从有限标签中学习：用于印度音乐分析的传导式图标签传播

Parampreet Singh[1],Akshay Raina[2], Sayeedul Islam Sheikh[3], Vipul Arora[4]

[1,2,4]电气工程系， [3]化学工程系[1,2,3,4]印度理工学院坎普尔分校，印度 [4]鲁汶天主教大学 电子邮箱：params21@iitk.ac.in[1],akshayy.rainaa@gmail.com[2], sayeedul21@iitk.ac.in[3],vipul.arora@kuleuven.be[4]

arXiv:2601.03626v1 [eess.AS] 7 Jan 2026

摘要：

监督式机器学习框架依赖大量已标注数据集，才能在真实任务中取得稳健表现。然而，音频和音乐领域缺乏大型已标注数据集，因为标注此类录音需要大量资源、耗时费力，而且通常要求具备领域专业知识。本研究探索使用标签传播（LP）这一基于图的半监督学习技术，以无监督方式自动为未标注数据集添加标签。我们基于音频嵌入构建相似度图，在传导式半监督设置下，将有限的标签信息从小型已标注子集传播至更大的未标注语料库。我们将该方法应用于印度古典音乐（IAM）的两项任务：拉格识别和乐器分类。对于这两项任务，我们整合了多个公共数据集以及从 Prasar Bharati[1] 档案馆获取的其他录音，以执行标签传播。实验结果表明，与传统基线方法（包括基于预训练归纳式模型的方法）相比，标签传播显著降低了标注开销，并产生质量更高的标注。这些结果凸显了基于图的半监督学习在普及数据标注、加速音乐信息检索领域进展方面的潜力。

关键词： 标签传播、音乐信息检索、印度古典音乐、半监督学习

1. 引言

大多数现代机器学习（ML）方法都需要大量监督，才能在真实世界的问题上取得稳健的性能。然而，已标注数据有限，且实例标注成本高昂，这严重制约了机器学习解决方案的可扩展性和部署。这凸显了对稳健系统的需求：此类系统能够为原始样本分配高质量标签。

这一挑战在音频或音乐领域尤为突出，因为为这些录音添加标注需要长时间聆听，并确保帧级标注的精确率

[1]Prasar Bharati是印度的公共广播机构，由Doordarshan电视网和全印度广播电台组成。该机构保存着大量印度古典音乐录音。

。尽管网上有大量音频/音乐数据集 [1–13], ，但大多数数据集都受限于标注和采集成本，难以扩展，这可能导致标注错误 [14]。这也反映出此类数据集规模有限。例如， [11] 发布了 TablaSolo 数据集，其中仅包含 38 首塔布拉鼓独奏（印度打击乐器）曲目。同样， [3,5] 是用于声音事件分类的数据集，录音时长分别只有约 3 小时和 9 小时。音乐数据库语料库 [6], 中的大多数数据集也存在类似的局限。

因此，在保持标签质量较高的同时，使用能够对音频或音乐录音进行自动标注的系统十分重要。在此类场景中，基于已标注数据集进行训练并对未标注数据集推断标签的传统机器学习系统并非有效解决方案，因为它们需要丰富监督 [10, 15, 16]。其中一种解决方案是标签传播（LP），这是一种半监督技术，旨在从少量标注的已标注数据集进行学习，并利用直推式学习将标签传播到大型未标注数据集上。这种方法具有巨大潜力，近年来也吸引了大量研究兴趣 [17–19]。它利用了两个假设：（1）彼此距离较近的数据点很可能具有相同的标签；（2）同一流形上的大多数数据点应具有相同的标签 [20]。未标注数据点依据其特征与已标注数据点特征之间的相似性进行标注。大多数标签传播算法都基于图，其中两个节点（实例）之间的边编码了它们之间的亲和性。即使数据有限，这也能有效利用数据的底层结构，通过图传播标签。已有一些关于图像标签传播的重要研究 [18, 20–23], ，但这种方法很少被用于音频/音乐数据集。

标签传播为大规模元数据扩展提供了一种有效方案，尤其适用于已标注数据稀缺、未标注数据丰富的领域。它不依赖特定任务，因此可应用于广泛的场景，非常适合为来自 YouTube、Spotify 或公共档案馆等平台的大规模音频和音乐语料库添加标注。通过利用数据中固有的结构，标签传播提供了一种可扩展且高效的人工标注替代方案。

在本研究中，我们采用标签传播框架来完成音乐信息检索中的两项关键任务：拉格识别 [10, 15, 24] 和乐器识别。我们结合多个公开数据集与来自 Prasar Bharati档案馆的大型未标注音频录音语料库，在多种音乐内容上开展标签传播。结果表明，所提出的方法能够生成高质量标注，其表现往往优于多种基线方法，包括全监督归纳学习方法。

2. 相关研究

已标注音频数据的稀缺一直是推动音频与音乐处理领域机器学习应用发展的重大瓶颈。尽管人们对将机器学习应用于音频与音乐处理的兴趣日益浓厚，但可用带标签数据集的规模和范围有限，往往阻碍了相关进展。许多广泛使用的资源，例如 TinySOL [9], 、TablaSolo [11], 和 IRMAS [25], ，都受到数据规模或样本数量的限制。AudioSet [1] 和 FSD50K [2] 等大型数据集覆盖范围更广，但通常存在标注较弱的问题，且对于专业音乐信息检索任务而言，标注细节不够充分。对于印度古典音乐而言，IAM 拉格识别数据集 [7], 、Saraga [8], 和 PIM [10] 等数据集提供了宝贵资源，但规模有限，需要大量人工标注，而且通常侧重于拉格等特定方面

识别，缺乏更广泛的适用性。

数据集规模、多样性和标注质量方面的这些局限性凸显了采用能够最大限度发挥有限已标注数据效用的方法的必要性。在此背景下，标签传播通过将标签从少量已标注子集自动扩展到规模大得多的未标注数据集，为解决音乐与音频机器学习研究中的一个关键瓶颈提供了有前景的方案。

标签传播 是一种半监督学习方法，利用数据流形的结构，将标签从少量已标注样本传播到规模更大的未标注数据集。Zhu 和 Ghahramani [21] 提出了早期工作，采用全连接层图构建标签传播算法，其中边权重由数据点之间欧几里得距离的高斯核确定。该方法在保持已标注数据不变的同时，迭代地传播标签。Zhou 等人 [20] 通过在基于图的框架中同时引入局部与全局一致性，扩展了这一思想。他们引入了归一化图拉普拉斯矩阵，并将标签传播表述为闭式解，从而实现了高效计算。近年来，Iscen 等人 [22] 提出了一种将深度学习与标签传播相结合的方法。他们利用神经网络的嵌入构建稀疏亲和矩阵，再通过扩散过程传播标签。该方法兼具深度网络的表征能力以及图所捕获的结构信息。

我们的工作借鉴了标签传播领域的这些进展，并将其应用于音乐领域，具体应对大规模未标注数据集带来的挑战。

3. 数据集

对于拉格分类和乐器分类这两项任务，我们结合使用了经过整理的数据集和公开数据集。下文将介绍与各项任务相关的数据来源和组成。

3.1. 拉格分类

对于拉格分类任务，我们使用了 [10], 中介绍的 PIM 数据集，其中包含印度斯坦古典音乐演出的标注音频录音。该数据集包含 501 多个经人工标注的音频文件，共计 23,365 个时长为 30 秒的音频片段，对应 141 个不同的拉格。数据集还包含其他元数据，包括拉格、主音、塔拉、加拉纳以及表演者标注。数据集中的拉格和主音标签均由人工标注并经过核验。该数据集是我们实验的主要已标注数据来源。我们采用标签传播（LP），将这些标注扩展到来自 Prasar Bharati档案馆的大规模未标注音频录音语料库中。

3.2. 乐器识别

针对乐器识别任务，我们整理了一个数据集，其中的录音主要来自 Prasar Bharati档案馆，并辅以若干公开数据集中的样本。Prasar Bharati的音频主要包含印度古典音乐现场演奏中常见的乐器，如西塔琴、塔布拉鼓、维纳琴、帕卡瓦杰鼓和长笛。这些音频面临的一项显著挑战是乐器类别分布不均衡——总时长约有 65% 集中在出现频率最高的五个类别中。这种不均衡可能会影响机器学习模型的性能，尤其是在标签传播过程中。为解决这一问题，

表 1：从各类乐器识别公共数据集中抽取的样本。数字表示每种乐器的音频样本数量。乐器包括：手风琴（Acc.）、铙钹（Cym）、架子鼓（D.K.）、吉他（Guit.）、管风琴、钢琴、塔布拉鼓（Tab.）、小号（Trum.）、西塔琴（Sit.）、长笛（Flu.）、小提琴（Vio.）

数据集	Acc.	Cym.	D.K.	吉他	管风琴	钢琴	Tab.	小号	Sit.	Flu.	Vio.
FSD50K	99	835	351	2185	339	844	96	632	-	-	-
AudioSet	-	-	-	-	-	-	949	-	851	2697	-
IRMAS	-	-	-	-	-	721	-	577	-	451	580
TinySOL	689	-	-	-	-	-	-	96	-	118	284
独奏塔布拉鼓	-	-	-	-	-	-	38	-	-	-	-

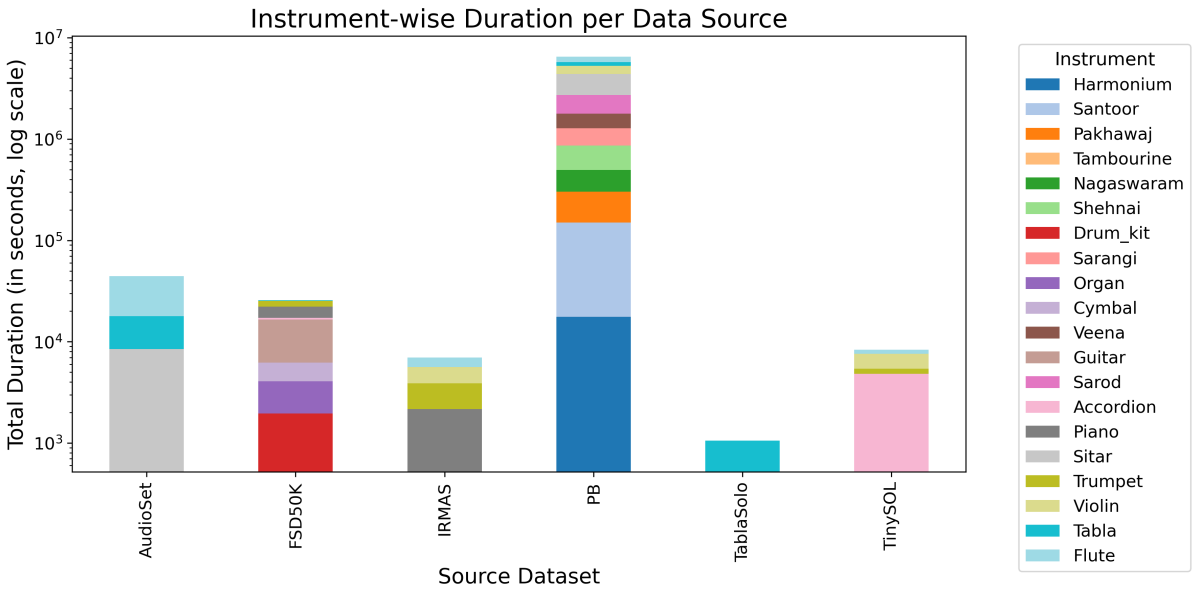


图 1：所用各数据集乐器样本按来源划分的时长（秒）。PB 代表 Prasar Bharati 音频。

我们通过纳入来自多个开源数据集的特定类别样本，对 Prasar Bharati的音频进行扩充，从而改善类别分布。具体而言，我们选取：TablaSolo [11]中的塔布拉鼓录音、FSD50K [2] 中的吉他、鼓组和塔布拉鼓录音、TinySOL [9] 和 IRMAS [25] 中的长笛和小提琴录音，以及 AudioSet [1] 中的长笛和鼓组样本。图 1 展示了按来源划分的乐器时长完整分布，其中包括 Prasar Bharati和外部数据集的贡献；从其他数据集获取的样本数量则列于表 1。合并后的数据集共包含 20 个乐器类别，如图 1 的图例所示。

我们将整个数据集划分为已标注数据集和未标注数据集，并从未标注数据集中整理出一个包含 200 段经人工标注和验证的音频录音的金标准集，用于评估。

4. 标签传播

可靠的已标注数据稀缺，因此有必要采用高效的传导式标签传播方法。我们利用未标注数据的伪标签来训练分类器，并利用从网络中获得的嵌入构建图 [22]。该方法分为两个步骤：（1）使用整个数据集训练网络（未标注数据点使用伪标签），

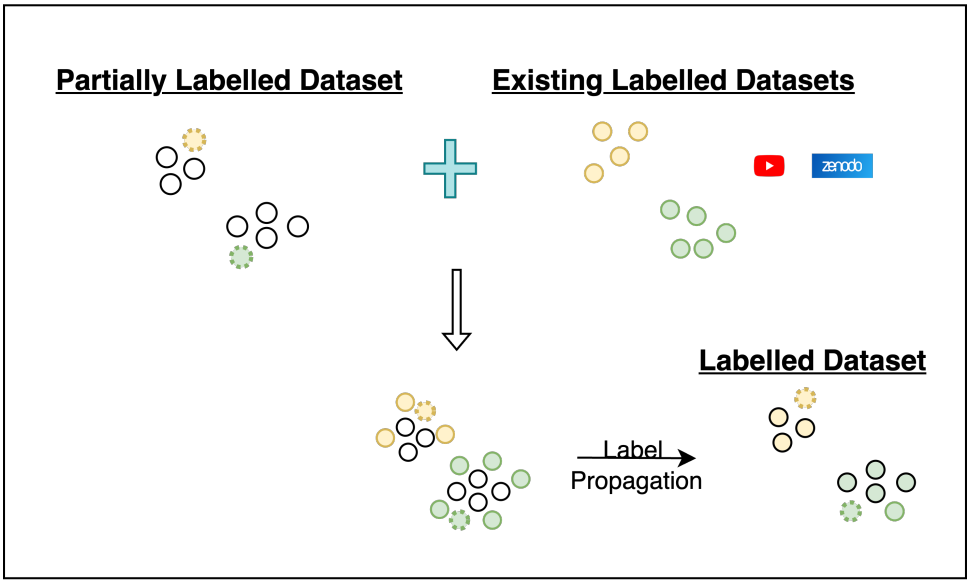


图2：展示对未标注音频样本进行自动标注的标签传播流程图。将一个小型已标注数据集与一组规模更大的部分标注或未标注样本相结合。随后利用相似度图进行标签传播，使标签信息从已标注子集传递至未标注数据，从而以传导式半监督的方式为整个语料库添加标注。

（2）使用网络生成的嵌入构建近邻图。图2中的流程图展示了标签传播的过程。首先，将一小组带有稀疏标签的音频样本（左侧）与来自 YouTube 和 Zenodo 等平台的外部已标注数据集（右侧）合并。这些数据集合并后形成一个同时包含已标注和未标注样本的统一数据集。通过应用标签传播，标注样本中的标签信息被扩展至未标注样本，从而得到一个完整标注的数据集（底部）。该方法显著减少了人工标注工作量，同时最大限度地利用现有数据来完成下游机器学习任务。

4.1. 分类器训练

设 $X = \{x_1, x_2, \ldots, x_l, x_{l+1}, \ldots, x_n\}$ with $x_i \in \mathcal{X}$为一组 n 数据点，其中前 l 个数据点带有类别标签 $y_i \in C$，其余 $u = n - l$ 个数据点未标注。标签空间记为 $C = \{1, 2, \ldots, c\}$。我们采用由特征提取器 $h_{\theta_1} : \mathcal{X} \to \mathbb{R}^d$ 和分类器 $g_{\theta_2} : \mathbb{R}^d \to \mathbb{R}^c$ 组成的深度网络。特征提取器将每个输入 x_i 映射为 d维嵌入 $z_i = h_{\theta_1}(x_i)$，分类器则输出各类别的置信度分数。整体模型表示为 $f_\theta(x) = g_{\theta_2}(h_{\theta_1}(x))$，其中 $\theta = (\theta_1, \theta_2)$。

训练目标包含多个组成部分：

$$L_s(X_l, Y_l; \theta) = \sum_{i=1}^{l} l_s(f_\theta(x_i), y_i), \tag{1}$$

$$L_p(X_u, \hat{Y}_u; \theta) = \sum_{i=l+1}^{n} l_s(f_\theta(x_i), \hat{y}_i), \tag{2}$$

$$L_u(X; \theta) = \sum_{i=1}^{n} l_u(f_\theta(x_i), f_{\hat{\theta}}(\hat{x}_i)), \tag{3}$$

其中， L_s 是针对已标注数据计算的监督损失（通常为交叉熵）。 L_p是利用标签 $\hat{}y_i$ 计算的伪标签损失，这些标签由标签传播预测得出。 L_u 是应用于 $X_l \cup X_u$ 的无监督损失项，旨在确保嵌入在输入的不同变换下保持一致。

4.2. 传导式标签传播

为了为无标签样本生成标签 $\hat{}y_i$ ，我们采用一种受扩散过程 [20]启发的传导式标签传播方法。其核心思想是根据所有数据点的嵌入构建相似度图，并沿图结构传播已知标签。

设 $W \in \mathbb{R}^{n \times n}$ 为一个对称邻接矩阵，其对角线元素为零，其中每个条目w_{ij} 表示嵌入 z_i 与 z_j之间的相似度。该矩阵通过对称的 k近邻（k-NN）图构建。我们定义 $W = M + M^T$，其中：

$$m_{ij} = \begin{cases} [(z_i^\top z_j)^\gamma]_+, & \text{if } i \neq j \text{ and } z_j \in \mathrm{NN}_k(z_i), \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

其中，$\mathrm{NN}_k(z_i)$ 表示嵌入空间中 z_i 的 k 个最近邻， γ 是锐度超参数， $[\cdot]_+$ 表示仅取正部（ReLU 函数）。

接下来，我们使用以下公式计算对称归一化亲和矩阵 S：

$$S = D^{-1/2} W D^{-1/2}, \tag{5}$$

where $D = \mathrm{diag}(W\mathbf{1}_n)$ 是度矩阵， $\mathbf{1}_n$ 是长度为 n的全 1 向量。现在，我们构造一个形状为 $n \times c$ 的标签矩阵 Y：如果 x_i 已标注且其类别为 j，则 $Y_{ij} = 1$；否则为 $Y_{ij} = 0$ 。因此，已标注样本采用独热编码，未标注行则全为零。标签传播通过闭式解为所有节点计算软标签：

$$P = (I - \alpha S)^{-1} Y, \tag{6}$$

其中， $P \in \mathbb{R}^{n \times c}$ 包含传播后的标签分布，而 $\alpha \in [0, 1)$ 控制标签扩散的强度。 P 的各行可解释为类别概率分数。最后，我们使用以下方式分配伪标签：

$$\hat{y}_i = \arg\max_j P_{ij}. \tag{7}$$

这会为每个示例 x_i 分配传播得分最高的类别。值得注意的是，对于大型 n ，计算矩阵逆 $(I - \alpha S)^{-1}$ 在计算上不可行，因为它不是稀疏矩阵。因此，我们按照 [22], 求解线性方程组：

	精确率	召回率	F1
语音	1.0	0.93	0.97
音乐	1.0	0.98	0.99

表 2：PANNs [26] 模型在 PIM 数据集已标注部分上的语音与音乐分类任务表现。该模型在 501 段标注了语音和音乐片段的印度斯坦古典音乐录音上进行评估。排除同时包含语音和音乐的模糊片段后，按 30 秒片段级计算指标。

$$(I - \alpha S)Z = Y \tag{8}$$

使用共轭梯度法，得到 $Z \approx P$。我们在联合优化前文所述损失函数的同时，迭代传播这些伪标签，确保总体损失在迭代过程中持续下降。收敛后，使用扩散过程得到的最终伪标签作为预测结果进行评估。

5. 实验

我们采用第 4 节讨论的 LP 方案，扩充乐器识别和拉格识别任务的元数据。首先，在已标注数据集 X_l 的 80% 上，以全监督方式训练深度神经网络。随后，利用训练好的网络为剩余 20% 的已标注录音和所有未标注录音 X_u 推断标签。为评估模型在未标注数据集上的表现，我们从 X_u 中抽取部分录音，进行人工标注和核验。最后，我们报告模型在留出的 20% 已标注数据以及未标注数据集中经人工标注的录音上的准确率。下面将详细介绍这两项任务的实验细节。

5.1. 乐器识别

对于乐器检测任务，我们首先对音频录音进行预处理，剔除所有时长短于 1 秒的文件。剩余文件被切分为 5 秒的音频片段，每个片段继承其来源录音的乐器标签。我们使用窗口大小为 1024、跳跃长度为 512、梅尔频带数为 64 的参数，从每个片段中提取梅尔频谱图，作为网络的输入特征。

作为基线，我们使用经过修改的 ResNet-18 [27]，并以全监督方式进行训练。对于我们提出的方法，我们采用第 4 节所述的标签传播（LP），利用一小部分带标签数据和更多无标签样本。两个模型均使用 Adam 优化器和随机梯度下降进行训练，最多训练 50 轮。评估时，我们保留了一个由 200 条音频录音组成、经人工标注的黄金测试集，并在留出的测试集上计算准确率以评估性能。

5.2. 拉格识别

对于拉格识别任务，我们使用总计 61,705 个音频片段，每段 30 秒。其中 13,075 个带有标签，来自 PIM [10]数据集，涵盖 41 个已知拉格类别，以及第 42 个其他类别。其余 48630 个为未标注录音

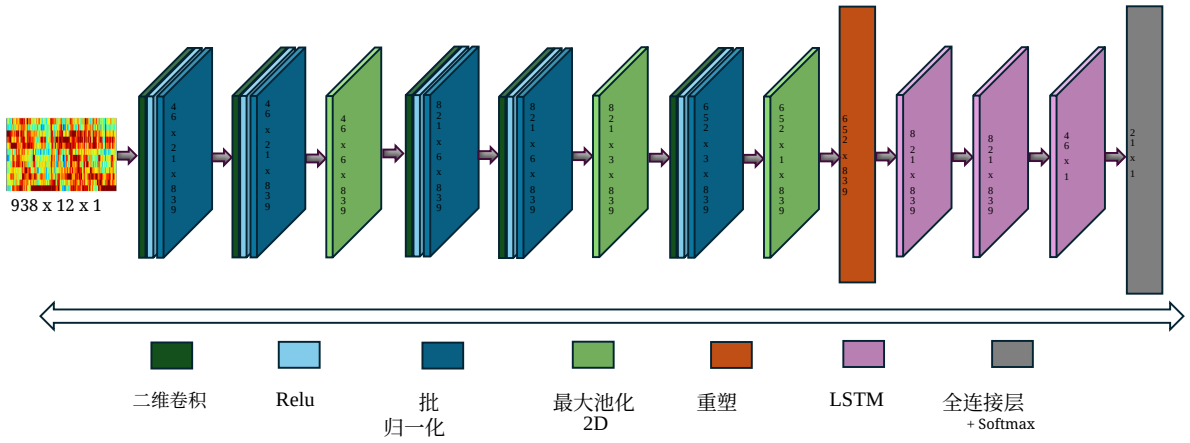


图3：用于拉格识别任务的模型架构。该模型由卷积层和长短期记忆层组成，用于捕捉音频数据中的时间依赖关系。

则来自 Prasar Bharati档案馆。为构建评估集，我们根据各拉格类别的代表性，从已标注数据集中为 42 个拉格类别分别至少选取一个完整音频文件。随后将这些录音切分成 30 秒的音频片段，并将所有生成的音频片段纳入评估集，共计 3,210 个评估片段。在 LP训练期间，我们会丢弃这些片段的真实标签，并将其与未标注数据集合并，从而构建直推式学习场景。

为滤除非音乐（语音）片段，我们首先为 PIM [10] 数据集中的 501 个录音标注语音和音乐的时间信息，并评估 PANNs模型 [26] 用于自动分段的性能。PANNs模型在 501 个印度斯坦古典音乐音频文件上进行测试，这些文件包含 23,224 个时长均为 30 秒的音频片段。由于标注存在歧义，包含重叠语音和音乐的音频片段将被排除。PANNs 的预测结果以片段级进行评估。对于模型最高预测结果不是音乐的情况，如果最高预测类别中包含相关的音乐标签（例如塔布拉鼓、西塔琴），且音乐 是预测概率第二高的类别，我们仍将其视为音乐。在 23,224 个音频片段中，21,691 个被分类为音乐，1,453 个被分类为语音，80 个被判定为模糊。如表 2 所示，PANNs模型在该任务中展现出稳健且可靠的性能，可用于标注来自印度广播公司录音的所有其他音频文件。

我们针对一个包含42个类别的分类任务训练模型，其中包括41个已知拉格（在已标注数据集中拥有最多音频文件的拉格）以及一个额外的其他类别，该类别包含所有其他剩余的拉格，以及使用PANNs模型从音频文件本身中提取的语音片段。对于主音归一化，我们使用 [10], 中提供的主音值；对于其余音频，我们使用 *CompIAM* 软件包 [28] 来计算主音值。正如 [10]中所述，该软件包已被证实对这一任务非常有用。

如图3所示，我们在分类任务中采用 CNN-LSTM 模型架构。首先，使用给定的标签以全监督方式训练模型，并在从已标注数据中提取的测试集上进行评估。接下来，我们采用相同的架构，仅预训练 10 个训练轮次，然后将其用于特征提取，并部署 LP 方法，使用有标签和无标签样本进行训练。在评估时，我们在音频片段级和音频文件级两个层面评估模型。在片段级，我们衡量每个音频片段的分类准确率，而

表3：标签传播与基线在乐器识别中的性能分析。准确率1 和准确率2 分别对应 X_l 和 X_u 子集上的准确率。

方法	准确率1（%）	准确率2（%）
基线	48.3	63.2
标签传播	**84.6**	**91.7**

表4：拉格分类任务的性能比较。CL：片段级，FL：文件级。

方法	精确率	召回率	F1分数
监督式（片段级）	0.63	0.57	0.60
监督式（文件级）	0.76	0.80	0.77
标签传播（片段级）	0.65	0.59	**0.62**
标签传播（文件级）	0.83	0.82	**0.82**

在音频文件级别，我们对来自同一音频文件的所有音频片段进行多数投票，以确定类别预测结果。

6. 结果与讨论

乐器识别。 表3展示了监督学习基线与LP方法之间的性能比较。准确率在两个数据集上进行报告：准确率1指已标注数据集中的一部分留出数据（X_l），准确率2指原始未标注数据集中的人工标注子集（X_u）。LP方法在准确率1和准确率2上分别达到84.6%和91.7%，而监督学习基线分别为48.3%和63.2%。这表明传播标签具有较高质量，并且LP后的模型预测结果相当可靠，即使是在训练期间未见过的数据上也是如此。

拉格分类。 表4展示了拉格分类的性能指标，分别在片段级（CL）和文件级（FL）进行评估。监督模型在片段级和文件级的F1分数分别为0.60（CL）和0.77（FL）。应用LP后，这些分数提升至0.62（CL）和0.82（FL）。虽然片段级的提升幅度有限，但文件级的改进十分显著（F1分数绝对提升5%）。对于我们的应用而言，这一点至关重要，因为该应用涉及为完整长度的音乐录音添加标签。尽管模型是在固定的拉格类别集上训练的，但它能够通过将此类实例归入通用的*其他* 类别，处理未见过或含义模糊的案例。

这些结果进一步印证了标签传播框架在为两项IAM任务生成有意义的伪标签方面的有效性。该技术可以轻松应用于下游的音乐信息检索分类任务。标签传播后的性能提升可用于音乐档案的自动编目或元数据生成等实际场景。总体而言，LP方法在各项任务中持续提升性能，展现出其作为大规模部分标注音乐数据集可靠且可扩展的标注策略的潜力。

7. 结论与未来工作

本文探讨了使用标签传播这一基于图的半监督学习技术，以应对音乐信息检索任务中已标注数据有限的挑战。我们聚焦于印度古典音乐（IAM）领域中的两项关键任务：拉格识别和乐器分类。通过在音频片段上构建相似度图，我们利用特征空间中的内在结构，以传导式半监督的方式，将标签从一个小规模的已标注子集传播到规模大得多的未标注语料库中。结果表明，标签传播是传统监督式方法的有效替代方案，尤其适用于IAM研究等领域；在这些领域中，专家标注成本高、耗时长，并且需要深厚的领域专业知识。该方法能够生成高质量的伪标签并实现可扩展标注，因此非常适合已标注数据稀缺但未标注数据丰富的场景。

本研究凸显了基于图的半监督学习在以最少人工标注开发稳健的音乐信息检索系统方面的广泛应用价值。未来的研究方向包括探索自适应或动态图构建，将时间和结构性音乐特征纳入传播过程，并扩展该框架以支持开放集识别，使系统能够自动检测并建模此前未见的标签，而不是将其归入笼统的“其他”类别。这些扩展将使系统更加贴近现实，更适用于新类别不断涌现的真实场景。

致谢

我们感谢印度 Prasar Bharati (PB) 提供音频录音和元数据。我们也感谢 Renu Chavan 和 Kajal Heer 在数据标注和验证方面所做的贡献。

参考文献

[1] J. F. Gemmeke、D. P. Ellis、D. Freedman、A. Jansen、W. Lawrence、R. C. Moore、M. Plakal 和 M. Ritter，《Audio Set：用于音频事件的本体和人工标注数据集》，载于 *2017*年 *IEEE* 声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2017年，第776–780页。[2] E. Fonseca、X. Favory、J. Pons、F. Font 和 X. Serra，《FSD50K：一个人工标注声音事件的开放数据集》，*IEEE/ACM*音频、语音与语言处理汇刊，第30卷，第829–852页，2021年。[3] K. J. Piczak，《ESC：用于环境声音分类的数据集》，载于第*23*届 *ACM*多媒体国际会议论文集，2015年，第1015–1018页。[4]J. J. Bosch、F. Fuhrmann 和 P. Herrera，《IRMAS：用于音乐音频信号中乐器识别的数据集》，2018年6月。 [在线]。获取地址：https://doi.org/10.5281/zenodo.1290750[5] J. Salamon、C. Jacoby 和 J. P. Bello，《用于城市声音研究的数据集与分类体系》，载于第*22*届 *ACM*多媒体国际会议论文集，2014年，第1041–1044页。

[6] X. Serra，《创建用于音乐计算研究的研究语料库：compmusic 项目案例》，载于 *AES* 第*53*届国际会议：语义音频；*2014*年*1*月*27*日至*29*日；英国伦敦。纽约：音频工程学会；*2014*年。文章编号 *1-1 [9* 页*]*。音频工程学会，2014年。

[7]S. Gulati、J. Serr`a、K. K. Ganguli、S. Sent¨urk 和 X. Serra，《印度古典音乐拉格识别数据集（音频）》，2016年8月。 [在线]。链接：https://doi.org/10.5281/zenodo.7278511

[8]B. Bozkurt、A. Srinivasamurthy、S. Gulati 和 X. Serra，《Saraga：印度古典艺术音乐研究数据集》，2018年5月。[在线]。可访问：https://doi.org/10.5281/zenodo.4301737[9]C. Emanuele、D. Ghisi、V. Lostanlen、F. L´evy、J. Fineberg 和 Y. Maresz，《TinySOL：孤立音乐音符音频数据集（5.0）》，2020年。 [在线]。可访问：https://doi.org/10.5281/zenodo.3685331[10] P. Singh 和 V. Arora，《印度古典艺术音乐拉格识别的可解释深度学习分析》，*arXiv* 预印本 *arXiv:2406.02443*，2024年。[11] S. Gupta、A. Srinivasamurthy、M. Kumar、H. A. Murthy 和 X. Serra，《塔布拉独奏录音中的音节化打击乐模式发现》，载于第*16*届国际音乐信息检索学会会议（*ISMIR*），2015年，第385–391页。[12]A. Shankar、G. Plaja-Roglans、T. Nuttall、M. Rocamora 和 X. Serra，《Saraga 视听：用于卡纳提克音乐分析的大型多模态开放数据集》，载于第*25*届国际音乐信息检索学会会议论文集。ISMIR，2024年11月，第61–69页。[在线]。可访问：https://doi.org/10.5281/zenodo.14877279[13] S. Kumar、P. Singh 和 V. Arora，《通过主动标注识别印度古典艺术音乐中的装饰音》，*arXiv*预印本 *arXiv:2505.04419*，2025年。[14] L. Schmarje、V. Grossmann、C. Zelenka、S. Dippel、R. Kiko、M. Oszust、M. Pastell、J. Stracke、A. Valros、N. Volkmann 等，《一条标注是否足够？——针对噪声及模糊标签估计的以数据为中心的图像分类基准》，神经信息处理系统进展，第35卷，第33 215–33 232页，2022年。[15] P. Singh、A. Mishra、A. Raina 和 V. Arora，《面向拉格识别的本体驱动层次学习》，载于*2025*年全国通信会议（*NCC*），2025年，第1–6页。[16] S. Kumar、P. Singh 和 V. Arora，《用于印度古典艺术音乐分析的置信度增强模型》，载于*2025*年*IEEE*国际声学、语音与信号处理会议研讨会（*ICASSPW*），2025年，第1–5页。[17] T. Xie、B. Wang 和 C.-C. J. Kuo，《Graphhop：一种增强型节点分类标签传播方法》，*IEEE*神经网络与学习系统汇刊，第34卷，第11期，第9287–9301页，2023年。

[18] T. Cai、R. Gao、J. Lee 和 Q. Lei，《亚群体偏移的标签传播理论》，载于第*38*届国际机器学习大会论文集，机器学习研究论文集系列，M. Meila 和 T. Zhang 编，第139卷。PMLR，2021年7月18日至24日，第1170–1182页。[19] S. Tankasala、L. Chen、A. Stolcke、A. Raju、S. Deng、C. Chandak、A. Khare、R. Maas 和 V. Ravichandran，《利用基于图的标签传播进行跨话语自动语音识别重评分》，载于*ICASSP 2023*，2023年。[20] D. Zhou、O. Bousquet、T. Lal、J. Weston 和 B. Sch¨olkopf，《利用局部与全局一致性进行学习》，神经信息处理系统进展，第16卷，2003年。[21] X. Zhu 和 Z. Ghahramani，《利用标签传播从有标签和无标签数据中学习》，*ProQuest*编号：提供给所有用户的信息，2002年。[22] A. Iscen、G. Tolias、Y. Avrithis 和 O. Chum，《用于深度半监督学习的标签传播》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2019年，第5070–5079页。[23] H. Zhu 和 P. Koniusz，《通过迭代图精炼进行基于原型标签传播的传导式小样本学习》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2023年，第23 996–24 006页。[24] P. Singh、A. Gupta、A. Mishra 和 V. Arora，《印度古典艺术音乐中未见拉格的识别与聚类》，载于第*26*届国际音乐信息检索学会会议论文集，2025年，第811–818页。[25] J. J. Bosch、J. Janer、F. Fuhrmann 和 P. Herrera，《音乐音频信号中主导乐器识别的声音分离技术比较》，载于*ISMIR*，2012年，第559–564页。[26] Q. Kong、Y. Cao、T. Iqbal、Y. Wang、W. Wang 和 M. D. Plumbley，《PANNs：用于音频模式识别的大规模预训练音频神经网络》，*IEEE/ACM*音频、语音与语言处理汇刊，第28卷，第2880–2894页，2020年。[27] K. He、X. Zhang、S. Ren 和 J. Sun，《用于图像识别的深度残差学习》，载于*IEEE*计算机视觉与模式识别会议论文集，2016年，第770–778页。[28]Gen´ıs Plaja-Roglans、Thomas Nuttall 和 Xavier Serra，《compiam》，2023年。[在线]。可访问：https://mtg.github.io/compIAM/