

人工智能驱动的教学分析：课堂活动的知情洞察

Oscar Canovas 计算机工程系
西班牙穆尔西亚大学 ocanovas@um.es

Felix J. Garcia-Clemente 计算机工程系
西班牙穆尔西亚大学 fgarcia@um.es

Federico Pardo 计算机工程系
西班牙穆尔西亚大学 federico.pardog@um.es

摘要——教师需要可靠的反馈来提升教学技能并分析课堂方法。然而，教师在进行授课时，往往难以同时掌握课堂活动的特征描述与分布规律。传统通过人工教练指导下的刻意练习来积累经验的方法已不再适用，因为培训教练、观察课堂和记录活动所需的时间周期过长。为应对这些挑战，研究者提出了运用人工智能技术分析音频记录的自动化方法，用于推断课堂氛围、建模教师话语并分类教学活动。尽管这些方法在推断教学活动方面已取得显著成果，但仍需更多支持工具和分析手段，帮助教师反思教学实践并追踪进步。本文提出了一种创新框架：通过深度学习实现说话人日记化处理，并采用机器学习算法对教学实践进行分类及不同教学风格分析。该方法整合多项非语言话语特征以提供信息性洞见，具体而言，我们定义了12项特征用于区分三种不同教学模式——讲授式、小组讨论及使用互动系统。研究表明，这些特征同样能有效分析教师在各类教学实践中的行为表现。

索引术语-音频分析、教学实践、说话人日志分析、教学分析、深度学习、机器学习

I. 介绍

及时反馈对教师成为优秀教育者至关重要。他们需要参与持续学习，反思自己的教学实践，并相应地调整教学方法。在教练指导下进行的刻意练习并伴随建设性反馈，往往既耗时又不切实际[1]。

认识到反馈对教师成长的重要性，关键在于帮助教师清晰深入地了解自身表现与进步。我们提出了一种旨在解决即时客观反馈缺失问题的方法。该方法着重为教师提供及时、准确且客观的反馈，以支持其专业发展并提升教学实践水平。现有研究表明，当教师获得师生对话比例的自动化反馈时，学生发言量会显著增加[2]。这表明即使教师的基本信息

课堂话语模式可以导致积极的变化朝着期望的方向发展。

本文提出的方案主要基于从日志化过程[3]中提取的特征。语音分段技术能够将音频录音分割为不同说话人的独立片段，从而便于深入分析诸如发言时长或言语重叠等非语言线索。通过提取12项非语言特征，我们得以深入了解教师话语的动态变化。这些非语言信号为课堂互动与动态提供了丰富的信息来源。例如，发言时长可以反映师生之间的对话分配比例，进而推导出更精准的参与度评估指标。言语重叠现象则能揭示言语中断或重复的情况，这直接反映了学生的参与程度及对话的活跃程度。

我们的目标之一是探索如何利用特征数据创建并验证一个模型，该模型能够复现人类编码员对课堂活动的识别。我们特别研究开发一种分类模型的可能性，该模型可区分三种类型的课堂活动：讲课、学生小组合作以及使用观众反馈系统。

鉴于不同课堂活动中存在多种教学风格，我们将通过分析提取的特征是否能有效区分个体教师或采用相似教学方法的教师群体。通过研究这些特征所捕捉到的独特模式与行为特征，我们旨在深入理解教育工作者在具体课堂活动中所采用的教学方法差异。

因此，本文提出以下研究问题：

- RQ1：从日记化过程中得出的特征能有效地识别不同的教学实践吗？
- RQ2：在特定的教学实践中，从日记化过程中得出的特征是否适合于区分不同教师的教学风格？

正如我们将要展示的，我们的主要贡献不仅是为教师提供一个人工智能驱动的机制来识别他们正在进行的课堂活动，而且还使他们能够分析他们的话语模式，并进行比较。

与其他教师的数据进行模式分析。这些信息为教师及时反思自己的教学实践提供了宝贵的见解。

二、相关工作

近年来，针对课堂氛围[4]和不同情境下话语分析的研究层出不穷。我们的研究重点在于教师话语的自动化分析，特别是通过录音资料进行深度挖掘。虽然用录音评估教师教学效果的做法并不新鲜，但自动分析这类录音的实践直到近年才真正兴起[5]。当前多数研究方案采用机器学习或深度学习技术来解析教学方法与风格差异，具体包括提取非语言特征或运用自然语言处理技术。本文提出的方案主要基于日志化过程提取的特征——即对发言者及其发言实例进行标注。这项任务可采用多种方法实现，从传统方式到先进的神经网络模型[3]都适用。

在课堂教学话语研究领域，多个科研团队致力于开发和验证自动化系统，旨在识别基础性话语结构（如讲座与小组活动）。例如，唐纳利等人[6]通过训练监督式机器学习模型对教学环节进行分类，取得了0.64至0.78的F1分数。此外，[7]研究证明了运用自动语音识别与分类模型实现课堂语音自动分割并识别教师提问实例的可行性。基于自动语音识别的其他研究则聚焦于教师与学生课堂对话的分段处理[8]，以及利用低声声学特征进行分析。多数方案特别关注教师角色[9]，以此来划分主动学习任务[10][12]。

然而，值得注意的是，这些工作主要集中在实现高分类准确率上，经常忽视提供可以作为描述性和信息性数据的话语特征[2][11]。要深入分析不同教学风格并为教师提供有价值的建议，不仅需要精准的分类标准，更需全面把握其背后的话语动态。信息性数据对于捕捉教学实践的细微差别、提供深度洞见至关重要。本文的研究首次运用信息特征，不仅实现了教学实践的分类，还对每种教学方法内部的教学风格进行了系统分析。

正如我们将在第四部分C节中介绍的，我们借鉴了前人研究[13][14]中为群体会议分析设计的部分话语特征。我们发现，某些轮替特征、参与率等要素，

这些作品中呈现的与沉默相关的特征也适用于教学实践的分析。

目前，有一些商业产品可用，例如初创公司TeacherFX¹，它主要专注于量化教师与学生的谈话比例，并提供与参与和互动相关的图表。然而，他们的方法在分析的特征数量方面是有限的，这阻碍了我们希望通过我们的研究实现的详细和全面的分析。

除了非语言分析方法外，传统的自动教师话语分析方法主要依赖自动语音识别（ASR）转录文本[15]。这种广泛应用的方法通过提取涵盖词汇、句子和语篇层面的高级语言特征，将其作为监督分类器的输入，从而以通用化方式检测目标语篇特征。训练好的分类器可作为后续微调的初始参考点[16]，在此过程中通过添加任务专用输出层，并利用有限的领域特定数据更新参数进行微调。值得注意的是，在某些研究中[17]，深度学习技术已被应用于数学课堂中特定对话策略的识别。虽然我们现在正在研究一个基于非语言特征和NLP（自然语言处理）技术的课堂分析混合系统，但后一种方法不在本文的讨论范围内。

三、方法

A. 数据集

本研究对计算机科学本科课程中的《计算机网络》和《计算基础》两门课程的音频资料进行了系统分析。为确保数据集的全面性，我们邀请了1名女教师和3名男教师分别录制授课过程，最终形成38份音频文件——每份对应一位教师所教授的独立课程。参与样本的教师平均从业年限达20.5年，这些音频文件累计时长约24小时，单个文件时长介于1至2小时之间。所有数据采集均获得教师、学生及机构伦理审查委员会的正式授权。

此外，每个音频会话都包含补充元数据。这些元数据包含的信息包括：音频文件对应的课程名称、录音日期、音频时长、录音时在场的学生人数、使用的录音设备、课堂采用的教学方法，以及授课教师的身份信息。

对录音的各个部分有四个标签：

- 讲座。这代表一个传统的课堂，老师讲解课程材料，学生

¹<https://teacherfx.com>

听老师讲课，做笔记，偶尔提问。

- Wooclap。当教室中使用观众响应系统[18]向学生提问并通过移动设备与他们互动时，会使用此标签。Wooclap²是所有情况下使用的平台。
- 小组工作。这表示学生以小组形式解决问题的课程。
- 其他。当音频片段与任何先前的描述不匹配时，会使用此标签，例如教师准备开始上课或在课间休息时。对于本工作，带有此标签的音频片段会被省略。

B. 人工编码

如前所述，我们已存储了各课堂所采用教学方法的相关元数据，但这些信息缺乏时间标记。因此，我们需要建立时间戳系统来将教学实践与对应音频片段关联起来。采用的标注方法包含三个要素：音频时间戳用于标识标签起始位置，另一个时间戳用于标记已标注片段的结束位置，以及对应的授课方法。本人工标注流程由一名独立编码员使用ELAN软件³完成。生成的数据作为训练各类机器学习模型的基准数据，这些模型用于测试教学实践的自动化分类。据编码员介绍，数据集包含13小时的讲座、3小时的Wooclap（在线课堂）和8小时的工作小组活动。

C. 实验设置

在一次线下培训中，教师们接受了课堂录制的专项指导。作为流程的一部分，他们首先在各自教室进行试录，既为测试做准备，也为了确保数字录音设备（TASCAM DR-07X）的最佳摆放位置。每位教师需录制至少十节涵盖不同教学方法的课程。我们从数据集中精选出音质优良的录音样本进行深入分析。

这些课程在传统的讲堂里进行，可容纳60-90名学生。教室配有桌椅，面向教学区，其中包括一张完全嵌入技术的桌子。为了确保录音设置的无缝性，一个手持式数码录音机被悄悄地放在老师的桌子上，距离老师和前排的学生至少1.5米远。

四、建议办法

我们的方法包含一个五阶段的流程，如图1所示。每个阶段都会为后续阶段提供输入，从而获取更有价值的信息。数据可视化

²<https://wooclap.com>

³<https://archive.mpi.nl/tla/elan>

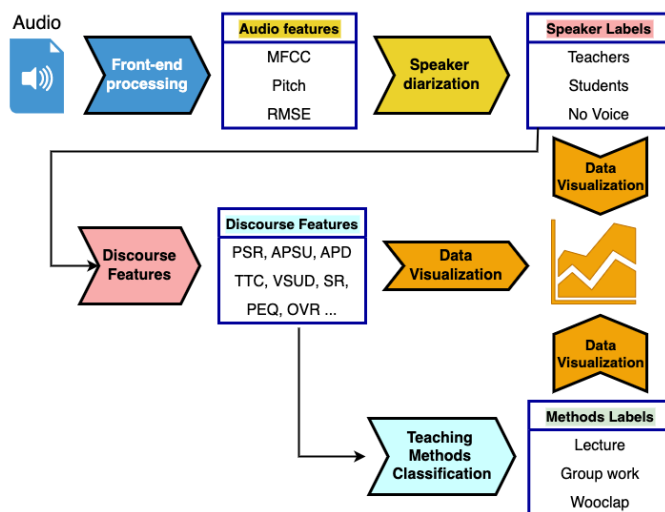


图1：处理阶段和数据元素的流程

stage专门为教师提供视觉信息，以进行可操作的教学分析。

A. 前端处理

音频处理的初始阶段需要从录音中提取低阶特征。关键特征包括梅尔频率倒谱系数（MFCC）[19]、基频以及每个有声帧的平均能量值。这些特征可通过Torchaudio⁴或Librosa⁵等专业工具获取。在学术研究领域，这类特征常被用作说话人对话语法技术的输入参数。

B. 扬声器日志

说话人识别是将音频录音分割成多个片段并标注说话人身份，从而确定“谁在何时发言”的过程。该系统包含语音活动检测（VAD）模型，用于识别音频中存在语音的时间段并过滤背景噪音；同时配备说话人特征提取模型，从VAD识别的语音片段中提取独特的说话人音频特征。在我们的方法中，虽然能够区分教师的讲话，但无法区分不同学生，将其他所有说话者统一标记为“学生”。要实现精准的说话人识别，可采用Pyannote-audio⁶等工具，这也是我们在实验中选用的解决方案。

C. 特征计算

参考前人研究[13][14]，我们从日常对话中提取出若干非语言特征，这些特征对话语特征描述具有重要价值。具体而言，针对教师与学生两种角色，我们分别定义了四个话语特征：

⁴<https://pytorch.org/audio/stable/transforms.html>

⁵<https://librosa.org/doc/main/feature.html>

⁶<https://github.com/pyannote/pyannote-audio>



图2：不同教学实践的时间线：（上）讲座；（中）工作小组；（下）Wooclap

- 参与说话比率（PSR）。指录音过程中每个角色的参与比例。
- 参与者发言语句（PSU）。当前录音片段中的语句数量。
- 参与者发言话语比率（PSUR）。指录音片段中每个角色的发言话语比例。
- 平均参与者发言时长（APSUD）。每个角色发言的平均时长。

此外，我们还定义了8个全局语篇特征：

- 平均延迟持续时间（ALD）。静默期的平均持续时间。
- 静默率（SR）。指记录节段中静默时间的周期比值。
- 平均暂停持续时间（APD）。同一参与者在两次发言之间保持沉默的平均时长。
- 参与度平等指数（PEQ）。该指标用于评估不同角色间的参与平衡情况，其计算方法参照文献[14]所述。数值越接近1，表明参与度分配越均衡。
- 投票计数（TTC）。学生和老师之间的对话中发生的转折次数。
- 极短语音比率（VSUR）。指总语音中极短语音（少于2秒）的占比。
- 重叠率（OVR）。不同参与者之间重复话语的比率。
- 重叠话语率（OVUR）。指话语重叠的比率。

D. 教学方法分类

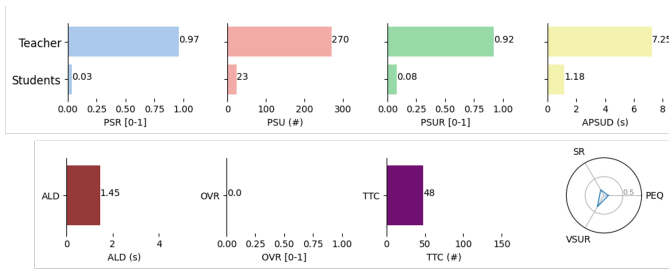
我们开展了一项包含三个类别的多分类任务：讲座、小组作业和伍克拉普。研究采用监督式机器学习算法，对输入录音进行精准标签预测。通过对比支持向量机（SVM）、k近邻（kNN）、随机森林、朴素贝叶斯、逻辑回归和梯度提升等模型，最终筛选出最适合本任务的算法方案。

由于教师可以在一节课中使用多种教学方法，我们的分类阶段旨在确定每种方法的起始和结束点。为此，分类过程涉及分析录音的小片段。我们检查了从60秒到300秒不等的重叠窗口大小。我们的初步结果表明，使用120秒或180秒的窗口可以获得粒度和精度之间的最佳平衡。更多关于实验结果的细节将在第五节中展示。

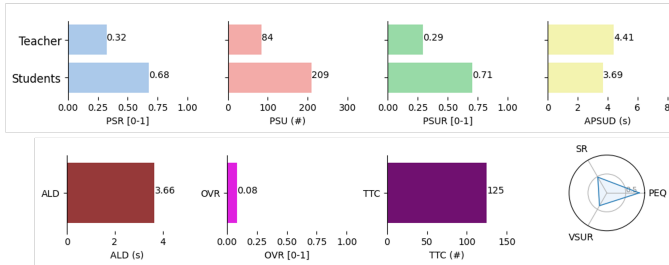
E. 数据可视化

有效的数据可视化对于为教师提供有关其教学活动的宝贵见解以及为他们提供可操作的分析以提高他们的教学实践至关重要。

我们的解决方案是为流程的每个阶段提供定制化数据可视化呈现，正如我们在[已移除以保护隐私]中展示的那样。具体而言，我们建议采用时间轴图来直观展示日志化处理的输出结果。如图2所示，时间轴可视化能清晰展现参与者的时间顺序，帮助教师理解时间关系。此外，时间轴还能有效分析不同交互环节之间的关联性。



(a) 讲座



(b) Wooclap会话

图3：不同教学实践的两个记录的特征可视化示例。

每种教学实践所展现的模式。在这里，T02指教师，SP-EAKER_0X指不同的学生。

为呈现教学特征，我们建议采用条形图和雷达图来辅助多维度特征对比。如图3所示，不同教学方法产生的某些特征值存在显著差异，这为探索这些特征在自动化评估教学实践中的应用潜力奠定了良好基础。

最后，在根据所采用的教学实践对音频片段进行标注后，我们可以通过箱线图等形式展示各教学方法中各类特征的相关信息。如第五章B节所示，我们还能对比不同教师在各项特征上的表现差异，从而识别出独特的教学风格，并为教师提供关于自身能力的宝贵参考。

V. 结果和讨论

A. (RQ1) 教学实践的确定

我们已从日志化过程中提取出12项独特的非语言特征，旨在评估这些特征在教学实践自动化分类中的适用性。为开展这项研究，我们采用无监督学习方法对数据进行探索。具体而言，我们运用了t-SNE（t分布随机邻域嵌入）等降维技术来降低数据维度，从而便于可视化分析。

通过将t-SNE算法应用于数据（包含三个主成分，且膨胀系数设为1），我们获得了如图4所示的可视化结果。观察发现，该结果验证了数据潜在的可分离性。

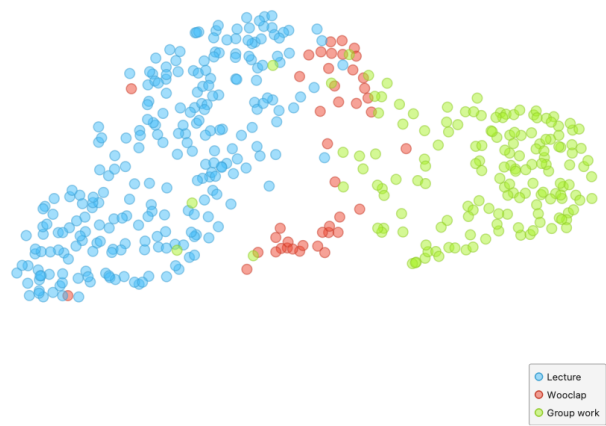


图4：t-SNE可视化显示了代表教学实践的三个不同集群。

数据中的聚类分布呈现出明显的分组特征，这为自动分类不同教学实践提供了良好前景，但对“Wooclap”标签的标注工作仍存在挑战。

在我们的研究中，我们彻底调查了几种机器学习技术，以确定它们在分类180秒音频样本的特征方面的有效性。

为确保评估的全面性，我们整合了所有参与实验的教师数据及各类教学方法。通过采用10折交叉验证法对每种技术的分类性能进行评估：具体操作是将数据集均分为十份，每次选取其中九份作为训练集，剩余一份作为验证集。整个验证过程重复十次，每次都某份数据作为验证集使用。

为了获得可靠且具有代表性的结果，我们对交叉验证过程中获得的性能指标进行了平均处理，这些指标来自三个不同的分类类别。最终计算得出的数值如表I所示。

表I：使用最佳参数的交叉验证中不同模型的AUC、F1、精确度和召回率。

模型	AUC	F1	准确	召回
k近邻	0.973	0.9288	0.9304	0.9279
SVM	0.9824	0.9373	0.9386	0.9364
随机森	0.9832	0.8945	0.8973	0.9067
幼稚贝斯取自父名	0.978	0.8743	0.9153	0.8559
对数几率回	0.9873	0.9342	0.9342	0.9343
梯Boostin	0.9801	0.9252	0.9246	0.9258

在被评估的模型中，SVM模型表现最佳，其F1分数接近94%。这表明SVM模型在基于教学实践进行精准分类时具有稳健性和有效性。

提取的特征。高F1分数代表精确性和召回率之间的平衡，突显了模型在准确识别相关实例的同时，最大限度减少误报和漏报的能力。

除了性能指标外，混淆矩阵（表II）还揭示了SVN模型的误分类模式。我们注意到哪些具体案例被错误分类，其中主要涉及与“Wooclap”方法相关的案例。在分类分析中，“Wooclap”类别的准确率达到了75.6%。通过检查误分类实例，我们发现这些错误几乎均匀分布在另外两个标签之间。这一发现与t-SNE表示法的分析结果相吻合，表明Wooclap会话的音频模式有时与讲座或小组工作存在相似性，从而导致误分类。

表II：SVM模型的混淆矩阵

实际值/预测值	讲座	沃克劳普	集团工作	Σ
讲座	97.3%	2.3%	0.4%	260
沃克劳普	11.1%	75.6%	13.3%	45
集团工作	1.8%	5.4%	92.8%	167
Σ	261	49	162	472

B. (RQ2) 差异化教学风格

在通过非语言特征验证音频片段教学方法分类的F1值显著优异后，我们接下来将重点分析其信息价值。正如我们在研究问题二中所述，我们将检验这些特征是否适用于区分特定教学实践中的不同教学风格。同时，我们还将探究这些差异对教师是否具有参考价值，并能否转化为可操作的分析方案。

由于涉及12项不同特征，若对所有特征进行详细分析将超出本文篇幅限制。因此我们聚焦于与静默相关的全局特征——“ALD”（平均中断时长），以及两项主要反映教师授课方式的特征：“APSUD教师”（指教师在音频片段中的发言平均时长）和“PSR教师”（指教师参与度比率）。这些特征将分别针对两种教学模式进行分析：传统讲授法与Wooclap教学法。之所以选择这些方法，是因为与小组工作相比，教师在其中起着更主要的作用。

首先，我们将考察这些特征的统计分布。图5展示了从数据集中生成的箱线图，代表了每位教师及其所选教学实践的所有样本。

从授课时长来看，图5b中的箱线图清晰展示了四位教师在不同音频片段中的平均发言时长（APSUD）。仔细观察可以发现，其中一位教师

该教师的发言时长分布明显高于其他教师。在分析的各个时间段中，其平均发言时长始终更长，显示出独特的说话模式，可能暗示着其教学风格与众不同。箱线图中显著的偏离值突显了考量教师个体特征的重要性，这些特征对课堂整体氛围的影响不容忽视。同一位教师的参与发言比例（PSR）也呈现出类似趋势。

为了进一步验证观察到的差异，需要进行统计分析。经确认数据分布为非正态分布，因此采用了非参数Kruskal-Wallis检验。表III提供了讲座数据分析的详细信息。

表III：四名教师在讲座中三个特征的Kruskal-Wallis检验结果

变量	χ^2	格	p	ϵ^2
醛类	81.5	3	<.001	0.315
APSUD教师	88.2	3	<.001	0.340
PSR教师	119	3	<.001	0.461

测试结果表明，教师在“APSUD教师”、“PSR教师”和“ALD”这三个变量上存在显著差异。效应量（ ϵ^2 ）显示，教师对这些变量的影响程度为中等至显著。

另一方面，从Wooclap课程数据来看，图5（d-f）中的箱线图揭示了一个有趣现象：教师群体呈现出两种相似分布的典型分组。第一组由教师1和教师4组成，在不同音频片段中，他们的平均中断时长与学员发言比例基本持平。第二组则由教师2和教师3构成，同样保持着相近的分布特征。这些发现表明，教学风格存在显著差异。因此，需要进行统计分析以确认观察到的这些分组的显著性，并确定其是否具有统计学意义。

表IV：Wooclap课程中四位教师三个特征的Kruskal-Wallis检验结果

变量	χ^2	格	p	ϵ^2
醛类	27.9	3	<.001	0.635
APSUD教师	6.53	3	.088	0.149
PSR教师	24	3	<.001	0.545

Kruskal-Wallis检验结果显示，在Wooclap课程中，教师群体在“PSR教师”和“ALD”变量上存在显著差异，而“APSUD教师”变量则未呈现显著差异。效应量（ ϵ^2 ）表明，教师对这些变量具有中等至较大的影响。

采用Dwass-Steel-CritchlowFligner检验进行两两比较，以确认教师群体中是否存在两个明显分组。如表V所示，对于“PSR教师”特征，T01与T04之间、T02与T03之间均未发现差异。

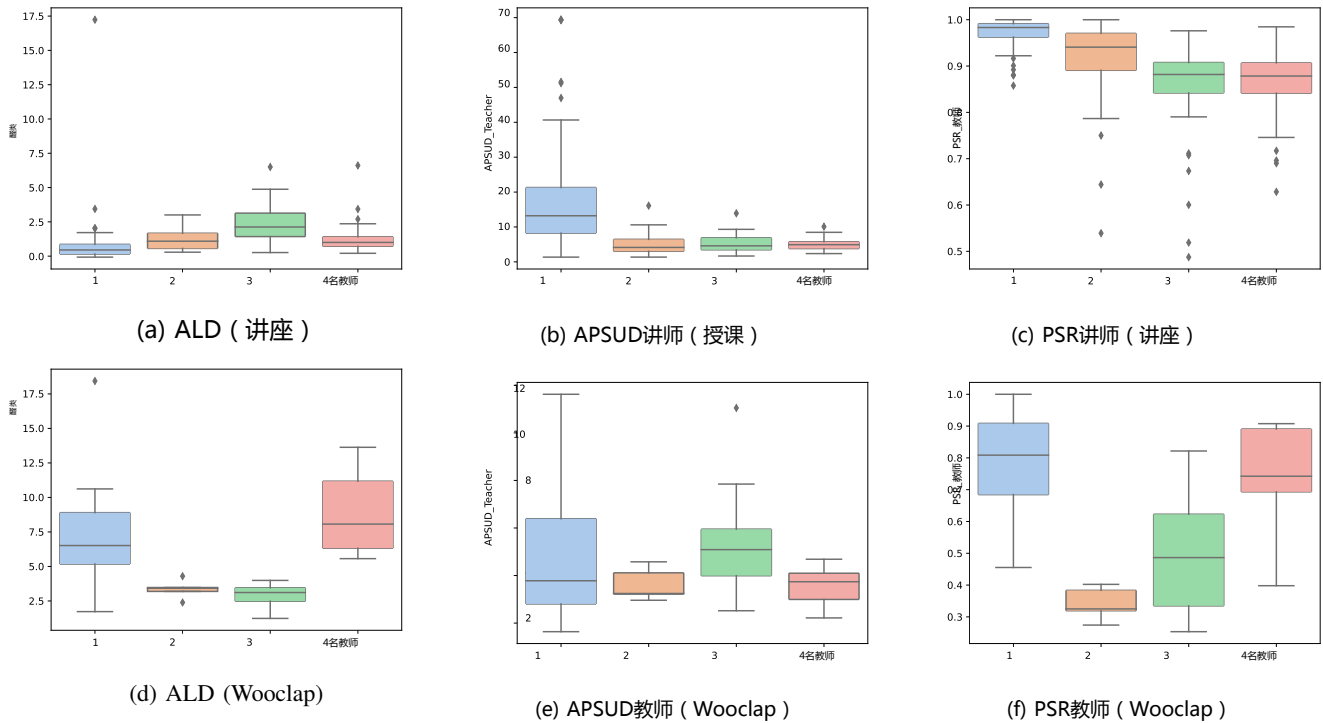


图5：参与教师在讲座（上）和Wooclap（下）环节中ALD、APSUD教师和PSR教师特征的分布情况。

表V：使用Dwass-Steel-Critchlow-Fligner检验的PSR教师成对比较

比较PSR教师	W	p
教师T01与教师T02	-4.475	0.008
教师T01与教师T03	-4.762	0.004
教师T01与教师T04	-0.616	0.972
教师T02与教师T03	1.913	0.529
教师T02与教师T04	4.391	0.010
教师T03与教师T04	5.179	0.001

C. 讨论

我们的研究对这两个问题都给出了肯定的答案。首先，我们成功证明了从日记化过程中提取的特征能有效识别不同的教学实践。此外，我们还证明了这些相同的特征适用于区分特定教学实践中不同教师的教学风格。此外，我们的提案为教师提供了每个分析阶段的宝贵的数据可视化。

我们认为，基于非语言话语特征的数据可视化能为课堂教学活动提供富有启发性的洞见。然而，教师可能需要一个参考基准或指导方案来提升教学效果。这个参考基准可以是专家定义的理想教学风格，也可以是通过整合教学风格、学习成果和学业成绩等数据集生成的AI模型。当无法获得理想教学风格时，该基准也可通过计算教学风格特征集的平均值来推导得出。

学术界。无论是新手教师还是有经验的教师，当他们对自己的教学风格感到担忧时，都需要一个可靠的参考点来指导他们的改进行动。我们提出的数据可视化机制正是为了满足这一需求而设计的宝贵工具，它能够促进决策过程。

例如，图5b中Lecture课程的APSUD特征箱形图显示，该特征的值对于教师1来说明显高于其他教师。这一观察结果表明，自动化推荐系统可能会建议教师1适当缩短发言时长。此外，这类建议可以形成集体指导方案。例如，假设教师4代表理想教学风格，根据图5中讲座环节箱线图的数据可视化分析，推荐系统可提出：除了缩短教师的发言时间外，还应增加课堂沉默时段的时长，并鼓励学生积极参与课堂讨论。

此外，基于我们精简数据集开展的统计分析证实了教学风格分组现象的重要性。这些分组具有多维度应用价值：首先，可为教师量身打造定制化AI模型，通过整合其教学风格中的个性化特征与偏好，实现智能化的个性化指导；其次，这些已识别的教学风格分组可作为后续研究的基础框架，为探索更多教学模式提供理论支撑。通过根据教师的教学风格进行分类，这使得

进行更有针对性和重点的调查，旨在更详细地探索与每种教学风格相关的具体教学方法、策略和结果。

六、结论和今后的工作

综上所述，本文提出了一种基于日志化处理技术的方法，通过从音频记录中提取12项非语言特征，为教师提供教学实践的深度洞察。研究结果表明，这些特征在区分不同教学方法和识别教师个人独特教学风格方面具有重要价值。该AI驱动机制使教师能够分析自身教学话语模式、与同行进行对比，并及时获得反馈以促进专业成长。

虽然我们提出的系统具有评估教师并促进特定教学实践遵循度的潜力，但需要特别指出的是，当前研究并未纳入评估教学质量或强制执行特定实践的标准衡量指标。相反，我们为教师提供信息性功能模块，以帮助他们反思自身的话语模式。进一步的分析，如将学生的表现或积极的感知与我们计算的特征相关联，可以实现人工智能驱动的推荐，增强我们系统的潜在影响。

另一个需要改进的领域是分类模型的准确性。教师话语具有内在的序列性特征，当前话语会受到前序话语的强烈影响。因此，一种改进方法是采用能够学习数据中长期时间依赖关系的深度序列学习模型。我们还将考虑通过词向量实现表征学习技术——这种技术通过将词语编码为捕捉其与其他词语相似性的向量来实现。

作为最终的研究方向声明，我们认识到解决样本多样性不足问题的重要性。为突破这一局限，我们计划开展新阶段研究，通过从更多元化的教师群体中收集补充数据。为扩大数据采集范围，我们还将纳入流媒体平台广泛传播的录音资料。这种方法将使我们能够捕捉到更具代表性的教学实践，并提高我们发现的普遍性。

致谢

作者感谢参与本研究并协助收集数据的评分员和教师。本研究由MCIN/AEI/10.13039/501100011033的NextGenerationEU/PRTR项目（欧盟）资助，项目编号TED2021-129300BI00；同时获得MCIN/AEI/10.13039/501100011033/FEDER的PID2021-122466OB-I00项目资助，项目编号为欧盟项目编号。

参考文献

[1] J. Archer, S. Cantrell, S. L. Holtzman, J. N. Joe, C. M. Tocci, and

J. 伍德,《更好的反馈促进教学提升:课堂观察改进实用指南》,约翰·威利出版社,2016年。

- [2] 王哲, 潘旭, K. F.米勒和K. S.科蒂娜,《课堂话语活动的自动分类》,《计算机与教育》第78卷,第115-123页,2014年。
- [3] T. J. Park, N. Kanda, D. Dimitriadis, K. J. Han, S. Watanabe, and S. Narayanan,《说话人日志化综述:深度学习最新进展》,《计算机语音与语言》第72卷,2022年。[Online]Available: <http://arxiv.org/abs/2101.09624>.
- [4] A.詹姆斯、Y. H. V.蔡、T.马斯奇克、A. M.努安兹、R.布尔、K.李和J.道威尔斯在《通过音频分析实现课堂氛围的自动化分类》一文中,于2019年发表于施普林格出版的第九届国际对话系统技术研讨会(第41-49页)。
- [5] M.E.Dale, A.J.Godley, S.A.Capello, P.J.Donnelly, S.K.D 'Mello, and S. P. Kelly, "Toward the automated analysis of teacher talk in secondary ela classrooms," Teaching and teacher Education, vol. 110, p. 103584,2022.
- [6] P. J. Donnelly, N. Blanchard, B. Samei, A. M. Olney, X. Sun, B. Ward, S. Kelly, M. Nystran, and S.K.D 'Mello, "Automatic teacher modeling from live classroom audio," in Proceedings of the 2016 conference on user modeling adaptation and personalization, 2016, pp.45-53.
- [7] P. J. Donnelly, N. Blanchard, A. M. Olney, S. Kelly, M. Nystrand, and S.K.D 'Mello, "Words matter: automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context," in Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 2017, pp.218-227.
- [8] S.K.D 'Mello, A.M.Olney, N.Blanchard, B.Samei, X.Sun, B. Ward, and S.凯利,《多模态捕捉师生互动以实现实时课堂中的自动对话分析》,载于2015年ACM国际多模态交互会议论文集,第557-566页。
- [9] D. Schlotterbeck, P. Uribe, R. Araya, A. Jimenez, and D. Caballero, "What classroom audio tells about teaching: a cost-effective approach for detection of teaching practices using spectral audio features," in LAK21: 11th International Learning Analytics and Knowledge Conference, 2021, pp.132-140.
- [10] 李浩、康宇、丁伟、杨思、黄继勇、刘振,《课堂活动检测的多模态学习方法》,收录于《2020年国际声学及信号处理会议(ICA-SSP)论文集》。IEEE出版社,2020年,第9234-9238页。
- [11] M. T. Owens et al., "Classroom sound can be used to classify teaching practices in college science courses," Proceedings of the National Academy of Sciences of the United States of America, vol.114, no.12, pp.3085-3090,2017.
- [12] 苏浩、卓博、吴翔、刘翔和孟浩,《基于讲座讨论录音的音频分类无监督方法》,收录于《国际语音学会议》2019年论文集,第3347-3351页。
- [13] I.巴塔查亚、M·弗利、N·张、T·张、C·库、C·米内、H·吉、C·里德尔、B·F·韦尔斯和R·J·拉德克,《用于无干扰群体会议分析的多模态传感器支持房间》,收录于《第20届ACM国际多模态交互会议论文集》(2018年,第347-355页)。
- [14] C. Lai, J. Carletta和S. Renals, "具有轮流特征的会议中参与者情感建模",见2013年情感社会语音信号研讨会论文集。
- [15] T. Nazaretsky, J. N. Mikeska和B. Beigman Klebanov,《用人工智能增强教师学习:论辩式讨论中教师关注学生观点的自动评估》,载于LAK23:第十三届国际学习分析与知识会议,2023年,第122-132页。
- [16] A. Suresh, J. Jacobs, M. Perkoff, J. H. Martin, and T. Sumner, "Finetuning transformers with additional context to classify discursive moves in mathematics classrooms," in Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022), 2022, pp.71-81.
- [17] A. Suresh, T. Sumner, J. Jacobs, B. Foland, and W.沃德,《自动化分析和反馈以改善数学教师的课堂话语》,载于人工智能

协会会议人工智能会议论文集，第33卷，第01期，2019年，第9721-9728页。

- [18] J. I. Castillo-Manzano, M. 卡斯特罗-努诺, L. López-Valpuesta, M. T.

Sanz-Díaz, 和 R. Yniguez, “测量ars对学业成绩的影响：一项全球元分析”，计算机与教育，第96卷，第109-121页，2016年。

- [19] B. Logan等人，《音乐建模的梅尔频率倒谱系数》，载于《Ismir》第270卷第1期，马萨诸塞州普利茅斯，2000年，第11页。