

课堂行为智能编码体系的构建方法与效果验证

□刘帅 彭常玲

摘要: 基于课堂行为的诊断评价是新时代教育评价改革的前沿领域。如何构建一种AI适配的课堂行为编码体系(AI-CBCS)是破解当前AI编码准确率偏低和大规模课堂诊断评价难以开展等难题的关键技术路径。AI-CBCS编码体系的构建方法包括三个环节:一是充分借鉴现有11种编码体系对课堂行为的界定,综合运用课堂观察和专家咨询等方法,最终形成包括讲授、表扬鼓励等在内的21种共识性课堂行为;二是通过将课堂行为拆解为具有高区分度的动作序列,来构建最大差异化的课堂行为体系,以提升课堂行为的区分度;三是通过评估课堂行为在教学过程中的相对重要性,对课堂行为设定编码优先级,以提升AI编码的准确率。最终形成的AI-CBCS包括15种课堂行为,由11种教师动作和15种学生动作序列表征。在此基础上,以SmartEdu平台上36节优课视频为例,基于AI-CBCS与传统编码体系进行多维比较分析,定量和定性分析结果表明,相较于现有编码体系,AI-CBCS编码体系能够更好地适配AI编码,且编码更加高效、准确,甚至能达到人工编码的水平,同时能更多元化反映真实的课堂状态。总体而言,AI-CBCS编码体系能够较好地解决原有编码体系对课堂行为编码交叉混淆的难题,为大规模课堂行为智能评价提供了技术和方法支撑。

关键词: 课堂评价; 课堂行为; 课堂行为编码体系; AI编码; 国家中小学智慧教育平台

中图分类号: G434 **文献标识码:** A **文章编号:** 1009-5195(2026)03-0080-11 **doi:**10.3969/j.issn.1009-5195.2026.03.010

基金项目: 全国教育科学“十三五”规划2020年度重点课题“中西部地区推进高考综合改革研究”(AEA200013)。

作者简介: 刘帅,博士,教授,博士生导师,湖南师范大学教育科学学院/交叉科学研究院,湖南省教育信息化与智能化高校重点实验室(湖南长沙 410081);彭常玲,博士研究生,湖南师范大学教育科学学院/交叉科学研究院(湖南长沙 410081)。

一、引言

2025年中共中央、国务院印发《教育强国建设规划纲要(2024—2035年)》,强调要“深化教育评价改革”“建立基于大数据和人工智能支持的教育评价和科学决策制度”(新华社,2025);2025年教育部办公厅发布《关于组织实施数字化赋能教师发展行动的通知》,提出要“推动数据支撑的教师评价改革”“探索应用人工智能技术帮助师范生诊断课堂教学行为”(中华人民共和国教育部,2025)。可见,课堂评价已成为新时代教育评价改革的重要领域,尤其是基于课堂行为的诊断评价,已受到学界广泛关注并成为领域研究热点(武法提等,2025)。

基于行为的课堂评价主要包括“课堂行为编码”与“基于编码的分析”两个关键环节(李红美等,2024)。编码是分析的前提,分析是编码的目的,基于行为编码分析课堂关键特征,有助于优化课堂教学实践(Zhang et al., 2023)。当前基于行为编码的

课堂评价主要可分为单模态与多模态两类。其中,基于单模态行为的课堂评价是指从视频、音频、文本或生理等单一数据源中编码课堂行为,以实现课堂的评价(柏宏权等,2024);基于多模态行为的课堂评价则强调综合使用教学视频、音频、文本、生理等多种模态数据进行课堂评价,整合更丰富的课堂行为以提升课堂评价质量(周德青等,2025)。

尽管基于多模态行为的课堂评价在精准性方面更具优势,但其存在数据质量要求高、采集设备不普及、数据融合效能低等挑战(王梦珂等,2024),导致其常态化应用非常受限。基于单一视频行为编码的课堂评价具有数据采集便捷、应用成本低、易于推广等特点,因而成为大规模智能化课堂评价的最佳选择。当前基于行为编码的课堂评价主要存在以下问题:一是人工编码稳定性差,分析效率低(赵丽等,2022);二是现有编码体系中课堂行为及其编码规则对AI而言过于抽象,运用AI技术进行编码的准确率低、可用性差。例如,“主动提问”与

“主动应答”两种课堂行为的表面动作（站+说话）相似（卢国庆等，2021；Jiang，2025），人类可以通过对复杂场景的还原与分析进行识别，而AI对课堂动作的分析并不能结合具体的课堂场景，因而可能导致编码行为出错。

为解决上述问题，本文构建了一种AI适配的课堂行为编码体系（AI Adapted Class Behavior Coding System，AI-CBCS）。首先，基于现有编码体系，综合多种方法重新筛选具有重要教学意义的课堂行为；其次，将课堂行为抽象为高区分度的动作序列组合，以构建最大差异化的课堂行为体系；再次，针对并发的课堂行为，根据行为相对于课堂的重要性设定编码优先级，以提升AI编码的准确率；最后，以国家中小学智慧教育平台（SmartEdu，网址为<https://basic.smartedu.cn/>）中的36节优课视频为例，运用AI-CBCS与传统编码体系对视频案例中的课堂行为进行编码分析，以检验其编码效果。

二、相关研究

1. 课堂行为编码体系

课堂行为编码体系是课堂行为量化研究的基础，直接影响课堂行为分析的有效性。1963年，Flanders首次提出弗兰德斯互动分析体系（FIAS）（Flanders，1963），其将课堂言语行为分为10类，其中教师言语行为占7类，学生言语行为占2类，最后一类是沉默或混乱。基于FIAS衍生出系列课堂行为编码体系。例如，顾小清等人（2004）提出基于信息技术的交互分析体系（ITIAS），以解决信息技术时代FIAS中人技交互分析不足的问题。该编码体系包括教师言语、学生言语、沉寂和技术等4大类及其下属的18小类课堂行为，但其定义的部分课堂行为具有较高的相似度，如将“主动应答”与“主动提问”两类行为都用“站+说话”表示，易导致编码混乱的问题。随着AI对课堂教学的介入越来越深，赵刚等人（2020）基于FIAS、ITIAS等提出基于视听觉特征的教学行为分析编码体系（VATBAS），由于该编码体系定义的课堂行为视听觉特征仍较为抽象，因而基于该编码体系利用AI对课堂行为进行识别的能力亦比较有限。

课堂行为编码体系不仅定义了各类行为，还设定了记录这些行为的编码规则。现有编码体系中只设定了笼统的抽样原则，如每隔3秒记录一次发生的主要课堂行为（赵呈领等，2016）。然而，该原则

中对“主要课堂行为”的定义较为抽象，人工编码者可能会依据个人理解和主观判断来选择主要课堂行为，而AI对“主要课堂行为”的识别必须建立在编码体系对该行为的清晰界定上。因而，基于现有课堂行为编码体系的局限性，往往会导致人工编码与AI编码结果的不一致。

2. 基于AI的课堂行为编码

从现有研究来看，基于AI的课堂行为编码主要可分为单模态与多模态编码两类。

（1）单模态课堂行为编码

该类研究多以教师动作、学生动作或师生动作识别为主进行编码体系设计。教师动作识别方面，Zhao等人（2021）提出将3D BP-TBR算法应用于课堂视频中教师向学生鞠躬、指向黑板、黑板写字、清理黑板等8种动作识别，其整体识别准确率为81.00%；Ye等人（2023）提出将HRNet姿态评估算法应用于课堂视频中教师教学、黑板写字、展示、看黑板等6种动作识别，其识别准确率达到90.30%；段欣然等人（2024）提出将YOLOv5和ST-GCN算法应用于课堂视频中教师举手示意、走动、指示引导、使用课本、书写板书等5种动作识别，其识别准确率达到88.40%。由此可见，当前研究中可识别的教师动作种类较少，准确率处于中等偏上水平。

学生动作识别方面，Nguyen等人（2025）提出一种融合Swin Transformer与Skeleton模型的学生动作识别算法，并对课堂视频中常见的4种学生动作（即举手、阅读、睡觉和写作）进行识别，识别准确率达到94.37%；陈晨等人（2024）采用DMS-YOLOv8算法对课堂视频中7种学生动作（即听课、睡觉、站立、玩手机、写字、阅读、举手）进行识别，识别准确率为88.20%；Ding等人（2025）提出WS-FC SlowFast算法并对课堂视频中14种学生基本动作（即抬头、低头、转头、躺、走、举手等）进行识别，识别准确率最高达到85.95%。可以看出，随着学生动作识别种类的增加，其识别准确率有随之降低的趋势。

相较于教师或学生单一主体的动作识别，同时识别师生动作的相关研究较少。Chen等人（2022）提出运用YOLOv4和物联网技术对所选取课堂视频中的13种师生动作进行识别，其教师动作识别准确率达到91.20%，学生动作识别准确率达到89.90%；Sharma等人（2024）提出STAR-3D算法并对课堂视频中20种师生动作进行识别，其整体识别准确率为

83.50%。可见, 现有算法对师生课堂动作的区分能力处于中等偏上水平, 尚有进一步的提升空间。

(2) 多模态课堂行为编码

多模态课堂行为智能编码方面, 张乐乐等人(2022)利用摄像头、眼动仪、智能手环、智能平板、知识图谱等设备或技术, 获取课堂教学过程中师生行为、师生情感、讨论交流、人机交互、随堂测验等多模态数据, 以实现多模态数据支持的课堂行为分析, 但其行为编码及分析结果准确率并未报告。殷宝媛等人(2024)采用生理仪器、智能录播系统、眼动仪和录音笔等采集生理、视频、眼动、音频和访谈等多模态数据, 以实现教师课堂管理行为的定量评价, 其行为编码及评价结果准确率也未报告。周德青等人(2025)标记了视频、音频、图像、文本等四类模态数据, 综合利用BERT、Wav2Wev2.0、YOLOv11等算法对教师(如讲授、提问、反馈等)和学生(如听讲、应答、提问等)行为进行识别, 其行为识别准确率亦未报告。总之, 相较于单模态, 多模态课堂行为编码需要可穿戴设备(如眼动仪)、生理仪器、智能学习设备等多种设备支持, 其数据质量要求高、采集设备不普及、数据融合效能低(王梦珂等, 2024), 导致其难以实现常态化与规模化应用。

综上所述, 当前关于多模态课堂行为智能编码仍难以实现常态化与规模化应用, 单模态(尤其是视频)课堂行为智能编码, 因其数据成本低且易于推广, 仍是当前研究的主流方向。目前基于课堂视频的课堂行为编码主要针对学生举手、站立、说话等简单动作识别, 可供参考的师生动作类别少且无法识别时序性课堂行为(如主动提问=举手+指示+站立+对话+坐下), 因而也难以以为真实课堂评价和教学优化提供更具有有效性的数据支持。

三、构建方法

1. 厘清共识性课堂行为

厘清课堂行为是建立课堂行为编码体系的重要步骤, 也是课堂行为编码的前提和保障。目前国外已经形成的比较典型通用的课堂行为编码体系包括FIAS (Flanders, 1963)、言语互动分类体系(VICS) (Amidon et al., 1968)、言语与非语言课堂互动分析体系(CAFIAS) (Silverman et al., 1990)、STEM课堂观察编码体系(COPUS) (Achen et al., 2015)等。我国关于课堂行为编码体系的研究虽然起步较晚,

但取得的成果较多, 主要是基于FIAS衍生出的系列编码体系, 如基于信息技术的交互分析体系(ITIAS) (顾小清等, 2004)、改进型弗兰德互动分析体系(iFIAS) (方海光等, 2012)、数字化课堂教学互动行为分析编码体系(OOTIAS) (韩后等, 2015)、数学课堂互动双编码体系(MFIAS) (武小鹏等, 2017)、化学课堂及信息技术的互动分析编码体系(CCITIAS) (阮雪丹等, 2019)、基于视听觉特征的教学行为分析编码体系(VATBAS) (赵刚等, 2020)、智慧课堂互动教学双编码分析体系(SCMIAS) (李红美等, 2024)等。本文将上述国内外的11个编码体系所包含的课堂行为进行整理分析, 最终形成24种课堂行为, 具体如表1所示。其中, “出现频次”表示该行为在所有编码体系中出现的次数, 频次越高表明该行为的认可度越高。本文依据出现频次将24种课堂行为分为三类, 分别为核心课堂行为10种(出现频次不少于9次, 覆盖80%以上编码体系)、次核心课堂行为7种(出现频次介于5~8次, 覆盖40%~80%的编码体系)和边缘课堂行为7类(出现频次小于5次, 覆盖不足40%编码体系)。核心课堂行为和次核心课堂行为共同构成了主要课堂行为。

教学环境或教学模式的变革可能会产生新的课

表1 11个典型编码体系的课堂行为梳理

分类	编号	课堂行为	出现频次	表述示例
教师行为	1	讲授	11	讲授、讲解、lecturing、presents information
	2	表扬鼓励	10	表扬鼓励、positive value assessment、praises
	3	批评/维护权威	10	批评或维护权威、criticizes、express anger
	4	采纳意见	9	采纳意见、accepts or uses student's ideas
	5	命令指示	9	指令、指示(组织与管理活动)、gives direction
	6	表达/接受情感	7	表达/接受情感、表扬鼓励、采纳意见
	7	提出开放问题	6	提出开放性问题、ask broad question
	8	提出封闭问题	6	提出封闭性问题、ask narrow question
	9	教师操纵技术	6	教师操纵技术、资源演示
	10	演示	3	教师演示实践、方法演示、资源演示
	11	课堂巡视	1	观察巡视、moving through the in-class
	12	个别指导	1	个别指导
	13	板书	1	板书

续表1

分类	编号	课堂行为	出现频次	表述示例
学生行为	14	主动应答	10	应答(主动)、unpredictably replies
	15	被动应答	10	应答(被动)、predictably replies
	16	主动提问	10	主动提问、student-initiated talk、student questions
	17	讨论合作	9	同伴讨论、交流讨论、group discussions、
	18	实践练习	7	做笔记或练习、学生操纵技术、实践或实验
	19	思考	7	思考、静思、有助于教学沉寂、individual thinking
	20	技术作用学生	5	技术作用学生、观察、watching video
	21	成果展示	3	演示、汇报展示、作品分享
	22	学生操作实验	2	学生操作实验、实践实验
课堂混乱	23	运算求解	1	运算求解
	24	混乱	9	无助于教学的混乱或沉寂、confusion or silence

堂行为，而这些行为最初大多出现在边缘课堂行为中。本文基于对当前 SmartEdu 平台上课堂教学实际情况的分析，将“个别指导”“课堂巡视”和“成果展示”3种行为纳入主要课堂行为范畴。对初步筛选出的20种主要课堂行为，本文从准确性和全面性等方面向专家咨询了修订意见，共有7位领域专家作答并提出了修改意见。通过梳理7位专家的反馈意见，最终得到21种具有重要教学意义的课堂行为，如表2所示。其优化主要体现在：根据对 SmartEdu 上课堂教学的观察，新增“个别指导”“课堂巡视”“成果展示”3种课堂行为；根据专家意见，良好板书不仅有助于学生理解和记忆教学内容，还能提升课堂视觉效果和教学效率，故

表2 完善后的课堂行为集

序号	课堂行为	行为来源	序号	课堂行为	行为来源
1	讲授	核心行为	12	实践练习	次核心行为
2	表扬鼓励	核心行为	13	思考	次核心行为
3	批评/维护权威	核心行为	14	提出开放问题	次核心行为
4	主动应答	核心行为	15	提出封闭问题	次核心行为
5	被动应答	核心行为	16	教师操纵技术	次核心行为
6	主动提问	核心行为	17	技术作用学生	次核心行为
7	采纳意见	核心行为	18	个别指导	课堂观察
8	命令指示	核心行为	19	课堂巡视	课堂观察
9	讨论合作	核心行为	20	成果展示	课堂观察
10	混乱	核心行为	21	板书	专家意见
11	表达/接受情感	次核心行为		—	

补充“板书”这一行为。

2. 构建最大差异化课堂行为体系

课堂行为通常由一个或多个动作序列组成，现有编码体系中的课堂行为表现出相似性高、特征重叠的特点，因而不利于使用 AI 进行精准识别和有效编码。为此，本文提出构建最大差异化课堂行为体系，通过将21种课堂行为拆解为高区分度动作序列，以最大化行为差异来提高编码精度。

首先，基于课堂“行为—动作”映射关系，以FIAS、ITIAS和VATBAS三个具有代表性的编码体系为例，分析现有典型编码体系中各课堂行为在教学场景中的外显形式，并构建行为中各主体（教师、学生）的时序动作特征。

其次，将课堂行为所对应的所有可能动作序列视为一个集合，计算与其他行为集合间的相似度，以此来衡量编码体系中课堂行为的区分度。相似度越接近0，表示编码区分度越大，说明该行为特征相对于其他行为的差异性越大，更易区分。课堂行为r的相似度 S_r 的计算公式为：

$$S_r = \sum_{i=1}^n \frac{|Re_r \cap Re_i|}{|Re_r \cup Re_i|} - 1 \quad (1)$$

式中，n代表课堂行为总类别数，r代表当前被计算相似度的课堂行为，i代表与当前行为r比较的课堂行为， Re_r 与 Re_i 分别代表课堂行为r与i对应的动作序列集合。

现有典型编码体系中，课堂行为相似度计算结果如表3所示。以FIAS中编号1对应的“表达/接受情感”课堂行为为例，首先依次计算编号1与FIAS中所有编号对应行为的交集与并集的比值（即交并比），如编号1与编号1的交并比为1，编号1与编号2的交并比为1/3，编号1与编号3的交并比为1，等等；其次计算所有交并比的和后再减1，即可得到编号1对应课堂行为的相似度为2.67。

为了更好地区分各课堂行为，本文通过如下三个步骤对课堂行为及其动作编码进行优化：

第一，合并相似功能的行为。如在教师行为中，考虑到表征课堂行为的动作序列存在高相似度及课堂行为在实际教学情境中的含义具有内在一致性，故将“表扬鼓励”“批评/维护权威”“表达/接受情感”“采纳意见”4种行为统一合并为“反馈”，并以课堂动作序列T5/T6表征。这一思路与赵刚等人（2020）将“表扬鼓励”与“批评/维护权威”合并为“反馈”的编码体系优化思路一致。类似地，在学生行为中，将

表3 FIAS、ITIAS和VATBAS中课堂行为与动作映射关系详表(部分)

编码体系	编号	课堂行为	原始定义或描述	动作序列(T表示教师,S表示学生)	相似度
FIAS	1	表达/接受情感	接纳学生积极或消极情感	站T1+说话T2	2.67
	2	表扬鼓励	对学生的语言、动作或行为进行表扬或鼓励	①点头T5;②鼓掌T6;③站T1+说话T2	1.20
	3	采纳意见	重复、补充或发展学生意见或想法	站T1+说话T2	2.67
.....					
ITIAS	1	表达/接受情感	接纳及澄清学生态度或情感	站T1+说话T2	2.67
	2	表扬鼓励	称赞或鼓励学生动作或行为	①点头T5;②鼓掌T6;③站T1+说话T2	1.20
	3	采纳意见	承认、修饰或重述学生说话	站T1+说话T2	2.67
.....					
VATBAS	1	主动提问	提出问题不具有固定答案;提出问题具有固定答案	①说话T2+站S1;②说话T2+举手S5	1.33
	2	主动/被动应答	学生举手并站起来回答问题;教师邀请学生,学生站起来回答问题	①举手S5+站S1;②指向学生T8+站S1	1.33
	3	成果展示	学生通过书本、模型等呈现内容、演示成果的行为;学生使用计算机或多媒体设备呈现内容、演示成果的行为	①走S3+操作某物S13;②站S1+板书S12;③站S1+操作某物S13	1.33
.....					

“主动提问”与“主动应答”2种行为合并为“主动应答”，以课堂动作序列S5+S1表征；由于“提出开放问题”与“提出封闭问题”2种行为需要通过自然语言处理读取内容确定，在仅采用视觉AI的情况下相似度较高且对行为判定影响较小，故将其合并为“提问”行为。总之，相似功能的行为合并有助于提升教师和学生课堂行为的可辨识度，进而提高课堂行为编码的准确性。

第二，整合技术行为。根据教育信息传播原理，信息技术是教师传播教学信息的载体和工具，应融入教学全过程。因此，对于教师借助信息技术进行讲授的行为，不应将其编码为“教师操作技术”行为，而应将其归为教师言语“讲授”行为。这样不仅能避免原有研究对技术类行为编码混淆、归类模糊的问题（李红美等，2022），而且也不会降低对教师行为整体分析的准确性。至此，最大差异化课堂行为体系由原来的21种课堂行为优化为15种。

第三，提取时序动作特征。考虑到每一种课堂行为在同一时刻对应多个动作，故将课堂行为通过高区分度的动作序列特征进行锚定。例如，课堂行为“反馈”包含“点头”“鼓掌”“站+说话”等动作序列，由于“站+说话”与其他动作序列相似度极高，故不用其识别“反馈”行为，由此可实现对该行为的高区分度动作表征和更准确的行为识别。

经过优化，本文初步构建了基于最大差异化的课堂行为编码体系（见表4），包括11种教师动作（即站、说话、板书、抬手、点头、鼓掌、弯腰、指向学生、操作某物、巡视、其他，用符号T1~T11表示）和15种学生动作（即站、坐、走、说话、举

手、低头、抬头、趴、写、翻书、围成簇、板书、操作某物、左顾右盼、其他，用符号S1~S15表示）。由于本文采集了3秒以内的所有动作，这些动作发生的时间间隔较短，因此可认定为“并发”，而无需考虑其顺序。该编码体系依据不同的行为主体（含教师与学生，分别用T与S表示）和言语方式（含言语与非言语，依次用L与N表示），将课堂行为分为TL、TN、SL、SN和课堂混乱（用O表示）5类；且每一类行为对应1~4个动作序列。在此基础上，本文以AI对课堂动作的识别能力为权重，对初步构建的课堂行为编码体系进行迭代优化，最终确

表4 AI适配的课堂行为编码体系

行为分类	编码	课堂行为	动作序列	相似度	质性表征
教师言语行为(TL)	1	讲授	T1+T2;T2+T9	0	直接影响
	2	提问	T2+S1;T2+S5	0	间接影响
	3	反馈	T5;T6	0	间接影响
	4	命令指示	T2+T8;T4+T8;T8+S1	0	直接影响
教师非言语行为(TN)	1	个别指导	T7	0	直接影响
	2	课堂巡视	T10	0	间接影响
	3	板书	T1+T3	0	直接影响
学生言语行为(SL)	1	主动应答	S5+S1;S5+S2	0	主动型
	2	被动应答	S1+S4	0	被动型
	3	讨论合作	S11	0	主动型
学生非言语行为(SN)	1	技术作用学生	T9+S7	0	被动型
	2	实践练习	S2+S9;S2+S13	0	主动型
	3	成果展示	S3+S12;S3+S13;S1+S12;S1+S13	0	主动型
	4	思考	S10;S6	0	主动型
课堂混乱(O)	1	混乱	T11;S8;S14;S15	0	—

立了AI适配的课堂行为编码体系（即AI-CBCS）（见表4）。由于该编码体系中的课堂行为可用确定的动作序列进行表征，故其各课堂行为的相似度均为0，换言之，使用AI编码时必可得到唯一的行为序列。同时，基于对课堂行为的外在表现及其背后动机的考量，还对各种行为进行了质性表征，将教师行为划分为直接与间接影响行为，将学生行为划分为主动与被动学习行为，以便用于后续分析。

3.设计动作—行为为编码规则

课堂行为的兼容性使得同一采样时间段内可能会同时出现2种及以上课堂行为，为避免编码混淆的问题，本文依据课堂行为在教学过程中的相对重要性，对课堂行为设定了编码优先级，以固化编码结果。已有研究指出，言语行为约占总体课堂行为的80.00%（高巍，2009），其中师生互动言语行为是言语行为中的关键行为（孙发勤等，2025）。由于传统课堂仍以教师讲授为主，学生处于相对被动地位，因此为保证课堂教学过程中能充分采集到学生言语行为数据，本文将学生言语行为作为第一优先级进行编码，其编码优先级依次为主动应答（SL1）、被动应答（SL2）、讨论合作（SL3）；将教师言语行为作为第二优先级进行编码，其编码优先级依次为提问（TL2）、反馈（TL3）、讲授与命令指示（TL1 & TL4）。由此，在采样时间窗口为j时，将编码结果定义如下：

时刻j的动作集合 b_j 表示为：

$$b_j = \{i | Re_i \subseteq T_j \cup S_j\} \quad (2)$$

其中 T_j 表示教师动作集合， S_j 表示学生动作集合。

课堂行为集 A_j 是固定间隔时间段 Δ 内（本文为3秒）所有动作的集合，表示为：

$$A_j = \{a_j a_{j+1} \dots a_{j+\Delta} | \exists a_j a_{j+1} \dots a_{j+\Delta} \in R_b^{verbal} \cap \prod_{t=1}^{\Delta} b_{j+t}\} \quad (3)$$

其中 $a_j a_{j+1} \dots a_{j+\Delta}$ 代表固定间隔时间段内可能的动作序列， R_b^{verbal} 为动作序列到课堂行为的映射规则集合（详见表4），由此 A_j 包含了时间段内所有组合成课堂行为的动作集合。

由于 $|A_j| \geq 1$ ，即在同一间隔时间段 Δ 内可能存在多个行为，因此需要对其行为进行排序。通过对 A_j 中的行为进行优先级排序，则可输出唯一的课堂行为 A_j^* ，优先级排序规则表示为：

$$A_j^* = \begin{cases} SL1 & \text{if } SL1 \in A_j \\ SL2 & \text{else if } SL2 \in A_j \\ SL3 & \text{else if } SL3 \in A_j \\ TL2 & \text{else if } TL2 \in A_j \\ TL3 & \text{else if } TL3 \in A_j \\ TL1 \text{ 或 } TL4 & \text{else if } TL1 \text{ 或 } TL4 \in A_j \\ TN \cup SN \cup O & \text{otherwise} \end{cases} \quad (4)$$

其中，SL代表学生言语行为集合，TL代表教师言语行为集合，TN代表教师非言语行为集合，SN代表学生非言语行为集合，O代表混乱行为集合。

最终的课堂行为 A_j^* 可表示为：

$$A_j^* = \begin{cases} A_j & \text{if } A_j \text{ 中出现学生行为} \\ \operatorname{argmax}|A_j^*| & \text{else if } A_j \text{ 中出现教师行为} \\ \operatorname{argmax}|A_j^*| & \text{otherwise} \end{cases} \quad (5)$$

由公式（5）可知，若当前课堂中出现学生行为，则按对应学生行为作为当前课堂行为编码；当没有学生言语行为而包含多个教师言语行为时，选择出现频次最高的那种教师言语行为作为本时段的课堂行为编码；当没有学生和教师言语行为而包括多个师生非言语行为与混乱行为时，则选择出现频次最高的那种非言语或混乱行为作为本时段的课堂行为编码。

四、实验与结果

1.数据集与评价指标

（1）数据集构建

由于目前国内外尚未有公开的符合本研究需要的课堂动作数据集供研究人员直接使用，因此本文通过对SmartEdu中的优课视频进行预处理和标注，以构建课堂动作数据集。具体构建过程如下：

一是从SmartEdu平台随机选取和下载36个已获得市级及以上奖励的中小学语文、数学、信息技术等学科的课堂视频作为本文的测试数据集，三个学科各为12、13、11节课。二是将下载的视频裁剪为单个1920×1080的短视频，共计得到323个短视频（总时长为1069.61分钟），并确保裁剪的短视频均衡覆盖所有主要课堂行为。三是针对裁剪出的短视频，采取每秒5帧进行分割（Ding et al., 2023），共计得到320883帧图像。四是采用图像标注工具VoTT对所得的320883帧图像进行典型动作标注（包括教师11种动作和学生15种动作），得到In-class_Action数据集。

（2）人工编码一致性检验

本文分别基于FIAS、ITIAS、VATBAS和AI-CBCS四类编码体系，依次对SmartEdu中优课视频（按3秒取样）进行人工编码，将其编码结果视为真实值，以便用于后续检验AI编码的准确性。具体来说，首先，从下载的三个学科视频中分别随机选取2个优课视频作为案例。其次，招募8名研究生（C1~C8）随机分为4组，每组由2名编

码者构成,使用特定一类编码体系对6个案例视频进行独立人工编码,其中C1与C2使用FIAS编码、C3与C4使用ITIAS编码、C5与C6使用VATBAS编码、C7与C8使用AI-CBCS编码。最后,采用组内相关ICC系数(计算见公式(7))分析每组编码员在6个案例视频中编码结果的一致性。结果显示,C1与C2的ICC系数为0.82~0.93($p<0.001$),C3与C4的ICC系数为0.78~0.90($p<0.001$),C5与C6的ICC系数为0.76~0.86($p<0.001$),C7与C8的ICC系数为0.83~0.92($p<0.001$)。根据Wisner等人(2025)提出的组内相关ICC系数参考值,以上各组ICC系数均处于一致性较好区间,表明本次人工编码具有较强的可靠性和准确性,可以作为后续检验AI编码效果的可靠真实值。

(3) 评价指标筛选

课堂行为智能编码的有效性主要体现为AI与人工编码的一致性,本实验中主要使用以下2项评估指标。

一是课堂动作识别准确率。本文将课堂动作识别准确率(Accuracy, ACC)(Pareek et al., 2023)定义如下:

$$ACC = \frac{TP+TN}{TP+FN+FP+TN} \quad (6)$$

其中,TP表示将特定课堂动作(如“站”)识别为该动作的数量,FN表示将特定课堂动作识别为非该动作的数量,FP表示将非特定课堂动作识别为特定动作的数量,TN表示将非特定课堂动作识别为非该动作的数量。

二是编码一致性。本文采用双向随机效应模型评估人工和AI编码的组内一致性(Intraclass Correlation Coefficient, ICC),该模型广泛用于N个评分者对同一组对象评分一致性的度量(Varela et al., 2021)。一般认为,当 $0.75 \leq ICC < 0.90$ 时,表示一致性较好;当 $ICC \geq 0.90$,表示一致性极好(Wisner et al., 2025)。ICC系数计算公式如下:

$$ICC = \frac{MS_{subject} - MS_{error}}{MS_{subject} + (k-1)MS_{error}} \quad (7)$$

其中, $MS_{subject}$ 表示通过各类行为占比计算课堂视频的均方, MS_{error} 表示同一课堂视频不同测量结果占比的均方,k表示每个视频的测量次数。

2. 实验设置

本实验基于湖南师范大学人工智能算力平台

(<https://aistation.hunnu.edu.cn/index.html>),使用针对公开数据集的自建高精度普适性动作识别模型(LTDRM)(Fu et al., 2024; Liu et al., 2025)完成。该模型基于动作的置信度、相邻动作的关联性和动作位置,利用模糊推理提取并更新动作特征。具体而言,首先,将15种课堂行为包含的26种课堂动作(其中教师11种动作、学生15种动作)的In-class_Action数据集,按照6:2:2随机分割成训练集、验证集和测试集。其次,基于训练集对LTDRM模型进行迁移训练,使该模型在迁移学习过程中加速修正模型权值,以提高其课堂动作检测与识别能力。最后,在验证集和测试集中,使用得分大于0.70的检测框作为最终的动作分类。

3. 定量分析

本文基于FIAS、ITIAS、VATBAS和AI-CBCS中课堂行为间的相似度,进一步计算出各体系下课堂行为的平均相似度(见表5),结果表明,AI-CBCS的平均相似度显著低于FIAS、ITIAS、VATBAS,表明AI-CBCS在课堂行为分类上具有更高的区分度,能够有效提升AI对课堂行为的识别水平。

表5 不同编码体系下的课堂行为平均相似度

	FIAS	ITIAS	VATBAS	AI-CBCS
课堂行为平均相似度	0.13	0.05	0.03	0

本文亦对验证集和测试集中的课堂动作识别准确率进行了计算(见表6),结果显示:LTDRM模型在验证集和测试集上对教师10种动作(除T11外)的整体识别准确率分别为90.46%、91.24%,对学生14种动作(除S15外)的整体识别准确率分别为94.25%、93.18%,均高于90.00%,说明基于LTDRM识别的课堂动作是高度可信的,可用于课堂行为的有效编码。

本文进一步采用ICC系数对三个学科6个案例的人工和AI编码的绝对一致性进行了评估。具体来说,在LTDRM模型提取的时序动作序列基础上,分别采用FIAS、ITIAS、VATBAS和AI-CBCS四类编码体系对6个案例进行人工与AI编码,分别统计2位人工编码者与AI对15种课堂行为的编码频次,并将其归化为各视频中行为总数的百分比,进而计算各类行为占比的ICC系数。由表7可知,基于AI-CBCS的6个样例视频的ICC系数均在0.84及以上,远高于FIAS、ITIAS和VATBAS等现有编

表6 验证集和测试集中课堂动作识别准确率

教师动作	准确率(%)		学生动作	准确率(%)	
	验证集	测试集		验证集	测试集
站T1	98.53	98.19	站S1	97.26	96.57
说话T2	91.67	91.15	坐S2	99.03	98.98
板书T3	94.41	94.22	走S3	83.65	82.43
抬手T4	93.05	92.51	说话S4	94.24	90.29
点头T5	65.57	78.48	举手S5	95.31	94.44
鼓掌T6	88.49	93.50	低头S6	95.40	94.62
弯腰T7	87.18	87.26	抬头S7	97.55	97.01
指向学生T8	90.73	90.07	趴S8	96.82	95.63
操作某物T9	91.43	89.61	写S9	84.17	85.91
巡视T10	98.84	97.77	翻书S10	94.72	93.05
整体	90.46	91.24	围成簇S11	93.55	92.20
			板书S12	98.74	98.51
			操作某物S13	94.93	95.29
			左顾右盼S14	93.63	93.14
			整体	94.25	93.18

(注:所测视频均为SmartEdu平台上的优课视频,T11和S15两种动作均未检测到。)

码体系,说明AI-CBCS能够较好地适配基于AI的课堂行为编码。

本文还对教师行为、学生行为和课堂混乱三类结构要素的编码一致性进行了定量分析。由表8可知:人工(C7与C8)与AI对6个案例视频的编码结果表现出高度一致,11种课堂行为的ICC系数大于0.80($p < 0.001$),2种课堂行为的ICC系数介于0.60~0.80($p < 0.05$),仅有2种课堂行为的ICC系数小于0.50($p > 0.05$);除“课堂混乱”外,AI-CBCS编码体系其他两个结构要素的ICC系数均为0.88($p < 0.001$),这表明AI-CBCS可以作为AI课堂行为编码的可靠体系。

4.定性分析

本文截取部分视频的课堂行为进行定性分析。图1展示了利用AI-CBCS体系对课堂行为进行智能编码的效果,图中的“标记动作T”“标记动作S”“标记行为”均为人工标注的“真实值”,其中标记动作T和S中的灰色框代表与真实值相似的其他课

表7 不同编码体系下人工与AI在案例视频中编码一致性检验结果

学科	学段	教学内容	视频时长	ICC系数			
				FIAS	ITIAS	VATBAS	AI-CBCS
语文	六年级	为人民服务	40:49	0.56***	0.42***	0.48***	0.86***
	八年级	雁门太守行	40:18	0.53***	0.48***	0.42***	0.89***
数学	五年级	组合图形面积	40:00	0.65***	0.52***	0.48***	0.88***
	八年级	三角形高、中线与角平分线	39:04	0.52***	0.49***	0.52***	0.84***
信息科技	五年级	模拟新冠肺炎接触传播	39:54	0.64***	0.50***	0.57***	0.91***
	七年级	模拟停车场之统计车位数	40:00	0.55***	0.48***	0.45***	0.90***

(注:***表示 $p < 0.001$ 。)

表8 AI-CBCS体系下人工与AI在6个案例视频中各种行为编码平均占比及ICC系数

行为分类	编码	课堂行为	行为占比(%)			各种行为ICC得分	行为分类结构ICC得分
			编码者C7	编码者C8	AI		
教师行为	言语行为	1 讲授	27.36	27.66	27.26	0.96***	0.88***
		2 提问	5.64	5.82	4.61	0.93***	
		3 反馈	1.81	1.66	0.16	0.69*	
		4 命令指示	2.77	2.61	1.24	0.81***	
	非言语行为	5 个别指导	2.27	2.23	1.08	0.78**	
		6 课堂巡视	3.93	4.05	4.29	0.97***	
		7 板书	1.44	1.72	0.82	0.87***	
学生行为	言语行为	8 主动应答	6.65	7.21	10.12	0.90***	0.88***
		9 被动应答	2.10	2.40	2.07	0.84***	
		10 讨论合作	3.95	3.71	3.07	0.99***	
	非言语行为	11 技术作用学生	8.22	8.42	7.68	0.98***	
		12 实践练习	24.89	24.59	27.62	0.93***	
		13 成果展示	7.32	7.16	8.99	0.99***	
课堂混乱	14 思考	0.64	0.76	0.90	0.40	—	
	15 混乱	0.00	0.02	0.00	0.00	—	

(注:*表示 $p < 0.05$; **表示 $p < 0.01$; ***表示 $p < 0.001$ 。)

堂行为,无色框表示因课堂场景切换导致画面不连续而无法识别;编码行为区域的无色框表示该时刻所识别动作无法编码为课堂行为,灰色框表示识别并编码出的课堂行为。由图1可知, AI识别的课堂动作与真实动作的定位区间相似,验证了AI模型对课堂动作识别的有效性;与现有编码体系相比, AI-CBCS的编码行为结果与真实值相差更小,亦证明了其编码有效性更好。

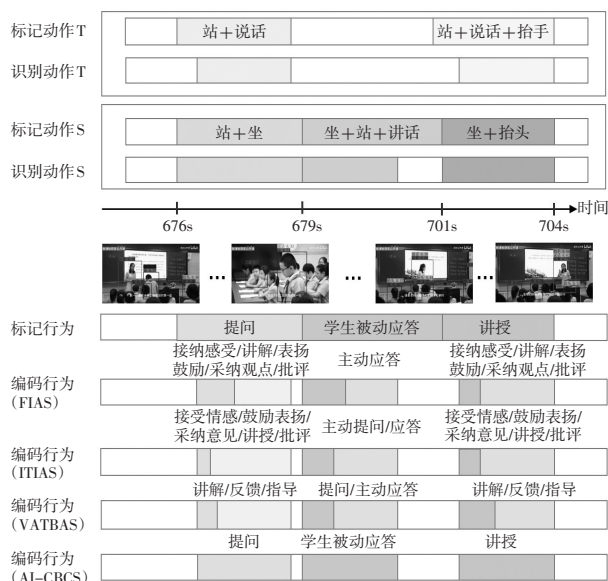


图1 对课堂行为片段编码效果的定性对比

本文还以优课案例中“为人民服务”为例,分别采用FIAS、ITIAS、VATBAS和AI-CBCS四类编码体系对该课例进行行为编码,并借助S-T行为分析法(Yuan et al., 2011)从教学模式、课堂结构、师生交互、教学风格、学生主动性、非言语行为水平等6个维度对案例课堂进行定性评价。由表9可知,基于AI-CBCS的课堂评价结果与人工评价结果(用“标准值”表示)基本一致,且其相较于FIAS、ITIAS、VATBAS三大编码体系体现出三方面的优势:一是AI-CBCS能对课堂中丰富的学生行为进行更细粒度的区分,有效解决现有编码体系难以对学生参与课堂活动的主动性和

被动性进行度量的问题;二是AI-CBCS能够有效弥补FIAS与ITIAS无法关注非言语行为(如“个别指导”“观察巡视”“板书”等)的不足;三是AI-CBCS提供了VATBAS所缺失的教师教学风格指标,即通过分析间接与直接影响的比例来探究教师教学风格。

五、总结

针对已有课堂行为编码体系对AI适配度差的问题,本文尝试构建了一种AI适配的课堂行为智能化编码新体系。首先,基于现有编码体系和当前课堂教学的实际情况,筛选出适配AI编码的有效课堂行为;其次,将课堂行为拆解为具有高区分度的动作序列,构建最大差异化的课堂行为体系,以提升课堂行为的区分度;再次,通过评估课堂行为在教学过程中的相对重要性,对课堂行为设定编码优先级,以提升AI编码的准确率;最后,以SmartEdu平台上36节优课视频为例,从多个维度检验了本文所构建的AI-CBCS的编码效果。定量与定性分析结果均表明,相较于现有编码体系, AI-CBCS体系下的AI编码与人工编码的一致性更高,且能够更加准确和多元地反映真实的课堂教学状态。

本文也存在一些不足,如教师和学生行为中的“其他”类动作数据缺失,“反馈”和“思考”两种课堂行为的AI匹配度仍然不高, AI编码过程中因课堂场景切换出现行为编码间断等问题。后续研究将致力于扩充数据集,纳入常规课堂视频以覆盖“其他”类动作,并将文本与表情等多种特征融入到AI-CBCS中,以提升“反馈”和“思考”课堂行为编码精度。例如,针对教师“反馈”,可同步提取课堂话语中的反馈性词汇及其面部表情,并与点头、鼓掌等动作进行跨模态对齐,构建“文本一表情一动作”综合表征的行为编码体系;又如,针对学生“思考”,可引入微皱眉、唇线抿紧等微表情来进行综合表征。此外,还需持续优化相关AI

表9 不同编码体系下的课堂评价结果的定性对比分析

编码体系	教学模式	课堂结构	师生交互	教学风格	学生主动性	非言语行为水平
标准值	混合型	教师主导	交互程度较低	直接控制型	学生主动参与为主	低
FIAS	混合/练习/讲授/对话型	教师主导/双主体	交互程度较低/高	直接/间接控制型	—	—
ITIAS	混合/练习/讲授/对话型	教师主导/双主体	交互程度较低/高	直接/间接控制型	—	—
VATBAS	混合/练习/讲授/对话型	教师主导/双主体	交互程度较低/高	—	—	高/中/低
AI-CBCS	混合型	教师主导	交互程度较低	直接控制型	学生主动参与为主	低

(注:“标准值”代表人工评价结果,“—”表示无评价结果。)

算法,增强其在复杂教学场景切换的鲁棒性,进而构建更加科学有效的课堂行为智能评价系统。

参考文献:

- [1]柏宏权,朱俊(2024).小学人工智能教师画像构建研究[J].电化教育研究,45(7):96-104.
- [2]陈晨,保文星,陈旭,等(2024).改进YOLOv8的学生课堂行为识别算法:DMS-YOLOv8[J].计算机工程与应用,60(24):222-234.
- [3]段欣然,王玫,韩天利,等(2024).基于视频理解的教学过程感知与分析[J].计算机科学,51(10):56-66.
- [4]方海光,高辰柱,陈佳(2012).改进型弗兰德斯互动分析系统及其应用[J].中国电化教育,(10):109-113.
- [5]高巍(2009).课堂教学师生言语行为互动研究[J].教育研究与实验,(5):43-49.
- [6]顾小清,王炜(2004).支持教师专业发展的课堂分析技术新探索[J].中国电化教育,(7):18-20.
- [7]韩后,王冬青,曹畅(2015).1:1数字化环境下课堂教学互动行为的分析研究[J].电化教育研究,36(5):89-95.
- [8]李红美,卞鹏,徐梦秋(2024).智慧教室环境下数学课堂互动教学双编码分析系统开发与行为模式挖掘[J].现代教育技术,34(3):105-115.
- [9]李红美,王玲玲,卞鹏,等(2022).面向教学行为大数据的课堂互动双编码通用分析框架及其变式应用[J].远程教育杂志,40(3):67-75.
- [10]卢国庆,谢魁,刘清堂,等(2021).基于人工智能引擎自动标注的课堂教学行为分析[J].开放教育研究,27(6):97-107.
- [11]阮雪丹,罗德炳,郑柳萍,等(2019).基于CCITIAS的专家型化学教师实验课堂行为个案研究[J].化学教育(中英文),40(13):53-58.
- [12]孙发勤,王雯(2025).大语言模型驱动的IPCRF框架:课堂问答数字化分析的创新与实践[J].远程教育杂志,43(2):103-112.
- [13]王梦珂,陈增照,石雅文,等(2024).智能技术支持的多模态互动教学评价框架设计与应用效果[J].现代教育技术,34(9):91-101.
- [14]武法提,高姝睿,赖松(2025).同步课堂中基于多模态数据的专注度分析[J].现代远程教育研究,37(1):93-101.
- [15]武小鹏,张怡(2017).数学学科视域下课堂互动双编码模型应用研究[J].数学教育学报,26(5):59-65.
- [16]新华社(2025).中共中央 国务院印发《教育强国建设规划纲要(2024—2035年)》[EB/OL].[2025-07-20].https://www.gov.cn/gongbao/2025/issue_11846/202502/content_7002799.html.
- [17]殷宝媛,王雪静,孙馨,等(2024).教师课堂管理行为多模态解码:行为特征、分类识别与时序发展[J].电化教育研究,45(10):101-109.
- [18]张乐乐,顾小清(2022).多模态数据支持的课堂教学行

为分析模型与实践框架[J].开放教育研究,28(6):101-110.

[19]赵呈领,梁云真,阮玉娇(2016).富媒体环境下职业院校课堂教学行为及特征研究[J].中国电化教育,(2):113-120.

[20]赵刚,朱文娟,胡碧灵,等(2020).多维度特征融合的课堂教学行为智能分析模式构建[J].电化教育研究,41(10):36-44.

[21]赵丽,贺玮,王洋(2022).人工智能支持的课堂教学行为分析:困境与路径[J].电化教育研究,43(1):86-92.

[22]中华人民共和国教育部(2025).教育部办公厅关于组织实施数字化赋能教师发展行动的通知[EB/OL].[2025-09-03].http://www.moe.gov.cn/srcsite/A10/s7034/202507/t20250704_1196586.html.

[23]周德青,李惠聪,穆肃(2025).云端课堂教学行为的智能识别与关联分析[J].现代远程教育研究,37(5):105-112.

[24]Achen, R. M., Lumpkin, A. (2015). Evaluating Classroom Time Through Systematic Analysis and Student Feedback[J]. International Journal for the Scholarship of Teaching and Learning, 9(2): 4.

[25]Amidon, E. J., Hunter, E. (1968). Abstracted from Verbal Interaction Category System (VICS) [J]. Classroom Interaction Newsletter, 3(2): 1-5.

[26]Chen, H., Guan, J. (2022). Teacher-Student Behavior Recognition in Classroom Teaching Based on Improved YOLO-V4 and Internet of Things Technology[J]. Electronics, 11(23): 3998.

[27]Ding, D., Zhao, Y., Zhang, J., et al. (2025). A New Method of Classroom Behavior Recognition Based on WS-FC SLOWFAST[J]. International Journal of Gaming and Computer-Mediated Simulations, 17(1): 1-28.

[28]Ding, H., Liu, C., He, S., et al. (2023). MOSE: A New Dataset for Video Object Segmentation in Complex Scenes[C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. France: Paris: 20167-20177.

[29]Flanders, N. A. (1963). Intent, Action and Feedback: A Preparation for Teaching[J]. Journal of Teacher Education, 14(3): 251-260.

[30]Fu, W., Luo, Z., Liu, S., et al. (2024). Spatiotemporal Correlation Based Self-Adaptive Pose Estimation in Complex Scenes[J/OL]. Digital Communications and Networks. <https://doi.org/10.1016/j.dcan.2024.03.007>.

[31]Jiang, S. (2025). TinyML Based Edge Intelligent English Classroom Quality Assessment Scheme[J]. Internet Technology Letters, 8(4): e70072.

[32]Liu, S., Luo, Z., Fu, W. (2025). Fcdnet: Fuzzy Cognition-Based Dynamic Fusion Network for Multimodal Sentiment Analysis[J]. IEEE Transactions on Fuzzy Systems, 33(1): 3-14.

[33]Nguyen, H., Tran, N., Nguyen, M., et al. (2025). Empowering Classroom Behavior Recognition Through Hybrid Spatial-Temporal Feature Fusion[J]. Applied Intelligence, 55: 863.

- [34]Pareek, P., Thakkar, A. (2023). RGB-D Based Human Action Recognition Using Evolutionary Self-Adaptive Extreme Learning Machine with Knowledge-Based Control Parameters[J]. *Journal of Ambient Intelligence and Humanized Computing*, 14(2): 939-957.
- [35]Sharma, V., Gupta, M., Kumar, A., et al. (2024). STAR-3D: A Holistic Approach for Human Activity Recognition in the Classroom Environment[J]. *Information*, 15(4): 179.
- [36]Silverman, S., Buschner, C. (1990). Validity of Cheffers Adaptation of Flanders Interaction Analysis System[J]. *Journal of Classroom Interaction*, 25(1/2): 23-28.
- [37]Varella, N. C., Almeida, R. S., Nogueira, L. A. C., et al. (2021). Cross-Cultural Adaptation of the Richards-Campbell Sleep Questionnaire for Intensive Care Unit Inpatients in Brazil: Internal Consistency, Test-Retest Reliability, and Measurement Error[J]. *Sleep Medicine*, 85: 38-44.
- [38]Wisner, K., Kim, S., Meek, P. M., et al. (2025). Test-Retest Reliability of the Verran Professional Governance Scale[J]. *Journal of Nursing Measurement*, 33(2): 210-216.
- [39]Ye, Y., Wang, J., He, P., et al. (2023). An Action Analysis Algorithm for Teachers Based on Human Pose Estimation[J]. *Computers and Electrical Engineering*, 111: 108915.
- [40]Yuan, J., Ou, H. H. (2011). S-T Analysis of Classroom Teaching[C]// *Proceedings of the 2011 6th International Conference on Computer Science & Education (ICCSE)*. Singapore: Curran Associates, Inc.: 136-140.
- [41]Zhang, Z., Xie, J., Xu, X. T., et al. (2023). Modeling the Information Behavior Patterns of New Graduate Students in Supervisor Selection[J]. *Information Processing & Management*, 60(3): 103342.
- [42]Zhao, G., Zhu, W., Hu, B., et al. (2021). A Simple Teacher Behavior Recognition Method for Massive Teaching Videos Based on Teacher Set[J]. *Applied Intelligence*, 51(12): 8828-8849.

收稿日期 2025-10-26 责任编辑 刘选

Construction and Validation of an Intelligent Coding System for Class Behavior

LIU Shuai, PENG Changling

Abstract: Diagnostic evaluation based on class behavior is a promising research direction in the reform of educational evaluation, which is difficult to implement in massive classes because of the low coding accuracy of current AI tools. Therefore, an AI-adapted class behavior coding system (AI-CBCS) is constructed as a critical technical path in this domain. There are three main stages in the construction of AI-CBCS. First, drawing on the definitions of class behaviors from 11 existing coding systems, 21 common in-class behaviors are selected using methods of expert consultation and class observation. Second, a class behavior system with maximum discrimination is constructed by decomposing the behaviors into high-distinguishable action sequences to improve the distinguishability of classroom behaviors. Third, the coding priority behaviors are determined by considering their relative importance in teaching process to improve the coding accuracy of AI-CBCS. The final AI-CBCS consists of 15 class behaviors, represented by a total of 26 actions, including 11 teacher actions and 15 student actions. In order to validate the effectiveness of the proposed system, 36 high-quality class videos from the SmartEdu platform are extracted as the sample, and a multidimensional comparative analysis is conducted between AI-CBCS and existing coding systems. Quantitative and qualitative results show that the AI-CBCS is better to adapt to AI-enabled coding, featuring higher efficiency and accuracy, and can even reach the level of manual coding. Meanwhile, it can reflect the actual class status in a more diversified manner. Overall, the AI-CBCS can effectively address the problem of overlapping and confused coding of class behaviors in the existing coding system, providing technical and methodological support for intelligent evaluation of large-scale class behaviors in future.

Keywords: Class Assessment; Class Behavior; Coding System for Class Behavior; AI Coding; National Smart Education Platform for Primary and Secondary Schools (SmartEdu)