 中华人民共和国教育部(2023). 教育部关于批准 2022 年国家级教学成果奖获奖项目的决定 [EB/OL]. [2024-09-04]. http://www.moe.gov.cn/srcsite/A10/s7000/202307/t20230724_1070571.html.
- [13] 孙众,吕恺悦,骆力明等(2020). 基于人工智能的课堂教学分析[J]. 中国电化教育, (10): 15-23
- [14] 肖思汉,刘畅(2016). 课堂话语如何影响学习:基于美国课堂话语实证研究的述评[J]. 教育发展研究, (24): 45-54.
- [15] 杨晓哲(2021). 基于人工智能的课堂分析架构:一种智能的课堂教学研究[J]. 全球教育展望, (12): 55-65.
- [16] 郑永和,严晓梅,王晶莹等(2020). 计算教育学论纲:立场、范式与体系[J]. 华东师范大学学报(教育科学版), (6): 4-6.

[17] 张治,徐冰冰(2022). 人工智能教育大脑的生态架构和应用场景[J]. 开放教育研究, 28(2): 64-72.

[18] 张敏霞,王陆(2004). 用 S-T 法支持教师案例研习与反思的探索[J]. 教师专业发展与教育信息化, (2): 1-3.

[19] 祝智庭,彭红超(2020). 全媒体学习生态:应对大规模疫情时

期上学难题的实用解方[J]. 中国电化教育, (10): 1-6.

[20] 祝智庭(2001). 现代教育技术——走向信息化教育[M]. 上海: 高等教育出版社: 2-20.

(编辑:李学书)

The Framework and Technical Implementation of Artificial Intelligence Classroom Analysis Based on Computational Education

ZHANG Zhi¹, CHENG Shuyi² & WANG Tianrong³

(1. Baoshan District Education Bureau, Shanghai Future Learning Research and Development Center, Shanghai 201999, China; 2. University of Shanghai for Science and Technology, Shanghai 200093, China; 3. Baoshan District Education Service Center, Shanghai Future Learning Research and Development Center, Shanghai 201999, China)

Abstract: Artificial intelligence-enabled technologies have provided new opportunities for computation-based educational research, propelling classroom teaching analysis into the computational era. Grounded in competency-oriented and student-centered classroom values, this study developed a five-dimensional observational perspective and a high-quality classroom analysis framework, which includes "active learning, deep learning, interactive learning, efficient learning, and diverse learning." The study also established an artificial intelligence classroom analysis technical framework within computational education. It developed an artificial intelligence classroom analysis system with seven major functional modules that included data collection, data processing, computation, analysis, scaffolding, resources, and evidence inquiry, in addition to 154 learning analytics indicators. Based on these, the study conducted classroom analysis with a sample of 10,109 classes from 161 schools in Baoshan District, Shanghai, creating a three-tier "micro-meso-macro" approach to classroom improvement and a three-step "diagnosis-reflection-research" trajectory for empowering teacher development. Additionally, it has formed three types of artificial intelligence classroom analysis application models: "theory-driven," "data-driven," and "problem-driven."

Key words: computational education; classroom analysis; artificial intelligence evidence-based practice; teaching evaluation