

基于师生说话人确认技术的课堂交互分析

温彦, 张玉浩, 渠嘉威, 于恩海

(山东科技大学 计算机科学与工程学院, 山东 青岛 266590)

摘要: 针对传统课堂交互分析依赖人工编码导致效率低下、主观性强等问题, 提出了基于师生说话人确认技术的课堂交互分析方法。在中小学真实课堂场景中, 该技术取得了91.8%的说话人身份确认准确率, 成功实现了对话轮次转换、话语时长分布及课堂互动模式等信息的自动化统计, 并进行了多维可视化分析。这种方法显著提升了课堂诊断的客观性和教研效率, 为教育数字化转型提供了关键技术支撑。

关键词: 课堂交互分析; 说话人确认技术; 多维可视化; 教育数字化转型

DOI: 10.11907/rjdk.251538

扫描二维码阅读全文:



中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)011-0035-07

Classroom Interaction Analysis Based on Teacher-Student Speaker Verification Technology

WEN Yan, ZHANG Yuhao, QU Jiawei, YU Enhai

(School of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: To address the issues of low efficiency and strong subjectivity caused by manual coding in traditional classroom interaction analysis, this study proposes a classroom interaction analysis method based on teacher-student speaker verification technology. It introduces the practical application in real primary and secondary school classroom scenarios, achieving 91.80% speaker identification accuracy. The method successfully automates the extraction of core parameters including turn-taking transitions, utterance duration distribution, and classroom interaction patterns, while enabling multidimensional visualization analysis. This approach has significantly enhanced the objectivity of classroom diagnostics and teaching research efficiency, providing key technical support for educational digital transformation.

Key Words: classroom interaction analysis; speaker verification technology; multi-dimensional visualization; educational digital transformation

0 引言

人工智能在当代极大地提升了生产力, 为人们的生活带来了诸多便利。如今, 各行各业都可见人工智能的应用, 教育领域也不例外。近年来, 诸如人工智能教育系统、智慧课堂、自适应学习系统等名词频繁进入视野, 预示着教育信息化已成为不可阻挡的发展趋势。中共中央、国务院印发的《中国教育现代化2035》重点部署了面向教育

现代化的十大战略任务, 其中第八项明确提出要“加快信息化时代教育变革, 统筹建设智能化校园”^[1]。教育信息化并非人工智能授课、人工智能辅导等教课形式的信息化, 其核心在于每个教育环节的信息化, 通过教育数据的联通与流程整合, 实现各环节间的协同作用, 最终提升整体教育质量^[2]。教育部《教育信息化2.0行动计划》明确指出, 需构建“基于人工智能的课堂观察与诊断系统”, 课堂的诊断观察是促进教育改革发展的关键抓手^[3]。利用信息技术分析课堂教学行为, 评价教学效果, 这一做法符合

收稿日期: 2025-08-07

基金项目: 山东省教育科学“十四五”规划重点课题(2023ZD038); 山东省本科教学改革研究项目(Z2023186); 山东科技大学青年教师教学拔尖人才培养项目(BJ20211116); 山东科技大学教育教学研究“群星计划”(QX2023M20)

作者简介: 温彦(1984-), 女, 博士, 山东科技大学计算机科学与工程学院副教授、硕士生导师, 研究方向为机器学习、智慧教育、人工智能; 张玉浩(2000-), 男, 山东科技大学计算机科学与工程学院硕士研究生, 研究方向为人工智能、智慧教育; 渠嘉威(2000-), 男, 山东科技大学计算机科学与工程学院硕士研究生, 研究方向为人工智能、智慧教育; 于恩海(2001-), 男, 山东科技大学计算机科学与工程学院硕士研究生, 研究方向为人工智能、智慧教育。本文通讯作者: 温彦。

新课程改革提倡的“立足过程,促进发展”方针^[4]。

话语是课堂教学过程中知识传播的基本载体,师生间的话语转换是课堂交互的主要形式之一^[5]。Molinari 等^[6]认为,教师在课堂语言互动中占主导地位,教学需要借助语言交流来实现。日本学者佐藤学也提出,学习是以语言为媒介,通过语言性实践构建有意义的内容。因此,通过分析师生话语来实现课堂交互分析,已成为主流方法。在此基础上,本文聚焦于课堂话语分析领域,利用声纹技术对师生说话人进行分析,成功实现了对特定时间片段中说话人身份的准确辨别。同时,结合 S-T 分析理论对教师与学生的课堂行为进行了精准分类,并综合了教学模式分布图、课堂时轴图以及 S-T 曲线等多维数据,从而实现了课堂教学互动的全面、系统性分析^[7]。

1 研究现状

1.1 课堂分析方法

常用的课堂分析方法具体包含弗兰德斯分析系统(Flanders Interaction Analysis System, FIAS)^[8]和学生-教师分析法(S-T 分析)^[7]。其中,FIAS 的核心逻辑是通过矩阵表格记录下教室内师生互动的状态并赋予编码;S-T 分析法则简化了记录内容。近年来,课堂分析研究向着智能化、信息化的方向发展,具体表现如下:张治等^[9]提出了五维观测视角和基于人工智能的高质量课堂分析框架;汤倩雯等^[10]构建了立足智慧课堂的多模态学习分析模型;吴军其等^[11]利用多模态数据测评,实现了对学生学习状态的更全面分析;张慧等^[12]选取典型课例,从教学风格、教学模式、教学结构等方面对比不同分析工具的优缺点;骆祖莹等^[13]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,深入剖析了整个教学活动过程。

1.2 说话人识别

说话人识别的概念最早由 Kersta 于 1962 年在《Nature》杂志上发表的论文中首次提出。随后的发展过程中,相关技术经历了从传统信号处理方法到深度学习方法驱动的转型。传统方法研究中,Davis 等^[14]提出的梅尔频率倒谱系数(MFCC 特征)与 Reynolds 等^[15]提出的通用背景模型(GMM-UBM 模型)为说话人识别技术的进一步发展奠定了技术基础。2010 年 Dehak 等^[16]正式提出了 i-vector 的概念,将说话人和会话变化统一建模为一个低维的 i-vector,并详细描述了其在说话人识别中的应用,成为了传统方法的里程碑。

近年来,深度学习发展迅速,因其得到的特征表现力空前强大,embedding 随之逐渐取代 i-vector 成为主流的语音特征。深度神经网络模型(DNN)中提取到的特征 d-vector^[17]初步展现出深度特征的优势,随后时延神经网络(TDNN)代替 DNN,通过时延神经网络实现说话人嵌入的

动态建模,提取 x-vector^[18]特征,准确率较 i-vector 提升 15%~20%;2020 年 Desplanques 等^[19]提出的 ECAPA-TDNN 模型通过引入通道注意力机制、密集连接结构和更深的时序建模层,构建了具有上下文感知能力的增强型说话人嵌入,表征能力逐渐强大。Sell 等^[20]提出的 d-vector 方法首次将深度神经网络应用于说话人表征学习,而 khardar 等^[21]开发的 x-vector 系统通过时延神经网络(TDNN)实现了更鲁棒的嵌入提取。基于 Transformer 的架构(如 WavLM^[22])通过预训练语音表征大幅提升了表征精度,同时,对比学习框架(如 ECAPA-TDNN)则优化了嵌入判别性^[19]。Fujita 等^[23]结合自注意力机制实现了端到端的说话人识别。该研究还探索了目标说话人语音提取(TSE)与聚类联合优化等新范式,这些方法正在推动该领域向更少监督、更强泛化的方向发展^[24]。2023 年,阿里巴巴团队提出的 CAM++ 模型^[25]在降低参数量、加快推理速度的同时提升了模型的准确率,对声纹识别的发展起到了推动作用。

2 师生课堂行为分析模型

在课堂教学评估领域,传统的课堂分析方法主要依赖“现场观察”和“人工记录”,这在实际应用中存在人力消耗大、主观误差高等问题。因此,可以借助人工智能技术,运用师生说话人确认技术,对课堂行为进行自动标注,从而降低人工成本,同时提高分析的效率和可靠性。

2.1 问题引入

声纹技术如今已相当成熟,但当前将声纹技术应用于课堂教学场景的实证研究尚处于探索阶段,相关领域面临如下挑战:①受限于课堂场景数据采集的隐私性和复杂性,现有公开可用的教学场景语音数据集规模有限,导致监督学习方法在课堂环境中的泛化性能受限;②通用说话人确认模型在具有混响效应、多人交互对话、非平稳背景噪声等特征的低质课堂声学环境中表现显著下降。针对上述问题,本文提出基于跨领域迁移学习的说话人确认模型,在降低训练成本的同时提高对课堂任务的适配性。

2.2 模型设计

本文中,教学分析的核心环节(包括行为分类编码、指标计算和课堂模式判断)均借鉴了 S-T 分析法。针对传统 S-T 分析法中依赖固定间隔的人工采样问题,本文设计了模型,对采集到的课堂音频数据进行智能化采样和分类。对于课堂交互分析任务,仅需区分教师行为、学生行为和静默片段。其中,可以利用说话人确认模型对教师行为进行区分,利用语音活动检测技术(VAD)对课堂静默片段进行区分。音频采样和分类的具体流程如图 1 所示。

2.2.1 音频采样与切分

根据 S-T 分析理论,对课堂音频进行固定间隔(30 s)采样,每个采样处切分 1.5 s 音频以进行后续的行为分类编

码工作,如图2所示。

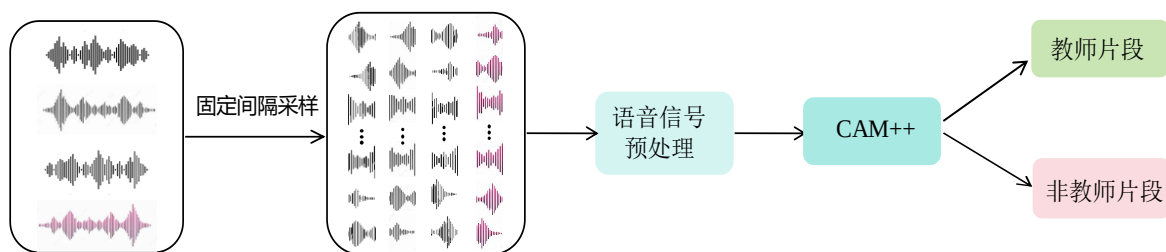


Fig. 1 Demonstration of audio sampling and classification process

图 1 音频采样和分类流程示意图

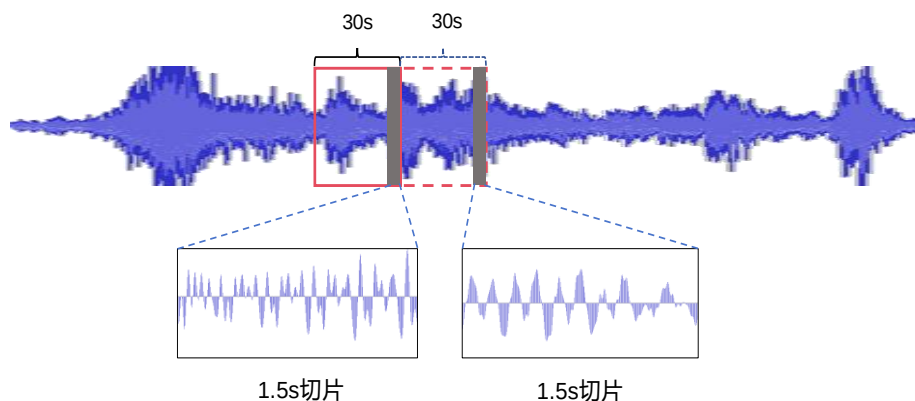


Fig. 2 Demonstration of audio sampling

图 2 音频采样示意图

2.2.2 语音信号预处理

在语音信号输入模型之前,需要进行语音信号预处理,以去除非语音部分并增强语音信号特征。语音信号预处理主要通过语音活动检测技术(Voice Activity Detection, VAD)来保留有效的语音部分,从而实现分离静默片段、减少数据量并提高语音数据质量的效果。同时,通过MFCC特征提取,可以增强语音特征的表征能力。MFCC整体流程如图3所示。

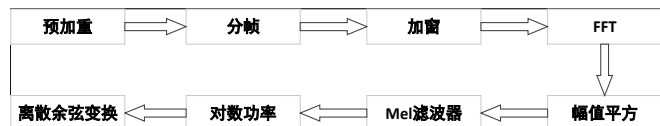


Fig. 3 Flowchart of MFCC

图 3 MFCC整体流程图

2.2.3 说话人确认模型

本文提出一种基于跨领域迁移学习的说话人确认优化框架,通过两个阶段训练策略实现模型的有效适配。首先,在CN-Celeb大规模通用中文说话人数据集上对轻量级CAM++^[25]神经网络进行预训练,以学习基础的声纹特征表示;随后,在该网络基础上新增两个线性层用于微调。采用分层微调策略,在自建的课堂数据集上对网络的高层参数进行调整。该方法创新性地将说话人确认模型与迁移学习相结合,通过参数效率优化缓解数据稀缺问题。

2.2.4 实验部分

(1)实验数据集。本文采用的真实课堂教学录音源自

各大视频平台,从中筛选出音质较佳、课堂场景真实自然的实录内容。所选数据涵盖了多种授课风格,并包含语文、数学、心理等不同课程类型,涉及小学、初中及高中多个学段。该数据集的多样性有助于提升模型的泛化能力。

本文通过录播平台对50节公开课进行了调研。调研发现,大多数课堂具有较高的互动性,尤其是传统三门主要学科——语文、数学和英语。然而,小部分课堂由于教师采用了较为传统的教学风格,互动成分较少甚至不存在,这些课堂未被纳入我们的主要研究范围。由于标注精力有限,难以对所有课堂进行深入分析,我们仅选取了其中具有代表性的20节课堂实录作为数据集。数据集中涵盖了小学、初中和高中3个阶段的不同科目(不包括英语课堂),并按照编号进行分类:编号为10-20的课堂用于预训练模型的微调;编号为1-10的课堂则作为测试集,用于验证模型的有效性。

(2)评估指标。等错误率(EER)为误拒绝率(FRR)和误接受率(FAR)相等时的错误率,如公式(1)所示:

$$EER = FRR = FAR \quad (1)$$

其中,误拒绝率(FRR)表示实际为正的样本中,被错误预测为负的比例;误接受率(FAR)表示实际为负的样本中,被错误预测为正的样本。具体如公式(2)-(3)所示:

$$FRR = \frac{FN}{TP + FN} \quad (2)$$

$$FAR = \frac{FP}{TN + FP} \quad (3)$$

(3)对比实验。将教师音频设定为正样本,学生音频设定为负样本。分别使用预训练的CAM++模型和微调后的CAM++模型对10节课堂进行教师说话人确认实验,结果如表1所示。

Table 1 Experiments results

表1 实验结果

模型	Accuracy/%	EER/%	FAR/%	FRR/%	正样本 数量	负样本 数量
预训练 CAM++	75.4	24.5	28.6	19.2	498	333
微调后 CAM++	91.8	9.3	10.5	8.1	498	333

由表1可知,经过微调后,CAM++模型的准确率达到91.8%,相比预训练模型提升了16.4个百分点;等错误率(EER)为9.3%,较预训练模型降低了15.2个百分点;误接受率(FAR)与误拒绝率(FRR)也分别较预训练模型下降了18.1和11.1个百分点。实验表明,通过分层微调的迁移学习策略,模型在课堂低质声学环境中的鲁棒性显著提升,可有效适配师生说话人确认任务。

3 基于师生说话人确认的智慧课堂交互分析

将标注后的语音片段映射为S-T行为编码,经过统计分析计算出课堂交互分析各指标,最终生成多维可视化分析报告。研究表明,该系统不仅能准确识别讲授型、对话型等4类基本教学模式,还可量化呈现师生互动频率、话语权分布等深度教学特征。

结合教育实践主流模式,本文聚焦以下4类预设课堂类型,其核心特征如下:

(1)练习型课堂。以学生自主练习与反馈为核心,学生静默练习时长占比高,教师语音以答疑为主。

(2)讲授型课堂。以教师知识传授为核心,教师语音占比高、行为转换少。

(3)对话型课堂。以师生高频互动为核心,学生主动表达占比高、行为转换频繁。

(4)混合型课堂。以适当的教师学生活动和师生互动为核心,教师知识传授、学生主动表达和师生互动三者均保持适中与均衡。

3.1 行为分类编码与指标计算

3.1.1 行为分类编码

S-T分析的核心在于建立一个教师行为(T行为)与学生行为(S行为)的二元分类框架。教师行为的主要形式包括讲授、提问、点评和反馈等,而学生行为则主要表现为回答、讨论和提问等。此外,静音片段中的学生行为也被纳入研究范围,如思考、书写和实验等。

3.1.2 指标计算

S-T分析的主要指标包含:教师行为占有率(Rt)和行为转换率(Ch),如公式(4)–(5)所示:

$$R_t = \frac{N_T}{N_T + N_S} \quad (4)$$

$$Ch = \frac{N_C}{N_T + N_S} \quad (5)$$

其中, N_T 代表课堂中教师行为个数, N_S 代表课堂中学生行为个数, N_C 表示行为转变次数。

3.2 教学模式分析

3.2.1 教学模式结构分布图

教学模式结构分布图是基于S-T分析法构建的可视化工具,旨在通过量化师生行为特征对课堂教学模式进行分类学划分。教学模式分布图可根据教学模式分类表提供的标准构建,教学模式分布表具体标准如表2所示。

Table 2 Categories of teaching models

表2 教学模式分类表

教学模式	标准条件
练习型	$R_t \leq 0.3$
讲授型	$R_t \geq 0.7$
对话型	$Ch \geq 0.4$
混合型	$0.3 < R_t < 0.7, Ch < 0.4$

教学模式分布图以教师行为占有率(R_t)为横轴,师生行为转换率(Ch)为纵轴,通过预设阈值($R_t = 0.7, Ch = 0.4$)将坐标系划分为4个象限,分别对应讲授型(高 R_t)、练习型(低 R_t)、对话型(高 Ch)与混合型(中 R_t 、高 Ch)四类教学模式,如图4所示。通过将课堂观测数据映射至图中,可直观识别教学结构的倾向性。例如,分布于“讲授型”区域表明教师主导性强,而分布于“混合型”则反映师生互动均衡。

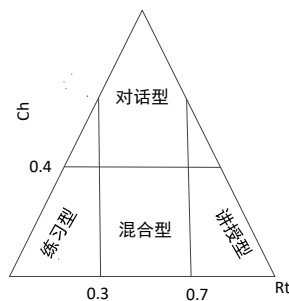


Fig. 4 Structure of the teaching model

图4 教学模式结构分布图

3.2.2 各课堂 R_t 、 Ch 指标计算及教学模式分析

基于前文的模型分析可以分别得到课堂中的教师行为时长和学生行为时长,代入指标计算公式(4)和(5),得出10节课堂 R_t 、 Ch 计算结果如表3所示。

Table 3 The values of R_t – Ch for each class表3 各课堂 R_t – Ch 计算结果表

课堂编号	1	2	3	4	5	6	7	8	9	10
R_t	0.52	0.27	0.64	0.34	0.49	0.37	0.57	0.69	0.71	0.56
Ch	0.60	0.33	0.42	0.28	0.61	0.33	0.42	0.36	0.32	0.53

根据 R_t 、 Ch 值在课堂模式分布图中标注出坐标位置,如图5所示。10节课堂的 R_t 、 Ch 坐标标注中,对话型课堂占据了最高比例,具体包括课堂1、3、5、7和10。其次,混合型课堂包括课堂4、6和8。在练习型和讲授型课堂中,

各有一节相应的课堂,分别是课堂2和课堂9。

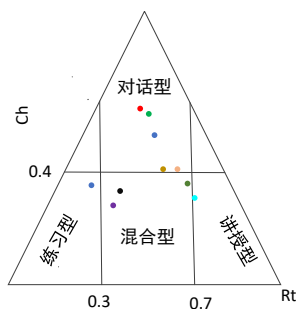


Fig. 5 Coordinates of Rt-Ch for each class

图5 各课堂 Rt-Ch 坐标图

4 可视化分析

根据实验结果以课堂2为例对课堂进行可视化分析,具体包括课堂整体分析、师生行为时轴图分析和S-T曲线分析,并从各可视化维度对课堂2做出全面整体的课堂交互分析。

4.1 课堂整体分析

课堂整体分析主要包括教学模式结构分布图分析和课堂说话人饼状图分析。其中,饼状图分析较为简单,根据课堂2实验结果,可以直观地表示出师生说话人以及静默片段占比图,如图6所示。

教学模式结构图分析需要结合公式(4)和(5),计算出课堂2的教师行为占有率 R_t 和行为转化率 Ch ,并在教学模式结构图中标出 R_t 、 Ch 对应的坐标点,从而直观地得出教学模式。通过计算可知,课堂2的教师行为占比为 26.8%,学生语言占比为 48.8%。由于学生行为包括学生语言和静默片段,因此课堂2的学生行为总占比为 73.2%,行为转化次数为 27 次。代入公式可得 $R_t = 0.27$, $Ch =$

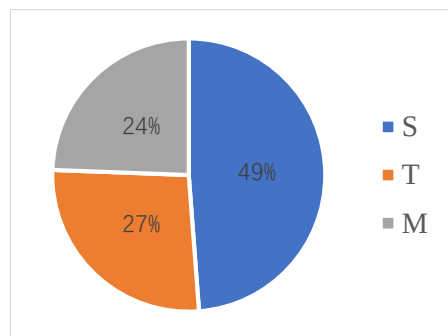


Fig. 6 Pie chart of teacher-student behavior proportion for class 2

图6 课堂2师生行为占比饼状图

0.33,最终分类为练习型(如图7所示),与实际课堂类型相符。

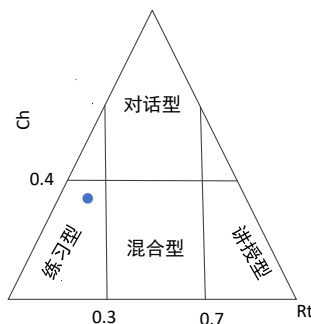


Fig. 7 Coordinates of Rt-Ch for class 2

图7 课堂2 Rt-Ch 坐标标注图

4.2 师生行为时轴图分析

师生行为时轴图可以直观展现课堂中教师和学生话语随课堂时间的分布情况。根据实验结果,按照片段序号将分类后的片段进行拼接,即可得到完整的说话人时间分布图,如图8所示。其中,不同的颜色分别用于表示教师话语和学生话语。

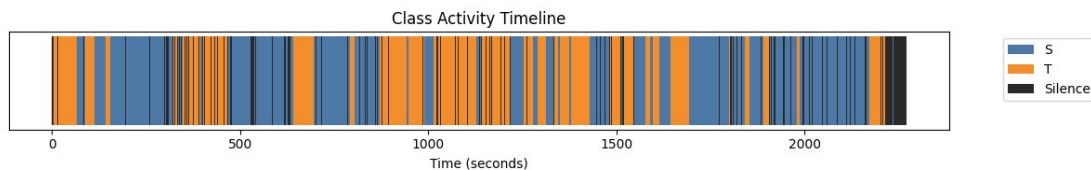


Fig. 8 Time-axis chart of teacher-student behavior

图8 课堂2师生行为时轴图

由图8可知,时轴图呈现出高度动态性和互动性,教师话语片段与学生话语片段交替频繁。学生话语占据了时轴的大部分,课堂2的时轴图中显著展示了学生的主导作用,许多“话语区块”由学生发起或主导,生生互动在某些时段成为时轴的主线。总体来看,话语权在教师与学生之间,以及学生与学生之间的流转更加频繁,课堂节奏更具弹性和互动性,沉默区间可能用于学生的思考或小组协作。

4.3 S-T 曲线分析

S-T曲线能够直观地展示教学过程中师生行为的时间

变化以及师生互动情况。S-T曲线图中,横轴表示学生话语的时间积累(表现为水平直线,直线的长度代表学生话语的时间长度),纵轴表示教师话语的时间积累(表现为垂直直线,直线的长度代表教师话语的时间长度),45°斜线代表静默时间段,如图9所示。

由图9可知,课堂2的S-T曲线展现出“频繁转折、多向延展”的动态变化。曲线由频繁交替的较短水平线段(教师引导/反馈)和显著延长的垂直线段(学生发言/讨论)共同构成。当学生进行观点陈述、小组汇报或辩论(学生主导)时,曲线呈现显著的垂直上升趋势(纵坐标快速增

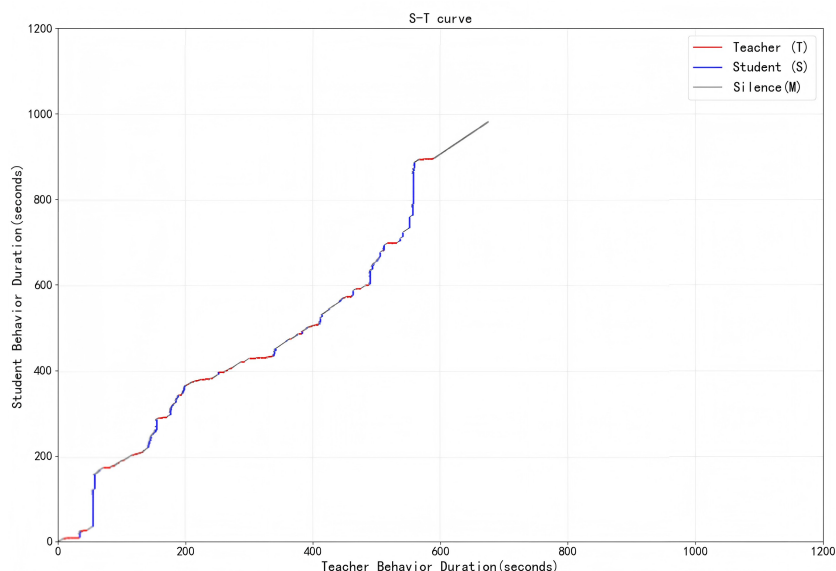


Fig. 9 S-T curve for class 2

图9 课堂2 S-T 曲线

加),形成多个明显的“上升台阶”;而教师介入讲解或反馈时,曲线则转为水平向右延伸(横坐标增加)。这种水平与垂直方向的频繁切换,直观反映了话语权在师生间的动态流转与共享。45°斜线(静默)出现频率较高,对应习题、思考或行为转换等非言语时段。整体曲线远离坐标轴,在图中部区域曲折行进,波动幅度大,呈现了高频率的师生及生生互动模式,学生话语对整体时间积累贡献显著。

5 结语

本文聚焦于课堂交互分析的智能化转型,旨在解决传统课堂交互分析中信息化程度不足的问题(如标注成本高、主观性较强等)。特别是在教育信息化需求的背景下,本文提出了一种将人工智能技术应用于课堂交互分析的方法。通过利用人工智能技术中的说话人确认技术和迁移学习思想,并结合传统的S-T课堂互动分析理论,成功地应用了说话人确认技术对课堂进行交互分析。这一方法在很大程度上降低了人工成本,并在一定程度上解决了课堂相关研究领域中的数据集匮乏问题。

在未来工作中,将进一步丰富课堂数据集的多样性,不再局限于中文课堂,还将引入英文、日语等多种语言的课堂数据。同时,将着力扩大样本规模,以避免当前小数据量引起的模型欠拟合问题。另一方面,将探索更多维度的智能化课堂分析方法。例如,基于计算机视觉的学生行为识别(如专注度、参与度分析)、基于语音情感识别的师生互动状态分析,以及利用知识图谱对课堂对话内容进行自动结构化,从而构建一个从行为、情感到认知的多模态课堂智能分析框架。

参考文献:

[1] CHEN T Z, TIAN H Q. The core mission of the new journey to china's ed-

ucation modernization and its major challenges[J]. Journal of Higher Education, 2023, 44(3): 26-32.

陈廷柱, 田华泉. 中国教育现代化新征程的核心使命和主要挑战[J]. 高等教育研究, 2023, 44(3): 26-32.

[2] LI Y X, CHEN A X, JIA J Y. Annual report on the frontiers and hotspots of China's educational technology research in 2024[J]. China Educational Technology, 2025(5): 61-67.

李雅琨, 陈昂轩, 贾积有. 2024中国教育技术研究前沿与热点年度报告[J]. 中国电化教育, 2025(5): 61-67.

[3] REN Y Q. Stepping into the new age of Chinese education informatization: interpretation of education informatization 2.0 action plan(1) [J]. e-Education Research, 2018, 39(6): 27-28, 60.

任友群. 走进新时代的中国教育信息化——《教育信息化2.0行动计划》解读之一[J]. 电化教育研究, 2018, 39(6): 27-28, 60.

[4] CAI H Y, XHENG Q H, HU T H, et al. Performance assessment based on information technologies: concept analysis, constituting elements, and classifying framework[J]. China Educational Technology, 2024(2): 1-7.

柴唤友, 郑勤华, 胡天慧, 等. 基于信息技术的表现性评价: 概念解析、构成要素及分类框架[J]. 中国电化教育, 2024(2): 1-7.

[5] HUANG W. On operational mechanism of classroom dialogue—based on the perspective of discourse analysis[J]. Educational Research, 2014, 35(7): 123-130.

黄伟. 课堂对话的运作机理——基于话语分析的视角[J]. 教育研究, 2014, 35(7): 123-130.

[6] MOLINARI L, MAMELI C, GNISCI A. A sequential analysis of classroom discourse in Italian primary schools: the many faces of the IRF pattern [J]. British Journal of Educational Psychology, 2013, 83(3): 414-430.

[7] CHENG Y, LIU Q T, WANG F, et al. Research on the application of an improved, video-based s-t analysis method [J]. e-Education Research, 2016, 37(6): 90-96.

程云, 刘清堂, 王锋, 等. 基于视频的改进型S-T分析法的应用研究[J]. 电化教育研究, 2016, 37(6): 90-96.

[8] XIONG X P, YANG H, REN S Z. Design of flanders interaction analysis system based on event recording—a case study of classroom observation in high school biology "photosynthesis"[J]. Chinese Journal of ICT in Educa-

- tion, 2021(2): 24-29.
- 熊旭萍, 杨昊, 任山章. 基于事件记录的弗兰德斯互动分析系统设计——以高中生物《光合作用》课堂观察为例[J]. 中国教育信息化, 2021(2): 24-29.
- [9] ZHANG Z, CHENG S Y, WANG T R. The framework and technical implementation of artificial intelligence classroom analysis based on computational education[J]. Open Educational Research, 2024, 30(5): 87-100.
- 张治, 程抒一, 王天蓉. 基于计算教育学的人工智能课堂分析框架和技术实现[J]. 开放教育研究, 2024, 30(5): 87-100.
- [10] TANG Q W, ZHANG H, WU Y A. Modeling a multimodal learning analysis in the context of smart classrooms[J]. Journal of Teacher Education, 2024, 11(5): 49-58.
- 汤倩雯, 张浩, 吴意安. 智慧课堂多模态学习分析模型构建[J]. 教师教育学报, 2024, 11(5): 49-58.
- [11] WU J Q, WU F Y, ZHANG M M, et al. Construction and application of collaborative learning engagement analysis model for smart classroom from a multimodal perspective [J]. e-Education Research, 2022, 43(7): 73-80, 88.
- 吴军其, 吴飞燕, 张萌萌, 等. 多模态视域下智慧课堂协作学习投入度分析模型构建及应用[J]. 电化教育研究, 2022, 43(7): 73-80, 88.
- [12] ZHANG H, LI F X, MA X F. Comparative study of classroom interaction analysis tools for "excellence courses"[J]. Chinese Journal of ICT in Education, 2020, (2): 79-82.
- 张慧, 李凤霞, 马秀芳. “优课”课堂互动分析工具的对比研究[J]. 中国教育信息化, 2020, (2): 79-82.
- [13] 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.
- 骆祖莹, 赵琦琦, 段福庆. 基于教师近场语音的课堂教学过程自动分析[J]. 现代教育技术, 2021, 31(8): 76-84.
- [14] DAVIS S B, MERMELSTEIN P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980, 28(4): 357-366.
- [15] REYNOLDS D A, QUATIERI T F, DUNN R B. Speaker verification using adapted Gaussian mixture models [J]. Digital Signal Processing, 2000, 10(1-3): 19-41.
- [16] DEHAK N, KENNY P J, DEHAK R, et al. Front-end factor analysis for speaker verification [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 19(4): 788-798.
- [17] SNYDER D, GARCIA-ROMERO D, POVEY D, et al. Deep neural network embeddings for text-independent speaker verification [C]//Interspeech, 2017: 999-1003.
- [18] SNYDER D, GARCIA-ROMERO D, SELL G, et al. X-vectors: robust DNN embeddings for speaker recognition [C]//2018 IEEE International Conference On Acoustics, Speech and Signal Processing (ICASSP), 2018: 5329-5333.
- [19] DESPLANQUES B, THIENPOND T, DEMUYNCK K. ECAPA-TDNN: emphasized channel attention, propagation and aggregation in TDNN based speaker verification [C]//Interspeech, 2020: 3830-3834.
- [20] GARCIA-ROMERO D, SNYDER D, SELL G, et al. Speaker diarization using deep neural network embeddings [C]//2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017: 4930-4934.
- [21] KHADAR K V A, SUNIL KUMAR R K, SAMEER V V. Speaker diarization based on X vector extracted from time-delay neural networks (TDNN) using agglomerative hierarchical clustering in noisy environment [J]. International Journal of Speech Technology, 2024, 27(3): 567-578.
- [22] CHEN S, WANG C, CHEN Z, et al. Wavlm: large-scale self-supervised pre-training for full stack speech processing [J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(6): 1505-1518.
- [23] FUJITA Y, KANDA N, HORIGUCHI S, et al. End-to-end neural speaker diarization with self-attention [C]//2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2019: 296-303.
- [24] MEDENNIKOV I, KORENEVSKY M, PRISYACH T, et al. Target-speaker voice activity detection: a novel approach for multi-speaker diarization in a dinner party scenario [C]//Proceedings of the Interspeech, 2020: 274-278.
- [25] WANG H, ZHENG S, CHEN Y, et al. CAM++: a fast and efficient network for speaker verification using context-aware masking [DB/OL]. <https://arxiv.org/abs/2303.00332>.

(责任编辑:陈维捷)