

课堂教学语言智能化评价： 效用、应用与优化

冯秀梅, 蒋雨辰, 王依婷

(华中师范大学 物理科学与技术学院, 湖北 武汉 430079)

摘要:课堂教学语言量化分析多以人工编码为主,存在内容复杂、工作量大等困难,难以满足大规模课堂评价的实际需求,近年来迅速发展的人工智能技术恰为这些问题带来了新的解决思路。研究以助力培养高质量师范生为导向,从效用评估、具体应用与智能优化三个层面出发,为实现智能化课堂教学语言评价提供技术路径与实证依据。首先,基于师范生教学语言特点构建编码系统,开发自动编码模型,验证其在教学语言编码任务中的可行性与有效性。其次,将自动编码模型应用于师范生实训视频与优质视频的教学语言对比分析,梳理教学语言存在的问题,提出具有普适性的优化建议。最后,结合大语言模型初步构建教学语言优化助理,并基于优化建议为师范生提供个性化的教学语言指导。研究表明,人工智能技术有助于摆脱课堂教学语言量化分析的困境,在自动化、个性化教育评价上具有潜力。

关键词:教育评价;课堂教学语言评价;人工智能;自动化

中图分类号:G40-057 **文献标识码:**A **文章编号:**2095-8129(2025)04-0051-11

基金项目:华中师范大学 2025 年度数智赋能教育教学创新专项教学改革项目“AI 赋能师范生教学能力培养的实践探索”(202525),项目负责人:冯秀梅。

作者简介:冯秀梅,教育学博士,华中师范大学物理科学与技术学院教授,博士生导师;蒋雨辰,华中师范大学物理科学与技术学院硕士研究生;王依婷,华中师范大学物理科学与技术学院硕士研究生。

针对课堂教学进行及时、全面、有效的评价能快速提升教师教学能力,促进高素质专业化教师队伍的建设。传统课堂教学评价多以定性课堂观察为主,对评审者的专业性要求较高。往往由多个专家同时评价以克服定性评价主观性较强的问题,但同时也消耗了大量的专家资源^[1]。事实上,课堂教学的结构化特征为量化评价提供了空间。自 20 世纪 60 年代弗兰德斯建立课堂语言互动分析系统(FIAS),按照固定时间单位对课堂教学进行切片,开创了对课堂定量结构化评价的先河以来^[2],多种改进的互动分析系统不断涌现,以适应信息时代

教学改革和特定学科特点的需要。如基于信息技术的互动分析编码系统(ITYAS)^[3]、凸显科学课堂操作行为特征的现代化学课堂互动分析系统(3C-FIAS)^[4]、中学物理弗兰德斯互动分析编码系统(PFIAS)^[5]等。按照固定时间单位切割课堂可能割裂课堂的结构,为弥补这种不足,量化评价也从课堂对话^[6]、教学活动^[7]和教学事件^[8]等视角对课堂进行分析,对课堂结构进行精细化理解。课堂量化评价通过分类统计课堂行为、关键事件和教学活动,并分析其演化特征,提升了评价的客观性,降低了对评价者专业性的要求,但同时也大大增加了

评价内容的复杂度和评价的工作量。目前,课堂教学量化分析以人工编码为主,对人工消耗过大,导致课堂中隐含信息没有被充分挖掘,无法满足大量课堂及时反馈的需求,不利于课程改革的快速推进。实际上,课堂量化评价面临的困境恰为智能化教学评价的引入提供了现实契机。

随着《新一代人工智能发展规划》^[9]等系列文件的出台,人工智能技术在课堂教学评价领域的应用成为研究热点。例如,基于 GRU 循环神经网络自动分析课堂教学事件^[10],以 Bi-LSTM 神经网络为核心实现教师课堂对话自动标注^[11],这些方法均能有效提高课堂教学结构的识别效率与分析精度。借助自动量化分析结果所揭示的课堂教学结构特征,得以深入探索教学存在的问题,提炼课堂教学改进策略^[10]。此外,大语言模型的发展为课堂教学内容智能优化提供了新路径,预训练—提示机制对用户输入的文本实现内容生成^[12],进一步支持课堂教学内容的智能优化^[13]。因此,本研究根据师范生提升教学技能的现实需求,使用人工智能技术对模拟课堂教学语言进行自动编码,对比分析师范生实训视频和优质视频语言结构上存在的差异。在此基础上,构建融合大语言模型的课堂教学语言智能优化助理,为师范生提供个性化的教学语言优化策略。

一、基于人工智能技术的课堂教学语言智能化评价研究设计

为利用人工智能技术对课堂教学语言进行自动化评价,需要从以下三个方面展开探讨。(1)人工智能技术在教学语言分类自动编码任务上的表现。本研究先构建课堂教学语言编码系统,基于该系统建立人工标注的语言行为数据集,利用该数据集中的部分数据进行训练,形成课堂教学语言自动编码模型,然后

利用该模型对另一部分数据进行编码,与人工标注的结果进行对照,从而评估自动编码模型的可行性和有效性。这是课堂教学语言智能化评价的基础。(2)师范生实训视频与优质视频在教学语言结构上存在的差异。使用自动编码模型对两类课堂的教学行为进行编码,分析师范生实训视频与优质视频在语言结构上的差异,以优质视频的语言为蓝本,对师范生实训课堂语言提出优化建议,探讨人工智能技术在课堂语言评价上的实际应用。(3)利用人工智能技术构建课堂教学语言智能优化助理。研究设计提示信息,根据师范生课堂语言优化建议,让通用大语言模型学习优质视频中的语言表述,形成语言优化助理,对师范生实训课堂中的语言表述进行内容优化。本研究以师范生课堂教学语言为研究对象,从效用评估、具体应用及智能优化三个方面推进实现课堂教学语言智能化评价,具体过程见图 1。

二、基于人工智能技术的自动编码模型有效性检验

(一)课堂教学语言编码系统

师范生的职前教学技能训练常常以模拟课堂的形式开展。模拟课堂的典型特征是无真实学生参与,师范生只能通过“教师提问—假想学生回答—教师反馈”或者“教师指令—假想学生参与—教师反馈”的形式来模拟师生交互过程。在模拟训练课堂中,语言对教学内容的呈现、课堂结构的推进等方面都起到关键作用,所以教学语言是模拟授课的核心关注点和训练重点^[14]。研究以 H 大学物理师范生 2019 级和 2020 级两个年级在教学技能训练课上录制的实训视频的课堂教学语言作为研究对象,视频总数 248 份,每个视频时长约 8 分钟。

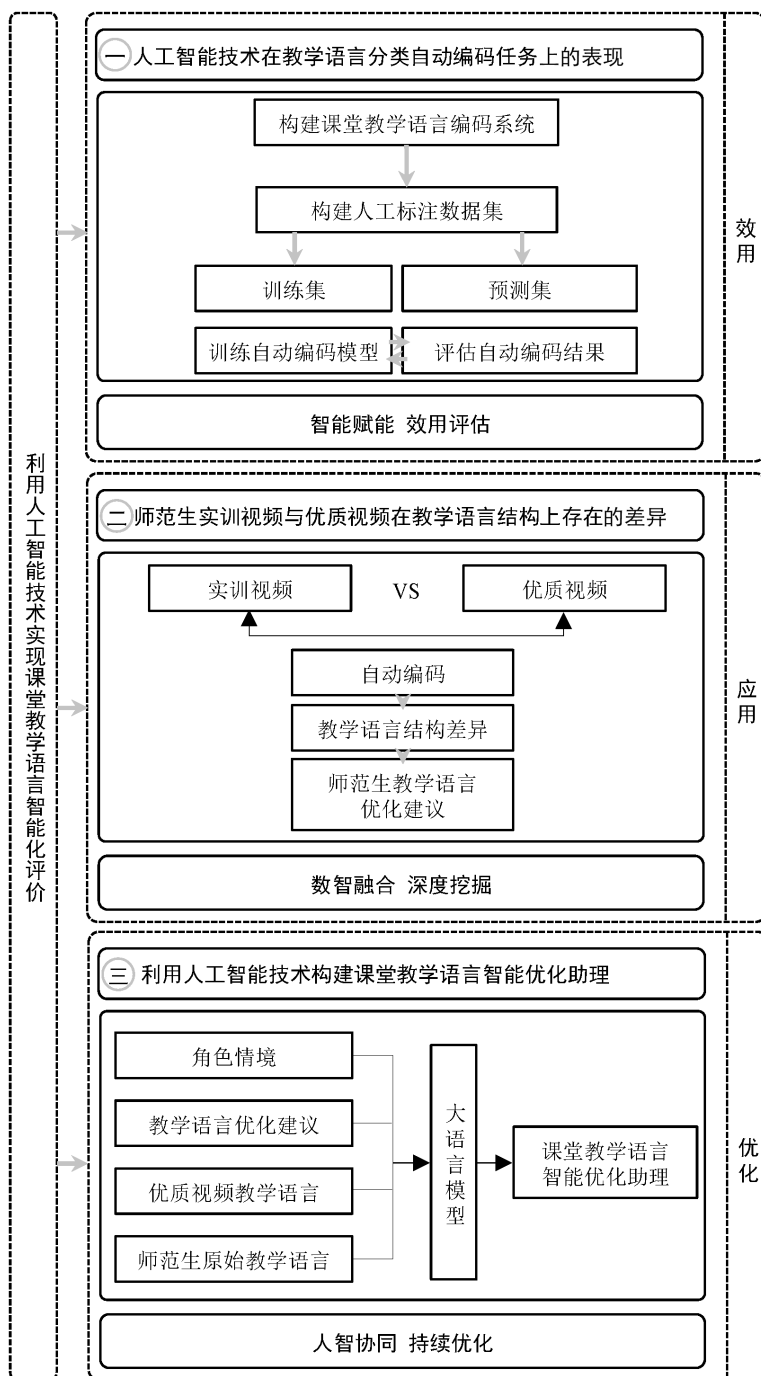


图 1 课堂教学语言智能化评价研究设计

课堂语言智能分析的编码系统以弗兰德斯互动分析系统为基础^[15],结合模拟课堂没有学生参与的特点,剔除学生语言和沉寂行为,重点关注课堂中师范生的教学语言行为。弗兰德斯互动分析系统将教师的语言分成 7 类,其中“接受学生情感”“鼓励或表扬学生”“赞同或使用学生观点”都是教师对学生的积极反馈,主要作用是对学生进行情感激励。精细的

语言分类固然能反映更多的课堂细节,但也会导致更多编码分歧,降低评价的信度,同时也会将更重要的课堂逻辑结构淹在课堂细节中^[16]。因此,本研究将这 3 种合并为反馈类别。最终课堂教学语言编码系统由反馈、提问、讲授、指令、批评或维护权威共 5 个类别组成,具体分类如表 1 所示。

表 1 课堂教学语言编码系统

编码	类别	说明	示例
1	反馈	教师对学生的行为进行评价,属于积极影响	这三位同学发表的言论都非常好
2	提问	教师对单个或所有学生提出问题	那么你知道电和磁之间有联系吗
3	讲授	教师表达观点,陈述知识,围绕相关知识进行讲解、讲评	气球从静止向上飞起,说明它一定是受到了一个向上的合力
4	指令	教师利用语言承转课程环节,或要求、建议学生作出行为、进行思考、解决问题	接下来我们就用实验来求证一下,老师这里邀请一位同学跟我一起做实验
5	批评或维护权威	批评、责备学生,维护教师权威	看来 WQ 同学课前要好好预习

(二)人工编码过程

依据编码系统构建人工文本标注数据集。利用语音转写工具将讲课视频中的语言转为文本,转录完成后对物理专业术语进行了人工矫正。除此之外,人工矫正还审核:(1)只保留师范生的教学语言,忽略教学中播放其他视频辅助教学等产生的辅助语言;(2)检查语义的完整性,以转录文字中的“。”“?”“!”标点符号为一句话结束的标识。

本研究以完整的一句话作为语义单元对文本进行分割,依据表 1 中 1~5 类语言行为,分析转换文本中的每一句话属于哪一类编码。让两位经过编码培训的研究人员独立编码 10 个相同文本,针对出现的分歧进行充分讨论。初次编码的分歧主要来源于讲授和指令两个类别的区分,经过讨论达成以下共识:(1)对实验现象和实验操作的描述,即使包含部分指令类动词,如观看、拿起等,仍归为讲授类。例如:首先看这个潜水艇,潜水艇在水里与水接触,那么潜水艇由于有重力,它对下面的水有一个压力的作用,力的作用是相互的,所以

水会给潜水艇一个向上的力。此类话语的主体是对实验现象进行讲解,不是要求、建议学生作出行为,因此归为讲授。(2)将推进课程进度的语言归为指令。例如:上课,同学们大家好,今天我们要讲的是磁场、磁感线。这里没有实际知识讲解,只是课程环节的承接句,归于指令类。因此,区分讲授和指令两个类别的关键点为语言是否围绕物理知识展开,并在表 1 的编码系统中对相应类别作详细说明。讨论完成后,两位编码人员独立完成剩余文本编码,并对其进行一致性检验,结果达到 0.94,这说明每个编码类别的语义表述清晰,符合内部可靠性要求,该人工标注数据集可作为后续自动编码模型训练与性能评估的基础语料。

2020 级人工文本标注数据集共 7 955 句,作为训练集;2019 级人工文本标注数据集共 905 句,作为测试集,具体类别数量如表 2 所示。因批评或维护权威类语句数量极少,无法形成有效训练样本,故不对此类别讨论自动编码效用。

表 2 数据集各类别统计数量表

类别	训练集(2020 级)/句	测试集(2019 级)/句
反馈	562	122
提问	2 443	283
讲授	3 672	270
指令	1 276	229
批评或维护权威	2	1
总计	7 955	905

(三)机器学习形成自动编码模型

机器学习是人工智能领域中的一种重要技术。要实现人工智能对教学语言自动编码,

首先要用特征向量准确有效地表示教学语言。特征向量是对文本词汇或语义特征的数学表达形式,作为模型可处理的数值化输入。机器

学习模型通过学习特征向量与语言类别之间的映射关系,实现自动分类编码的功能。为了对机器学习结果进行评估,需将人工标注的编码数据集分为训练集和测试集,训练集用于训练机器模型,而测试集用于证明模型的性能。

机器学习的文本分析中常用 n 元语法 (n-gram) 模型,将文本切分成固定长度的词序列来捕捉词语之间的关系,筛选出高频组合作为有效特征,用于构建文本的特征向量表示^[17]。当 $n=1$ 时,称为一元语法 (unigram) 模型;当 $n=2$ 时,称为二元语法 (bigram) 模型;当

$n=3$ 时,称为三元语法 (trigram) 模型。当 n 大于 3 后,数据稀疏和计算代价的问题会比较显著,在实际研究中较少使用。本研究根据准确率 (accuracy),即预测结果正确的值除以总的测试数,以及 Kappa 值,即预测结果和标准标注结果之间的一致性,来综合评定具体选择哪个语法模型。如表 3 所示,随着语法模型的复杂度提高,特征向量数、准确率和 Kappa 值均有所提升,可以据此猜想:特征向量数越多,准确率越高,模型表现越优。

表 3 n 元语法模型选择指标数值

指标	一元语法模型	一二元语法模型	一二三元语法模型
准确率	0.909	0.913	0.914
Kappa 值	0.862	0.869	0.871
特征向量/个	2 095	7 759	11 985

为验证这个假设的正确性,采用一二三元语法模型,以准确率为纵轴,特征向量数为横轴,1 000 为间隔作图,发现随着特征向量数的增加,准确率呈现先增后减的趋势,在 5 000 个特征向量时达到最高点,准确率为 0.916 8,后

续总体呈下降趋势,如图 2 所示。因此,选择 5 000 特征向量能在节省算力的同时,保证较高的准确率。综合上述分析,训练集选择一二三元语法模型进行特征向量提取,得到 5 000 个特征向量。

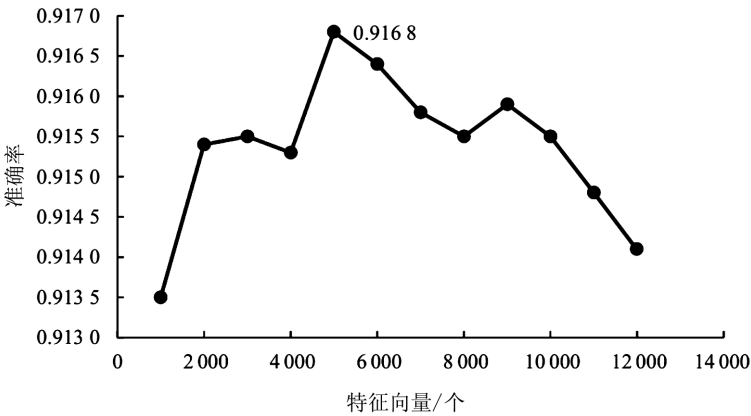


图 2 准确率随特征向量数变化趋势

在完成特征提取后,需选择合适的机器学习算法进行训练,并对模型性能进行评估。逻辑回归 (logistic regression) 是一种经典的文本分类算法,自然语言处理领域中被广泛应用于分类任务^[18]。因此,文章选择训练集中提取的 5 000 个特征向量,并基于逻辑回归算法进行训练,最终在训练集上分类准确率为 0.916 8, Kappa 值为 0.872 9,均大于 0.81,自动编码一致性

较高^[19]。为了解训练集中机器学习的编码标准,按照影响值 (influence) 选取对分类结果影响最大的前 10 个特征向量,结果如表 4 所示。当教学语言开头位置出现“对”时,模型倾向于归为反馈语言;语言包含“?”时多归为提问语言;包含“叫作”时多归为讲授语言;包含“一下”时多归为指令语言。

表 4 四个类别下高度影响分类结果的特征向量

反馈	提问	讲授	指令
(开头)对	?	叫作	一下
觉得	? (结尾)	。	请
谢谢	什么	叫	今天
听到	呢	等于	思考
是的	呢?	注意,	(开头)请
说	呢? (结尾)	也	好,
(开头)很好	什么?	物体	接下来
(开头)很	什么? (结尾)	力	时间到
(开头)我	的?	读数	(开头)注意
同学	的? (结尾)	就是	尝试

表 4 显示的高影响特征向量较好地代表了各类别下的典型课堂教学语言特征,整体贴合人工编码标准的描述。但机器学习与人工编码结果仍存在一定差异,原因在于提取的特征向量难以涵盖所有的语言结构,也不能全面体现上下文之间的联系,造成部分编码结果出现偏差。例如:这位同学第二次接触的时间,老师可以说大概在 0.16 秒。这是对某同学参与活动后的结果进行陈述,并非评价反馈,但模型编码时,遇到“同学”等词后会更倾向于将其归类到反馈,从而产生错误编码。少量情况下,句子开头的“对”仅为口头禅性质的语气词。例如:对,弹弓的原理其实和橡胶水枪的作用原理是一样的。该句前文无学生回答,不构成反馈,但模型仍受“(开头)对”特征影响,产生错误编码,以上问题需要通过进一步的算法优化和数据集修正来解决^[20]。

(四)自动编码结果评估

1. 模型量化评估指标

机器模型在训练集上的表现固然重要,但在新的测试集上的泛化能力更为关键,能反映模型在实际应用中对未见数据的适应性和准确性。在文本分类领域,精确率(precision)、召回率(recall)和 F_1 值指标被广泛用于评估分类模型的泛化性能^[21]。TP(true positive)表示预测值和真实值都是反馈类别的数量;FP(false positive)表示预测值为反馈类,但真实值不是反馈类的数量;FN(false negative)表示预测值为其他类别,但真实值为反馈类的数量。精确率,简称为 P 值,指某类别下预测值正确的数

量占这类全部预测结果的比率,即 $P =$

$\frac{TP}{TP+FP}$;召回率,简称为 R 值,指某类别下预

测值正确的数量占这类全部真实结果的比率,

即 $R = \frac{TP}{TP+FN}$; F_1 值,指精确率和召回率的

调和平均值, $F_1 = \frac{2 \times P \times R}{P + R}$,只有当 P 和 R 同

时较高时才能得到较高的 F_1 值, F_1 越大说明模型的性能越好。图 3 以反馈类别为例,展示上述 TP、FP 和 FN 的含义。

		预测值			
		反馈	提问	讲授	指令
真实值	反馈	TP	FN	FN	FN
	提问	FP			
	讲授	FP			
	指令	FP			

图 3 模型评估参数:以反馈类为例

2. 自动编码结果评估

测试集必须是新的数据集,以 2019 级人工文本标注数据集作为测试集,使用训练模型进行自动编码。其混淆矩阵结果如图 4 所示,纵坐标表示人工文本标注的真实答案,横坐标表示机器学习预测结果,对角线的值为机器学习自动编码分类与人工文本标注一致的数量。对角线整体呈现深蓝色,可见机器学习自动编

码效果较好,绝大部分分类结果与人工标注结果一致。

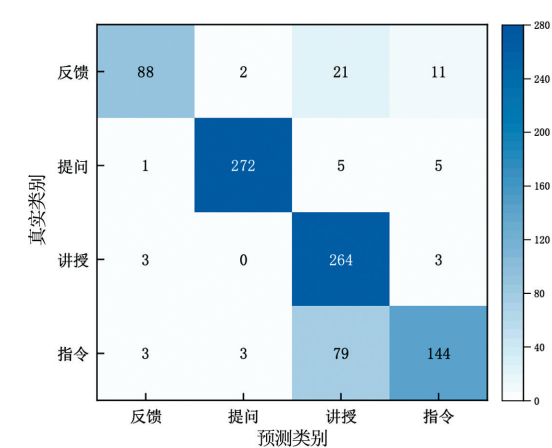


图 4 机器学习混淆矩阵

各类的精确率、召回率和 F_1 值如表 5 所示。四个类别中,提问的 F_1 值最大,达到 97.14%,相较而言,指令类别的结果较差,但也达到 73.47%。模型平均精确率、召回率和 F_1 值已达到 80% 以上,说明模型对师范生课堂教学场景具备良好的适应性,能够较为精准地对课堂教学语言进行分类,使用该模型对课堂教学语言进行自动分类编码将节省大量时间和人力。

表 5 机器学习各项指标结果

类别	精确率/%	召回率/%	F_1 值/%
反馈	92.63	72.13	81.10
提问	98.19	96.11	97.14
讲授	71.54	97.78	82.63
指令	88.34	62.88	73.47
平均值	87.68	82.23	83.59

三、自动编码模型的应用

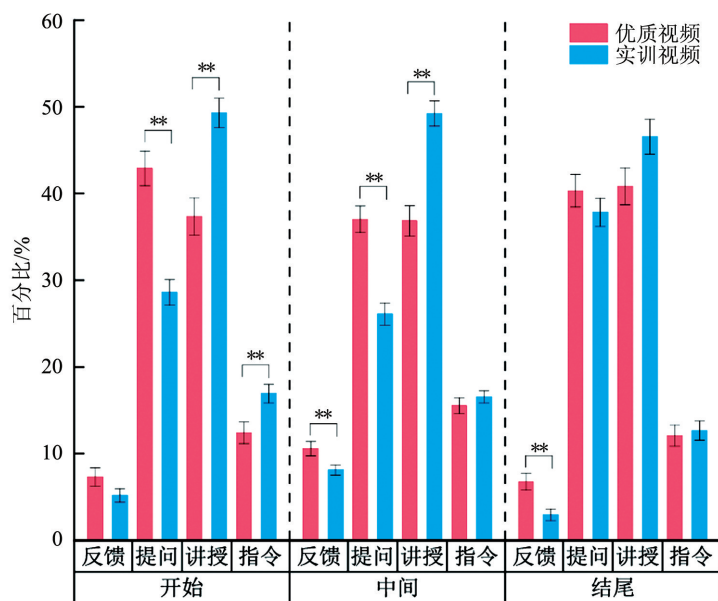
本研究进一步将自动编码模型应用于课堂教学语言的实时分析任务,以验证其适用性与实用价值。选择全国物理师范生教学技能竞赛“格致杯”和“华夏杯”中获奖的 59 个视频作为优质视频集。这两类竞赛的授课形式与师范生教学技能训练课一致,均为 8~10 分钟的模拟授课,台下没有真实的学生。获奖的优质视频均经过各校专业教师的多次指导和选手的反复打磨,并在全国近千名物理师范生模拟授课教学竞赛中脱颖而出,其课堂教学语言

可作为师范生模拟课堂的典范。将优质视频转录为文本,并用已训练完成的自动编码模型对其教学语言进行自动标注,4 个语言类别的精确率、召回率和 F_1 值在 0.7 以上,自动编码分类效果较好。因此,本研究以优质视频作为师范生模拟授课的优秀范本,分析两类课堂不同阶段的语言分布和转换规律差异,探讨师范生课堂教学语言的改进途径,为优化教学语言提供实证依据。

(一)教学语言类型分布特征

为了深入了解师范生实训视频和优质视频语言分布差异,将课堂分为开始(20%)、中间(60%)和结尾(20%)3 个阶段,计算每位师范生在不同课堂阶段下的 4 类教学语言所占百分比。从图 5 中的结果可以看出,实训视频和优质视频在课堂各阶段的教学语言结构差异显著。在课堂开始和中间阶段,优质视频中的提问语言比例显著高于实训视频,且各个阶段比例均超过 30%。在中间和结尾阶段,优质视频中的反馈语言比例显著更高。而实训视频在 3 个阶段的讲授语言占比均高于 40%,明显高于优质视频。上述结果表明,优质视频更加注重课堂互动与启发,而实训视频的教学语言较为固化,依赖讲授,缺乏有效交互。

同时也发现,在课堂开始阶段,师范生实训视频的指令语言占比显著更高。通过分析具有代表性的一组实训视频和优质视频中课堂教学语言类别随时间的变化,发现实训视频中的指令语言集中在课程开始阶段,主要用于引入本节课的知识内容。而优质视频中的指令语言贯穿课堂前、中、后期,分布较为均匀。以该组具体教学内容为例,推动课堂环节的指令包括“分发传感器—观察传感器结构—组织小组讨论—鼓励学生发表观点—观看实验视频—总结课堂知识—抛出课后任务”,以指令语言承转课堂关键环节,一方面指导学生参与课堂活动,另一方面也显示了严密的教学逻辑。而师范生在实训视频中则采用教师演示实验或播放实验视频的方式,学生仅扮演观看者的角色,课堂参与度较低。



注: * * 表示组间差异显著, $p < 0.01$ 。

图5 不同课堂阶段教学语言比例差异

(二)教学语言转换特征

为深入探究课堂教学语言的动态转换特征,采用语言转换概率作为核心分析指标。具体计算方式如下,针对每一类教学语言(反馈、提问、讲授、指令),统计其后续语言类别的频次,并以总频次为基准计算其转向后续各类语言的概率(以百分比形式呈现)。

师范生实训视频与优质视频在教学语言转换上存在较大差异。实训视频各类语言均大概率转换为讲授,其中“反馈→讲授”的概率为43.98%，“提问→讲授”的概率为36.67%，“指令→讲授”的概率为44.86%。结果表明实训课堂的教学以知识讲解为主,互动较少。

而优质视频各类语言更倾向于转换为提问,例如反馈和指令分别有43.22%和38.23%的概率转换为提问。并且优质视频中“反馈→反馈”(11.00%)的语言转换概率高于实训视频(6.93%),“提问→提问”(41.55%)的语言转换概率也高于实训视频(31.93%)。这表明其更倾向于构建连续的互动链条,维持教学流畅性和学生的参与度。

(三)师范生课堂教学语言优化建议

基于课堂教学语言结构特征的差异,利用大语言模型对师范生教学语言提出优化建议,具体如下。

第一,减少讲授依赖,强化互动意识。语

言形式的讲授虽然能够在较短时间内给学生传授大量的系统的科学文化知识,但易出现满堂灌的僵化教学模式,削弱学生的主动思考与探究能力,不利于综合素养的培养。由于缺乏教学经验,师范生在授课时仍以知识点的完整讲解为核心目标,忽视学生视角的互动需求,这导致在课堂上出现话语霸权的现象,学生参与度不足。而优质视频中的师范生更善于控制讲授时间,避免长时间单向讲解,通过合理的课堂节奏安排减少不必要的讲授内容,适时融入提问、反馈和学生活动,以提高课堂互动性和学生的学习投入。因此,师范生应着力于减少讲授语言的依赖,合理设计互动环节,通过精练讲解、构建问题链、采用多元教学策略等方法,使课堂更具启发性和参与感,提升教学效果。

第二,学习提问策略,增强引导反馈。课堂提问是教学过程中必不可少的重要环节之一,有目的、有技巧的课堂提问配合高质量反馈,能促使学生主动思考、调整认知,发挥课堂主体作用。但师范生课堂的反馈和提问比例整体较低,限制了课堂的互动性,使学生在大部分时间处于被动接受知识的状态。师范生急需优化提问的时机与方式,学习优质视频中的教学策略,将提问贯穿课堂始终,并结合反馈形成有效互动。在对学

应与处理的基础上引出下一个提问,形成连贯的问题链,推动课堂发展。

第三,把握课堂节奏,优化实验设计。准确简洁的课堂指令是把控课堂节奏、确保教学环节流畅衔接的重要工具。师范生的指令主要集中于课程开始阶段,未能与提问、反馈等语言相互衔接,难以有效推动教学流程。师范生可以通过设计小组实验等探究式教学活动,在实验过程中提供精准明确的指令,以组织实验操作、引导思考讨论的形式,确保学生能够在实践中深化理解,并将课堂结构清晰地整合起来,使教学过程更具流畅性和节奏感。因此,师范生在教学实践中应优化指令的分布与表达,使其不局限于知识导入,还能在教学全过程中发挥组织和引导作用,提升课堂互动性和学生的自主探究能力。

四、人工智能技术优化师范生教学语言

师范生在完成模拟授课后,通常需要专家对师范生的模拟课堂进行点评,帮助他们快速成长,但由于专家资源的紧缺,不能做到为每名师范生的教学语言提供持续个性化指导。基于课堂教学语言优化建议,利用大语言模型构建的课堂教学语言智能优化助理,能打破时间地点和资源的限制,支持师范生在课堂实践中优化教学语言表达^[22]。

为验证其可行性,选择师范生 A 讲解“平抛运动”课题时导入的教学语言作为优化对象,原文如下:同学们好,在上节课中我们提出了一个关于平抛运动的猜想,那就是平抛运动可以分解为水平方向上的匀速直线运动和竖直方向上的自由落体运动,并且通过观看演示实验验证了我们的猜想。那么我们这节课将亲自做实验,通过实验来验证我们的猜想。这是老师上节课最后留下来的第二种验证平抛运动的方法,平抛运动的闪光照相。我们通过记录做平抛运动的小球,在不同的时间点和位置得到了平抛运动的图像,然后我们再通过平抛运动水平方向上的位移和竖直方向上的位移,看是否与我们预期符合,来验证我们的猜想。那么具体要怎么做?请同学们观看一个演示实验,并且找一下演示实验仪都由哪些部

分组成。以上师范生 A 的导入片段能代表师范生教学语言存在的普遍问题,即讲授内容过长,指令表达不够清晰,针对学生的提问少且指向性不强,这导致学生不明白观看演示视频的具体目的,教学语言需要进一步优化。

在使用大语言模型时,提问的方式将极大影响输出的结果^[23]。因此,逐步优化提问的方式,最终形成的提示信息包含角色情境、教学语言优化建议、优质视频教学语言和师范生原始教学语言四个部分。在角色情境部分说明此时的角色情景和应完成的优化教学语言任务,再将前文基于教学语言结构数据对比分析所得到的优化建议输入到提示信息中。优化建议是面向两个群体分析提出的,虽有普遍适用性,但针对性较差。因此,在保留优化建议的基础上,在提示信息中增加相同课题的优质视频教学语言作为模板,让大语言模型自动学习其语言结构和内容后,优化师范生原始教学语言。最终优化结果如下:同学们好!上节课我们提出了一个关于平抛运动的猜想,即平抛运动可以分解为水平方向上的匀速直线运动和竖直方向上的自由落体运动。通过观看演示实验,我们初步验证了这个猜想。今天,我们将亲自做实验,通过实验进一步验证这一猜想。我们将使用平抛运动的闪光照相法来记录小球在不同时间点的位置,并生成平抛运动的图像。然后,我们会分析小球在水平方向和竖直方向的位移,看看是否与我们的预期相符,从而验证我们的猜想。现在,请大家完成以下步骤。(1)观看演示实验:请大家集中注意力,观察实验中使用的各个仪器部件。(2)提问环节:思考并回答,哪些仪器部件对记录小球运动轨迹至关重要?为什么?你认为实验中的关键步骤有哪些?(3)实验指导:分成小组,每组两人。讨论并设计一个简单的实验,用来记录平抛运动的轨迹。使用实验室中的仪器进行实验,记录数据并生成图像。请大家先观看演示实验,现在开始。大语言模型的优化结果使整个教学语言更加精练并富有逻辑,点明学生观看演示实验时的具体要求,并明确说明在观看过程中所应思考的问题,切实增加了指令和反馈类型的教学语言。同时模

仿优质视频的教学内容,在观看实验的过程中加入了分组讨论,并要求设计实验,增加了更为丰富多样的自主活动,这也能为师范生的课堂设计提供有效的灵感。未来可以考虑建立优质视频教学语言模板库,只需专家将优化要求和建议输入大语言模型中,即可实时为每位师范生提供个性化的教学指导。师范生则能在不断交互过程中,反思个人教学语言存在的问题,迅速调整语言表达和教学方法,从而逐步提升教育教学能力。

五、总结与讨论

在人工智能技术融合教育蓬勃发展的背景下,本研究立足于当前教学评价存在的问题和需求,从效用、应用和优化三个方面探讨课堂教学语言智能评价的可行性。

当前智能课堂教学评价的开发和实施仍处于起步阶段,构建专业化、智能化的评价框架,使其针对性地指向智能化课堂教学评价是现实急需^[24]。本研究基于弗兰德斯互动分析系统,把握模拟课堂教学特点,形成课堂教学语言编码系统。该编码系统定义清晰,可操作性强,切片方式确保了教学语言的完整语义,可针对课堂教学语言进行类别判定与质量分析。训练得到的自动编码模型结果表现优异,具有较好的泛化能力,初步实现黄焕等对于课堂话语分析应用探索的设想^[16],能协助实现多维数据课堂教学行为指标体系中话语数据的自动量化分析^[25],共同探索和丰富在不同框架理论下课堂教学智能化评价的实践结果^[26]。这将大大节省研究人员人工编码所需的人力和时间,切实解决当前量化分析教学课堂存在的问题,满足大规模课堂评价需求^[27]。

在对比分析中,利用自动编码结果分析师范生实训视频与优质视频在教学语言结构和内容上的差异,揭示师范生在教学语言方面存在的具体问题,提出针对师范生教学语言优化的改进策略。这一过程不仅为编码模型在教育场景中的具体应用提供了范式,也为智能化教学助手的构建提供了方向指引和实证依据。基于大语言模型的智能化教学助手有望成为师范生专业发展的重要支持工具,推动教育领

域的持续创新与优化,助力精准化、个性化的教师培养^[28]。本研究初步验证了大语言模型作为师范生教学语言优化助手的可行性,展示了其在提升教学语言质量、促进教学改进中的应用潜力,为个性化教育与智能化教学评价提供了新的思路和实践路径。

本研究充分利用人工智能在大数据处理、语言学习和文本生成方面的技术优势,实现了课堂教学语言的高效自动编码,并为师范生提供个性化反馈建议,为后续构建更完善的智能化教学评价体系奠定了基础。然而研究仍存在一定局限性,对于需要更精细颗粒度分析课堂特征的场景区,还需丰富编码类别,并在每一个类别下进行水平划分,以实现对课堂教学的质量水平判定。此外,研究仅聚焦于教学语言的分析优化,暂未收集其他模态教学信息。未来可以补充其他模态信息,丰富当前数据集,构建课堂教学智能化评价平台,为师范生课堂教学技能的发展提供立体化的智能支持。

参考文献:

- [1] 王金羽,詹逸思,冯起,等.教育质性研究中人机协同文本挖掘技术的运用——以某高校教学评估中文文本数据为例[J].清华大学教育研究,2022(2):56-63.
- [2] FLANDERS N A. Intent, action and feedback: a preparation for teaching[J]. Journal of Teacher Education, 1963, 14(3): 251-260.
- [3] 顾小清,王炜.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,2004(7):18-21.
- [4] 高瑛,林康立,马宏佳.研究现代化学课堂教学行为的量化工具——3C-FIAS[J].化学教育,2016(5):18-24.
- [5] 陈珍国,邓志文,于广瀛,等.基于FIAS分析模型的翻转课堂师生互动行为研究——以中学物理课堂为例[J].全球教育展望,2014(9):21-33.
- [6] HOWE C, HENNESSY S, MERCER N, et al. Teacher-student dialogue during classroom teaching: does it really impact on student outcomes? [J]. Journal of the Learning Sciences, 2019, 28(4/5): 462-512.
- [7] 穆肃,左萍萍.信息化教学环境下课堂教学行为分析方法的研究[J].电化教育研究,2015(9):62-69.
- [8] 孙众,吕恺悦,施智平,等. TESTII 框架:人工智能支持课堂教学分析的发展走向[J].电化教育研究,2021(2):33-39,77.
- [9] 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. [2017-07-20]. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

- [10] 孙众,吕恺悦,骆力明,等. 基于人工智能的课堂教学分析[J]. 中国电化教育,2020(10):15-23.
- [11] JACOBS J, SCORNAVACCO K, CLEVENGER C, et al. Automated feedback on discourse moves: teachers' perceived utility of a professional learning tool[J]. Educational Technology Research and Development, 2024, 72(3):1307-1329.
- [12] 罗恒,廖小芳,茹琦琦,等. 生成式人工智能支持的教师评语研究:基于初中数学课堂的实践探索[J]. 电化教育研究,2024(5):58-66.
- [13] 郑永和,周丹华,张永和,等. 计算教育视域下的 ChatGPT:内涵、主题、反思与挑战[J]. 华东师范大学学报(教育科学版),2023(7):91-102.
- [14] 叶立军,李燕,斯海霞. 初中数学新老教师课堂教学语言比较研究[J]. 数学教育学报,2015(4):40-43.
- [15] FLANDERS N A. Analyzing teacher behavior[M]. Reading-MA: Addison-Wesley Publishing Company, 1970:34.
- [16] 黄焕,刘清堂,朱晓亮,等. 不同教学风格的课堂话语特征分析及应用研究[J]. 现代教育技术,2013(2):27-30,36.
- [17] 凤丽洲,杨贵军,徐雪,等. 基于 N-gram 的双向匹配中文分词方法[J]. 数理统计与管理,2020(4):633-643.
- [18] 顾鑫,曹丹华,吴裕斌,等. 基于逻辑回归的多任务域快速分类学习算法[J]. 计算机工程与应用,2017(15):47-56,205.
- [19] LANDIS J R, KOCH G G. The measurement of observer agreement for categorical data[J]. Biometrics, 1977, 33(1):159-174.
- [20] 杨晓哲,王晴晴,蒋佳龙. 基于人工智能的课堂师生对话分析:IRE 的自动分类与分水平构建[J]. 电化教育研究,2023(10):79-86.
- [21] SOKOLOVA M, LAPALME G. A systematic analysis of performance measures for classification tasks[J]. Information Processing & Management, 2009, 45(4):427-437.
- [22] 宋宇,许昌良,穆欣欣. 生成式人工智能赋能的新型课堂教学评价与优化研究[J]. 现代教育技术,2024(12):27-36.
- [23] WAN T, CHEN Z. Exploring generative AI assisted feedback writing for students' written responses to a physics conceptual question with prompt engineering and few-shot learning[J]. Physical Review Physics Education Research, 2024, 20(1):010152.
- [24] 吴立宝,曹雅楠,曹一鸣. 人工智能赋能课堂教学评价改革与技术实现的框架构建[J]. 中国电化教育,2021(5):94-101.
- [25] 杭留一,叶海智,张瑾,等. 基于多维数据的课堂教学行为评价研究[J]. 现代教育技术,2022(4):68-77.
- [26] 赵明明,周森. 如何分析课堂发言:人工智能与统计方法的结合[J]. 北京大学教育评论,2019(4):35-44,184-185.
- [27] 骆祖堂,张丹慧. 课堂教学自动评价及其初步研究成果[J]. 现代教育技术,2018(8):38-44.
- [28] 李英哲,刘剑玲. 新课标背景下 ChatGPT 在初中数学备课中的创新功能与价值定位[J]. 中国电化教育,2024(3):109-114.

Intelligent Evaluation of Classroom Teaching Language: Effectiveness, Application, and Optimization

FENG Xiumei, JIANG Yuchen, WANG Yiting

(College of Physics Science and Technology, Central China Normal University, Wuhan 430079, China)

Abstract: Quantitative analysis of classroom teaching language traditionally relies on manual coding, presenting significant challenges such as complexity and labor-intensive workloads, which hinder its applicability to large-scale classroom evaluations. Rapid advancements in artificial intelligence technology have recently offered promising solutions to these issues. Therefore, this study aims to foster the development of high-quality pre-service teachers by providing a technical pathway and empirical evidence for intelligent classroom language evaluation from three key dimensions: effectiveness evaluation, practical application, and intelligent optimization. Initially, an encoding system specifically tailored to the instructional language characteristics of pre-service teachers was developed. Then, an automated encoding model was constructed and tested for feasibility and effectiveness in language coding tasks. Subsequently, this automated encoding model was applied to analyze instructional language differences between pre-service teachers' training videos and exemplary teaching videos, systematically identifying problems and proposing generalized optimization recommendations. Finally, an instructional language optimization assistant was preliminarily established using large language models, offering personalized language guidance to pre-service teachers based on identified optimization strategies. The findings demonstrate that artificial intelligence technologies effectively address the challenges of quantitative instructional language analysis and show significant potential in automation and personalized educational evaluation.

Key words: educational evaluation; evaluation of classroom teaching language; artificial intelligence; automation

责任编辑 谭小军 李 双