

# 多模态学习分析：“多模态”驱动的智能教育研究新趋向\*

王一岩, 王杨春晓, 郑永和

(北京师范大学 教育学部, 北京 100875)

**摘要:** 智能感知技术的发展为学习分析领域的变革提供了多元的数据支持, 实现智能技术与教育研究的深度融合, 推动多模态学习分析的研究进展。该文立足于人工智能时代“多模态”的价值内涵, 从多模态信息感知、多模态数据表征、多模态计算方法三个方面实现了“多模态”意蕴的多元分解。论述了多模态学习分析的历史沿革和研究意义, 从多源性、层次性、时序性、情境性四个层面分析了多模态学习分析的潜在特征。人工智能时代多模态学习分析研究的开展需要从面向智能精准测评的学习者建模、面向智慧学习空间的教育情境建模、面向教育生态治理的教育系统建模三个方面加以重视, 实现多模态学习分析与真实教育场景的深度融合。未来多模态学习分析研究的开展需要从多学科融合的基础理论体系建构、多模态数据集的创建共享、学生认知结构的量化表征、基于多模态学习分析的智能服务体系以及数据安全和伦理道德规范等方面加以重视。

**关键词:** 多模态; 多模态感知; 多模态数据; 多模态机器学习; 多模态学习分析

**中图分类号:** G434

**文献标识码:** A

近年来, 随着眼动仪、脑电仪、功能磁共振等智能感知设备的发展, 以及人工智能领域自然语言处理、计算机视觉、语音识别、生理信息识别等智能技术的成熟, 为智能教育带来了新的发展契机, 有助于实现面向智慧学习空间的多模态数据采集, 挖掘教育过程的潜在规律, 为认知诊断、情感计算、交互分析、情境感知等研究的开展带来多元化的数据支持, 实现多模态数据驱动的教育现象的解释和教育规律的发现, 创设智能技术支持的智能教育服务模式, 推动教育研究逐渐向数据化、科学化、精准化的方向发展。本文重点关注多模态的理论与方法与学习分析领域的深度融合机制, 从“多模态”的价值内涵出发论述多模态学习分析的价值意蕴、应用目标和研究要点, 理清多模态学习分析对智能教育时代学习分析研究开展的多元化应用前景, 以期为后续相关研究的开展提供借鉴意义。

## 一、多模态

### (一)从“模态”到“多模态”

模态(Modality)是“物质媒体经过社会长时间塑造而形成的意义潜势, 是用于表征和交流意义的社会文化资源”<sup>[1]</sup>。关于“模态”的代表性观点主要有以查理斯为代表的符号系统说和以克瑞斯为代表的交互方式说<sup>[2]</sup>。前者建立在社会符号学的基础上, 认为“模态”是“可以被具体的感知过程解释的社会符号系统”<sup>[3]</sup>, 如: 声音、图像、文字等; 后者认为“模态”是“人类通过感觉器官建立的与外部环境之间的交互方式”, 如视觉、听觉、触觉、嗅觉、味觉等。近年来, 随着人工智能领域前沿研究的快速发展, 赋予其新的价值意蕴, 相关学者从数据感知模式和数据表征方式的角度将“模态”定义为机器对外界信息的感知模式或信息通道<sup>[4]</sup>, 主要包括三个方面: 其一, 数据表征模式。按照人工智能的细分研究领域进行划分, 将“多模态”定义为融合文本、语音、视频、生理信息的数据表征模式; 其二, 数据采集机制。将“多模态”与多传感器相关联, 将每种传感设备采集到的数据视为一种“模态”, 如: 生理信息识别中基于眼动、脑电、红外等智能

\* 本文系教育部科学事业费重大项目“教育科学研究战略基地培育”(项目编号: 28500/211700001)阶段性研究成果。

① 郑永和为本文通讯作者。

传感设备采集的生理信息数据；其三，数据特征主体。体现为对特征主体局部信息的数据化表征，如计算机视觉领域对人的眼睛、嘴巴等表情特征的描述和对头部姿态、骨架等肢体特征的描述均可视为一种“模态”。对智能教育领域“多模态”意义的诠释要建立在“模态”概念深入解析的基础上，结合人工智能背景下数据感知融合与智能分析的发展现状，对“多模态”的意义加以阐述，构建基于多模态数据感知与融合的智能教育应用机制，增强不同模态信息之间的互补性，构建多模态的符号表征体系，实现基于多模态数据融合的精准确分析。

人是多模态学习的总和，通过视觉、听觉、味觉、嗅觉等感官系统建立与复杂环境的交互机制，进而经由大脑皮层完成对外界信息的完整意义建构。人工智能作为一门模拟人类智能对外部信息进行加工、处理和意义建构的技术科学，其基本业务逻辑与人类对外部信息的感知、获取和处理过程较为相似，均包含外部世界的符号表征体系以及智能实体(人类智能和机器智能)对外部世界的信息感知通道和意义建构模式。因此将智能教育领域的多模态划分为多模态的“符号表征体系”、多模态的“信息感知通道”和多模态的“意义建构模式”三个层面。多模态符号表征体系是对真实教学场景的多元分解，包含构成完整教学情境的“学习者—教师—教学内容—教学媒体—教学活动—教学环境”等要素，从学习分析的角度可以进一步拆解为与学习者学习过程和认知发展相关的“行为—认知—情感—交互”等层面，进而构建面向真实教学场景的符号表征体系，包含“语音—表情—动作—文本—呼吸—心跳—体温—眼动—脑电—皮肤电—激素水平”等方面；多模态信息感知通道是机器对外界环境的感知系统，旨在利用智能传感设备实现对教学情境的多元感知，通过多通道信息的感知与融合生成面向智能分析的多模态数据，服务于智慧教学分析的实际需要；多模态的意义建构模式指利用人工智能神经网络模拟人类加工处理信息的计算过程，对相关信息进行有机整合，提取相关符号表征状态潜藏的语义信息，根据先前研究经验和现有符号表征模式，生成与教学活动的开展有关的信息、知识和智慧，从而完成对多模态信息的完整意义建构，指导相关教学研究工作的开展。

## (二)“多模态”意蕴的多元分解

### 1. 基于多媒体学习认知理论的多模态信息感知

多媒体学习认知理论的双通道假设认为，“学习者的认知发展是大脑对言语信息和非言语信息双重编码的结果，借助言语系统和非言语系统的内部

连结形成对事物完整的意义建构”<sup>[5]</sup>。多媒体学习认知理论认为学习的发生是学习者利用视觉通道和听觉通道对外界信息进行加工和处理的过程，将学习发生过程分解为“感知注意、信息组织和信息整合”三个阶段，重视对语言表征信息和视觉表征信息的选择性加工以及与先验知识的主动整合<sup>[6]</sup>。图文理解整合模型进一步强化了学习者知觉和认知层面之间的复杂作用关系<sup>[7]</sup>，强调知识建构的过程是通过不同通道对图、文信息进行获取和加工，基于言语信息和视觉信息的整合形成命题表征和心理模型，再由命题表征和心理模型的交互整合以完成知识信息的整体建构<sup>[8]</sup>。认知神经科学领域的相关研究表明，学习者对于外部信息的感知是视觉、听觉、触觉等信息感知通道与大脑共同作用的结果，外部符号表征体系通过不同感知通道信息的相互补充在学习者大脑进行神经成像，形成学习者对事物的完整感知。大量研究表明，“视—听”融合的多模态信息加工机制对于人类感知和理解世界具有重要意义，基于多模态的信息整合，实现不同感知通道的内部连结和信息互补，进而完成完整的意义建构。基于多媒体学习认知理论的多模态信息感知对于多模态学习分析的开展具有良好的现实意义，有助于构建人类智能和机器智能之间的桥梁<sup>[9]</sup>，帮助机器更好地模拟人类对外界多模态信息的感知和加工模式，构建面向智慧学习空间的多模态符号表征体系，利用智能感知技术构建面向多模态特征的信息加工通道，进而挖掘多模态数据潜藏的语义信息，完成对数据潜在信息的完整意义建构。

### 2. 基于人工智能技术的多模态数据表征

当前阶段，人工智能领域的发展已经由基于神经网络和遗传算法的计算智能阶段迈向以文本、语音和图像等符号系统表征的语义信息的识别和加工为核心的感知智能阶段<sup>[10]</sup>。以计算机视觉、语音识别、自然语言处理、生理信息识别、平台数据采集技术为代表的人工智能前沿研究领域的发展为教育数据的智能感知提供了多元的技术支持。基于计算机视觉技术对智慧教育场景中学生的表情、动作等图像信息进行智能化的采集，用以分析学生的课堂表现，对学生课堂学习过程中的专注度、情感状态等信息进行智能识别；基于语音识别技术，对学生课堂发言中的话语信息进行采集，根据学习者的语音语调信息和话语信息，分析学习者的认知发展状况和情感状态；基于自然语言处理技术，从语义层面分析对学习者的信息进行深入的挖掘分析，提取其中潜在的观点和情感信息，基于语义网络和知识图谱分析学习者的认知发展状况；基于

脑电感应、眼神追踪等生理信息识别技术,对学习者的眼动、脑电、皮肤电、激素分泌等数据进行采集,为学习分析相关研究的开展提供多样化的生理数据支持;基于平台数据采集技术,对学习者在智慧教学平台的学习过程进行精准监测,形成面向学习者个体的在线学习流数据,根据学习者的检索、浏览、观看、测试数据分析学习者的学习偏好和知识掌握程度,为其提供智能化的认知诊断和学习资源推荐服务。通过对学生多源异构数据的采集分析,形成面向学习者的多模态数据集,利用智能化分析方法对学生深层次的认知和情感状态进行精准测评,实现面向学习者的精准化学习分析。

### 3. 基于机器学习技术的多模态计算方法

多模态数据为我们感知和理解世界提供了多元的信息支持,对于多模态数据的采集和处理也是数据科学家们研究的关键问题,衍生出了多模态机器学习<sup>[11]</sup>这一旨在利用机器学习技术处理和关联多种模态信息的前沿研究领域。其中多模态融合是近年来多模态机器学习领域的一个重要研究方向,在多模态情绪感知<sup>[12]</sup>、多模态交互分析<sup>[13]</sup>等前沿研究领域具有广泛应用,能够为多模态学习分析研究的开展提供良好的技术解决方案。多模态融合旨在将从不同的单模数据中提取的信息集成到一个紧凑的多模态表示中,利用多模态数据的信息互补机制对复杂环境中的多元信息进行完整表征。根据其融合阶段和融合机制的不同可以分为以下三种模式:数据级融合是将从文本、视频、语音等多种模态信息中提取的特征组合起来成为单一的特征矩阵,然后输入到机器学习的分类器中,在模型训练过程中通过相同模态和不同模态之间的特征组合,以充分挖掘多模态数据的潜藏信息,缺点在于需要在模型训练之前把不同模态的特征调整为同一格式,在此过程中容易造成数据的失真;特征级融合是指将不同模态的数据映射到高维空间,再从模型的中间层选取适当的位置进行融合操作,主要包括基于简单操作的方法、基于注意力的方法和基于张量融合的方法,由于其操作的灵活性,逐渐成为多模态融合领域研究的重要方向;决策级融合指对不同模态特征采用最适合的模型分别进行训练,对所得到的结果进行加权处理以获得最佳的决策支持<sup>[14]</sup>。

## 二、多模态学习分析

### (一)多模态学习分析的发展沿革

关于多模态学习分析的研究由来已久,2012年相关学者提出将文本、视频、音频等多模态数据运用于多模态交互研究领域,为多模态学习交互领域

相关研究的开展提供了新的研究方向,也奠定了多模态学习在交互分析领域的核心价值;2012年的多模态交互国际会议(ICMI)首次以工作坊的形式组织相关学者参与“多模态学习分析”的研讨活动,确立了多模态学习分析的前沿地位<sup>[15]</sup>,指出该研究领域旨在将多模态分析技术与学习科学研究相结合,帮助研究者更好地理解学习的发生机制;2016年,学习分析与知识国际会议(LAK2016)设立面向多模态学习分析的数据挑战工作坊,组织领域内相关学者参与到多模态数据分析的实践工作中,探究多模态数据支持的学习分析研究的发展方向<sup>[16]</sup>;次年,LAK2017邀请相关学者做了关于多模态学习分析的主题报告,对基于多种传感设备的多模态生物识别技术的进行了系统阐述,并对其在学习者外在行为表征和内在生理信息的数据采集机制进行了详细论述,提出利用线上、线下一体化的学习分析解决混合学习环境中的关键问题<sup>[17]</sup>;2019年美国教育研究协会(AERA)年会提出利用文本、图表、视音频等多模态数据构建基于多模态叙事的研究范式,为教育研究寻求证据公平<sup>[18]</sup>。

### (二)多模态学习分析的意义解读

多模态学习分析的概念最早由Stefan Scherer等人在第十四届多模态交互国际会议上提出<sup>[19]</sup>,认为多模态学习分析是多模态教学与学习、多模态数据、计算机支持的分析三个概念的复合体,旨在利用三个概念之间所形成的三角关系来描述或模拟复杂学习环境中的学生学习<sup>[20]</sup>。人工智能时代对于多模态学习分析研究范畴的界定,需要立足于多模态学习分析研究开展的实际需要,搭建学习分析理论和人工智能技术之间的桥梁,利用智能传感设备和智能分析技术探究学习生态的全貌,深化学习分析的研究底蕴,促进学习分析研究朝着数据化、科学化、智能化的方向发展。

随着人工智能领域的发展和技术的成熟,赋予了学习分析新的价值意蕴,使得相关研究的开展不再局限于传统的学习平台数据的采集分析层面,加强对不同教学场景的关注,实现多元学习空间的深度融合,探究学习者行为、认知、情感状态的变化,揭示深层次的学习发生机制。在理论层面,借助多模态数据实现多学科、宽领域的建模分析,参照“脑科学、认知神经科学、数据科学”等领域的研究成果,帮助机器更好地模拟人类的感知、计算和意义建构模式,探究学习者外在行为表征模式与内在认知发展之间的潜在作用关系,推动学习分析领域基础理论体系的变革;在技术层面,将自然语言处理、计算机视觉、语音识别、生理信

息识别技术和平台采集技术相结合,扩展学习分析的数据感知通道,实现基于文本、语音、视频、生理信息、在线学习行为等数据的多元采集与融合,为多模态学习分析的开展提供了多元的数据支持,利用多模态机器学习技术实现对多模态数据的融合分析,借助多模态数据间的信息互补机制,拓宽学习分析的广度和深度;在应用层面,多模态数据为学习分析研究的开展提供了广泛的证据支持,实现多模态数据驱动的行为分析、认知诊断、情感计算、交互分析和情境感知,实现多模态数据赋能的学习现象的解释和学习规律的发现。

### (三)多模态学习分析的特征分解

#### 1.多源性:教学信息的多源互补

教学信息的多源互补是多模态学习分析相关研究开展的前提,通过对多模态教学数据的采集分析,强化不同模态数据之间的信息互补机制,有效还原教学生态的全貌。多模态学习分析的多源互补体现在信息加工层面的多源互补和数据感知层面的多源互补两个方面。前者表现为“视—听”结合的双通道编码,能够帮助学习者形成完整的意义建构,如:中国传统文化中的“察言观色”,表示通过观察说话人的语调、表情能够帮助倾听者更好的理解言语信息表达的核心思想;后者体现在数据分析层面的视觉和听觉信息互相补充,能够对模型的训练提供多元的数据支持,如:在多模态情绪识别任务中,基于语音和视频的多模态数据能够取得比单模态数据更高的精确度<sup>[21]</sup>。通过对多场景、多学科、多时序的教学信息进行智能化的采集分析,实现教学信息的多源互补,利用量化学习的思想对教学场景中的信息进行数据化表征,实现教学生态全貌的有机分解,为智能化学习分析研究的开展提供数据支持。

#### 2.层次性:教学生态的层次分解

多模态学习分析的层次性体现在利用多模态数据对一定学习周期内“教育主体、教育客体、教育情境、教育资源”进行精准监测,对与教学过程和教学结果相关的潜在特征进行有机分解,理清其中的层次作用机理,构筑基于多模态学习分析的智慧教学过程层次分解机制。(1)教学情境的层次分解。利用多模态数据构建智慧教育视域下智慧学习空间物理环境、社会环境、心理环境的层次分析机制,对影响学习者的环境要素进行多元分析,为智慧学习环境的创建提供支持;(2)学习者模型的层次分解。通过对智慧学习环境下学习者的行为、生理、心理信息的监测,探究学习者的外在表征和内隐状态之间的潜在作用关系,分析学习者的认知、情

感、交互发展状况,构建基于多层次数据表征的精准化学习者模型;(3)交互模式的层次分解。通过对多模态数据的精准分析,构建面向智慧学习空间中“人—机—物”的交互分解机制,挖掘深层次的教育发展规律,服务于智能化教学分析研究的开展;(4)教学过程的层次分解。构建面向智慧课堂“教师行为、学生行为、教学媒体、教学内容、教学活动”的序列化分析模型,研究教学过程的各构成要素对教学过程和教学结果的影响机理。

#### 3.情境性:教学情境的智能感知

教学情境的智能感知是学习分析领域需要解决的关键问题,也是多模态学习分析相关研究的重要组成部分。按照Korhonen等人的情境划分方法,可以将情境分为用户情境、任务情境、环境情境、社会情境、时空情境、设备情境、服务情境和网络连接情境八种类型<sup>[22]</sup>。教育情境的复杂性为智能教育相关研究的开展带来了极大挑战,如何对教育情境进行精准建模,实现教育情境的可计算是未来一段时间内智能教育领域相关研究人员需要解决的关键问题<sup>[23]</sup>。利用多模态学习分析的思想,借助可穿戴设备和智能感知技术实现对智慧学习空间中复杂教育情境的多元分解,实现对学习者行为、生理、心理等方面的全方位监测,通过对特定情境下学习者相关表现的智能监测和精准分析,构建面向智能化教学场景的感知计算模式,挖掘复杂教育情境下学习者的行为序列模式和交互发生机制,用以分析教学情境要素对学习者的深层次知识建构和认知发展的潜在作用机理。

#### 4.时序性:教学分析的时空融合

多模态学习分析的时序性是指在构建多模态学习分析机制的基础上,重视与学生学习过程相关的周期性变化规律,构建面向时空融合的多元化教学分析模型。重视完整学习周期中学习者知识、能力、情感的发展变化情况,分析学习者认知和情感的迁移和变化规律,挖掘深层次的学习发生机制,实现基于时间序列的学习者建模分析;强化基于时间序列的教学场景数据的采集和分析,构建多模态、时序性的教学分析模型,挖掘教学内容、教学过程、教学活动相关信息与学生学业发展之间的作用机理,实现面向学习全周期的纵向挖掘分析,构建智能化学习支持服务体系,助力于智能教学工作的开展。

### 三、多模态学习分析的应用场景

#### (一)面向智能精准测评的学习者建模

学习者模型的构建是学习分析领域的重要研究

内容,也是多模态学习分析需要解决的关键问题。利用多模态学习分析技术构建面向个体学习者的全空间、多场景、时序性的建模分析,对智慧学习空间中学习者的表情、动作、呼吸、心跳、语言、皮肤电等数据进行全方位的采集分析,借助多模态数据实现面向学习者“认知—行为—情感—交互”外在表征模式的可计算。结合教育学、心理学、脑科学、认知神经科学的相关研究内容对学习者的内部潜在特征进行系统解释,探究学习者“认知发展状况—行为表达模式—情感发生机制—学习交互模式”的协同进化机理,在更深层次上探究学习者的学习发生机制。

#### 1.实现面向学习者的智能化认知诊断

认知诊断是通过一定学习周期内学习者的学业诊断数据进行持续性的采集分析,对学习者的认知发展状况进行连续性的隐式建模,从而实现对个体认知结构、认知过程和信息加工模式的精准测评<sup>[24]</sup>。利用多模态学习分析技术实现对学习者多时空、多学科、多领域的全方位建模分析,构建面向学习者个体的学科知识图谱和能力图谱,实现对学习者认知发展状况的全方位的精准测评,探究学习者学业发展的影响因素,为学习者制定智能化的学业提升服务。

#### 2.实现面向学习者的智能化行为分析

学习行为是学习者在某种动机指引下为获得某种特定学习结果而选择各种各样的手段去实现学习结果的活动的总和<sup>[25]</sup>。课堂教学环境下学习行为的发生是教师、教学活动、教学环境等情境要素和学习者认知、情感等潜在要素共同作用的结果,如何利用多模态数据实现学习行为的可计算,分析线上、线下、课内、课外等多种学习情境中学习者的行为表征模式,挖掘相关要素对学习者的行为表达模式的影响机理,是多模态学习分析需要解决的关键问题。需要实现基于多通道感知的学习行为建模与理解,对学习者的学习行为进行量化表征,研究学习者认识、情感要素对学习者的外在行为表征的作用机理,挖掘深层次的学习行为发生机理。

#### 3.实现面向学习者的智能化情绪感知

心理学的研究表明,个体情绪的发生机制和表达机制均具有一定的复杂性,强调外界刺激和内部生理、心理和认知状态对于个体情绪变化的共同作用。因此单一模态的情感数据难以对个体的情绪状态进行准确表征,需要利用多种智能传感设备实现对学习者语音、表情、文本、生理信息等多模态数据的精准化采集,对学习者的情感状态进行精准测评,进而实现对学习者情绪外在表征模式和内在发

生机理的全方位建模分析,构建基于个体特征、外界刺激、内部需要的多层次情感表达模型,对学习者的情感发生机制进行多学科、多层次、宽领域的深度诠释。

#### 4.实现面向学习者的智能化交互分析

重视对智慧教育环境下学习者交互模式的挖掘分析,通过对交互参与者的语音、表情、姿态、脑电等可表征的符号信息系统的监测,判断交互主体的心理状态,分析学习过程中学习者与教师、教学资源、教学媒体、教学环境之间的交互机制,挖掘复合交互情境对学习者的认知、行为、情感的影响机理<sup>[26]</sup>,有助于优化教学资源和教学媒体的服务模式,为教师提供个性化的教学建议,促进教学绩效的提升。

多模态学习分析旨在利用学习者的外在表征信息实现对学习者内隐状态的智能化测评,为学习者学习发生机制研究的开展提供多元化的数据支持。后续研究的开展需要关注智慧教育环境下“认知—行为—情感—交互”的协同进化机理,在实践研究开展层面挖掘学习者认知、行为、情感、交互的关联关系,挖掘学习者认知发展和情感变化对学习者的行为表达的影响机制;在理论层面运用脑科学、认知科学、教育学、心理学等多学科的理论知识对学习者的学习发生机制进行精准解读,推动学习科学领域相关研究的开展。

#### (二)面向智慧学习空间的教育情境建模

面向智慧课堂的智能化教学情境感知是多模态学习分析要解决的关键问题,如何利用多模态的思想实现对智慧教育情境的多元分解,模拟智慧课堂的演化模式,对于智慧课堂环境下学习现象的解释和学习规律的发现具有重要意义。

#### 1.实现教学场景构成要素的多元分解

构建面向不同教学构成要素的多模态数据集,通过对“教师—学生—媒体—资源—环境”的全方位、细粒度测评分析,实现课堂教学构成要素的有机分解,利用大数据的分析方法挖掘其中潜在的教育发展规律,实现细粒度的教学场景感知,帮助教师优化教学设计模式,提升课堂教学的成效。

#### 2.实现教育情境信息的多元感知与融合

利用智能感知技术实现对智慧学习空间中“用户情境、时间情境、物理情境、社会情境、心理情境、活动情境”<sup>[27]</sup>的精准刻画,综合考量学生个体、教学媒体、教学活动、教学交互等要素构建的复合教育情境,分析相关构成要素对教师教学策略和学生学习成效的影响机理,实现面向教学全过程的全方位建模分析。

### 3.实现基于教育情境感知的智能教育服务

对特定教学情境下与学习相关的媒体、资源、环境、事件进行精准刻画,分析相关构成要素对教师教学策略和学生学习成效的影响机理,实现面向教学全过程的全方位建模分析,综合考量学生的认知发展状况、情感发生机制、学习交互模式和教师的专业素养、教学方式以及选用的教学媒体和教学内容对学习者的学习成效的影响机理,进而优化智能教育服务模式,创设智能技术驱动的智慧教育新形态。

#### (三)面向教育生态治理的教育系统建模

从系统科学的观点来看,大到区域教育生态系统,小到智慧课堂生态系统,均能够被视作一个系统。如何从系统科学的角度探究各级各类教育系统的的发展规律,实现智能化的教育生态治理,是智能教育领域急需解决的关键问题。多模态学习分析为此问题提出了新的解决方案,利用大数据的思想通过数据化、标签化的方法对系统的构成要素、系统结构进行精准刻画,进而模拟系统的运作模式和演化机理,在最大程度上还原系统的全貌,实现教育系统的精准测评。

##### 1.智慧课堂生态系统建模

智慧课堂生态系统是智能技术与智慧教学深度融合的产物<sup>[28]</sup>。旨在以大数据和人工智能技术为依托,以培育面向未来社会发展的优秀人才为导向,强调现代教育理念与新兴智能技术的协同作用,研究智慧课堂的构成要素、系统结构和运行模式,探究智慧课堂内部的数据流动机制、信息治理机制和生态反馈机制,强调智慧课堂环境下各构成要素的交互作用机制,创设智慧课堂教学文化新形态,促进智慧课堂教学生态的系统变革。

多模态学习分析为智慧课堂生态系统的建构提供了多元的理论和技术支持。利用量化学习<sup>[29]</sup>的思想实现智慧课堂生态系统构成要素的多元分解,基于多模态数据的感知与融合,实现对智慧课堂生态系统构成要素的精准化建模分析;构建时空融合的多模态数据表征体系,模拟智慧课堂的数据流动机制,理清相关构成要素对学习生态变革的作用机理,实现多模态数据驱动的智慧课堂教学生态的精准治理;利用智能技术实现智能化的教育情境感知和交互作用模式,模拟智慧课堂的运行模式,理清智慧课堂生态系统中“人一机一物”的交互作用机制,为智慧课堂生态系统的建构提供理论支持。

##### 2.区域教育生态系统建模

从生态学视角来看,教育活动的发生和学生个体的发展依托于区域、学校、家庭、社会为学生成

长创设的各级各类教育环境,大到区域的社会经济文化发展状况,小到班级的文化氛围都会对学生的成长产生影响。智能教育时代区域教育生态系统的构建需要利用多模态学习分析的思想对区域教育发展过程中全时空、多领域的数据进行采集和融合,构建深度分析模型,评估区域教育发展过程中环境要素和人的要素之间的作用机理,利用教育生态学的方法揭示教育生态系统的发展规律和演变机理,构建面向区域教育生态系统的精准体系,保障教育生态系统的良性发展。

多模态学习分析为破解生态学视域下区域教育生态系统的发展规律和演化机理奠定了基础,利用数据科学的思想在最大程度上还原区域教育生态系统全貌,运用生态学的理论和方法对区域教育发展过程中“教育系统与教育环境、教育投入与教育产出、教育主体和教育客体”的作用机理进行精准分析,推动数据驱动的区域教育生态变革。实现对区域人口、环境、资源、教育经费投入、教师人才结构以及学生家庭社会经济文化状况等方面的全方位建模分析,研究相关构成要素与区域教育产出之间的协同作用关系,探究复杂教育系统的演化机理,实现多模态数据驱动的区域教育生态治理,为智能化教育决策的制定提供支持。

#### 四、多模态学习分析的发展进路

多模态学习分析旨在利用“多模态”的思想和方法对智能教育领域的关键问题进行深度诠释,进而挖掘教育规律、优化学习过程、创新智能服务,推动学习分析研究体系的系统变革,探究“多模态”驱动的智能教育研究新趋向。本文结合智能教育领域已有研究从多模态学习分析的基础理论体系建构、多模态数据集的创建共享、学生认知结构的量化表征、基于多模态学习分析的智能服务体系以及数据安全和伦理道德规范等方面对多模态学习分析的研究进路加以论述,以期为后续相关研究的开展理清方向。

##### (一)多学科融合的多模态学习分析理论体系建构

多模态学习分析研究的开展旨在从系统科学的角度对学习者的学习情境、学习过程等要素进行有机分解,探究教育生态系统的发展规律和演化机理,利用数据科学的方法实现对教育现象的解释和教育规律的发现。其研究内容涉及教育学、心理学、数据科学、脑科学、认知神经科学等多个领域。因此,需要从多模态学习分析研究的实际需要出发,构建多科学融合的智能教育理论体系,为后续研究的开展理清方向。其一,借助多媒体学习认知理论探究



人类对外界信息的感知、注意和理解模式,构建人类智能和机器智能之间的桥梁,帮助机器更好地模拟人类对外界信息的感知和理解模式,运用多模态的数据采集和分析方法,实现面向教学主体和教学过程的全方位、多层次的建模分析,构建基于多模态数据的感知、加工、理解和应用模式。其二,借助量化学习理论从数据层面剖析学习发生的表征形态,运用数据科学的思想对学习过程进行全方位的建模分析,突出大数据“建模、分析、预测”的技术功用,构建面向智慧教育发展的数据化认知方式,为学习者提供个性化的学习支持服务,构建面向未来教育发展的精准教育服务模式。其三,借助经典教育理论诠释教育发展规律。教育是一个复杂系统,多模态学习分析研究的开展涉及学习者建模、教育情境建模、教育系统建模等多个层次,相关研究内容涉及学习者行为、认知、情感、交互以及智慧课堂教学生态系统和区域教育生态系统发展规律的诠释,需要立足于多模态学习分析研究的实际需要,运用多学科的研究思想对教育规律加以诠释,构建多学科融合的理论体系,实现复杂教育系统中教育规律的发现和解释。

#### (二)面向真实教学场景的多模态数据采集与分析

从长远来看,要推动多模态学习分析研究的开展并在领域内形成规模化研究体系,需要立足于我国教育教学实践研究开展的实际情况,参照国外先进教学经验,面向真实的教学场景,利用智能感知设备进行多模态数据采集,构建跨越时空的多模态数据融合机制,实现对真实教学场景中相关要素的精准刻画。同时,加强多模态学习分析数据集的开放力度,鼓励领域内相关学者积极参与多模态学习分析的相关研究,形成规模化的研究体系,共同推动多模态学习分析研究的开展。

多模态数据的融合分析旨在对同一时空情境下学习者的语音、文本、视频、眼动、脑电等数据进行精准化的采集和融合分析,是多模态机器学习领域研究的重点,也是破解多模态学习分析领域技术壁垒的关键。需要探究多模态数据的对齐和融合机制,借由多模态数据之间的信息互补,提升学习分析的准确度,充分借鉴人工智能领域成熟的技术模型,构建面向单一模态的数据感知通道,再借由多模态融合算法实现多模态信息的深度融合,进而提升多模态学习分析的准确度。

#### (三)基于多模态数据的学生认知结构量化表征

人工智能领域相关研究的最终目标是在最大程度上帮助机器模拟人的感觉、思维模式,借助算力的提升和算法的优化实现智能感知与预测。智慧

学习空间中关于人的行为、认知、情感、交互的量化表征是多模态学习分析相关研究开展的重点,如何在保证数据真实有效的前提下对学习者的行为模式、认知状态、情感态度进行准确表征,挖掘学习者认知情感状态的协同进化机理,对多模态学习分析研究的开展带来了极大挑战。需要在认知神经科学和量化学习理论的指导下,实现面向真实教学场景的实践探索,利用多模态学习分析技术对学习者和学习过程进行准确表征,挖掘其中潜藏的教育发展规律;结合教育学、心理学、脑科学、认知神经科学的研究内容,对智慧学习空间中的学习发生机制进行深层次的挖掘分析,构建面向智慧学习空间的符号表征体系和信息加工通道,利用智能分析技术模拟其中的意义建构模式,实现学生认知结构的量化表征。

#### (四)基于多模态学习分析的智能教育服务模式

多模态学习分析的研究愿景是通过多模态数据的采集、分析、应用、反馈,赋予传统学习分析新的价值意蕴,拓展教育科学研究的广度和深度,创设多模态学习分析支持下的智慧教育新形态。智能技术的引入能够为教学过程的改进和教学结果的优化提供全方位支持,促进智能技术与真实教学场景的有机融合,构建基于智能化数据“采集、分析、应用、反馈”的数据闭环,强调智能技术支持下的人与机器的协同发展,强化“教学智慧、学习智慧、数据智慧”的深度融合,为教学决策的制定和学习计划的生成提供支持,有助于创新教育科学研究范式、优化智能教育服务模式,实现数据驱动的教育系统精准治理。(1)实现系统精准学习测评。借助智能诊断技术,对学生的“认知发展状况—行为表达模式—情感发生机制—学习交互模式”等方面进行精准化的建模分析,为学习者提供个性化的资源推荐和学习路径规划服务;(2)推动智慧课堂精准教学。通过对教师教学行为、教学内容、教学资源、教学媒体以及学生表现的全方位建模,分析课堂教学的构成要素对学生学习表现的影响机理,为教师提供智能化的教学设计策略,创设数据支持的智慧课堂教学模式<sup>[30]</sup>;(3)创设课堂精准治理体系。实现面向教师、学生、教学环境、教学媒体、教学资源的时序性、细粒度建模分析,挖掘物理环境、社会环境、心理环境等要素对学习过程和学习结果的影响机理,分析教学成功的构成要素,形成数据启发式的教学决策模式,优化智慧课堂精准管理的体制与机制,实现面向智慧课堂的精准治理;(4)重构智能教育评价模式。教育改革发展新时期,完善教育评价体系,变革教育评价机制,是现代教育教

学快速发展的现实需求,也是智能时代优化教育供给模式、完善教育服务体系的重要依据。需要从宏观层面构建智能时代教育评价的逻辑框架,实现面向“教师、学生、学校、区域”的全方位测评,实现数据驱动的精准化教育评价模式,推动评价方法的科学化;(5)变革教育科学研究范式。多模态数据的采集为教育研究的开展提供了多元的数据支持,对教育实证研究的开展带来了新的发展契机,强化数据科学与教育研究的深度融合,实现数据驱动的“循证式”教育科学研究<sup>[31]</sup>,探究深层次的教育发展规律,为教育教学实践活动的开展提供有益指导。

#### (五)面向多模态学习分析的数据安全和伦理道德规范

教育数据伦理是大数据时代教育领域关于道德重构的哲学研究<sup>[32]</sup>,关注大数据从业人员的道德信念和行为规范,通过制定相应的法律法规和行业标准,保障数据采集分析工作的开展能够真正推动行业的良性发展。多模态学习分析研究开展需要在脑科学、认知科学、数据科学的指导下,借助多种智能传感设备对学生进行多模态的数据采集,探究学习者的认知、行为、情感发展状况,进而为教师、学生、管理者提供智能化的教育支持服务。所需数据的多样性、时序性、层次性将在极大程度上揭示个体学习发生的内在机理,对人工智能时代的数据安全和伦理道德问题带来极大挑战。如何保证学习数据不被不法分子获取,保障学习者的隐私不被披露,需要在政策层面加强教育大数据的立法工作,制定行业发展的职业道德规范,构建面向实践层面的数据脱敏机制,明确教育主体的数据权力,形成行业工作人员可以遵循的行为规范,使得数据安全问题不会阻碍人工智能时代教育科学研究的开展。

#### 五、总结和展望

多模态学习分析对于学习分析研究的开展具有多元的价值意蕴,能够在极大程度上推动认知诊断、情感计算、交互分析、场景感知、学习者建模等研究领域的发展,为智能教育领域研究的开展提供理论和技术支持,有助于在更深层次上探究人工智能时代的学习发生机制,促进智能教育领域研究体系的变革。在智能技术快速发展的时代背景下,强化认知神经科学理论与智能感知技术的深度融合,推动多模态学习分析在教育研究实践中的广泛应用,促进其理论体系和技术模型的完善,将在极大程度上拓展学习分析的研究范畴,促进智慧教育生态的变革。后期将在现有研究的基础上,强化对情感计算、情境感知、交互分析等具体应用领域的

探索,挖掘多模态学习分析的深层次价值内涵和技术服务模式,加深对多模态学习分析机制的理解,推动多模态学习分析领域的长足发展。

#### 参考文献:

- [1] Kress G, Leeuwen T V V. Multimodal discourse: the modes and media of contemporary communication [M]. London: Edward Arnold, 2001.
- [2] 王慧君, 王海丽. 多模态视域下翻转课堂教学模式研究[J]. 电化教育研究, 2015, 36(12): 70-76.
- [3] Kress G. Multimodality: A Social Semiotic Approach to Contemporary Communication [J]. Journal of Pragmatics, 2010, 43(14): 3624-3626.
- [4] Lahat D, Adali T, et al. Multimodal data fusion: an overview of methods, challenges, and prospects [J]. Proceedings of the IEEE, 2015, 103(9): 1449-1477.
- [5] Clark J M, Paivio A. Dual Coding Theory and Education [J]. Educational Psychological and Review, 1991, 2(3): 149-210.
- [6] Mayer R E. Applying the science of learning to medical education [J]. Medical Education, 2010, 44(6): 543-549.
- [7] Schnotz W, Bannert M. Construction and interference in learning from multiple representation [J]. Learning and Instruction, 2003, 13(2): 141-156.
- [8] 李松清, 赵庆柏等. 多媒体学习中图文加工的认知神经机制[J]. 心理科学进展, 2015, 23(8): 1361-1370.
- [9] 张琪, 李福华等. 多模态学习分析: 走向计算教育时代的学习分析学[J]. 中国电化教育, 2020, (9): 7-14+39.
- [10] 肖睿, 肖海明等. 人工智能与教育变革: 前景、困难和策略[J]. 中国电化教育, 2020, (4): 75-86.
- [11] Baltrušaitis T, Ahuja C, et al. Multimodal Machine Learning: A Survey and Taxonomy [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(2): 423-443.
- [12] Soleymani M, Garcia D, et al. A survey of multimodal sentiment analysis [J]. Image and Vision Computing, 2017, 65: 3-14.
- [13] 田阳, 陈鹏等. 面向混合学习的多模态交互分析机制及优化策略[J]. 电化教育研究, 2019, 40(9): 67-74.
- [14] Wollmer M, Weninger F, et al. YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context [J]. IEEE Intelligent Systems, 2013, 28(3): 46-53.
- [15] 钟薇, 李若晨等. 学习分析技术发展趋向——多模态数据环境下的研究与探索[J]. 中国远程教育, 2018, (11): 41-49.
- [16] 李香勇, 左明章等. 学习分析的研究现状与未来展望——2016年学习分析和知识国际会议述评[J]. 开放教育研究, 2017, 23(1): 46-55.
- [17] 吴永和, 李若晨等. 学习分析研究的现状与未来发展——2017年学习分析与知识国际会议评析[J]. 开放教育研究, 2017, 23(5): 42-56.
- [18] 吴怵, 彭晓玲等. 教育研究的去芜存菁之路: 从多模态叙事到证据公平——美国AERA 2019年会述评[J]. 远程教育杂志, 2019, 37(4): 13-23.
- [19] Scherer S, Worsley M, et al. 1st international workshop on multimodal learning analytics [A]. Proceedings of the 14th ACM international conference on multimodal interaction [C]. New York: ACM, 2012: 353-356.
- [20] Worsley M, Abrahamson D, et al. Multimodal learning analytics [A]. Proceedings of the 12th International Conference of the Learning

- Sciences, ICLS [C]. New York: ACM, 2016. 1346–1349.
- [21] Hossain M S, Muhammad G. Emotion recognition using deep learning approach from audio-visual emotional big data [J]. Information Fusion, 2019, 49: 69–78.
- [22] Korhonen H, Arrasvuori J, et al. Analysing user experience of personal mobile products through contextual factors [A]. Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia [C]. Limassol, Cyprus: ACM, 2010. 1–10.
- [23] 田阳, 万青青等. 多空间融合视域下学习环境及学习情境探究[J]. 中国电化教育, 2020, (3): 123–130.
- [24] 黄涛, 王一岩等. 智能教育场域中的学习者建模研究趋向[J]. 远程教育杂志, 2020, 38(1): 50–60.
- [25] 李松林. 课堂教学行为分析引论[J]. 教育理论与实践, 2005, 25(4): 48–51.
- [26] Andrade A. Understanding student learning trajectories using multimodal learning analytics within an embodied–interaction learning environment [A]. Proceedings of the Seventh International Learning Analytics & Knowledge Conference [C]. Vancouver: ACM, 2017. 70–79.
- [27] 黄志芳, 赵呈领等. 基于情境感知的适应性学习路径推荐研究[J]. 电化教育研究, 2015, 36(5): 77–84.
- [28] 李逢庆, 尹苗等. 智慧课堂生态系统的构建[J]. 中国电化教育, 2020, (6): 58–64.
- [29] 刘三女牙, 李卿等. 量化学习: 数字化学习发展前瞻[J]. 教育研究, 2016, 37(7): 119–126.
- [30] 刘邦奇. 智能技术支持的“因材施教”教学模式构建与应用——以智慧课堂为例[J]. 中国电化教育, 2020, (9): 30–39.
- [31] 管珺琪, 孙一冰等. 智慧教室环境下数据启发的教学决策研究[J]. 中国电化教育, 2019, (2): 22–28+42.
- [32] 刘三女牙, 杨宗凯等. 教育数据伦理: 大数据时代教育的新挑战[J]. 教育研究, 2017, 38(4): 15–20.

#### 作者简介:

王一岩: 在读博士, 研究方向为智能教育、情感计算。

王杨春晓: 在读博士, 研究方向为教育认知神经科学、科学教育。

郑永和: 教授, 博士生导师, 研究方向为智能教育、科学教育、教育信息科学与技术、科技与教育政策。

## Multimodal Learning Analytics: a New Research Trend in Intelligence Education Driven by Multimodality

Wang Yiyan, Wang Yangchunxiao, Zheng Yonghe

(Faculty of Education, Beijing Normal University, Beijing 100875)

**Abstract:** The development of intelligent perception technology provides multiple data support for the reform of learning analysis field, realizes the deep integration of intelligent technology and education research, and promotes the research progress of multimodal learning analytics. Based on the value connotation of “multi-modality” in the era of artificial intelligence, this article realizes the multi-decomposition of “multi-modality” meaning from three aspects: multimodal information perception, multimodal data representation, and multimodal calculation methods. This paper discusses the historical evolution and research significance of multimodal learning analytics, and analyzes the potential characteristics of multimodal learning analytics from four levels: multi-source, hierarchical, sequential, and situational. The development of multimodal learning analytics research in the era of artificial intelligence needs to pay attention to three aspects: learner modeling for intelligent and accurate evaluation, educational context modeling for smart learning space, and educational system modeling for educational ecological governance, so as to strengthen the deep integration of multimodal learning analytics and real education scenarios. In the future, the development of multimodal learning analytics research needs to pay much attention to the following aspects: the construction of the basic theoretical system of multidisciplinary integration, the creation and sharing of multimodal datasets, the quantitative representation of students’ cognitive structure, the intelligent service system supported by multimodal learning analytics and the problem of data security and ethics based on multimodal learning analytics.

**Keywords:** multimodality; multimodal perception; multimodal data; multimodal machine learning; multimodal learning analytics

收稿日期: 2020年8月10日

责任编辑: 邢西深