

# 音乐信息检索领域下的音乐流派分类研究综述

宋 扬, 王海龙, 柳 林, 裴冬梅

(内蒙古师范大学 计算机科学技术学院, 内蒙古 呼和浩特 010022)

**摘 要:** 以现阶段的音乐流派分类(music genre classification, MGC)研究进展及现状为探索内容, 介绍音乐流派分类的基本概念以及分类流程, 综述音乐流派中的机器学习、深度学习的音乐分类方法, 梳理总结相关音乐流派数据集。针对现阶段音乐流派分类方法进展及存在的不足, 对未来的发展进行了分析与展望。

**关键词:** 音乐信息检索; 音乐流派; 音乐分类; 分类方法

**中图分类号:** TP 391 **文献标志码:** A **文章编号:** 1001-8735(2022)04-0418-08

doi:10.3969/j.issn.1001-8735.2022.04.014

随着数字媒体的快速发展, 数字音乐无处不在。音乐用户的消费习惯已逐渐从实体音乐转向在线音乐平台。根据全球唱片协会发布的报告, 2020 年全球唱片市场同比增长 7.4%, 音乐行业总收入为 216 亿美元, 其中, 流媒体收入增长 19.9%, 达到 134 亿美元, 占全球唱片收入的 62.1%<sup>[1]</sup>。海量的音乐资源激发着音乐用户产生各种复杂的音乐检索需求, 音乐流派分类在音乐信息检索(music information retrieval, MIR)中占据主要地位, 同时在音乐组织管理、检索、音乐推荐中都具有重要的应用, 如何建立分类性能更好的音乐流派分类模型是 MIR 领域亟待解决的关键问题。

传统的音乐流派分类方法的步骤大致可分为特征提取和机器学习两个部分<sup>[2]</sup>。音乐流派分类的特征提取是从音乐信号中提取出能够反映音乐流派信息的某种特征, 机器学习能够将提取的某种特征进行学习, 以此来对音乐流派进行精确的分类<sup>[3-4]</sup>。但传统的音乐流派分类方法提取出的特征往往缺乏通用性, 同时, 传统的音乐流派分类方法对音乐相关知识的依赖度也较高。针对上述问题, 深度学习可以看作是提取器和分类器的结合, 可以直接从音乐信号中提取出有效特征并进行学习<sup>[5]</sup>。但目前基于深度学习的分类方法侧重于全局特征的学习, 而音乐流派是由多种高度多样化中间特性组成, 具有不一样的抽象层次<sup>[6]</sup>。因此, 现有的音乐流派分类方法仍存在一定的局限性, 不能满足音乐流派分类的要求<sup>[7]</sup>。综上, 在回顾国内外音乐流派分类发展的基础上, 对当前主流技术和前沿进展进行总结分析, 指出未来研究趋势, 并对音乐流派分类存在的问题及未来发展方向进行了总结与展望。

## 1 音乐流派

### 1.1 基本概念

音乐流派(music genre)是由流行乐、古典乐、爵士乐等音乐作品的独特节奏、音色、曲调、地域文化等元素形成的不同类别的音乐风格, 具体概念见表 1。

音乐流派作为音乐网站使用的分类检索项之一, 也是数字音乐数据库管理和存储的重要分类属性, 因此, 音乐流派的自动分类对音乐发展具有重要研究意义和应用价值。音乐流派的分类作为 MIR 领域的一个重要分支, 其研究不断深入, 是多媒体应用中不可或缺的一部分, 主要包括检索、欣赏、创作、教学和数字娱乐<sup>[8]</sup>。

收稿日期: 2021-11-25

基金项目: 国家重点研发计划资助项目(2020YFC1523300); 内蒙古自治区自然科学基金资助项目(2020MS06030, 2021MS06025)。

作者简介: 宋 扬(1997-), 女, 内蒙古兴安盟人, 在读硕士研究生, 主要从事音乐信息检索研究。

通讯作者: 王海龙(1975-), 男, 内蒙古乌兰察布人, 副教授, 主要从事分布式并行应用研究, E-mail: 9177385@qq.com。

表 1 音乐流派说明  
Tab. 1 Music genre description

音乐流派类别	概念
古典乐(Classical)	西方中世纪开始、在欧洲主流文化背景下创作的西方古典音乐
乡村音乐(Country)	曲调简单、节奏平缓,带有叙事性、乡土气息元素的西方民族特色流行音乐
迪斯科音乐(Disco)	由唱片播放出来的跳舞音乐,属于电子音乐的一种
嘻哈音乐(HipHop)	由浓郁的鼓点、旋律组成的说唱音乐
爵士乐(Jazz)	具有摇摆特点的 Shuffle 节奏为基础的即兴音乐
摇滚乐(Rock)	以灵活大胆的表现形式和富有激情的音乐节奏表达情感的音乐类型
蓝调音乐(Blues)	充满压抑、不和谐的情绪,以歌曲直接陈述内心想法表现形式的音乐类型
雷鬼音乐(Reggae)	融合了节奏蓝调的抒情曲风和拉丁音乐元素的音乐类型
流行音乐(Pop)	那些结构短小、形式活泼、情感真挚,并被广大群众所喜爱,广泛传唱或欣赏,流行一时的甚至流传后世的器乐曲和歌曲
金属乐(Metal)	以重金属为主的重型音乐

1.2 音乐流派分类流程

音乐流派分类已成为近年的研究热点,有效识别音乐流派不仅有助于音乐信息的检索,同时也往往作为特定类型的音乐操作,如语音识别,音乐检索等的预处理<sup>[3]</sup>。其本质上属于模式识别的范畴,一般包括两个主要部分:特征提取和分类识别。

音乐流派分类主要流程如下:首先,对输入的音乐信号进行预处理;其次,提取音乐特征;再经过有监督或无监督分类训练,得到分类器;最后,使用测试数据并验证分类结果。

(1) 预处理。在音乐流派分类研究中,来自不同数据源的音乐文件可能具有不同的存储格式和采样频率,且原始音乐信号中可能存在干扰噪声等无用信息。因此,预处理是音乐流派分类研究的第一步和基本环节。主要包括统一信号格式、采样量化、成帧、加窗、预加重等,为后续的特征提取步骤提供标准化、可靠的音乐样本。

(2) 特征提取。特征提取是整个分类过程的关键部分,因为特征的选择和提取在很大程度上决定了分类效果,主要目的是找出更贴合音乐信号特征的关键特征参数。在音乐流派分类领域,常用的特征参数包括短时能量、过零率、短时谱、线性预测系数(LPC)和 Mel 倒谱系数。

(3) 训练分类器。训练分类器是决定分类最终性能的关键模块。通过构建训练样本集对分类器进行训练,找到合适的参数,最终得到训练好的分类器。

(4) 分类结果测试。用于测试分类器的效果。待预测的新音乐样本进入分类器,利用训练好的分类器对待预测的样本进行分类和预测,判断其是否属于该类别,最后统计分类准确率。

2 音乐流派分类方法

在音乐流派分类任务中,分类是否准确的决定因素是提取特征的精准度以及分类器建模的实用性。以往的分类方式是提取和分类分为两个部分。从原始音乐信号中手动提取特征,然后选择合理的分类器对提取的特征进行分类建模。这种传统方式虽然具有一定的实用性,也取得了一定成效,但在提取特征过程中效率低,操作较难,而且提取出的特征不具备普遍性。随着深度学习的发展及应用,越来越多的研究者将音乐流派分类和深度学习模型相结合,制作音乐声谱图进行输入。

2.1 传统分类方法

传统音乐流派分类方法从特征提取和分类两部分来研究音乐流派分类内容。提取出的特征能够反映不同音乐类型,反之,分类效果也能够体现特征选择的准确性。

2.1.1 特征提取 音乐本质上是频率、振幅、相位的振动波在不同时间点上的组合,音乐特征可以从多种

角度分类,有不同的描述方法<sup>[9]</sup>。特征提取的作用是将音乐中能够表征音乐本身的特征提取出来构成特征向量。如果从音乐理解的角度来划分,音乐特征具有低级和中级之分;如果从变换域的层次来看,音乐特征可分为频域、时域和倒谱域三种;按照音乐的帧长来划分可以分为长时、短时特征。Weihs 等<sup>[10]</sup>认为音乐有长时、短时、语义和成分四种特征。Scaringella 等<sup>[11]</sup>将音乐特征分为音色、节奏和音调。Fu 等<sup>[12]</sup>将音乐特征分为声学特征和感知特征两种。音乐特征的提取是分类任务中必不可少的一步,特征的质量是影响分类准确率至关重要的因素,只有提取出能够准确表征音乐流派的音乐特征,才能提高分类的整体性能。

由于长时特征的提取过程较为烦琐,多个研究方法通过结合短时特征来得到长时特征,本文着重介绍常见的短时特征,见表 2。

2.1.2 机器学习分类方法 在传统的音乐流派分类任务中,特征提取结合有效的分类器可以很好提升分类效果,比较常见的传统分类器有支持向量机(SVM)<sup>[13]</sup>、高斯混合模型(GMM)<sup>[14]</sup>和 K 近邻(KNN)<sup>[15]</sup>等。

国外音乐流派分类的研究起源于 2002 年,Tzanetakis 等<sup>[14]</sup>将音乐流派分类视为一项模式识别任务,提出了一种基准数据集 GTZAN 共计 1000 首音乐,包含 10 种音乐流派。分别提取了音色、节奏和音高三个特征输入到 GMM 和 KNN 进行分类,并对这些特征的性能进行了比较研究,获得了 61% 的准确率,成为最具开创性的工作之一。这些特征后来成为音乐流派分类研究的共同特征,实验结果被广泛引用,许多类似的研究也将该实验结果作为参考对比。

2002 年以来,Laurier 等<sup>[16]</sup>选择多种音频特征作为特征集,采用 SVM 进行分类。Lee 等<sup>[17]</sup>以长期调制谱分析和 MFCC 特征为基础,采用特征融合的方式来识别音乐流派的类型,可以观察出音乐变化情况以及节奏信息,以确保形成的特征具有紧凑性和有效性。随后,学者们又把目光放到了音乐信号频谱特征时间演化的研究上,认为捕捉频谱特征的难度会比其他方法难度低。Lim 等<sup>[18]</sup>根据音色特征找到相应的光谱时间特征,建立音乐流派分类模型,利用 SVM 分类器进行特征选择进而实现分类。Nanni 等<sup>[19]</sup>提取了节奏直方图、频谱描述符以及光谱图,划分成不同的子窗口,根据每一组特征进行相应的描述和训练,最后将融合后的音乐特征输入到 SVM 进行分类,从而获得分类结果。

国内音乐流派分类的研究起步较晚。杨翠丽等<sup>[20]</sup>对多数投票机制进行了优化和完善,通过计算每一个片段的百分比,结合音乐的节奏特征来划分整个音乐样本。俞晓文等<sup>[21]</sup>从输入特征出发,利用旋律特征找到音高和节奏,并将其作为输入特征,使用前馈神经网络作为音乐特征的分类器,证明了这种提取音乐特征的方式具有良好的实用性。甄超等<sup>[22]</sup>基于声学特征提出了特征交互正向特征选择算法,在标签语料库中通过文档主题生成模型(LDA)来构建新的分类模型,从而将这些标签按照不同的类别划分,这种声学特征与标签流派分类融合的方式,能够有效提高分类的效果和精确度。李镛等<sup>[23]</sup>在二维光谱中综合运用尺度不变特征变换(SIFT)和空间金字塔匹配(SPM),找到了相应的特征向量,输入到 SVM 来达到分类效果,通过实验研究证明,采用听觉频谱的分类方式远优于 MFCC 特征。肖晓红等<sup>[24]</sup>将音色特征、节奏特征通过 bagging 训练两组隐马尔可夫模型进行分类,达到了良好的效果,同时也证明了融合节奏特征可以有效提高分类准确率。杜威等<sup>[25]</sup>利用层次分析法分析了音乐信号中的光谱特征。首先通过多个 SVM 构建相应的分类树,计算每一个叶节点的得分,并且将这—个分值看成是特征,结合 KNN 算法来确保精确度。

以上音乐流派分类方法主要分为特征提取和构建分类器两部分。然而目前在特征提取、分类方面仍存在一些不足:在特征提取方面,特征提取方式有限,大多通过人为提取特征,且提取的特征可能只适用某一流派音乐,无法在其他流派中发挥同样的效果;在分类方面,由于音乐流派分类中特征向量的维数往往是

表 2 常见的短时特征

Tab. 2 Common short-term characteristics

特征	分类
时域特征	过零率(ZCR)
	短时能量(STE)
	自相关系数(AF)
频域特征	频谱质心(SC)
	频谱衰减(SR)
	频谱流量(SF)
	频谱带宽(SB)
	频谱平坦度测量(SFM)
	频谱波峰因数(SCF)
	幅度谱包络(ASE)
	倍频频谱包络(OSC)
	小波系数直方图(DWCH)
倒谱域特征	线性预测倒谱系数(LPCC)
	梅尔频率倒谱系数(MFCC)
	傅里叶倒谱系数(FCC)



几十维甚至几百维,一个实用、有意义的音乐分类系统需要处理八个甚至十个以上的音乐流派类别。在实现的过程中,随着分类需求以及分类数量的增加,计算复杂度也将逐渐增加,这为进一步提高分类精度带来了困难。

近几年深度学习在计算机视觉和自然语言处理领域取得良好成绩,音乐流派分类的解决方案也逐渐将重心转移到深度学习上,以卷积神经网络(CNN)为代表的深度学习方法,为解决音乐流派分类问题带来了新的思路 and 方向。

## 2.2 基于深度学习分类方法

20 世纪 60 年代是深度学习的萌芽阶段;21 世纪初期,Hinton<sup>[26]</sup>提出的深度置信网络(DBN)推动了深度学习的发展。随着深度学习应用领域的不断开展,在音乐流派分类任务中也得到应用<sup>[27-28]</sup>。在深度学习中,重要特征的获取都是自发性的,可以很好地完成音乐流派分类任务。

在国外相关研究中,传统的机器学习方法逐渐被深度学习所取代。如卷积神经网络(CNN)具有很强的特征学习能力,已经在大量分类任务中证明了它的有效性。Costa 等<sup>[29]</sup>将 CNN 用于音乐流派分类任务,将结果与手动特征、SVM 分类器做比较,在提取特征时结合了 CNN 和手工提取两种方式,从而提高音乐分类的准确率。Li 等<sup>[30]</sup>将 CNN 作为特征提取器,结合多数投票机制对提取的特征进行选择,在 GTZAN 数据集上取得显著成果<sup>[31]</sup>。Lee 等<sup>[32]</sup>提出了一种不同于典型帧级特征表示的 CNN 结构,通过降低音频信号的频率来提高训练的速度,从而缩短训练所需要的时间,应用迁移学习的方式对模型进行扩展,样本层 CNN 在每一层中学习的滤波器可视化,以识别分层学习特征。Song 等<sup>[33]</sup>在散射变换特征的基础上,提出新的深度递归神经网络自动标注算法。散射特征使得带有门循环单元(GRU)的循环神经网络(RNN)能够实现变换谱的功能,而且这种变换的方式相对于 MFCC 和 Mel 谱图来说效率更高。

国内相关研究中,何丽等<sup>[34]</sup>在音乐流派分类中使用了 LSTM 网络,利用 GTZAN 数据集提取 Mel 谱质心、倒谱系数和谱对比度三个特征,并且通过 LSTM 网络来进行特征训练。陆欢<sup>[35]</sup>采用倒谱系数提取音频的 MFCC 特征矩阵,以其特征值作为 CNN 的输入对音频信号进行训练,验证了 CNN 分类的有效性。郑旦<sup>[36]</sup>提出了基于深度置信网络的多特征融合音乐分类算法,对音乐信号从多个角度进行特征选取并融合,由融合得到的训练集在改进的深度置信网络中进行训练,通过调整 RBM 中可见层与隐藏层单元之间的权值得使得模型达到最优性能。杨汶雯等<sup>[37]</sup>针对目前深度学习在音乐分类中准确性较低的问题,从音乐特征角度进行改进,根据声学、信号学的基本理论,提出音乐特征指标体系,基于自适应粒子群优化神经网络的算法对音乐流派进行分类。

由上可知,CNN 和 RNN 广泛应用于音乐流派分类任务中。CNN 保证了一定程度的平移、缩放,可以有效地利用时间和频率不变性,但未考虑全局结构。而 RNN 考虑全局结构,在综合局部特征方面更加灵活。因此,在序列数据中 RNN 被广泛用于建模时间依赖关系<sup>[38]</sup>。它们在训练时间和参数数量方面均表现良好,然而顺序计算的耗时性仍然是一个显著的限制。针对 RNN 中存在的问题,LSTM 通过改变单元结构和添加存储单元来判断是否有必要记住信息。与 RNN 不同,LSTM 的每个存储单元包括输入门、遗忘门和输出门,是网络中信息传输的接口。因此,LSTM 比 RNN 更适合处理具有相对较长时间间隔的序列和预测任务中的关键延迟元素。

目前音乐流派分类结合了机器学习和深度学习两种方法。深度学习<sup>[39]</sup>采用监督或半监督的高效算法,有效地降低了传统机器学习算法人工特征获取的误差,能够提升音乐流派的分类准确率。但深度学习的成本高且模型设计较为复杂,还需进一步结合音乐信息加以改进,从而提升分类的整体性能。不同类型分类方法对比见表 3。

## 3 音乐流派数据集

音乐流派方面的公共数据集数量相对较少,本文总结了音乐流派分类领域部分常用的数据集,具体信息见表 4。

(1)GTZAN 数据集由 Tzanetakis 和 Cook 收集<sup>[14]</sup>,涉及的流派类别共 10 个(蓝调、古典、乡村、迪斯科、嘻

哈、爵士、金属、流行、雷鬼和摇滚)。每一个流派类别均从 CD、收音机、MP3 压缩音乐文件中收集 1000 段音乐摘录,每 100 段的音频时长大约 30 秒,每一个项目的内存有 22.05 kHz,为 16 位的单声频道。

表 3 不同类型分类方法对比

Tab. 3 Comparison of different types of classification methods

对比项目	KNN	SVM	GMM	HMM	CNN	RNN	LSTM
原理	一个样本在特征空间中的 K 个最相邻的样本中的大多数属于某一个类别	在特征空间中寻找间隔最大化的分离超平面的线性分类器	多个单高斯分布进行混合	用来描述一个含有隐含未知参数的马尔可夫过程	多层神经网络,每层由多个二维平面组成,而每个平面由多个独立神经元组成	处理时序数据的神经网络模型	特殊的 RNN,主要通过三个门控逻辑实现(遗忘、输入、输出)
音乐分类适用范围	非线性分类、容量较大的训练数据	小样本分类、非线性分类和高维空间分类	基频和能量相关特征的音乐分类	连续、大量的、与节奏特征相关的音乐分类	输入频谱图分类、低级声学特征的音乐分类	结合频谱特征和韵律特征的时间序列分类	时间序列有关的分类
优点	易实现,无需估计参数	具有较好的泛化推广能力	拟合功能强、鲁棒性较强	时间序列的建模能力较强、系统的扩展性好	共享卷积核、泛化能力更强、特征分类效果好	可对序列内容建模能力强	解决长序列训练过程中的梯度消失和梯度爆炸问题
缺点	计算量较大、可解释性较弱、需大量内存	求核函数的困难、且多分类问题难解决	对数据的依赖性过强、初始值和阶数较难确定	模型复杂度高、拟合功能一般、鲁棒性差	容易出现梯度消散问题	容易出现梯度消散或梯度爆炸问题	计算量大、耗时

(2)ISMIR2004<sup>[9]</sup>是一个音乐信息检索竞赛的音乐流派分类数据集,数据集共有 1458 个样本,分为 6 种不同的流派:古典、电子、爵士乐、金属/朋克、摇滚/流行和世界。1458 首音乐作品平分为训练集和测试集。

(3)Extended Ballroom<sup>[40]</sup>数据集是在 Ballroom 数据集的基础上发展而来,作为 Ballroom 数据集的拓展集,Extended Ballroom 数据集共有 4180 个曲目,超出 Ballroom 数据集五倍多。其中,样本包括恰恰舞、狐步舞、牛仔舞、斗牛舞、快步舞、伦巴舞、萨尔萨舞、桑巴舞、慢华尔兹舞、探戈舞、维也纳华尔兹舞、华尔兹舞和西海岸摇摆舞。

(4)Million Song 数据集(MSD)中的音乐信息来自世界各个音乐社区,可以从多个音乐平台进入,包含超过 100 万首音乐。MSD 提供每首歌曲的元数据和预先计算的音樂特征。受到版权内容的限制且音乐评价来源复杂,因此并没有规范的音乐流派分类标签。

(5)MagnaTag ATune 数据集中歌曲的片段数据达 2.5 万首,每一个片段的时长大约都在 29 秒左右。歌曲片段可分为电子、爵士乐、嘻哈、摇滚、布鲁斯等共 69 个流派类别。数据由在线游戏提供,每段剪辑由两名玩家播放,只有两人同时标记一个音乐片段时,标签才会生效。

## 4 音乐流派分类性能评价

### 4.1 评价指标

为有效判断模型质量,以往研究使用的评价标准为 AUC<sup>[41]</sup>,表示预测的正例排在负例前面的概率。其具体解释以 FPR(假阳率)为横轴,以 TPR(真阳率)为纵轴,分类阈值在不断调整过程中会形成一系列点对,并且将这些点对一一对应到坐标轴上,用直线将这些点连接起来,即形成 ROC 曲线,曲线下的面积代表 AUC 值。这一评价标准在二分类的领域中具有较好的应用,因此可以将多分类转化为二分类问题,最后求得结果的平均值。如果将分类问题都看成二分类问题,样本就只有两种结果,一种是符合该类别,反之则不

表 4 数据集相关信息

Tab. 4 Data set related information

数据集	音乐总数量/首	音乐类别数目/类
GTZAN	1000	10
ISMIR2004	1458	6
Extended Ballroom	4180	13
Million Song	100 万	—
MagnaTag ATune	2.5 万	69

符合。该指标也可用于二分类的评价任务。

一般评估指标包括精确率、召回率等多个方面的内容。为能够对这些指标有一个更好的认识,本文列出了多分类问题的混淆矩阵,见表5。

使用表中的定义,单个类别的精确率(precision,  $P$ )定义为 TP 与 (TP+FP) 的比值,表示在所有被预测为正样本中,实际标签为正样本的概率;真阳率即召回率(recall,  $R$ )定义为 TP 与 (TP+FN) 的比值,表示实际标签属于该类别的样本中正确预测的样本的比例;假阳率  $F_{PR}$  定义为 FP 与 (FP+TN) 的比值,表示实际标签不属于该类别的样本中,但是在分类模型中则被判断为该类别的样本比例。

#### 4.2 结果分析

早期的音乐流派分类是基于传统机器学习方法,通过特征提取输入到分类模型中,分类效果并不理想。

深度学习极大促进了音乐流派分类的发展。音乐流派分类方法在 GTZAN、ISMIR2004 和 Extended Ballroom 数据集上的结果见表6。

表5 多分类问题混淆矩阵

Tab. 5 Confusion matrix for multi-class problems

真实结果	预测结果		
	属于该类	不属于该类	总数
属于该类	true positive (TP)	false negative (FN)	实际属于该类 (TP+FN)
不属于该类	false positive (FP)	true negative (TN)	实际不属于该类 (FP+TN)
总数	预测属于该类 (TP+FP)	预测不属于该类 (FN+TN)	TP+FN+FP+TN

表6 模型结果对比

Tab. 6 Comparison of model results

年份	分类方法	数据集	准确率/%	文献
2002	GMM&KNN	GTZAN	61.00	[13]
2013	SVM	ISMIR2004	81.00	[42]
2016	HMM	GTZAN	66.80	[22]
2019	LSTM	GTZAN	68.33	[30]
2019	CNN	GTZAN	88.00	[31]
2021	CNN	Extended Ballroom	97.20	[43]

### 5 存在的问题及展望

#### 5.1 存在的问题

音乐流派分类发展至今已取得一定的成果,众多学者为音乐流派分类提供了新思路和方法,但仍有许多难题有待解决。

(1)目前的大多数公开数据集都是使用古典、摇滚、金属等通用的流派对音乐进行分类,但不同地区的文化和语言有所差异,通用音乐流派并不适用于民族音乐等特定领域的流派分类,因此建立适用于特定领域的数据集较为困难。

(2)在特征选择阶段,目前还不清楚哪些特征对音乐流派分类效果的影响最为显著,无法直接找到与分类目标有明显关联的深层音乐特征,且对于样本数量较少的数据集,提取的特征数量有限,最终导致分类结果不理想。

(3)相比机器学习,深度学习对特征的学习更加深入,但其对数据的依赖性较强,在面对小规模样本数据集时,无法达到理想的分类效果,给分类工作带来一定局限性。随着研究的深入,如何运用深度学习方法更准确地识别音乐流派是当前工作的难点之一。

#### 5.2 展望

针对现有研究的局限性,今后音乐流派分类可以从以下几个方面进行进一步的研究:

(1)在音乐流派数据集的建立方面,可以将特定领域的音乐流派与通用流派相区分,通过构建适用于特定领域的音乐流派数据集,进而推动音乐流派数据集的多元化建立。

(2)在音乐流派分类方面,可以通过对深度学习网络结构进行改进、引入新的参数,从而对无标注的音乐数据进行更加精准的分类。在此基础上,将深度学习与其他方法相融合,有望进一步提升分类效果。

(3)未来音乐流派分类可以灵活利用音乐歌词、专辑封面等其他模态信息,为音乐流派分类带来更为理想的效果。

## 参考文献:

- [1] International Federation of the Phonographic Industry. IFPI global music report 2020[EB/OL]. (2020-05-04) [2021-11-25]. <https://www.ifpi.org/ifpi-issues-annual-global-music-report/>.
- [2] 李伟,李子晋,高永伟. 理解数字音乐:音乐信息检索技术综述[J]. 复旦学报(自然科学版),2018,57(3):271-313.
- [3] STRUM B L. A survey of evaluation in music genre recognition[C]//International Workshop on Adaptive Multimedia Retrieval. Cham:Springer,2012:29-66.
- [4] CORRÊA D C, RODRIGUES F A. A survey on symbolic data-based music genre classification[J]. Expert Systems with Applications,2016,60:190-210.
- [5] BISHARAD D, LASKAR R H. Music genre recognition using convolutional recurrent neural network architecture[J]. Expert Systems,2019,36(4):e12429.
- [6] LEE J, NAM J. Multi-level and multi-scale feature aggregation using pretrained convolutional neural networks for music auto tagging[J]. IEEE Signal Processing Letters,2017,24(8):1208-1212.
- [7] NANNI L, COSTA Y M G, LUMINI A, et al. Combining visual and acoustic features for music genre classification[J]. Expert Systems with Applications,2016,45:108-117.
- [8] TSUNOO E, TZANETAKIS G, ONO N, et al. Beyond timbral statistics: Improving music classification using percussive patterns and bass lines[J]. IEEE Transactions on Audio, Speech, and Language Processing,2010,19(4):1003-1014.
- [9] WU M J, JANG J S R. Combining acoustic and multilevel visual features for music genre classification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM),2015,12(1):1-17.
- [10] WEIHS C, LIGGES U, MÖRCHEN F, et al. Classification in music research[J]. Advances in Data Analysis and Classification,2007,1(3):255-291.
- [11] SCARINGELLA N, ZOIA G, MLYNEK D. Automatic genre classification of music content: A survey[J]. IEEE Signal Processing Magazine,2006,23(2):133-141.
- [12] FU Z, LU G, TING K M, et al. A survey of audio-based music classification and annotation[J]. IEEE Transactions on Multimedia,2010,13(2):303-319.
- [13] 孙辉,许洁萍,刘彬彬. 基于多核学习支持向量机的音乐流派分类[J]. 计算机应用,2015,35(6):1753-1756.
- [14] TZANETAKIS G, COOK P. Musical genre classification of audio signals[J]. IEEE Transactions on Speech and Audio Processing,2002,10(5):293-302.
- [15] SUDARMA M, HARSEMADI I G. Design and analysis system of KNN and ID3 algorithm for music classification based on mood feature extraction[J]. International Journal of Electrical & Computer Engineering,2017,7(1):486.
- [16] LAURIER C, HERRERA P, MANDEL M, et al. Audio music mood classification using support vector machine[J]. In MIREX Task on Audio Mood Classification,2007(1):2-4.
- [17] LEE C H, SHIH J L, YU K M, et al. Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features[J]. IEEE Transactions on Multimedia,2009,11(4):670-682.
- [18] LIM S C, LEE J S, JANG S J, et al. Music-genre classification system based on spectro-temporal features and feature selection[J]. IEEE Transactions on Consumer Electronics,2012,58(4):1262-1268.
- [19] NANNI L, COSTA Y, AGUIAR R L, et al. Ensemble of deep learning, visual and acoustic features for music genre classification[J]. Journal of New Music Research,2018,47(4):383-397.
- [20] 杨翠丽,郭昭辉,武港山. 基于改进投票机制的音乐流派分类方法研究[J]. 计算机工程,2008,34(9):213-215.
- [21] 喻晓雯,张楠,张勇. 音乐作品风格流派的神经网络识别方法研究[J]. 计算机工程与应用,2011,47(27):246-248.
- [22] 甄超,宋爽,许洁萍. 多模态音乐流派分类研究[J]. 计算机科学与探索,2011,5(1):50-58.
- [23] 李镔,李秋颖,关欣. 基于听觉图像的音乐流派自动分类[J]. 天津大学学报,2013,46(1):67-72.
- [24] 肖晓红,张懿,刘冬生,等. 基于隐马尔可夫模型的音乐分类[J]. 计算机工程与应用,2017,53(16):138-143.
- [25] 杜威,林浒,孙建伟,等. 一种基于分层结构的音乐自动分类方法[J]. 小型微型计算机系统,2018,39(5):888-892.
- [26] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science,2006,313(5786):504-507.
- [27] YU Y, LUO S, LIU S, et al. Deep attention based music genre classification[J]. Neurocomputing,2020,372:84-91.



- [28] YANG R, FENG L, WANG H, et al. Parallel recurrent convolutional neural networks based music genre classification method for mobile devices [J]. IEEE Access, 2020, 8:19629-19637.
- [29] COSTA Y M G, OLIVEIRA L S, SILLA C N. An evaluation of convolutional neural networks for music classification using spectrograms[J]. Applied Soft Computing, 2017, 52:28-38.
- [30] LI T L, CHAN A B, CHUN A H. Automatic musical pattern feature extraction using convolutional neural network [J]. Lecture Notes in Engineering & Computer Science, 2010, 1:1-5.
- [31] STURM B L. The state of the art ten years after a state of the art: Future research in music information retrieval [J]. Journal of New Music Research, 2014, 43(2):147-172.
- [32] LEE J, PARK J, KIM K, et al. Sample CNN: End-to-end deep convolutional neural networks using very small filters for music classification[J]. Applied Science, 2018, 8(1):150.
- [33] SONG G, WANG Z, HAN F, et al. Music auto-tagging using deep recurrent neural networks[J]. Neurocomputing, 2018, 292:104-110.
- [34] 何丽,袁斌. 利用长短期记忆网络进行音乐流派的分类[J]. 计算机技术与发展, 2019, 29(11):190-194.
- [35] 陆欢. 基于卷积神经网络的音乐流派分类[J]. 电子测量技术, 2019, 42(21):149-152.
- [36] 郑旦. 基于深度置信网络的多特征融合音乐分类算法[J]. 电子设计工程, 2020, 28(4):132-136.
- [37] 杨汶雯,石梦荧. 基于深度学习的音乐特征提取及流派分类[J]. 长江信息通信, 2021, 34(5):9-11.
- [38] 苗北辰,郭为安,汪镭,等. 隐式特征和循环神经网络的多声部音乐生成系统[J]. 智能系统学报, 2019, 14(1):158-164.
- [39] 毛勇华,桂小林,李前,等. 深度学习技术应用研究[J]. 计算机应用研究, 2016, 33(11):3201-3205.
- [40] NG W W, ZENG W, WANG T. Multi-level local feature coding fusion for music genre recognition[J]. IEEE Access, 2020, 8: 152713-152727.
- [41] 陈景霞,王丽艳,贾小云,等. 基于深度卷积神经网络的脑电信号情感识别[J]. 计算机工程与应用, 2019, 55(18):103-110.
- [42] DHANITH C, LAKSHMAN J. Automatic music genre classification of audio signals with machine learning approaches [J]. Journal on Computing, 2013, 3(2):1-12.
- [43] LIU C, FENG L, LIU G, et al. Bottom-up broadcast neural network for music genre classification[J]. Multimedia Tools and Applications, 2021, 80(5): 7313-7331.

## A Survey of the Research on Music Genre Classification in the Field of Music Information Retrieval

SONG Yang, WANG Hai-long, LIU Lin, PEI Dong-mei

(College of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010022, China)

**Abstract:** The paper explores the progress and current status of music genre classification (MGC) research, introduces the basic concepts of music genre classification and process of it; reviews machine learning and deep learning methods for music genre classification; and summaries relevant music genre datasets. It also analyses and makes prospect on the future development of music genre classification methods in light of the progress and shortcomings of the current stage.

**Key words:** music information retrieval; music genre; music classification; classification method

【责任编辑 张颖娟】