计算机科学
COMPUTER SCIENCE

课堂师生交互智能分析技术研究综述

崔家郡，康璐，马苗

引用本文

崔家郡，康璐，马苗. 课堂师生交互智能分析技术研究综述[J]. 计算机科学，2024，51(10)：40-49.

CUI Jiajun, KANG Lu, MA Miao. Survey on Intelligent Analysis Techniques for Classroom Teacher-Student Interaction Research [J]. Computer Science, 2024, 51(10): 40-49.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240400084

课堂师生交互智能分析技术研究综述

崔家郡[1]　康　璐[1]　马　苗[1,2]
1 陕西师范大学计算机科学学院　西安 710119
2 现代教学技术教育部重点实验室(陕西师范大学)　西安 710062
(cuijiajun@snnu.edu.cn)

摘　要　随着教育信息化的普及与不断发展,视频、图像、语音、文本等海量课堂数据被记录下来。对这些多模态数据进行有效分析,挖掘课堂师生交互信息,不仅能够帮助教师及时发现教学中存在的问题,及时调整教学内容以提高教学质量,而且是落实“以人为本”教学理念,推进现代教育走向智能化、个性化和数字化的重要手段。文中首先论述国内外师生交互行为的传统分析方法;然后从视频、图像、语音、文本及多模态等不同角度分类梳理课堂师生交互智能分析技术的研究现状,归纳总结课堂师生交互智能分析的核心要素、数据形式、关键技术、结果呈现和应用场景;最后分析课堂师生交互的多模态智能分析技术的优势与不足以及面临的挑战和未来趋势。

关键词:智能教育技术;课堂师生交互;多模态数据;“以人为本”教学理念
中图分类号　TP391

Survey on Intelligent Analysis Techniques for Classroom Teacher-Student Interaction Research

CUI Jiajun[1],KANG Lu[1] and MA Miao[1,2]
1 School of Computer Science,Shaanxi Normal University,Xi'an 710119,China
2 Key Laboratory of Modern Teaching Technology of Ministry of Education(Shaanxi Normal University),Xi'an 710062,China

Abstract　With the popularization and continuous development of education informatization,a huge amount of classroom data such as video,image,voice,text are recorded. How to effectively analyze these multimodal data and mine the classroom teacher-student interaction information can not only help teachers find the problems in teaching and adjust the teaching content in time to improve the quality of teaching,moreover,it is an important link to implement the concept of “human-centered” education and promote modern education towards intelligence,personalization and digitalization. The paper firstly discusses the traditional analysis methods of teacher-student interaction behaviors at home and abroad. Then,it classifies and analyzes the current research status of intelligent analysis techniques for classroom teacher-student interaction from different perspectives,such as video,image,voice,text and multimodal. Next,a technical process for classroom teacher-student interaction intelligent analysis is proposed,including core elements,data forms,key technologies,results presentation and application scenarios. Finally,the advantages and disadvantages of the current multimodal intelligent analysis technology for classroom teacher-student interaction are summarized,as well as the challenges and future directions.

Keywords　Intelligent educational technology,Classroom teacher-student interaction,Multimodal data,“Human-centered” teaching philosophy

1　引言

课堂是教书育人的第一阵地,其教学效果是影响人才培养质量的重要因素。2019年中共中央、国务院印发的《中国教育现代化2035》提出教育现代化的总体目标和主要任务,强调高等教育要坚持以人为本,提高质量,推动教育教学改革。2023年教育部办公厅印发《基础教育课程教学改革深化行动方案》(教材厅函[2023]3号),进一步明确教师应创新教学设计和教学方法,变革教与学方式,尊重学生的主体地位,发挥教师的主导作用,注重启发式、互动式、探究式教学,引导学生主动思考、积极提问和自主探究。

传统的教学模式以教师为中心,以课堂讲授为重点,向学生单向灌输知识,教学效果不尽人意[1],而“以人为本”教学理念是时代发展的必然要求。教育部发布的首个高等教育教学

到稿日期:2024-04-15　返修日期:2024-07-03
基金项目:国家自然科学基金(62377031);陕西省重点研发计划(2023-YBGY-241)
This work was supported by the National Natural Science Foundation of China(62377031) and Key Research and Development Program of Shaanxi Province(2023-YBGY-241).
通信作者:马苗(mmthp@snnu.edu.cn)

质量国家标准《普通高等学校本科专业类教学质量国家标准》标志着“以学生为中心”教育理念已从国家教育政策层面转变为全国高校开展专业教学和质量建设的主要指导原则。因此，如何尊重学生的主体地位，发挥教师的主导作用，及时分析课堂中的师生交互情况，探究如何通过两者的积极互动和有效沟通促进教师更好地了解学生需求，使学生更好地理解教学内容，具有重要的现实意义。

课堂中的师生交互是体现教师和学生在教与学的过程中进行信息传递、沟通的重要形式，一直受到国内外研究人员的关注，并提出了很多方法和理论。

本文综述课堂师生交互智能分析技术，第 2 章论述传统的课堂师生交互方法，重点论述多模态下的课堂师生交互智能分析技术；第 3 章介绍课堂师生交互智能分析技术的核心要素、数据形式、关键技术、结果呈现和应用场景；第 4 章总结课堂师生交互智能分析技术面临的挑战以及未来趋势。

2 课堂师生交互的分析方法

课堂师生交互的研究可分为传统课堂师生交互分析方法和基于 AI 技术的智能分析方法两类，前者主要通过观察、问卷、访谈的形式，由人工完成；而后者借助视频分析、图像处理、自然语言处理技术等智能手段，由计算机自动完成或计算机辅助完成。

2.1 传统课堂师生交互

传统的师生交互主要聚焦在课堂中的师生问答，由教师提出问题、学生回答问题和教师给出评价等环节组成。因此，对课堂中教师和学生的言语行为进行编码分析成为师生交互研究的主要手段。

传统课堂师生交互分析方法分为 FIAS 系统分析法及其改进方法和 S-T 分析法两类。

1）FIAS 系统分析法及其改进。20 世纪 60 年代，Flanders 等提出的弗兰德斯课堂互动分析系统（Flanders Interaction Analysis System，FIAS）是最基础的传统分析方法，其主要贡献在于为师生在课堂中的语言行为以及少数非语言行为设计了一套专用编码模式，如表 1 所列。

表 1 弗兰德斯课堂互动分析系统

Table 1 Flanders classroom interaction analysis system

分类		编码	内容
教师语言	间接影响	1	表达情感
		2	鼓励表扬
		3	采纳意见
		4	提问
	直接影响	5	讲授
		6	指令
		7	批评
学生语言		8	应答
		9	主动
沉寂或混乱		10	无有效语言

通过对师生交互行为的编码、统计以及分析，最终获得直观的行为矩阵，从而实现对课堂交互行为的系统分析。

在 FIAS 系统的基础上，1998 年 Edmund 等深入探索了教师提问与学生反馈之间的关系[2]；2010 年，Moskowitz 对编码规则进行细化，提出 Flint 互动分析系统，为课堂互动分析提供更精准的工具[3]；2013 年，Howe 等对国外近四十年来课堂交互分析领域的丰富研究进行了系统的梳理与总结，同时总结了如何分析课堂中的师生互动、生生互动等交互行为[4]。

在国内，研究人员对 FIAS 系统进行了深入的系列研究，并不断完善，以适应国内课堂环境的实际需求。代表性工作有：1978 年，Lin 率先将 FIAS 系统引入国内的课堂教学分析，并展开细致研究，在国内课堂互动分析领域具有重要意义[5]；2003 年，Ning 等就 FIAS 系统在初中物理课堂中的应用提出以定量的结构性的分析作为研究的系统结构线索，以质的研究提供意义理解和丰富情境的细节，建立数量结构与意义理解的联系的方法论，为课堂互动分析的实践注入了新活力[6]；次年，Gu 等推出 ITIAS 系统，细化学生行为，扩充到 18 种编码，更精确地描述课堂交互过程[7]；2012 年，Fang 等在 FIAS 和 ITIAS 的基础上提出了 iFIAS 系统，将 FIAS 的 10 类编码增至 14 类编码来分析课堂中的师生交互行为[8]；2024 年，Li 等以 5 节数学优课为例，构建智慧教室环境下数学课堂互动教学双编码分析系统，并采用肯德尔和谐系数验证其有效性[9]。

2）Student-Teacher 分析法，简称 S-T 分析法，是国内外教育教学中用来评估和分析学生与教师互动关系的方法，可帮助教师了解学生的需求和行为模式，帮助学生更好地理解教师的教学方法和期望。

该方法由学生行为分析和教师教学策略分析两部分组成，前者用来评估学生的参与度、学习态度、课堂行为和学习习惯，包括观察学生在课堂上的活跃程度、提问频率、小组合作表现以及如何响应教师的指导；后者用来评估教师的教学方法、课堂管理和提问技巧、反馈方式和课堂互动，包括教师如何组织课堂活动、使用教学资源、引导学生思考和解决问题，以及如何适应不同学生的学习需求等。在此基础上，通过师生行为占有率将教学模式分为练习型、讲授型、对话型和混合型 4 种，清晰地呈现课堂互动的结构与特点，为优化教学策略提供科学依据[10]。

Student-Teacher 分析法的代表性工作有：2012 年，Liu 等系统介绍 S-T 分析法的作用及意义，并以网易视频公开课作为案例进行实证对比研究[11]；2013 年，Liu 等以昆明市某初中的地理公开课为研究对象，利用 S-T 分析法分别获取师生课堂行为占有率，判断此课堂类型为对话型教学模式[12]；2014 年，Liu 等以洛阳市举办的全国高中化学优质课为研究对象，利用 S-T 分析法分别获取师生课堂行为占有率，判断课堂类型的教学模式[13]。

综上，传统的课堂师生交互方法大多依赖人工编码、统计和分析获取有效的师生交互信息，不足之处主要体现在以下几点。1）主观性：受到观察者或访谈者主观意识的影响，可能导致结果的不够客观准确。2）真实性：能获取到感兴趣的特定信息，难以获得师生全面的真实体验和感受。3）复杂性：需要较多人力和时间，实施过程受资源和条件限制。4）延时性：无法提供即时反馈，教师难以及时调整教学策略以适应学生需求。5）差异性：收集到的数据常需要复杂的数据处理，研究人员的统计和分析能力不同，导致结果不同。6）片面性：以言语行为交互分析为主，但难以全面理解学生的学习过程。

2.2 课堂师生交互智能分析技术

此类分析方法借助视频分析、图像处理、自然语言处理等人工智能技术，由计算机自动或辅助完成，下面分别从视频、图像、语音、文本及多模态5个角度展开论述。

2.2.1 基于视频的师生交互的智能分析

该技术通过分析课堂教学视频来智能获取师生在课堂上的互动行为和模式，其数据源包括教室内的专业设备录制的课堂教学视频、在线直播平台的教学视频、师生自行录制的课堂视频、教育资源共享平台的教学视频等。

视频数据中常见的师生交互信息有：教师走到同学中间，指导学生讨论、学习和答疑；教师邀请学生示范，学生走向黑板完成课上练习题；学生和教师共同作用于交互式电子白板或平板；教师授课，学生点头或学生回答问题，教师点头；教师授课，学生鼓掌或学生回答问题，教师鼓掌；学生回答问题后，教师摆手等。

相关的智能分析技术包括人体动作识别、时序动作检测、情绪分析等，其中，人体动作识别技术可从视频序列中检测、跟踪、识别出教师或学生，并对其行为进行描述和理解；时序动作检测技术可在给定的一段未分割长视频中检测出师生的动作片段，包括开始时间、结束时间和动作类别；视频情绪分析技术可对视频中师生的情感状态进行自动分析，识别和理解其中蕴含的情感信息。

其主要代表性工作包括：2019年，Li等[14]利用视频分析法针对“练习—反馈”活动样本进行教学交互行为分析，发现平板电脑提高了活动效率，但没有体现在支持个性化学习方面的优势；同年，Chen[15]利用运动目标检测和人体行为特征分类技术设计了一个课堂行为智能图像识别分析系统，可以识别学生的“举手、放下、起立、坐下等”动作和教师的“走动、板书等”动作，以对教学交互行为进行分析；2022年，Xu[16]搭建了面向视频流的学生课堂行为智能识别流程与方法，能够准确识别学生的6种课堂行为并分类量化，根据量化结果分析学生的学习投入度，从而进一步分析教学交互行为质量；2022年，Li等[17]提出基于DNN的多模态学习情绪分析方法，该方法结合视频和语音，实时检测学生的学习情绪，采用PAD情绪量表将学习情绪与学习状态对应起来，教师可以根据学生学习情绪的变化来判断学生学习的参与度和交互情况，从而及时调整教学方法和策略；2023年，Ma等[18]基于信息技术的互动分析编码系统获得课堂师生互动的量化数据，并结合视频图像追问数据背后的师生互动特征，研究发现师生互动主要表现为“教师引发-学生应答-教师反馈”模式；2023年，Liu[19]利用时序动作检测技术设计了R-C3D网络，实现了高中信息技术课堂视频教学环节分割与标注辅助，以及教学环节智能辅助检测、标注和分割。

此类分析方法的不足之处在于分析结果依赖于视频信息，易受以下因素制约：1）录播设备的位置，如高度、角度、成像距离等直接影响到视频质量，师生与成像设备的距离与角度会带来目标的尺度差异和部分遮挡；2）视频录制时的光线条件、背景噪音等都可能干扰到视频成像质量；3）当前研究多聚焦于挖掘帧序列信息，忽视了视频中的声音信号，导致分析结果具有片面性。

2.2.2 基于图像的师生交互的智能分析

该技术通过分析课堂教学图像来智能获取师生在课堂上的互动行为和模式，其数据源包括源自课堂视频的图像、利用拍摄设备直接采集的图像等。

图像数据中常见的师生交互信息有：教师的肢体动作，如起立、请坐、竖大拇指的手势；教师讲授时，学生抬头；教师讲授时，学生注视教师、黑板或PPT等；师生的表情交互，如微笑、皱眉等；教师讲授时，学生的肢体语言，如摇头、举手、记录等；师生距离，判断教师是否亲近学生。

相关的智能分析技术包括目标检测、人脸识别、手势识别、表情识别、人体姿态估计等，其中，目标检测和人脸识别技术可用于检测指定图像中的教师和学生，确定他们的身份和位置；手势识别技术可从图像中分析和识别出师生手部动作及类型，如表示竖大拇指等；表情识别技术可在指定图像中识别出师生表情，进而对其表情进行分类；人体姿态估计技术能够捕获人体手臂、头部、躯干等各身体关节的位置信息，可用于估计师生的动作姿态。

主要代表性工作包括：2018年，Zhou等[20]利用目标检测技术从帧中智能获取人脸数目、轮廓特征、身体动作幅度等信息，实现课堂教学视频中S-T行为智能识别；2022年，Zhou等[21]利用人体姿态估计技术识别学生的典型课堂交互行为，及时反映学生的学习状态；同年，Pabba等[22]利用表情识别技术分析学生的面部表情并将其分类为“无聊”“困惑”“专注”“沮丧”“打哈欠”“困倦”等，研发了学生群体参与度监控的实时系统来实现师生交互行为评价；2023年，Zhou等[23]在AlphaPose的基础上，利用历史帧中的关键点补充当前帧中缺失关键点的思路，通过人体姿态估计技术识别学生课堂行为，分析师生交互行为；同年，Zhang等[24]关注课堂学生举手行为，建立专用数据集并提出举手动作检测算法，通过多分支扩张卷积扩大感受野以减少误检率；Tang等[25]在微课技能培训中利用人脸识别技术确定师生身份，用于后续的师生交互情况分析。

此类分析方法的不足之处在于分析结果依赖于图像信息，易受以下因素制约：1）成像时机非常重要，只有在合适的时间采集图像，才能捕获到更多有价值的课堂教学中的瞬时交互情况；2）完整的课堂教学过程包括课堂师生各时刻的言语、动作、表情等信息，仅通过若干个离散的图像或单张图像无法体现完整的师生交互特点，导致交互分析结果片面化；3）分析结果主要依据师生的肢体动作和表情，忽略了师生交互的言语内容；4）成像设备与师生位置的不同易导致尺度变化、互相遮挡、光照不佳等。

2.2.3 基于语音的师生交互的智能分析

该技术通过分析课堂教学语音来智能获取师生在课堂上的互动行为和模式，其数据源包括利用录音笔等直接对课堂进行录音和对课堂视频进行提取语音信号等。

语音数据中常见的师生交互信息有：1）师生交谈时的对话，即从教师言语到学生言语的转变或从学生言语到教师言语的转变；2）师生交谈时声音的响度、音调、音色等特征；3）师生交谈时幽默、诙谐、严肃等话语风格，启发式、命令式等交互形式，以及停顿、迟疑等回应时间。

相关的智能分析技术包括语音识别、声纹识别、语音情感分析、声纹分割聚类等，其中，语音识别技术可将师生语音转换为文本序列的识别技术；声纹识别技术，也称说话人识别，可通过分析说话人的声音特征来识别师生身份；语音情感分析技术可通过分析语音信号的声学特征来识别师生情感状态；声纹分割聚类技术，也称说话人分割聚类或说话人日志，是一种处理包含多人交替说话的语音的技术，可用来确定一段语音中每个时间点是教师或哪个同学在说话。

主要代表性工作包括：2018 年，Fan[26] 利用声纹识别和语音识别技术识别教师指令，分析课堂语言，研发了课堂师生互动行为分析系统；2020 年，Yang[27] 利用声纹识别和语音情感分析技术对教师课堂语音情绪进行分类、识别、切割，并提出教师课堂情绪互动模型来分析教师的情绪状态、情绪强度和学生的专注率、举手率和疲倦率间的关系，其研究发现，教师积极情绪表达方式越丰富、强度越大，学生的举手行为越多，特别是“积极语音＋肢体”的复合情绪下，学生举手效果最佳；2021 年，Chen[5] 利用声纹分割聚类等技术分析课堂交互结构、课堂情绪转化率等关键数据，对课堂交互进行了量化分析；2021 年，Luo 等[28] 以教师近场语音为研究对象，提出基于时长的 S-T 分析方法，并研发了全自动可视化的课堂教学分析工具，深入剖析教学活动过程；2022 年，Zheng 等[29] 构建了基于音、视频的课堂互动行为分析框架，提出融合 LSTM 和 TDNN 的特征提取网络和基于双头注意力机制的时间池化网络，优化声纹分割聚类算法，提高了课堂互动行为分析质量；同年，He 等[30] 通过语音情感分析技术对教师和学生信息进行情感分类，实现语音情感识别，并将之用于在线学习时的师生情感互动。

此类分析方法的不足之处在于分析结果依赖于语音信息，易受以下因素制约：1）真实课堂环境获取的数据易受背景噪音、拾音效果和语速变化等影响；2）部分方言、少数专业术语可能难以识别；3）语音信号无法体现表情、动作、眼神等非言语信息的师生交互。

2.2.4 基于文本的师生交互的智能分析

该技术通过分析课堂教学文本来智能获取师生在课堂上的互动行为和模式，其数据源包括：通过语音识别技术得到的课堂音频转录文本、线上学习中互动讨论区的文字、对黑板板书进行识别得到的文本等。

文本数据中常见的师生交互信息有：可产生对话文本的师生问答、师生在互动讨论区进行的文字交流，以及教师对学生的反馈与评价。

相关的智能分析技术包括文本分类、文本情感分类、语义相似度计算等自然语言处理技术，其中，文本分类技术可将课堂文本数据自动归类到预定义的类别中；文本情感分类技术可对课堂文本中表达的情感或观点进行自动识别和分类，确定文本传达的情感倾向，如正面、负面或中性；语义相似度计算技术可用于评估两个课堂文本片段在意义上的相似程度，不限于文本字面的相似性，更侧重于深层的语义信息。

主要代表性工作包括：2015 年，Han 等[31] 对互动分析编码系统 ITIAS 进行改进，提出课堂教学互动行为分析编码体系 OOTIAS，用于对教师言语行为、学生言语行为、教师使用技术行为、学生使用技术行为进行量化统计，真实地反映各类课堂交互行为；2020 年，Huang[32] 基于在线课程评价文本构建了在线课程教师教学言语行为分析框架，利用相似度计算方法量化在线课程教学言语与课件内容之间的相似度，判断教师授课是否单纯地“读 PPT”，为交互式教学的量化探究打下基础；2022 年，Chen 等[33] 基于 IRF 话语结构分析框架总结小学课堂中师生言语互动行为比率、师生言语互动行为序列、师生言语对话结构，为提高师生言语互动质量提供技术支撑；2023 年，Shin 等[34] 利用自然语言处理技术对在线代数学习中的教学话语标记进行分析，揭示教师与学生的互动和交流情况；2024 年，Zhao 等[35] 提出基于师生课程交互数据挖掘的情感生成模型，细化在线体育教学中的情感话语，对师生多轮对话中的情感波动进行建模，完成师生情感预测任务。

此类分析方法的不足之处在于分析结果依赖于文本数据，易受以下因素制约：1）文本只能记录师生的言语内容，忽略了语调、表情、动作等非言语交互信息，易导致结果的片面性；2）课堂教学是长时、动态的过程，师生之间的交互行为往往随着课堂情境不断变化，而文本数据只能记录某一时段的交互内容，导致信息存在片面性。

2.2.5 基于多模态的师生交互的智能分析

综上，单一模态的课堂师生交互分析，如语言交流或面部表情，忽视了其他模态在交互中的重要作用。这种局限性导致无法完整把握师生交互的复杂性和多样性。因此，近年来部分研究者开始着眼于基于多模态数据的师生交互行为研究。

主要代表性工作有：2018 年，Susanne[36] 利用多模态分析技术对教师和学生学习模式的设计理论进行研究；2021 年，Wu 等[37] 利用文本分类、人体姿态估计等技术实现课堂教师行为的自动采集、计算、分析与评价，对教师言语行为进行量化计算，辅助教师反思自身在课堂语言中存在的问题，提高课堂教学质量；同年，Yu 等[38] 提出师生多模态融合模型，在模型级别融合了骨架和 RGB 信息，用于识别课堂师生动作；2022 年，Tong 等[39] 从“教学活动、技术使用、位置移动、身体姿态”4 个维度构建 ATMB 智慧课堂教学互动多模态分析框架，从情境性和时序性两方面对智慧课堂教学互动进行多模态分析；同年，Wang 等[40] 利用声纹识别和语音情感分析判别教师声音是积极、消极还是中性，利用表情识别计算教师微笑的次数和长度，利用眼睛凝视估计判断教师视线的方向和学生的注意力，并提出一种师生互动多模态分析框架以分析课堂中师生的语言和非语言互动；2022 年，Li 等[41] 结合多模态话语分析理论和 BSC 的协同、互动特征，构建了课堂师生多模态互动分析框架；同年，Dubovi[42] 使用面部表情、眼动追踪和皮肤电活动等多模态数据，在学生进行基于虚拟现实的模拟学习过程中研究学生的参与度；2023 年，Vivante 等[43] 以以色列中小学科学教师的 PD 课程视频为研究对象，利用言语、姿态、手势、表情等多模态数据分析了 10 名教师约 21.5 h 的课程参与情况。

3 课堂师生交互智能分析关键技术

在教育数字化的时代背景下，为充分利用智能教育技术

自动分析师生教学过程中的互动方式和效果，我们在文献[36-43]的基础上，总结梳理了课堂师生交互智能分析的主要框架，包括核心要素、数据形式、关键技术和结果呈现4个部分，如图1所示。

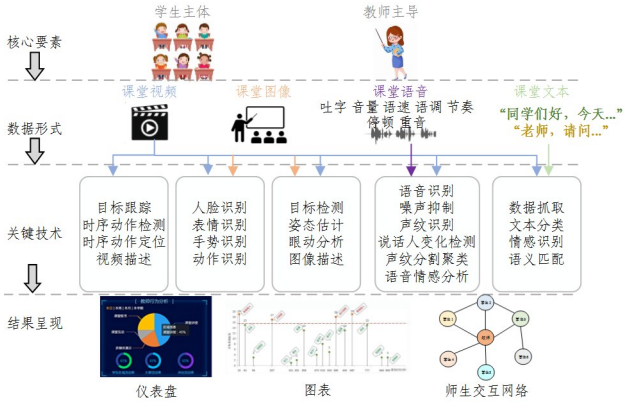

图1 师生交互智能分析技术主要过程

Fig. 1 Process of teacher-student interaction intelligent analysis technology

该框架中，核心要素为学生主体和主导教师，研究对象为两者的多模态交互信息，可用于探究师生交互的视频、图像、语音和文本多模态数据转化、数据形式，关键技术和结果呈现。

3.1 核心要素

教育规划纲要明确要求：把育人为本作为教育工作的根本要求，要以学生为主体，以教师为主导，充分发挥学生的主动性。因此，课堂师生交互研究的核心是教师和学生，分析重点是教与学过程中两者的互动内容与形式。

在教师方面，重点关注教师在课堂教学场景中的言行举止，包括如何运用口头语言、肢体动作、面部表情，甚至眼神变化等多种形式来激发学生的学习动机、传授科学知识、提高学生技能，体现教师如何实现教育育人的目的。

在学生方面，重点关注学生在课堂学习环境中的综合表现，包括课堂反应、参与度以及学习效果等，刻画学生的学习规律与学习状态。

与之不同，课程师生交互聚焦教师和学生之间的互动，包括两者间的知识传递、情感交流，以及认知协同等。

3.2 数据形式

通过数据采集获取的视频、图像、语音和文本等多模态信息，真实记录了教师和每位学生在课堂上的每一个动作、每一句话，甚至是微弱的表情和情绪变化，能够构建一个多模态的课堂师生交互专用数据集，为多模态师生交互智能分析提供数据支持，其数据形式、典型特征及相关智能分析技术如表2所列。

表2 课堂师生交互及智能分析

Table 2 Intelligent analysis on classroom teacher-student interaction

	数据形式	交互方式	典型特征	相关智能分析技术
课堂师生交互及智能分析	视频	动作序列	①头部：点头、摇头、抬头、低头等 ②手部：摆手、鼓掌、指向等 ③肢体：起立、坐下、走向等 ④其他：推搡、谩骂等	目标检测 人脸识别 目标跟踪 时序动作检测 时序动作定位 姿态估计 视频描述
		表情变化	①积极：微笑、好奇、兴奋等 ②中性：平和、淡漠等 ③消极：皱眉、沮丧、愤怒、悲伤、急躁、疑惑、冷漠等	
		言语	①讨论、争论、问答等 ②语速、音色、音调、响度等 ③话语内容及情绪等	
		目光	注视、跟随、移动等	
	图像	表情	①积极：微笑、好奇、兴奋等 ②中性：平和、淡漠等 ③消极：皱眉、沮丧、愤怒、悲伤、急躁、疑惑、冷漠等	目标检测 人脸识别 表情识别 手势识别 动作识别 姿态估计 眼动分析 图像描述
		手势	起立、请坐、竖大拇指等	
		头势	抬头、低头、转头等	
		目光	注视、跟随、疑问等	
		其他	师生距离、身体朝向等	
	文本	言语	①讨论、争论、问答等 ②话语内容及情绪等	数据抓取 文本分类 情感分类 语义匹配
		讨论区	师生间、生生间的文字交流等	
	语音	言语	①讨论、争论、问答等 ②语速、音色、音调、响度等 ③话语内容及情绪等	语音识别 噪声抑制 声纹识别 说话人变化检测 声纹分割聚类 语音情感分析
		话语切换	师生间话语权的转换次数	
		发言时长	发言者身份及时长	

3.3 关键技术

根据图1可知，课堂教学中师生互动形式不同，采集的数据形式不同，涉及的智能分析技术也有所区别，下面分类论述相关的智能计算技术。

3.3.1 人体姿态估计与时序动作检测

1)人体姿态估计

人体姿态估计(Human Pose Estimation)是计算机视觉的重要任务之一，其目标是捕获手臂、头部、躯干等人体各关节的

关键点,以其位置变化和状态特征估计人体的动作姿态。

根据姿态数据的维度,该技术分为 2D 姿态估计和 3D 姿态估计两类,前者是为每个关键点预测一个二维坐标;后者则增加了深度信息,可预测一个三维坐标。目前研究较多的是多人 2D 姿态估计,研究思路有 top-down 和 bottom-up 两种:前者先检测图像中的人体目标,将其从原图中裁剪出来后输入网络中进行人体骨骼关键点检测;后者首先找出图像中人体的所有骨骼关键点,然后对骨骼关键点进行分组,最后获得每个人的姿态。一般认为,top-down 精度更高,而 bottom-up 速度更快。相关数据集有 LSP 数据集、MPII 数据集和 COCO 数据集等。2D 姿态估计常见评价指标有正确关节点百分比(Percentage of Correct Keypoints,PCK)、正确肢体百分比(Percentage of Correct Parts,PCP)、检测关节点百分比(Percent of Detected Joints,PDJ)等;3D 姿态估计常见评价指标有平均每关节位置误差(Mean Per Joint Position Error,MPJPE)、平均每关节角误差(Mean Per Joint Angle Error,MPJAE)等。目前姿态估计技术的挑战在于目标遮挡、光照变化和背景干扰等,其常见模型、主流方法及最新研究进展见文献[44-46]。

2)时序动作检测

时序动作检测(Temporal Action Detection)是视频分析的常见任务之一,其目标是在给定的长视频中检测人体的动作片段,包括开始时间、结束时间和动作类别。

与目标检测类似,该技术包括 two-stage 和 one-stage 两类,前者精度高但计算量大、速度慢,后者计算量小、速度快,但精度不高。相关数据集有 THUMOS2014 数据集、ActivityNet 数据集等。该技术的评价指标包括 mAP(mean Average Precision),IOU(Intersection over Union)等。目前时序动作检测的挑战在于瞬时动作和微弱动作的捕捉和不同专用支撑数据集不足或缺失问题,其常见模型、主流方法及最新研究进展见文献[47-48]。

3.3.2 视频描述与图像描述

1)视频描述

视频描述(Video Captioning)是计算机视觉和自然语言处理领域的一个交叉研究方向,其目标是生成能够准确描述视频内容的文本序列。

按照视频描述结果,该技术可分为单语句描述和密集语句描述,前者用一句话描述视频内容,后者用多句话描述视频内容,更加细致。主流方法有端到端视频描述和多模态视频描述等,前者直接从视频生成文本,包括特征提取、特征编码和解码生成文本等环节;后者涉及视频、音频和文本等多模态信息的融合,更为复杂但文本描述更准确。相关数据集有 MSR-VTT 数据集、Charades-Captions 数据集等。该技术的评价指标包括 BLEU-n(Bilingual Evaluation Understudy),ROUGE-L,METEOR(Metric for Evaluation of Translation with Explicit ORdering),CIDEr 和 SPICE(Semantic Propositional Image Caption Evaluation)等。目前视频描述的研究聚焦于解决时空信息的深度理解、感兴趣目标分析及上下文推理等问题,其常见模型、主流方法及最新研究进展见文献[49-51]。

2)图像描述

图像描述(Image Captioning)是自然语言处理和计算机视觉领域的交叉分支,其目标是理解图像中的内容,并将其转换为简洁、准确的文本序列。

按照图像描述结果,该技术可分为单语句描述和密集语句描述,前者用一句话描述图像内容,后者用多句话描述图像中不同区域的内容,更加细致。相关数据集有 Flickr30k 数据集、COCO 数据集等。该技术的评价指标包括 BLEU-n(Bilingual Evaluation Understudy),ROUGE-L,METEOR(Metric for Evaluation of Translation with Explicit ORdering),CIDEr,SPICE(Semantic Propositional Image Caption Evaluation)等。目前图像描述的研究聚焦于视觉内容中目标关系的深度理解及逻辑推理等,其常见模型、主流方法及最新研究进展见文献[52-53]。

3.3.3 动作识别、手势识别与表情识别

1)动作识别

动作识别(Action Recognition)是计算机视觉领域的一个常见任务,其目标是从视频或图像中对人体动作或姿态进行识别。

根据使用的特征信息,该技术可以分为基于骨架的动作识别和基于时空的动作识别,前者使用人体关键点检测技术来提取人体骨架信息,并基于骨架序列进行动作识别;后者结合外观和运动信息使用 3D 卷积神经网络或特定的时空特征提取方法来识别动作。相关数据集有 UCF101 数据集、HMDB51 数据集等。该技术的评价指标包括准确率(Accuracy)、平均准确率(Mean Average Precision,mAP)等。目前动作识别的难点源于各类场景中动作的多样性和复杂性,个体差异及严重遮挡、视角变化等因素影响,其常见模型、主流方法及最新研究进展见文献[54-57]。

2)手势识别

手势识别(Gesture Recognition)是计算机视觉和机器学习领域的一个任务,其目标是从图像或视频中识别出人类的手势。

按照是否关注时序信息,该技术可分为静态手势识别和动态手势识别两类,前者识别单帧图像中的手势,用于简单的手势命令识别;后者需要识别连续的手势动作,理解手势的动态变化过程。目前这项任务的主要数据集有 MNIST,NVGesture 等。该技术的评价指标包括准确率 Accuracy、召回率 Recall 等。目前手势识别的难点源于手势的一致性、多样性、个体差异性及角度变化等,其常见模型、主流方法及最新研究进展见文献[58-59]。

3)表情识别

表情识别(Facial Expression Recognition)是计算机视觉和情感分析领域的一个重要任务,其目标是从给定的图像中识别出人脸表情,进而对表情进行分类。

根据表情识别的结果,可将其分为“积极、中性、消极”3 类和“高兴、吃惊、悲伤、愤怒、厌恶、恐惧和中性”7 类表情识别任务。此外,微表情识别用于识别短暂且细微的面部表情变化。相关数据集有 CK+数据集、FER2013 数据集等。常见评价指标有 Accuracy,F1 Score 等。目前表情识别的难点

在于表情细微差别判断及光照、视角、成像质量等，其常见模型、主流方法及最新研究进展见文献[60-61]。

3.3.4 语音识别与声纹分割聚类

1)语音识别

语音识别(Speech Recognition)是计算机科学和信号处理领域的一个分支，其目标是将人类的语音转换成机器可以理解的文本。

按照语音片段的连续性，该技术可以分为孤立词识别和连续语音识别两类，前者用于识别独立的、不连续的单词或短语；后者用于识别连续的语音流，包括句子和段落。相关数据集有 TIMIT 数据集、VoxCeleb 数据集等。该技术的评价指标包括词错误率(Word Error Rate，WER)、字符错误率(Character Error Rate，CER)等。目前语音识别的研究聚焦于更多类型的方言识别、强噪音干扰等问题，其常见模型、主流方法及最新研究进展见文献[62-63]。

2)声纹分割聚类

声纹分割聚类(Speaker Diarization)是语音处理的重要技术，其目标是从多人说话的一段语音中，准确地识别出说话人个数并将相同说话人的语音片段合并在一起。

现有的声纹分割聚类框架大致分为基于“分割-声纹-聚类”的分步联合框架和端到端的解决方案。相关数据集有 AMI(Audiovisual Meeting Indexing) 数据集、AISHELL-3 数据集等。该技术的评价指标包括 SER(Speaker Error Rate)，DER(Diarization Error Rate)等。目前声纹分割聚类研究聚焦于背景噪声抑制、说话人聚类算法优化等，其常见模型、主流方法及最新研究进展见文献[64-65]。

3.4 结果呈现

经过智能技术对课堂师生数据进行处理后，需要对多模态交互数据进行结果的可视化，涵盖师生姿态、手势、表情、文本以及语音片段等多个维度，以反映师生间的互动情况。常见的结果呈现方式包括仪表盘、图表信息或互动网络，下面分别论述。

1)仪表盘是一种常见的数据可视化工具，可用于实时监控和评估师生交互状态，即通过整合多种师生交互的数据形式，以性能指标和颜色变化直观地展示当前课堂的活跃度、学生的参与度以及教师的反馈速度等信息[66]。

2)图表信息是目前展示课堂师生交互信息的常用方式，即将师生的姿态、手势、表情、文本和语音等复杂数据转化为易于理解的表格和图例。例如，用柱状图展示不同时间段内师生交互的频次或强度，用折线图反映师生交互随时间变化的趋势[66]。

3)互动网络是社会交互网络在课堂师生交互中的具体应用，即将课堂中的教师和每个学生都视为互动网络中的一个节点，节点间的连线表示任意两者间的交互，用边的权重表示交互的密切程度，以此形成师生互动网络。

构建好师生互动网络后，可以根据互动网络的节点数量、节点权值、边的数量等信息，进一步划分课堂交互类型，如零交互、部分稠密交互、师生稠密交互、均衡充分交互等，如表 3 所列[5]。

表 3 课堂交互的不同类型

Table 3 Different types of classroom interaction

课堂交互类型	交互风格的特点
零交互	教师与学生之间不互动，填鸭式教学
部分稠密交互	教师与部分学生交互较多，而与其他大部分学生几乎不交互
师生稠密交互	师生交互为主，生生互动较少
均衡充分交互	师生之间与生生之间的交互均衡，学生言语总时长与教师言语时长基本相等

3.5 应用场景

课堂师生交互智能分析技术的应用场景广阔，涵盖了教育教学的各个层面，具有重要的实用价值和广泛影响。下面给出部分应用场景示例。

1)智能教室环境建设

课堂师生交互智能分析技术可以与各种智能设备相结合，如智能黑板、智能摄像头等。这些设备能够实时捕捉师生的交互行为，并通过算法进行分析和处理，为教师提供个性化的教学建议，同时也能够为学生创造更具互动性、更生动的学习环境。

2)个性化在线课程

课堂师生交互智能分析技术在一对一等个性化在线课程中发挥着重要作用。教师可以根据学生的个体差异和学习需求，量身定制个性化的学习计划和教学策略。个性化在线课程可以实时捕捉和分析师生的语音、面部表情、手势等多种模态的互动信息，学生能够更好地理解课程知识，教师能够更好地了解学生的学习状态。

3)智能学习助手

基于课堂师生交互智能分析技术可研发智能学习助手，使其通过多模态交互方式实现与学生之间的实时互动和反馈。学生可以通过语音、文字或手势等方式与智能学习助手进行交互，提出问题、表达观点或分享学习成果。智能学习助手则能够理解和分析学生的输入，给予及时的反馈和指导，帮助学生解决问题、深化理解。

4)教学评价与反馈系统

课堂师生交互智能分析技术可以用于构建教学评价与反馈机制。通过对师生的交互行为进行全面、客观的分析，为教师提供关于教学质量、教学方法等方面的反馈，帮助教师不断提升自己的教学水平；为学生提供关于学习进度、学习效果等方面的反馈，帮助学生更好地调整学习策略。

4 面临的挑战及未来趋势

4.1 面临的挑战

课堂师生交互智能分析技术面临着多方面的挑战，部分来自智能技术发展的瓶颈，部分源于实际应用场景限制、数据安全与隐私保护等，具体表现为高质量的数据集不足、智能分析方法的复杂性，以及数据安全与隐私保护的问题。

1)高质量数据集不足

多模态数据的收集和处理是一个复杂的过程。在实际应用中，数据质量和标注的准确性直接影响到分析结果的可靠性，而师生的语音和面部表情可能受到环境噪声、光线条件等多种因素的影响，导致数据质量下降；另一方面，数据标注

需要专业人员完成,人力与时间成本高。

2)智能分析方法的复杂性

真实课堂情况复杂多变,这就要求智能分析方法具有较好的准确性、实时性和鲁棒性,而多模态的师生互动智能分析技术涉及视觉、语音、文本等多模态信息。如何高效地实现各模态信息的语义对齐及信息融合还有一定难度,其决策过程往往难以解释,分析结果有时难以被用户接受和信任。

3)数据安全与隐私保护

在师生交互过程中,课堂师生交互智能分析会涉及师生的个人隐私信息,如身份信息、面部表情、语音内容等。因此,在技术应用过程中需要严格遵守相关法律法规,确保用户的隐私和数据安全。

4.2 未来趋势

1)多模态数据的深度融合。未来的多模态智能技术更加注重跨模态数据的融合与协同,尤其是以沟通内容为核心的视觉、听觉、文本等多种模态的数据语义对齐和特征深度融合。深度融合可以发现更加丰富多元的交互信息,准确理解师生交互的行为及表现。

2)大模型带动课堂师生交互智能技术的创新。随着语言大模型(Large Language Model,LLM)、视觉大模型(Visual Foundation Model,VFM)的技术突破,课堂师生交互智能分析技术的实用性不断提升,短期内有望实现更加全面、精准和人性化的师生交互行为分析手段。

3)个性化与适应性。随着大数据和个性化学习的发展,多模态智能技术将能够根据每个师生的特点和需求进行定制化的分析。通过分析不同师生的交互模式、习惯偏好等,可以提供更符合个体需求的教学建议和反馈,实现更高效的个性化教学。

4)实时性与动态性。未来的多模态智能技术将更加注重实时性和动态性分析,通过实时监测和分析师生的交互行为,可以及时发现教学中存在的问题和不足,并提供及时的反馈和建议,帮助教师和学生调整教与学的策略,提升教学效果。

5)隐私保护与伦理规范。随着人工智能技术的广泛应用,隐私保护和伦理规范将成为未来发展的重要议题。技术开发者将更加注重用户隐私的保护,通过加密、匿名化等手段确保数据的安全性。同时,相关的伦理规范也将不断完善,确保技术的使用合理合规。

结束语 本研究综述课堂师生交互智能分析技术,论述了国内外师生交互行为的传统分析方法,从视频、图像、语音、文本及多模态等不同角度,分类梳理课堂师生交互智能分析技术的研究现状,归纳总结课堂师生交互智能分析的核心要素、数据形式、关键技术、结果呈现和应用场景,最后分析课堂师生交互的多模态智能分析技术的优势与不足以及面临的挑战和未来趋势,体现了智能教育技术在课堂教学中的应用进展。我们相信,随着人工智能、大数据、云计算、物联网、大模型与教育教学的深度融合,智能教育技术在我国教育强国建设中必将发挥重要作用。

参 考 文 献

[1] LU Y Y,CHEN Z Z,CHEN R,et al. Research on the application framework of intelligent technology to promote teachers' classroom teaching behavior evaluation[J]. Modern Educational Technology,2022,32(12):76-84.

[2] EDMUND J,JOHN B. Interaction analysis:theory research and application[J]. Behavior Theories,1967(100):402.

[3] MOSKOWITZ G. Interaction analysis—a new modern language for supervisors[J]. Foreign Language Annals,2010,5(2):211-221.

[4] HOWE C,ABEDIN M. Classroom dialogue:a systematic review across four decades of research[J]. Cambridge Journal of Education,2013,43(3):325-356.

[5] CHEN Y S. Multidimensional classroom interaction analysis based on speech recognition [D]. Wuhan:Central China Normal University,2021.

[6] NING H,WU J H. Establishing a connection between quantitative structure and meaning understanding—an improved application of Flanders interactive analysis technology[J]. Educational Research,2003,24(5):23-27.

[7] JIN J F,GU X Q. Analysis and research on classroom teaching behavior in the information technology environment[J]. China Education Technology,2010(9):82-86.

[8] FANG H G,GAO C Z,CHEN J. Improved Flanders interactive analysis system and its application[J]. China Education Technology,2012(10):109-113.

[9] LI H M,BIAN P,XU M Q. Development of a dual-coding analysis system for interactive teaching of mathematics classroom in a smart classroom environment and mining of behavioral patterns [J]. Modern Educational Technology,2024,34(3):105-115.

[10] SHAN Y J. Using S-T analysis method to analyze classroom teaching problems in educational technology professional courses [J]. Modern Educational Technology,2008(10):29-31.

[11] LIU F,LIU Y,HUANG C Y. Comparative analysis of teaching process based on S-T analysis method—taking NetEase video open courses as an example[J]. China Education Info,2012(11):58-60.

[12] LIU Y,GOU L. Application of S-T analysis method in analysis of geography classroom teaching model[J]. Education of Geography,2013,(6):59-60.

[13] LIU L X,WANG P,HE A N,et al. Research on the application of S-T analysis method in high school chemistry teaching analysis[J]. Education in Chemistry,2014(1):27-30.

[14] LI M F,HE F. Analysis of teaching behavior in the "Practice-Feedback" link in the smart classroom-video analysis based on the 19th provincial first-class mathematics lesson example[J]. Modern Educational Technology,2019,29(6):62-68.

[15] CHEN J T. Research on intelligent image recognition analysis of classroom behavior [D]. Hangzhou:Zhejiang University,2019.

[16] XU J H. Research on intelligent recognition of student classroom behavior based on video streaming [D]. Fuzhou:Fujian Normal University,2022.

[17] LI M,LIU M,JIANG Z,et al. Multimodal emotion recognition and state analysis of classroom video and audio based on deep neural network[J]. Journal of Interconnection Networks,2022,22(Supp4):2146011.

[18] MA L D, CAO Y Z, WU S R. Research on teacher-student interaction in high-quality information-based mathematics classrooms in primary schools based on video case analysis[J]. Journal of Beijing Institute of Education, 2023, 37(5): 44-54.

[19] LIU J L. Research on auxiliary tools for segmentation and annotation of video teaching sessions in high school information technology classrooms [D]. Wuhan: Central China Normal University, 2023.

[20] ZHOU P X, DENG W, GUO P Y, et al. Research on intelligent recognition of S-T behaviors in classroom teaching videos[J]. Modern Educational Technology, 2018, 28(6): 54-59.

[21] ZHOU J, RAN F, LI G, et al. Classroom learning status assessment based on deep learning[J]. Mathematical Problems in Engineering, 2022, 2022: 1-9.

[22] PABBA C, KUMAR P. An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition [J]. Expert Systems, 2022, 39(1): e12839.

[23] ZHOU Y H, ZHANG H Y, WANG Y B, et al. Classroom behavior recognitionalgorithm based on AlphaPose[J]. Information Technology and Informatization, 2023(12): 204-208.

[24] ZHANG G, WANG L, CHEN Z. Hand-raising gesture detection in classroom with spatial context augmentation and dilated convolution[J]. Computers & Graphics, 2023, 110: 151-161.

[25] TANG J, ZHANG P, ZHANG J. Design and implementation of intelligent evaluation system based on pattern recognition for microteaching skills training[J]. International Journal of Innovative Computing, Information and Control, 2023, 1(19): 153-162.

[26] FAN Z J. Design and implementation of classroom teacher-student interactive behavior analysis system [D]. Wuhan: Huazhong University of Science and Technology, 2018.

[27] YANG N Y. Research on the characteristics and effects of teachers' classroom emotional work [D]. Wuhan: Central China Normal University, 2020.

[28] LUO Z Y, ZHAO Q Q, DUAN F Q. Automatic analysis of classroom teaching process based on teachers' near-field speech[J]. Modern Educational Technology, 2021, 31(8): 76-84.

[29] ZHENG Z W, HUANG Y T. Research on classroom interaction behavior analysis algorithm based on audio and video[C]// Proceedings of the International Conference on Control and Computer Vision. 2022: 127-133.

[30] HE Y, GONG Y Q. Improving the quality of online learning: a study on teacher-student interaction based on network multimodal data analysis[C]// Proceedings of the International Conference on Big Data and Education. 2022: 311-318.

[31] HAN H, WANG D Q, CAO C. Analysis and research on classroom teaching interaction behavior in 1∶1 digital environment [J]. E-Education Research, 2015, 36(5): 89-95.

[32] HUANG Y. Analysis of teachers' teaching speech behavior supported by online course data mining [D]. Wuhan: Central China Normal University, 2020.

[33] CHEN M, LI M Y. Research on classroom teacher-student verbal interaction behavior based on IRF discourse structure[J]. Educational Information Technology, 2022, (Z2): 3-8, 30.

[34] SGIN J, BALYAN R, BANAWAN M P, et al. Pedagogical discourse markers in online algebra learning: unraveling instructor's communication using natural language processing[J]. Computers & Education, 2023, 205: 104897.

[35] ZHAO Y, KONG X, ZHENG W, et al. Emotion generation method in online physical education teaching based on data mining of teacher-student interactions [J]. PeerJ Computer Science, 2024, 10: e1814.

[36] SUSANNE K. Contradictory xplorative Assessment. Multimodal Teacher/Student Interaction in Scandinavian Digital Learning Environments[J]. Journal of Education and Training Studies, 2018, 6(2): 133-148.

[37] WU L, CAO Y, DU Q, et al. The analysis path of classroom teacher behavior supported by artificial intelligence[J]. Artificial Intelligence in Education and Teaching Assessment, 2021, 11: 235-245.

[38] YU B, LIU Y, CHAN K C C. Multimodal fusion via teacher-student network for indoor action recognition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(4): 3199-3207.

[39] TONG H, YANG Y J. Research on smart classroom teaching interaction based on multi-modal data[J]. E-Education Research, 2022, 43(3): 60-68.

[40] WANG M, LUO L, CHEN Z, et al. Intelligent multimodal analysisframework for teacher-student interaction[C]// 2022 International Conference on Intelligent Education and Intelligent Research(IEIR). IEEE, 2022: 65-70.

[41] LI X J, LIU Q T, WU L J, et al. Dynamic collaborative analysis of teachers and students' multimodal interactive behaviors in hybrid synchronous classrooms [J]. E-Education Research, 2022, 43(8): 43-50.

[42] DUBOVI I. Cognitive and emotional engagement while learning with VR: the perspective of multimodal methodology[J]. Computers & Education, 2022, 183: 104495.

[43] VIVANTE I, VEDDER-WEISS D. Examining science teachers' engagement in professional development: a multimodal situated perspective[J]. Journal of Research in Science Teaching, 2023, 60(7): 1401-1430.

[44] XU Y, ZHANG J, ZHANG Q, et al. ViTPose: simple vision transformer baselines for human pose estimation[J]. Advances in Neural Information Processing Systems, 2022, 35: 38571-38584.

[45] ZHENG C, WU W, CHEN C, et al. Deep learning-based human pose estimation: a survey[J]. ACM Computing Surveys, 2023, 56(1): 1-37.

[46] LI J N, WANG D K, ZHANG S L. Deep-learning-based 2D human pose estimation: present and future[J]. Chinese Journal of Computers, 2024, 47(1): 231-250.

[47] SHI D, ZHONG Y, CAO Q, et al. Tridet: temporal action detection with relative boundary modeling[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 18857-18866.

[48] HU K, SHEN C, WANG T, et al. Overview of temporal action detection based on deep learning[J]. Artificial Intelligence Re-

view,2024,57(2):26.
[49] HUANG X K,ZHANG J Y,WANG X Y,et al. Survey of dense video captioning[J]. Computer Engineering and Applications,2023,59(12):28-48.
[50] TANG P J,WANG H L. From video to language:survey of video captioning and description[J]. Journal of Automation,2022,48(2):375-397.
[51] SHEN X,LI D,ZHOU J,et al. Fine-grained audible video description[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:10585-10596.
[52] XU H,ZHANG K,TIAN Y J,et al. Review of deep neural network-Based image Caption[J]. Computer Engineering and Applications,2021,57(9):14.
[53] SHI Y L,YANG W Z,DU H X,et al. A review of deep learning-based image description[J]. Chinese Journal of Electronics,2021,49(10):2048-2060.
[54] RAHMANI H,BENNAMOUN M,KE Q,et al. Human action recognition from various data modalities:a review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,45(3):3200-3225.
[55] BI C Y,LIU Y. A survey of video human action recognition based on deep learning[J]. Journal of Graphics,2023,44(4):625-639.
[56] LIN L,ZHANG J,LIU J. Actionlet-dependent contrastive learning for unsupervised skeleton-based action recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:2363-2372.
[57] WU T,CAO M,GAO Z,et al. Stmixer:a one-stage sparse action detector[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:14720-14729.
[58] XIE Y G,WANG Q. Summary of dynamic gesture recognition based on vision[J]. Computer Engineering and Applications,2021,57(22):68-77.
[59] WANG R P,WU S H,ZHANG M H,et al. Review of vision-based neural network 3D dynamic gesture recognition methods[J]. Computer Science,2024,51(4):193-208.
[60] BAO H,HU W X. Technical features of facial expression recognition and analysis of educational application scenarios[J]. Electronic Engineering & Product World,2023,30(3):39-43.
[61] LIN S N,ZHAO R,ZHANG W. A review of deep learning-based research on student expression recognition[J]. Information Technology & Informatization,2023(11):188-194.
[62] MA H,TANG B R,ZHANG Y,et al. Survey on speech recognition[J]. Computer Systems & Applications,2022,31(1):1-10.
[63] WANG A H,ZHANG L,SONG W Y,et al. Review of end-to-end streaming speech recognition[J]. Computer Engineering and Applications,2023,59(2):22-33.
[64] WANG W C. Research on speaker clustering and identification methods based on deep convolutional network [D]. Guangzhou:South China University of Technology,2021.
[65] BECCARO W,RAMÍREZ M A,LIAW W,et al. Analysis of oral exams with speaker diarization and speech emotion recognition:a case study[J]. IEEE Transactions on Education,2023,67(1):74-86.
[66] LUO Y Y,QIN X D,XIE Y P,et al. Intelligent data visualization analysis techniques:a survey[J]. Journal of Software,2024,35(1):356-404.

CUI Jiajun, born in 2000, postgraduate. His main research interests include intelligent educational technology and so on.

MA Miao, born in 1977, Ph.D, professor, Ph.D supervisor. Her main research interests include image processing, video analysis and smart education.

(责任编辑:何杨)