

多模态AI技术在河北传统音乐非遗数字化采集与标注中的应用研究

赵明¹ 李鑫²

1.燕山大学 河北秦皇岛 066000

2.秦皇岛职业技术学院 河北秦皇岛 066000

摘要：在信息化高速发展的今天，我们应当思考如何更好地保护及传承非遗文化，在众多非遗中，河北地区传统民歌作为中华文化的重要组成部分之一，我们应该用现代化的方式对其进行存储，从而更好地保护它。本文的研究目标在于利用多种人工智能方法对河北地区的传统民歌进行采集标注。分别从理论上论证、实践上的对比分析及试验结果三个方面来说明这种技术手段对音视频及文本等多种媒体信息的优势特点，并且对于智能标签系统的设计、知识图谱的应用等创新性的设计尤为突出。

关键词：多模态AI技术；河北传统音乐；非物质文化遗产；智能标注

一、多模态AI技术在河北传统音乐非遗数字化采集与标注中的应用意义

数字技术可以提供一种新的非物质文化遗产保护模式：通过运用多模态人工智能技术，融合声音、图像、文字等多种信息载体，对非物质文化遗产进行完整采集及精确描述。比如运用AI音频识别技术就可以精准地区别出传统音乐的旋律、节拍及音色。视频分析法能实现对动作表情的实时捕捉，可加速非遗类数据采集，并通过智能标记系统进行分类存储和快速检索，为后续研究保护提供基础资料。

当前非遗数字化保护工作已由单纯的信息采集向多形态并存发展以及智能应用方向迈进，在这个过程中，人工智能扮演着关键的角色，借助声纹识别、图像识别、机器学习等多种人工智能技术手段，它能帮助人们更好地挖掘非遗中蕴含的更多方面的信息，并进行进一步分

析。例如，在研究河北地区传统音乐文化的过程中，借助多模态人工智能同时观看音频轨道中的曲调变化及舞台画面中的人物运动状态，从中发现其中潜在关联，这为非物质文化遗产的发展提供理论依据。

但与此同时，运用多模态人工智能进行非遗数字化还面临数据完备性及准确性、标注标准性及智能化、应用中的伦理困境等问题亟待解决。本文试图就基于多模态人工智能实现河北传统音乐类非遗的数字化采集和智能化标注开展研究，并提出创新实践方案。这就给我们非遗保护提供了新的思路及方法。通过研究发现，在应用多模态人工智能（AI）技术的过程中能够有效提升对非遗的信息采集效率以及精确度，为实现非遗的活态保存提供一定的技术支持；同时在研究过程中也发现该技术所存在的伦理以及信息偏见等问题，并提出相应的解决办法。本课题的研究既为运用高新技术保护非遗提供了理论指导，也对实践应用层面的措施进行了有效引导，有利于实现传统与科技的深度融合。

二、多模态AI技术概述

人工智能技术指的是集合AI技术中的多个技术种类包括计算机视觉技术、语音识别技术和自然语言处理技术等来共同分析与整合声音、图像以及文字等多种类型的信息资源的技术手段，在非遗信息化环境中可实现由单一数据采集到全面信息整合转变的过程。由此为非遗保护提供全方位的技术支撑。

音频信息是重要的多模态AI组成部分，其基于深度学习算法提取传统音乐中的曲调、节奏、音色特征，可以

课题项目：本文系2025年度河北省社会科学发展研究课题“音乐人类学与AI技术融合视域下河北传统音乐非遗传承模式的智能优化研究”（HBSKFZ25QN166）的研究成果；

作者简介：

- 赵明（1986.05-），女，满族，黑龙江五常人，博士学位，音乐学讲师，主要从事音乐学及美育方面的教学工作；
- 李鑫（1981.08-），男，汉族，黑龙江鸡西人，硕士学位，讲师，主要从事教育学。

作为音乐类文化遗产数字化的基础性手段。例如针对河北梆子剧种研究，可以将该戏曲特有的唱腔以及伴奏进行精确表征，进而为后续的乐谱分析及保存提供数据支持。

通过影像记录方式，可以实现对非遗传承人肢体动作与面部表情的即时性捕捉，将传统技艺的活态表演转化为可量化的数字资源，为非遗项目的数字化保护与传承奠定基础。例如，在对北戴河渔歌号子进行表演解析时，借助多模态视频分析技术，可以精确跟踪领唱者与合唱者在模拟拉网、摇橹等劳动情境中的身体姿态变化。这种数字化的动作复原，不仅有助于深入探究号子节奏与劳动强度、呼吸律动之间的内在关联，更能揭示其背后蕴含的海洋文化特征与坚韧的劳动精神，从而为非遗技艺的科学化传承提供数据支撑。

用文本挖掘的方法对涉及非物质文化遗产的各种资料进行智能标注和分析，比如诗词歌赋、戏曲曲艺、民间故事、地方志书、学术论文等，在自然语言处理的基础上将非物质文化遗产信息实现结构化存储及检索，为研究带来极大的便利性。例如在研究河北省民歌时，这样就通过文本挖掘的方式找到了其中蕴藏的历史文化信息及相关的资料，可以对我们研究了解这些音乐起到很大的帮助作用。

多模态人工智能的应用为我们实现了非物质文化遗产跨模态关联与深度挖掘，在音频、视频以及文本信息相互结合的基础上，多模态人工智能可以发现非物质文化遗产内部的主要交互关系，为非物质文化遗产活态传承提供依据。例如在河北省传统音乐的研究中，我们借助于人工智能对声高变化和表演动作变化进行并行分析以识别其协同关系，并基于此产生了戏剧表现方式创新的设计想法。

三、河北传统音乐非遗的数字化采集

在数据多样性以及复杂性的环境下，河北民间音乐采集正面临尴尬的局面，在这种情况下AI技术为我们提供了多方位的声音信息采集方式，并可结合图像、文字等方式帮助非遗保护。其次，在采集声音资料方面，可以借助于专业的音响设备以及智能化的声音技术，将传统音乐实现有效的记录。

例如，在我们研究冀中笙管乐的过程中，使用音频采集的方式能够获得器乐表演形态以及曲调变化等对我们深入了解和保护音乐的基本资料；但音频采集所面临的环境背景噪音干扰以及录音设备限制的问题，通过引入多模态人工智能技术，则比如采用降噪技术、自适应录音技术等可提升声像档案信息质量。

动作捕捉技术能为非遗项目提供即时性、活态化的信息记录与量化处理。以昌黎民歌的表演研究为例，动作捕捉相机虽能追踪歌手在演唱时的肢体律动与细微表情，为传统文化的数字化传承提供可视化辅助，但其在实际应用中极易受光照环境及拍摄角度的制约。相比之下，多模态AI技术通过融合图像处理与智能摄像机感知，能够有效突破复杂环境限制，显著提升视频信息采集的精度与质量，实现对昌黎民歌表演形态的全方位、高保真记录。

可以通过文本抓取的方式，在非遗相关的文本书籍、歌词、剧本等进行智能分析与标注，运用自然语言处理技术将非遗相关信息进行结构化存储，并快速检索到这些信息。例如：当我们要了解河北省传统民歌歌词的内容时，文本收集能够帮助我们了解其蕴含的历史背景和文化暗示，为我们的音乐文化阐释提供强有力的支持。但文本收集面临不同类型的文献以及丰富的语言环境，而多元模式人工智能借助于文本分级技术和实物辨识技术提高了对文本信息的处理效率。结合多种模式人工智能技术的应用，可以对非物质文化遗产进行全方位搜集以及深层次链接。将声音、影像及文字信息相结合，多模式人工智能可帮助人们发现非遗之间的内在联系，并为非遗的活态保护提供依据。

（一）多模态数据采集与智能分类

河北省的传统艺术形式中有着大量民歌、戏曲、器乐表演等形式，需要用多种人工智能技术对这些传统音乐文化进行全面采集。采用高清摄像机、三维测量设备、声音记录设备进行表演过程记录；采用传感器设备对细微的演奏动作进行捕捉。人工智能可自动识别各类音乐形式，如区分河北梆子中出现的板式变化及与评剧不同的唱法等。借助机器学习的方法分析海量的影音材料建立河北省传统音乐数据库。在采集过程中，人工智能能够主动规划收集流程，在音频频率上自主调整机器配置参数以保证音频质量，比如在冀中地区交响乐收集的过程中，这种人工智能就能自动区分不同乐器的搭配方式，从而提升录音效果。这就是一种机器学习分类法。这样工作效率就很高了，并且对我们以后的研究提供可靠的数据。

（二）AI辅助的音乐特征提取与标注

传统乐谱书写具有较高的专业性并且耗费大量精力，AI则可以极大优化这一过程。采用深度神经网络对音轨频谱图、节奏特征、曲调结构进行识别提取基本特征。本文使用机器学习算法对河北民歌进行旋律线提取，并给它加上了调性、节拍形式等等。同样对于戏剧来说，AI也可以很好地识别歌唱中的延长音以及装饰音，并进

行自我打板眼及声类划分。

例如,在对河北梆子的音乐分析过程中,人工智能能够准确发现它所具有的“大慢板”的节奏特点。利用自然语言处理的方法将其学术化标签标注出来并发现一些听觉上不容易被发现的旋律特征,为研究提供新思路。这种智能标签化的方式还保护了古乐的形式。

(三) 智能内容分析与价值评估

人工智能技术可以更好地对河北传统音乐文化内涵及艺术特色进行发掘分析。应用人工智能技术评估其音乐史的价值以及音乐文化的艺术特点。也可以通过对典型歌曲的主题分析,得出该作品所反映的时代背景及其人民的生活情感表达方式;还可以从戏剧表演中分析其技艺传承性价值。AI还能辨别音乐中有哪些创新的部分并判断其与传统的关联性,在此基础上运用大数据分析建立价值判断模型来指导保护行为。这样的方法可以对保护工作做出明确的规划。

(四) 多模态AI驱动的音乐知识图谱构建

构建河北传统音乐知识库需要收集不同类型的资料信息,并运用AI技术将这些文字、图像、音频以及视频等整合成一个有机的数据库系统。首先应用NLP方法提取相关文献中的人物、故事、流派等信息。其次,利用计算机视觉技术对乐器图像及使用场景进行分析,识别乐器的外观特征;最后,采用音频信号处理技术对乐曲进行分析,检测出乐曲中的音符以及和弦信息。最后,利用知识图谱对以上资料进行整合,建立实体之间的关联,如将冀东地秧歌的演员、伴奏乐器、乐曲以及相关历史结合起来形成一张图谱,并可以基于该图谱开展检索及进一步的研究工作,从而实现传统音乐的有效推广。

四、河北传统音乐非遗的智能标注

具体而言,是指应用人工智能进行智能标注,并在机器标签中引入人工干预,从而达到非遗资料化保存及快速检索的目的。在音乐本体标注方面,可采用声纹标识记录作品的节奏、曲调、音阶等属性;运用机器学习方法可对海量非遗音频进行自动识别分类。可利用多模态AI中的迁移学习策略及多任务网络框架,协同分析音频信号与演奏者的指法动作,从而优化标签的准确性与专业性。

针对具有较强表演性的传统音乐非遗,可以使用视频标记对表演者的动作姿态、表情及情绪进行详细描述,结合肢体运动与面部表情感知技术,实现艺术魅力的完整复现。例如,在昌黎民歌的数字化保护中,利用多模态技术对表演者在演唱“秧歌调”时的扭动姿态、标志

性的面部神情进行同步记录与标注。这不仅能保存其独特的艺术形态,更能为昌黎民歌“声舞一体”的技艺传承与评价体系提供科学的数字化依据。

视频标注的优点是基于多模态的人工智能技术可以通过使用3D姿态估计以及情感计算算法来提高视频标注质量。虽然视频标注存在着动作复杂性、面部表情微妙标注难等问题,但是这都是它的优点。借助NLP技术中的文本标注能力可以智能地将非遗相关的图书、歌曲或者戏剧等进行标签化处理,在最短的时间内实现对非遗相关内容的检索,比如在河北省民歌中挖掘出其中隐含的信息和历史人文信息等内容。这对于我们进一步探索音乐世界有极大的帮助。基于上述对文字标记存在的问题:语言的变化性以及专业的领域的知识,使用多种模式人工智能技术进行知识图谱构建、语意分析,可以将应用深化拓展。通过这种综合式的人工智能技术,我们可以做到非物质文化遗产的信息智能标注、知识关联,将音频、视频及图片等内容转换成数据标签形式,并统一梳理管理,而多模态的智能技术也能够分析非遗内容之间的关联性,为其后续的发展起到一定帮助作用。

结束语

通过融合多模态人工智能技术,对河北地区传统文物开展数字化采集与标注工作,为文化遗产保护开辟了创新技术路径。在音频特征、影像资料及文本数据等多元信息维度上实现精度增强与品质升级,最终达成对我国传统文物的动态化长效保存目标。综上所述,在强调人工智能技术应用中存在的一些道德隐患及信息偏差的同时,本文提出相应的解决方案。相信伴随着人工智能的发展,其对非遗数字保存领域的影响会越来越大,并能推动相关交叉领域的研究与发展空间进一步扩大。

参考文献

- [1]张莹莹.非物质文化遗产进入美术课程资源系统的研究[D].首都师范大学,2023.
- [2]赵云海,刘瑞.数字化时代非物质文化遗产知识产权保护实践反思[J].文化遗产,2023(02):10-18.
- [3]胡红.“非遗”大潮下的传统表演艺术回眸——以音乐、舞蹈为例[J].四川戏剧,2024(06):91-93.
- [4]岳永逸.本真、活态与非遗的馆舍化——以表演艺术类为例[J].民族艺术,2024(06):79-87.
- [5]曾心昊.论我国表演类非物质文化遗产的博物馆化保护[J].文物鉴定与鉴赏,2023(09):126-130.