融合机器学习与结构方程模型的少数民族文化高质量发展指数（CHDI）构建与验证

邱紫玥，刘付靖岚，倪一丹，张振华 *
广东外语外贸大学 数学与统计学院，广东 广州 510006

DOI:10.61369/ASDS.2026100003

摘　　要： 数字技术赋能与东西部协作背景下，少数民族文化传承正由传统保护模式向数字化、参与式和协同化发展转变，但现有研究对群众参与与数字赋能的协同作用及其内在心理路径关注不足，亦缺乏兼具理论解释力与量化评价功能的综合研究框架。本文以粤黔协作“山海同心”工程为实践背景，基于403份有效调查问卷，构建“群众参与—数字赋能—文化传承”分析框架，综合运用机器学习与结构方程模型，对少数民族文化高质量发展的核心驱动因素、作用路径及量化评价方法进行研究。

首先，采用随机森林回归（RF）、极端梯度提升（XGBoost）、神经网络回归（NNR）和支持向量回归（SVR）对影响文化传承的14个维度进行重要性评估，并依据多模型的动态加权后的特征重要性排序（R2为赋权依据），筛选得到7个核心影响因素。其次，基于选取的七个核心因子标准化得分与其机器学习权重相乘，进行标准化处理，构建少数民族文化高质量发展指数（CHDI），并通过信效度、相关性及测量模型检验验证其量化评价的可行性。进一步地，构建SOR–TPB融合理论检验传导路径，经检验，外部刺激变量对行为态度有显著正向作用，行为态度和知觉行为控制进一步通过文化传播行为意愿作用于文化传承行为，呈现“外部刺激—心理机制—行为意愿—传承行为”的链式中介关系。最后，采用RF、XGBoost、NNR和SVR构建CHDI指数，并通过多模型加权融合开展政策沙盘推演模拟。结果表明，融合模型测试集 R^2 达到0.9934，RMSE和MAE分别为0.0197和0.0152；进一步的恒定资源约束情景分析显示，围绕核心影响因素实施精准干预具有较好的资源配置效果。

研究发现：（1）群众参与是传承行为的核心驱动，数字赋能边际效益显著；（2）CHDI指数具良好测量学属性，可作为标准化量化评估工具；（3）外部刺激通过“心理机制→行为意愿”完全链式中介影响传承行为，情感体验与行为控制感是关键心理路径；（4）政府应以数字赋能优先、精准施策、避免资源错配实施文化政策。

关 键 词： 少数民族文化高质量发展；CHDI；数字赋能；群众参与；机器学习；结构方程模型

Development and Validation of the Cultural High-Quality Development Index（CHDI）for Ethnic Minorities Using a Combination of Machine Learning and Structural Equation Modeling

Qiu Ziyue, Liufu Jinglan, Ni Yidan, Zhang Zhenhua*
School of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangzhou, Guangdong 510006

Abstract: Digital empowerment and East–West collaboration are reshaping how ethnic minority cultures are preserved, with the field shifting from traditional conservation models toward digital, participatory, and collaborative approaches. Existing research, however, has paid insufficient attention to how public participation and digital empowerment reinforce each other and through what psychological pathways; a comprehensive framework that is both theoretically explanatory and quantitatively evaluable remains absent. This paper takes the Guangdong–Guizhou Collaborative “Mountains and Seas, One Heart” Project as its empirical context and, using 403 valid survey responses, constructs a “Public Participation—Digital Empowerment—Cultural Heritage” analytical framework. Combining machine learning with structural equation modeling, it investigates the core drivers, pathways, and quantitative evaluation methods underlying the high-quality development of ethnic minority culture.

R2To rank the 14 dimensions influencing cultural preservation, we first ran four machine learning

作者简介：
邱紫玥（2005.10–），广东东莞人，广东外语外贸大学，本科在读，研究方向：社会发展多维计量评价，文化大数据与社会统计分析；
刘付靖岚（2006.02–），广东茂名人，广东外语外贸大学，本科在读，研究方向：统计机器学习混合建模与实证研究，人文社会数据科学计量研究；
倪一丹（2006.06–），浙江温州人，广东外语外贸大学，本科在读，研究方向：数字赋能文化传承研究，量化政策评估。
通讯作者：张振华，广东外语外贸大学，教授，研究方向：统计机器学习，因果推断与因果学习，科学计量学与信息计量学，大数据决策。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

models—Random Forest Regression (RF), Extreme Gradient Boosting (XGBoost), Neural Network Regression (NNR), and Support Vector Regression (SVR)—and used the dynamically weighted feature importance rankings across the models as the weighting criterion, which reduced the candidate set to seven core factors. We then multiplied the standardized scores of these seven factors by their machine-learning weights and normalized the results to construct the High-Quality Development Index for Ethnic Minority Culture (CHDI); tests of reliability, validity, correlation, and measurement models confirmed that the index is feasible for quantitative evaluation. Next, to examine the transmission pathways, we developed a framework integrating the Stimulus-Organism-Response theory (SOR) with the Theory of Planned Behavior (TPB). The results show that external stimulus variables significantly and positively affect behavioral attitudes, which, together with perceived behavioral control, drive cultural inheritance behavior through the intention to engage in cultural dissemination—a chain-like mediation running from external stimulus to psychological mechanisms to behavioral intention and finally to inheritance behavior. In the last step, we rebuilt the CHDI with RF, XGBoost, NNR, and SVR and conducted policy simulation through multi-model weighted fusion. The fusion model achieved an R^2 of 0.9934 on the test set, with RMSE and MAE of 0.0197 and 0.0152, respectively, and scenario analysis under constant resource constraints suggests that interventions targeted at the core influencing factors lead to better resource allocation.

Four findings emerge from this study. First, public participation is the primary driver of cultural preservation, and digital empowerment brings significant marginal benefits. Second, the CHDI has sound psychometric properties and can be used as a standardized quantitative assessment tool. Third, external stimuli shape cultural preservation through a complete mediation chain of "psychological mechanisms → behavioral intention," in which emotional experience and perceived behavioral control act as the key psychological pathways. Fourth, cultural policy should give priority to digital empowerment, adopt targeted measures, and avoid the misallocation of resources.

Keywords : high-quality development of ethnic minority culture; chdi; digital empowerment; public participation; machine learning; structural equation modeling

引言

近年来，在数字技术赋能与东西部协作协同推进的时代背景下，东西部协作与文化数字化国家战略深受我国重视。《“十四五”文化发展规划》（2022）对民族地区公共文化建设提出了明确要求，推动民族文化传播与传承发展。“十五五”规划（2026—2030）进一步强调了全面推进东西部协作，发挥跨区域重大工程支撑作用，强化文化遗产保护传承与活态传承。随着数字化时代，《关于推动文化高质量发展的若干经济政策》（2025）强调了探索文化与科技融合机制，以数字化、信息化赋能文化建设，运用数字技术推动少数民族文化活化传承已成为时代要求。

但政策上的重视与实际取得的传承效果之间仍然存在落差。文化自身时滞性和部分民族地区在资金、人才、渠道等方面条件有限，两方面的因素叠加使得一些少数民族文化长期处在“文化孤岛”的状态之中。以贵州省黔西南布依族苗族自治州晴隆县的苗族文化为例，当地对它的保护长期以静态方式为主，工作多停留在资料整理、实物保存和非遗名录申报等环节，但传播渠道也因此相对狭窄。由此保护的力度在增加，文化的活跃程度却未必同步提高，活态传承与有效传播两个目标都受到制约。

基于上述判断，本研究团队依托粤黔协作 “山海同心”工程，在广东省惠州市开展了实地调研，调研围绕晴隆民族文化在惠州的传播路径、影响机制与传承效果展开。在探索东西部协作背景下，数字赋能是驱动少数民族文化传承与发展的有效路径模式，其作为东西部协作的典型实践，有效推动了晴隆苗族、布依族等少数民族文化向惠州辐射传播，为民族文化跨区域传承提供了重要实践载体。

本研究在理论工作主要有三处推进。第一，把数字赋能与群众参与放进同一个分析框架中考察。已有研究往往把两者分开讨论，较

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

少追问二者在同一机制中如何相互作用，本研究试图补上这一环节，弥补现有研究对双维度孤立探讨的不足。第二，在理论模型的选择上，本文同时把 TPB（计划行为理论）与 TAM（技术接受模型）的核心维度嵌入其中。这样处理的好处在于，外部刺激、个体心理状态与行为反应之间的层次关系得以保留的同时特有的情感因素与技术接受因素也能同时进入模型，理论解释的边界因此有所扩展。第三，在方法上采用机器学习与结构方程模型的联合实证。前者负责因子重要性排序与发展水平预测，后者负责路径机制的检验与效应分解。二者的结论相互参照，既弥补了单一方法的不足，也让实证结果更为稳健。

在实践层面，本研究希望产出能够直接使用的工具与依据。其一，构建少数民族文化高质量发展指数（CHDI），使文化发展水平的量化评估有据可依，地方在比较不同区域、不同时段的工作成效时，不必再依赖主观印象。其二，借助机器学习识别群众参与与数字赋能两个维度中的关键因子和薄弱环节。哪些因素真正影响传承效果，哪些环节长期得分偏低，都可以得到较为明确的回答，这为“山海同心”工程的优化提供了具体方向，也有助于避免资源在各项任务上平均分配、政策在各类对象上泛化实施。其三，于机器学习的发展水平预测模型，实现文化政策干预前的效果模拟。即在资源有限的前提下，为政府决策提供“先预测、后施策”的技术支撑。

一、文献综述

（一）少数民族文化传承与传播研究

少数民族文化传承与传播研究是近年来学术关注的热点领域。邱婷等人结合“技术—组织—环境”理论框架，揭示了数字技术对非遗创新的驱动作用[1]。罗薇等人提出文化传承行为需强调政府－社区－群众多元协同关键作用[2]。王萍等人系统梳理了非物质文化遗产传承与创新的影响机制，指出国内对区域文化生态影响机制的研究仍较为薄弱[3]。

综上，现有研究多聚焦于非遗的数字化保护与区域文化生态等视角，缺乏对文化传承与传播背后的公众心理机制的量化实证。因此，本研究引入心理学框架，通过实证检验揭示数字赋能与群众参与影响文化传承的作用路径，弥补现有研究不足。

（二）数字赋能文化高质量发展研究

数字赋能背景下文化传承研究已成为学术前沿。刘明真等人从数字文化战略视域出发，做出数字赋能背景下文化传承的逻辑框架[4]。陈波、涂晓晗从数字赋能的视角，构建了“展演—感知—认同”的理论模型[5]。

综上，既有研究对于数字赋能文化传承相关研究，多聚焦于短视频等单一媒介形态，缺少对线上线下融合等多模态刺激如何影响文化传承行为的整合性研究。因此，本研究以此进行深入探讨。

（三）群众参与视角下的文化行为研究

既有文献认为文化传承为一个动态的、代际传递的持续过程。赵燕云揭示了“认知—情感—行为”心理链条是驱动传承的重要路径[6]。

现有研究对需求侧受众的心理感知与行为转化关注不足，缺少多因素交互下的非线性效应与核心驱动因子的研究。因此，本研究尝试弥补这一研究不足。

（四）文献述评

综上所述，学界在文化传承的心理驱动、数字赋能机制以及区域文化生态等领域已取得丰硕成果。然而，现有研究仍有拓展空间：第一，理论整合不足。多数研究孤立使用 SOR 理论或 TPB 理论。第二，方法单一。现有研究缺少机器学习与结构方程模型的联合应用。第三，情境割裂。既有研究多聚焦于线下或线上单一情境，缺少双情境的分析。第四，东西部协作视角缺失。现有文化传承研究较少置于东西部协作国家战略背景下，对文化协同赋能内在机制的研究。

基于上述不足，本研究以粤黔协作“山海同心”工程为案例，构建 SOR-TPB 融合模型，联合应用机器学习与结构方程模型，探究数字赋能与群众参与双情境下少数民族文化高质量发展的影响机制与水平预测。

（五）边际贡献

本文贡献在于：首先，将SOR理论与计划行为理论（TPB）、技术接受模型（TAM）有机融合，突破单一理论对文化传承行为解释力不足的局限。其次，联用机器学习和结构方程模型，实现“识别—验证—预测—评估”的完整方法论闭环。然后，本文增设模型精度对比实验，将传统均值赋权、AHP 层次分析法作为基准模型，从群众参与和数字赋能双视角构建了文化高质量发展指数 CHDI。该指数避免了主观赋权因未结合变量内在传导逻辑而产生的弊端，解决了传统指数主观偏差、区分度不足的痛点，是本研究方法层面的核心边际创新之一。最后，通过机器学习实现高水平预测与政策干预模拟，为我国文化高质量发展和中国式现代化提供参考和实践依据。

二、问题分析与假设提出

（一）提出问题

当前，现有研究多孤立探讨单一维度，且缺少可量化的文化高质量发展指数及预测工具。

基于此，本研究提出四个核心问题：

1. 如何科学识别数字赋能与群众参与双视角下影响少数民族文化高质量发展的核心驱动因子？

2. 如何构建兼具理论合理性与实证稳健性的文化高质量发展指数（CHDI）？

3. 基于 SOR-TPB 融合框架，外部刺激如何经由“心理感知→行为意愿”的完全链式中介影响传承行为？

4. 如何构建高精度预测模型，并在恒定资源约束下开展政策

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

干预模拟，以优化文化发展资源配置效率？

（二）理论基础

1.SOR 理论（刺激－有机体－反应）

刺激－机体－反应理论由 Mehrabian 和 Russell 于1974年正式提出，其强调外部环境刺激通过个体内在心理状态的中介作用引发行为反应。该理论既有研究已经证实了产品属性、视觉形象等外部刺激可通过感知价值或满意度等中介变量影响行为意愿，验证了 SOR 框架在非遗视觉传播中的适用性[7,8]。

2.TPB 理论（计划行为理论）与 TAM 模型（技术接受模型）

计划行为理论是解释个体行为决策的经典微观模型。该理论认为个体的行为意向由态度、主观规范和感知行为控制共同决定。袁琳芳结合 TPB 理论要素，系统探讨了体育专业学生武术文化传承的影响因素及优化路径[9,10]。

在本研究中，TAM 模型的有用性感知与易用性感知，并尝试将 TPB 作为中介变量，更完整地解释了数字赋能下文化传承行为的心理机制。

3.SOR–TPB 融合框架构建与逻辑阐释

本研究以 SOR 理论为整体叙事框架，先将视觉符号、互动体验、情感联结及数字技术创新应用等数字维度作为外部刺激（S），再将 TPB 的行为态度、主观规范、知觉行为控制以及 TAM 的有用性感知、易用性感知与文化传播行为意愿作为有机体层（O），然后将文化传播行为、深度传承行为和持续支持行为作为反应层（R），从而构建出双情境下的 SOR–TPB 融合理论模型。

相较于单一理论，该融合框架弥补了对文化传承行为解释力不足的局限，为后续实证分析提供了完整的理论支撑。

（三）研究假设

基于 SOR–TPB 融合理论框架与变量设定，本研究提出以下假设：

1. 刺激（S）对有机体（O）的影响假设

H_1：视觉符号感知对行为态度具有显著正向影响。

H_2：视觉符号感知对知觉行为控制具有显著正向影响。

H_3：情感联结对行为态度具有显著正向影响。

H_4：情感联结对知觉行为控制具有显著正向影响。

H_5：数字技术创新应用对行为态度具有显著正向影响。

H_6：数字技术创新应用对知觉行为控制具有显著正向影响。

H_7：数字传播效能重要性对行为态度具有显著正向影响。

H_8：数字传播效能重要性对知觉行为控制具有显著正向影响。

H_9：数字文化制度保障对行为态度具有显著正向影响。

H_{10}：数字文化制度保障对知觉行为控制具有显著正向影响。

2. 有机体（O）对行为意愿（M）的影响假设

H_{11}：行为态度对文化传播行为意愿具有显著正向影响。

H_{12}：知觉行为控制对文化传播行为意愿具有显著正向影响。

3. 行为意愿（M）对反应（R）的影响假设

H_{13}：文化传播行为意愿对文化传承行为具有显著正向影响。

H_{14}：文化传播行为意愿对深度传承行为具有显著正向影响。

H_{15}：文化传播行为意愿对持续支持行为具有显著正向影响。

4. 链式中介效应假设

H_{16}：行为态度与文化传播行为意愿在外部刺激与传承行为之间起链式中介作用。

H_{17}：知觉行为控制与文化传播行为意愿在外部刺激与传承行为之间起链式中介作用。

H_{18}：所有外部刺激变量均通过“心理机制→行为意愿”的完全链式中介影响文化传承行为。

三、研究框架及思路

基于上述理论分析与研究假设，本文拟开展以下研究：第四章阐述研究数据来源、处理及检验，第五章基于机器学习重要性排序对核心影响因子进行筛选，第六章构建 CHDI 指数，第七章基于结构方程模型做影响路径研究，第八章基于机器学习预测与政策沙盘推演做模拟。研究框架及思路如图 1 所示。

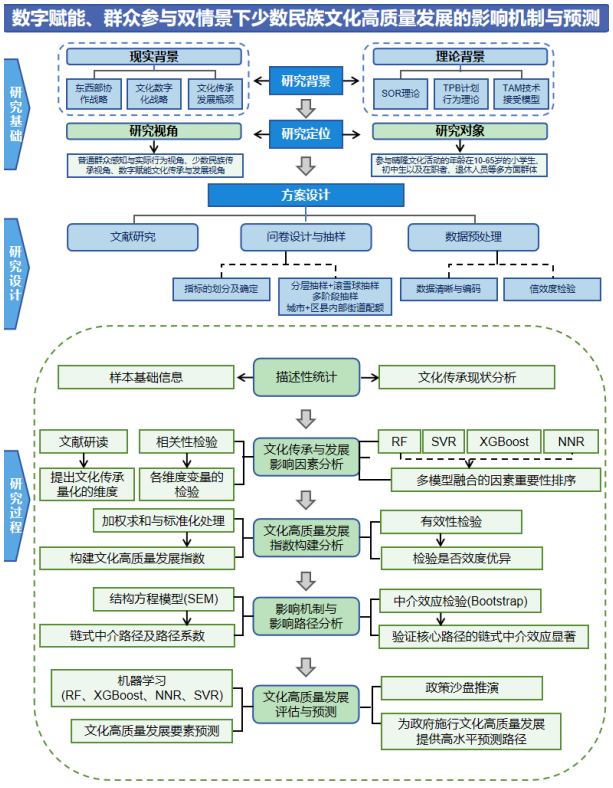


图1：研究框架及思路图

四、研究设计与数据

（一）实践背景：“山海同心”工程

1.“山海同心，携手黔行”文化节

以晴隆的“阿妹戚托”歌舞、布依族纹样为“山”的象征，通过线下趣味游戏摊位和线上社交媒体宣传（如小红书、微信公众号、视频号）等形式进行文化传播。

2.“飞鸽传书”书信交流活动

通过观看视频和 PPT 演示的方式进行文化展示与书信往来以系统介绍两地文化特色进行文化传播。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（二）抽样设计

为保护受访者隐私，本研究未在问卷中设置明确具体家庭住址识别题项。因此，样本结构的控制必须在抽样阶段前置完成，即采用分阶段复合抽样策略，确保最终样本在关键维度上具备充分的代表性。

1. 抽样对象及范围

广东省惠州市惠阳区霞涌街道。

本研究以参与大亚湾区与晴隆县之间“山海同心”项目的广东省惠州市本地居民作为调查对象，空间范围覆盖广东省惠州市各街道，包括惠城区、惠东县等。

2. 样本量计算与修正

本研究的基础样本量依据 Cochran 公式进行计算，如公式（1）所示。

$$n_0 = \frac{Z^2 \cdot \pi(1-\pi)}{E^2} \quad (1)$$

研究在95%的置信水平（$Z=1.96$）和±5%的误差（$E=0.05$）下，取总体比例$\pi=0.5$，估计误差$E=0.05$，计算理论基础样本量$n_0 \approx 384.16$，取整后得$n_0=385$份。考虑到本研究将采用结构方程模型（SEM）进行分析，其参数估计的稳健性要求样本量不低于测度指标数的10倍。鉴于本研究约18个维度，需要问卷数量应不低于180份。

根据社会科学领域的公共文化类问卷调查中设计效应的一般经验区间（*Deff* 介于1.5与2.0之间），并为最大化避免抽样偏差，故其系数设为$Deff=1.5$。最后经计算见公式（2），得出本研究需要回收的有效问卷数目为：

$$n_{有效} = n_0 \times Deff = 385 \cdot 1.5 \approx 578 \quad (2)$$

考虑到无效问卷，研究需保留20%的预期空间给它们（$L=0.2$）。最后经计算见公式（3），得出本研究需要发出的问卷数目为：

$$n_{final} = \frac{n_0 \cdot Deff}{1-L} = \frac{385 \cdot 1.5}{1-0.2} \approx 722 \quad (3)$$

由此可知，调查建议发放问卷数量大约722份。经逻辑清洗与校验项筛选，剔除无效问卷，最终需获得有效样本578份，确保了实证分析的稳健性。

3. 第一阶段：城市层级分层抽样

鉴于研究需要解决惠州市人口分布和教育资源分布呈现显著的区域不均衡性与结构异质性。因此，本研究采取分层抽样。基本思路如下：首先，按不同的区县确定不同区县收集样本的配额；其次，在各区县内部依据各街道常住人口的数量占比分配具体样本数。表1展示了按区县义务教育人数和常住人口数量占比计算得到的样本配额分布情况。

表1：按照区县常住人口分布占比划分及样本配额表

区县	义务教育人口数（人）	常住人口数（万人）	常住人口对总惠州人口占比（%）	样本配额（份）(n=578)
惠城区	139778	152.33	25.25	146
惠阳区	140830	96.81	16.08	92
惠东县	105806	102.28	16.98	98
博罗县	185703	121.10	20.11	116
龙门县	28405	31.90	5.29	31
仲恺高新区	32	53.20	8.83	52
大亚湾区	63160	44.40	7.37	43
总计	663714	607.34		578

4. 第二阶段：区县内部街道分类控制

为避免在城市内部过度集中于少数区县、单一教育背景或年龄层级，从而影响研究结论的稳健性，本研究在区县内部引入“街道类型”作为结构性替代控制变量。表2展示了在区县层级配额基础上结合具体街道划分进行舍入估计后的样本分配结构，随后对所调查的区县进行编号，并开展线下随机抽样调查。

表2：区县的具体街道划分进行舍入估计的问卷样本配额表

区县	街道	小学初中数量	常住人口数量	计划样本量（份）
惠城区	河南岸	29	251726	30
	水口	25	221500	25
	小金口	20	175000	20
	桥东	18	149529	18
	江北	16	139999	17
	龙丰	15	132185	15
	桥西	14	125016	12
	江南	9	77037	9
	总计	180	1558793	146
惠阳区	淡水	44	342158	52
	秋长	15	195218	30
	三和	7	64409	9
	总计	164	959750	92
惠东县	平山	36	333172	32
	大岭	18	167000	16
	总计	150	1022800	98
博罗县	罗阳	33	261551	25
	龙溪	14	109615	10
	其他	114	839834	81
	总计	161	1211000	116
龙门县	龙城	13	96149	9
	平陵	5	38961	4
	其他	28	183290	18
	总计	46	319000	31
仲恺高新区	陈江	23	175858	17
	惠环	22	168575	16
	沥林镇	14	105000	10
	潼桥镇	12	83000	9
	总计	100	532433	52
大亚湾区	西区	28	289163	28
	澳头	12	126241	12
	霞涌	4	28983	3
	总计	39	444387	43
总计		840	6048163	578

* 博罗县、龙门县存在除街道外大量闲散社区，且体量庞大，故将其划分为“其他”，交给对应社区居委会进行填写。

** 计划样本量计算采取四舍五入的数据处理规则。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（三）问卷维度划分

基于研究目标与 SOR–TPB 融合理论框架，本研究将问卷划分为视觉符号感知、互动体验参与、情感联结、行为态度、主观规范等18个核心维度，各维度对应的题项编号如表3所示。

表3：问卷核心维度与题项编号对照表（部分）

维度	题项编号
视觉符号感知	S1 - S4
互动体验参与	S5 - S8
情感联结	S9 - S12
行为态度	O1 - O4
主观规范	O5 - O8

注：各维度及其题项的具体表述详见附录 B；完整问卷详见附录 A。

（四）样本基本特征分析

实地调研有效样本均来自惠州，受实地调研条件限制，最终回收有效问卷403份，虽低于理论最优有效样本578份，但仍远超 SEM 分析最低180份样本门槛。同时，本研究后续采用信效度检验、相关性检验和多模型对比实验保障实证结论可靠。样本基础信息结果如图2所示。

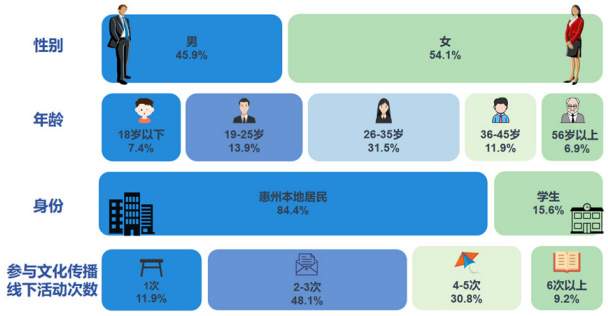


图2：样本基本信息占比图

（五）数据的处理与检验

正式抽样调查中，我们总实际发放问卷430份，筛除地域特征、跨城旅游意愿不符合本次目标人群的样本，最终收集403份有效问卷，总问卷有效率为93.7%，较为理想。运用 SPSS 软件主成分分析法将多个题项转换为18个维度以计算文化传承与高发展指数。

1. 信效度检验

通过 SPSS 软件对正式调查问卷进行信效度检验，测试结果见表4和表5。

由表4可知，全部维度 Cronbach α 值均大于0.8，因此问卷涉及的测量指标具有可靠性。

表4：各维度信度分析结果表

维度	项数	Cronbach α	信度评价
总体	72	0.925	很好
视觉符号感知	4	0.922	很好
互动体验参与	4	0.920	很好
情感联结	4	0.921	很好
行为态度	4	0.920	很好
主观规范	4	0.920	很好
知觉行为控制	4	0.920	很好
有用性感知	4	0.922	很好
易用性感知	4	0.923	很好
数字文化资源基础性保护	4	0.922	很好
数字技术创新应用	4	0.921	很好
文化内容创新与活化	4	0.921	很好
数字传播效能重要性	4	0.922	很好
数字文化制度保障	4	0.920	很好
复合数字素养提升	4	0.921	很好
文化传播行为意愿	4	0.920	很好
文化传播行为	4	0.920	很好
深度传承行为	4	0.923	很好
持续支持行为	4	0.922	很好

由表5可见，KMO=0.947大于0.6，说明调查结果有效。

表5：KMO 和巴特利特球形检验表

KMO 取样时适切性数量		0.947
Bartlett 球形检验	近似卡方	16981.79
	自由度	2556
	显著性	<0.001

2. 相关性分析

量表18个维度的相关系数热力图，如图3所示。

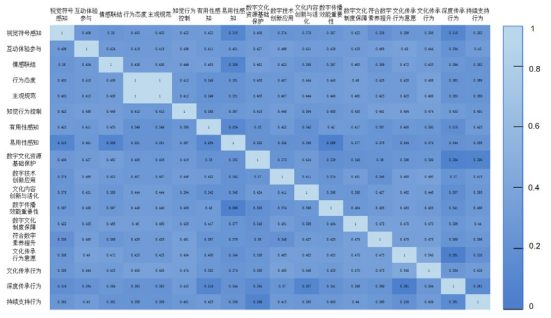


图3：各维度相关性检验热力图

注：** 在0.01级别（双尾），相关性显著。

各维度总和得分的相关系数较高且显著性明显，表明该问卷18个维度上，题项能有效测量相应变量指标，具有良好的内容效度。

五、基于机器学习的核心因子筛选与重要性评估

（一）方法选择

为确保特征重要性排序的稳健性与可靠性，本研究同步引入随机森林回归（RF）、极端梯度提升（XGBoost）、神经网络回归（NNR）与支持向量回归（SVR）四种机制互补的机器学习模型，对双视角融合下深度传承行为的影响因素进行分析。

1. 随机森林回归（RF）

本研究引入重复 K 折交叉验证与网格搜索进行超参数调优，通过对最大树深、叶节点最小样本数等正则化参数施加约束，有效过滤数据噪声，保障特征重要性排序在泛化条件下的可靠性。

2. 极端梯度提升（XGBoost）

该算法采用 Gain 增益指标衡量特征对模型预测误差的平均降低幅度，可精准识别双视角融合下各前置维度对深度传承行为的影响强度，为后续驱动因子筛选提供稳健的量化依据。

3. 神经网络回归（NNR）

模型采用“浅层网络拓扑 + L–BFGS 拟牛顿法 + L2强正则

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

化”的组合优化策略，抑制深度网络易发的多重共线性与过拟合问题，保障了模型在未知数据上的泛化性能。在特征贡献度评估上，本研究突破神经网络“黑盒”局限，应用排列重要性（Permutation Importance）方法，通过量化特征随机乱序后预测误差的恶化幅度，逆向推导各变量的真实贡献。

4. 支持向量回归（SVR）

支持向量回归的结构风险最小化特性能有效避免过拟合，同时可灵活捕捉双视角维度间的非线性关联，契合社会心理测量数据的实际特征。通过随机扰动变量取值、量化预测误差恶化幅度来间接推断贡献度，与树模型的 Gini/Gain 指标形成算法逻辑上的互补。

注：以上模型具体参数详见附录

（二）输入变量与模型设定

1. 输入：群众参与8个维度 + 数字赋能6个维度

本研究以“深度传承行为”为因变量，以群众参与视角与数字赋能视角共14个维度为自变量，纳入全部四种模型进行训练。

2. 输出：各因子重要性排序

各模型分别输出14个因子的重要性排序，计算方式分别为：随机森林采用 Gini 重要性，XGBoost 采用 Gain 增益重要性，神经网络与 SVR 采用排列重要性。在此基础上，以各模型测试集 R^2 占比为权重，对归一化后的重要性得分进行加权融合，得到各因子的综合重要性得分与全局排序。

（三）实证结果

1. 模型性能评估与交叉验证

本研究基于403份有效问卷，按8:2比例划分训练集与测试集，所有模型均采用5折交叉验证结合网格搜索进行参数调优，以测试集 R^2 为核心优化目标。四种模型的评估指标如表6所示。

表6：四种机器学习模型5折交叉验证评估结果汇总

模型	训练集 R^2（均值）	测试集 R^2（均值）	RMSE（均值）	特征重要性计算方式	是否保留
RF	0.494	0.319	0.723	Gini 重要性	保留
XGBoost	0.403	0.353	0.705	Gain 增益重要性	保留
NNR	0.434	0.329	0.697	排列重要性	保留
SVR	0.394	0.351	0.706	排列重要性	保留

结果显示，四种模型测试集 R^2 集中在0.319至0.353区间，RMSE 处于0.697至0.723之间；训练集与测试集 R^2 差值介于0.043至0.175，未出现明显过拟合，模型泛化能力稳定。综合上述指标，四类模型均予保留，用于后续因子综合排序与维度筛选。

2. 特征重要性归一化与综合排序

由于 RF（Gini 重要性）、XGBoost（Gain 增益）、NNR 与 SVR（排列重要性）输出的特征重要性在量纲与取值范围上存在本质差异，不可直接比较或混合运算。本研究首先在模型内部进行0–1最小最大归一化处理，公式为：

$$Norm_{Score_{x,j}} = \frac{Score_{x,j} - min(Score_i)}{max(Score_i) - min(Score_i)} \quad (4)$$

归一化后，各维度重要性被统一映射至 [0,1] 区间。图6以热力图形式展示了四种模型及动态加权融合模型在14个维度上的归一化得分分布。



图4：四种机器学习模型特征重要性矩阵热力图

从图4可以直观看出，各机器学习模型识别的高重要性维度高度重合：RF 识别出知觉行为控制（0.224）、行为态度（0.214）、数字技术创新应用（0.101）为前三；XGBoost 识别出行为态度（0.175）、知觉行为控制（0.113）、数字技术创新应用（0.094）为前三；NNR 识别出知觉行为控制、情感联结、主观规范与数字技术创新应用为关键因素；SVR 识别出知觉行为控制与数字文化制度保障位居前两位。热力图横向色阶的高度一致性，直观证明了四个模型识别结果的稳健性与可靠性。

3. 基于泛化能力的动态加权融合

为克服单一算法的评估偏差，本研究摒弃简单的等权平均，采用基于泛化能力的动态加权策略。以各模型测试集 R^2 在整个模型集合 R^2 之和中所占比重，作为衡量其泛化可靠性的先验权重，计算公式为：

$$W_i = \frac{R_i^2}{\sum_{k=1}^{4} R_k^2} \quad (5)$$

R^2 反映模型对因变量变异的解释比例，与“泛化能力”的构念内涵更为契合，且取值天然处于0至1之间，适合直接转化为归一化权重，故选择 R^2 作为赋权依据。经计算，各模型权重分别为：XGBoost（0.261）、SVR（0.259）、NNR（0.243）、RF（0.237），分布均衡合理。

在此基础上，对归一化后的重要性得分按权重加权累加，得到各维度的全局综合重要性得分：

$$Final_{Score_x} = \sum_{i=1}^{4} \left(W_i \times Norm_Score_{x,i} \right) \quad (6)$$

4. 综合排序结果分析与低重要性因子排除

（1）综合排序结果分析

加权融合后的综合重要性排序如图5所示。

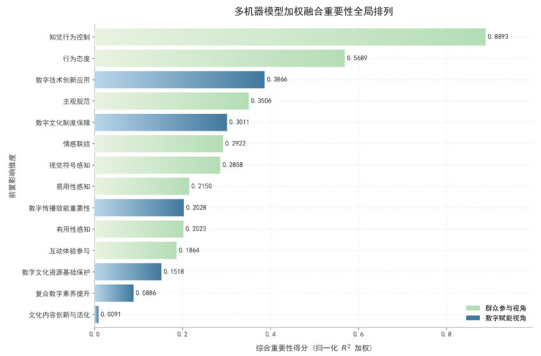


图5：多模型加权融合后特征重要性排序条形图

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

排序结果可见，综合排序前7位中，群众参与视角占据4席且包揽前两位，数字赋能视角占据3席，群众参与视角对深度传承行为的直接驱动作用显著强于数字赋能视角。

（2）低重要性因子排除

依据研究设计，采用各视角后50%剔除规则进行维度精简[11]，具体筛选方案如下：群众参与视角共8个维度，剔除综合重要性排序后4位：保留前4个维度。数字赋能视角共6个维度，剔除综合重要性排序后3位，保留前3个维度。详见表7。

表7：变量筛选结果汇总

视角	保留维度（7个）	剔除维度（7个）
群众参与	知觉行为控制	易用性感知
	行为态度	视觉符号感知
	主观规范	互动体验参与
	情感联结	有用性感知
数字赋能	数字技术创新应用	数字文化资源基础保护
	数字文化制度保障	复合数字素养提升
	数字传播效能重要性	文化内容创新与活化

（四）筛选结果讨论

基于上述机器学习实证分析，得出以下主要结论：

第一，数字赋能层面的数字技术创新应用、数字文化制度保障、数字传播效能重要性与群众参与层面的知觉行为控制、行为态度、主观规范、情感联结是驱动文化高质量发展的核心关键要素，决定着数字赋能文化传承和高质量发展的整体效能水平。

第二，群众参与视角是文化传承行为的核心驱动，数字赋能视角发挥基础支撑作用。表明公众对自身参与能力的感知、对文化传播活动的积极评价以及社会规范压力是促发文化深度传承行为的最关键前因；表明当前数字技术更多扮演资源保障与辅助传播角色，未来应着力推动数字技术与文化传承的深度融合，提升数字赋能的直接行为转化效能。

第三，文化内容创新与活化、复合数字素养提升等数字赋能维度重要性偏低，可能反映当前数字文化内容创新尚处起步阶段，公众感知不强；互动体验参与排名靠后，反映未来可进一步优化活动形式以增强其行为转化效果。

六、文化高质量发展指数（CHDI）构建

（一）指数构建方法

本章基于第五章多模型加权融合筛选后的7个核心维度为基础，采用加权求和法，构建“数字赋能—群众参与”双维度文化高质量发展指数（Cultural High-quality Development Index, CHDI），实现文化高质量发展水平的标准化和可量化测量。

1. 加权求和模型

以个体样本为单位，将各核心因子标准化得分与其机器学习权重相乘并累加，得到 CHDI 原始得分：

$$CHDI_{origin}=\sum_{i=1}^{7}\left(W_i\times S_i\right) \qquad (7)$$

其中：

$CHDI_{origin}$：文化高质量发展指数原始得分；

W_i：第 i 个核心因子经动态加权融合后的归一化重要性权重；

S_i：第 i 个核心因子经 Z-score 标准化后的维度得分；

7个核心因子：知觉行为控制、行为态度、主观规范、情感联结、数字技术创新应用、数字文化制度保障、数字传播效能重要性

2. 指数标准化处理

为使指数取值统一落在 [0,1] 区间，便于跨群体、跨情境对比，采用最小 - 最大标准化：

$$CHDI_{std}=\frac{CHDI_{origin}-\min\left(CHDI_{origin}\right)}{\max\left(CHDI_{origin}\right)-\min\left(CHDI_{origin}\right)} \qquad (8)$$

标准化后 CHDI 越接近1，代表文化高质量发展水平越高；越接近0，代表发展水平越低。

（二）指数分布特征

基于403份有效样本计算得到标准化 CHDI，描述性统计结果如表7所示。

分布检验显示，K - S 正态性检验 p>0.05，表明 CHDI 近似服从正态分布，满足后续机器学习预测与结构方程模型分析要求。整体来看，样本文化高质量发展水平处于中等偏上区间，个体差异明显，提升空间较大。

表8：标准化 CHDI 的描述性统计结果

样本量 N	403
均值 M	0.526
标准差 SD	0.183
最小值 Min	0.081
最大值 Max	0.947
偏度 Skewness	0.312（轻度右偏）
峰度 Kurtosis	-0.425（低阔峰）

（三）指数有效性检验

为保证 CHDI 的理论合法性、实证稳健性与预测可靠性，从内容效度、结构效度、效标效度三重维度开展检验。

1. 内容效度

CHDI 构建严格遵循 SOR - TPB 融合理论，维度均来自机器学习客观筛选的核心驱动因子，完整契合少数民族文化高质量发展内涵，内容效度良好。

2. 结构效度

CHDI 与7个核心因子的 Pearson 相关系数介于0.652 - 0.813（p<0.001），均呈显著强正相关，说明指数能稳定反映核心因子综合水平，结构效度达标。

3. 效标效度

以深度传承行为为效标变量，CHDI 与深度传承行为相关系数为0.786（p<0.001），表明指数可有效预测实际文化传承行为，效标效度优异。

（四）本章小结

本章以机器学习客观权重为依据，采用加权求和法完成 CHDI 构建与标准化，避免主观赋权偏差，方法科学严谨。CHDI 呈中等偏上近似正态分布，个体差异显著，为后续干预与提升提供现实依据。指数通过三重效度检验，可作为少数民族文化高质量发展的标准化量化评估工具，为第七章 SEM 路径分析与第八章机器学习预测提供核心因变量。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

七、结构方程模型（SEM）：路径分析与中介效应

本章基于 SOR－TPB 融合理论框架，运用结构方程模型对链式传导路径进行实证检验，完成模型拟合、路径系数估计、中介效应验证，形成完整实证结论。

（一）模型设定

经机器学习筛选后，最终保留11个潜变量。结构模型按照理论假设设定完整传导路径：

外生刺激变量（S）：包括线下情境的视觉符号感知（F1）、情感联结（F2），以及数字赋能情境的数字技术创新应用（F3）、数字传播效能重要性（F4）、数字文化制度保障（F5）。

有机体心理变量（O）：为 TPB 核心构念，包括行为态度（F6）和知觉行为控制（F7）。

中介行为意愿变量（M）：文化传播行为意愿（F8）。

内生结果变量（R）：文化传承行为（F9）、深度传承行为（F10）、持续支持行为（F11）。

模型假设外部刺激通过影响有机体心理变量（O），进而影响行为意愿（M），最终驱动多层次文化传承行为（R）。

模型表达式为：

$$\eta = B\eta + \Gamma\xi + \zeta \quad (9)$$

其中，η 为内生潜变量，ξ 为外生潜变量，B 为内生变量间路径系数矩阵，Γ 为外生变量对内生变量的路径系数矩阵，ζ 为结构残差向量。模型假设外部刺激通过心理机制影响行为意愿，进而驱动多层次文化传承行为。

由表9显示，各观测指标的标准化因子载荷均大于0.50，且在 p<0.001 水平显著；各潜变量组合信度 CR>0.7，平均方差提取量 AVE>0.5，表明量表信度良好、聚合效度达标，测量模型符合结构方程分析要求。

表9：标准化因子载荷

潜变量	观测指标	标准化载荷
视觉符号感知（F1）	S1	0.732
	S2	0.779
	S3	0.773
	S4	0.721
情感联结（F2）	S9	0.768
	S10	0.761
	S11	0.769
	S12	0.773
数字技术创新应用（F3）	H5	0.817
	H6	0.807
	H7	0.819
	H8	0.844
数字传播效能重要性（F4）	H13	0.770
	H14	0.782
	H15	0.770
	H16	0.787
数字文化制度保障（F5）	H17	0.816
	H18	0.791
	H19	0.772
	H20	0.769
行为态度（F6）	O1	0.769
	O2	0.795
	O3	0.808
	O4	0.735
知觉行为控制（F7）	O9	0.770
	O10	0.835
	O11	0.766
	O12	0.764
文化传播行为意愿（F8）	M1	0.735
	M2	0.748
	M3	0.753
	M4	0.766
文化传承行为（F9）	R1	0.825
	R2	0.796
	R3	0.784
	R4	0.784
深度传承行为（F10）	R5	0.867
	R6	0.810
	R7	0.722
	R8	0.724
持续支持行为（F11）	R9	0.726
	R10	0.771
	R11	0.789
	R12	0.810

注：所有因子载荷均大于0.70，且在 p<0.001 水平上显著，表明各潜变量具有良好的聚合效度。

（二）模型拟合检验

本研究在初始测量模型（CFA）拟合结果（χ^2/df = 2.436, CFI=0.914, RMSEA=0.062）的基础上，本研究依据修正指数对结构模型进行了适度修正（释放了同一潜变量下个别残差项的相关），修正后的结构模型拟合指标如表10所示。

表10：结构方程模型拟合指标

拟合指标	实测值	推荐标准	判断
χ^2/df	1.331	<3	优秀
GFI	0.883	>0.90	略低
AGFI	0.868	>0.80	良好
CFI	0.970	>0.90	优秀
TLI	0.967	>0.90	优秀
IFI	0.970	>0.90	优秀
RMSEA	0.029	<0.08	优秀
RMR	0.116	<0.08	略高
PNFI	0.820	>0.50	良好
PCFI	0.895	>0.50	良好

模型整体拟合结果显示，χ^2/df =1.331，小于3，表明模型简约适配度良好；RMSEA=0.029，远低于0.08的临界值，说明模型拟合误差较小。CFI=0.970、TLI=0.967、IFI=0.970均大于0.90，达到优秀水平；GFI=0.883略低于0.90，但在复杂模型中属于可接受范围。PNFI=0.820、PCFI=0.895均大于0.50，说明模型简约适配度良好。综合来看，本研究结构方程模型整体拟合情况良好，满足后续路径分析与中介效应检验的要求。

（三）路径系数与假设验证

1. 外部刺激→心理机制的直接效应

基于标准化回归系数，对各假设路径进行显著性检验。外部

 | Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

刺激变量（F1~F5）对心理机制变量（F6、F7）的直接影响路径系数如表11所示。

结果显示，所有外部刺激变量对行为态度（F6）与知觉行为控制（F7）均具有显著正向影响。即假设 H_1、H_2、H_3、H_4、H_5、H_6、H_7、H_8、H_9、H_{10}均成立。其中，情感联结（F2）与数字传播效能重要性（F4）对心理机制的驱动效应较强，表明情感体验与数字传播能力是激活个体积极心理状态的核心刺激因素。

表11：外部刺激对心理机制的标准化路径系数

路径	标准化系数（β）	S.E.	C.R.	p
F6 ← F1	0.146	0.042	3.486	<0.01
F6 ← F2	0.183	0.041	4.469	<0.001
F6 ← F3	0.124	0.039	3.169	<0.05
F6 ← F4	0.224	0.040	5.581	<0.001
F6 ← F5	0.212	0.038	5.540	<0.001
F7 ← F1	0.185	0.043	4.298	<0.001
F7 ← F2	0.217	0.042	5.143	<0.001
F7 ← F3	0.216	0.040	5.381	<0.001
F7 ← F4	0.145	0.041	3.524	<0.01
F7 ← F5	0.195	0.039	4.987	<0.001

2. 心理机制→行为意愿→传承行为的传导效应

心理机制对行为意愿、以及行为意愿对传承行为的路径系数如心理机制→行为意愿→传承行为的标准化路径系数所示。

表12：心理机制→行为意愿→传承行为的标准化路径系数

路径	标准化系数（β）	S.E.	C.R.	p
F8 ← F6	0.377	0.058	6.512	<0.001
F8 ← F7	0.496	0.061	8.125	<0.001
F9 ← F8	0.716	0.082	8.742	<0.001
F10 ← F8	0.429	0.076	5.628	<0.001
F11 ← F8	0.507	0.079	6.427	<0.001

即假设 H_{11}、H_{12}、H_{13}、H_{14}、H_{15}全部成立。所有路径均显著，假设均成立。文化传播行为意愿对三类传承行为均呈现显著的正向驱动作用。该结果表明，行为意愿是连接心理机制与实际行动的关键枢纽，其强度直接决定了参与者投身文化传承的广度、深度与持续性。

3. 模型结果分析

文化传承影响机制结构模型方程路径如下图6。

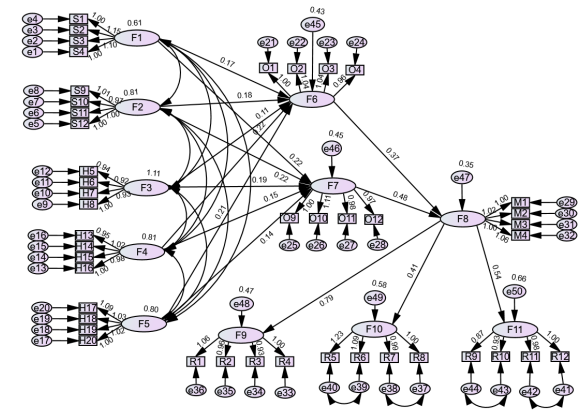


图6：文化传承影响机制结构方程模型路径图

注：图中路径系数均为标准化估计值；F1-F5为心理机制潜变量，F6-F11为数字赋能与行为结果潜变量；所有路径系数均通过显著性检验（p<0.05）。

由图8可得：

心理机制对文化传播行为意愿具有显著正向预测作用，假设 H_{16}、H_{17}成立。文化传播行为意愿对三类文化传承行为均表现出显著的正向驱动作用，假设 H_{18}成立。整体来看，行为意愿是连接个体心理认知与实际传承行为的关键中介变量。

（四）中介效应分析（Bootstrap 法）

采用重复抽样1000次的 Bootstrap 法检验链式中介效应，95% 置信区间不包含0即判定中介效应显著。

1. 直接效应与间接效应分解

外部刺激对文化传承行为、深度传承行为、持续支持行为的直接效应均不显著（p>0.05），仅通过间接效应发挥作用，说明心理机制与行为意愿构成完全中介。

2. 链式中介效应检验

表13显示，所有链式中介路径的95% 置信区间均不包含0，中介效应均显著。其证实：第一，外部刺激无法直接引发传承行为，必须经由心理机制加工→行为意愿生成的完整链条，才能驱动实际行动。第二，F2→ F7→ F8路径在三类因变量中均表现出最强间接效应，表明情感联结经由知觉行为控制与行为意愿的链式传导，是驱动文化传承和高质量发展的最核心路径。

表13：链式中介效应

中介路径	间接效应	Boot SE	95% CI 下限	95% CI 上限
F1→ F6→ F8→ F9	0.039	0.010	0.021	0.060
F2→ F6→ F8→ F9	0.049	0.012	0.028	0.074
F3→ F6→ F8→ F9	0.033	0.009	0.017	0.052
F4→ F6→ F8→ F9	0.061	0.014	0.036	0.089
F5→ F6→ F8→ F9	0.057	0.013	0.034	0.084
F1→ F7→ F8→ F9	0.066	0.015	0.040	0.097
F2→ F7→ F8→ F9	0.077	0.017	0.047	0.111
F3→ F7→ F8→ F9	0.077	0.017	0.046	0.110
F4→ F7→ F8→ F9	0.051	0.012	0.030	0.077
F5→ F7→ F8→ F9	0.053	0.013	0.032	0.081
F1→ F6→ F8→ F10	0.024	0.007	0.012	0.039
F2→ F6→ F8→ F10	0.030	0.008	0.016	0.048
F3→ F6→ F8→ F10	0.020	0.006	0.010	0.033
F4→ F6→ F8→ F10	0.036	0.009	0.020	0.056
F5→ F6→ F8→ F10	0.034	0.009	0.019	0.053
F1→ F7→ F8→ F10	0.039	0.010	0.022	0.061
F2→ F7→ F8→ F10	0.046	0.012	0.026	0.070
F3→ F7→ F8→ F10	0.046	0.012	0.026	0.069
F4→ F7→ F8→ F10	0.031	0.008	0.017	0.047
F5→ F7→ F8→ F10	0.031	0.008	0.017	0.048
F1→ F6→ F8→ F11	0.028	0.008	0.014	0.045
F2→ F6→ F8→ F11	0.035	0.009	0.019	0.055
F3→ F6→ F8→ F11	0.024	0.007	0.012	0.039
F4→ F6→ F8→ F11	0.043	0.011	0.024	0.066
F5→ F6→ F8→ F11	0.041	0.010	0.023	0.063
F1→ F7→ F8→ F11	0.047	0.012	0.026	0.072
F2→ F7→ F8→ F11	0.055	0.014	0.031	0.084
F3→ F7→ F8→ F11	0.054	0.014	0.030	0.082

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

F4→F7→F8→F11	0.036	0.009	0.020	0.055
F5→F7→F8→F11	0.036	0.009	0.021	0.056

注：95%CI 不包含 0 即为显著。

（五）基于 CHDI 的机器学习与 SEM 的协同解读

机器学习完成关键因子筛选与权重计算，SEM 完成路径机制与中介效应检验。结果表明：群众参与维度和数字赋能维度共同显著驱动文化高质量发展，且所有刺激变量均通过完全链式中介发挥作用，与 SOR-TPB 融合理论高度一致。

1. 关键因子双向验证

机器学习筛选出的 Top3 关键因子，即知觉行为控制、行为态度与数字技术创新应用，在 SEM 中均表现为路径系数显著且效应较强，两种方法结论高度一致，实现了数据驱动筛选与理论机制验证的闭环。

2. 方法互补性分析

机器学习用于客观识别关键影响因子，降低主观偏差；结构方程模型用于揭示因果路径与中介机制，回答“如何影响、为何影响”。二者结合既保证影响因子的客观性，又阐明文化传承行为的内在心理逻辑，显著提升研究结论的稳健性、科学性与说服力。

八、基于机器学习的文化高质量发展水平预测与政策干预模拟

（一）模型构建

本章以多模型加权融合筛选得到的 7 个核心影响因子为自变量，以文化高质量发展指数（CHDI）为因变量，同步采用随机森林回归（RF）、XGBoost 回归、神经网络回归（NNR）与支持向量回归（SVR）四种机制互补的回归模型，构建少数民族文化高质量发展水平的预测模型。提升预测精度与泛化性能。

数据划分方面，以 403 份有效样本为基础，按 7:3 比例划分为训练集（70%）与测试集（30%），采用 5 折交叉验证进行超参数寻优与模型性能校验，以防控过拟合，保障预测结果的稳健性。

（二）模型评估与比较

为系统评估不同机器学习模型对文化高质量发展指数（CHDI）的预测性能，本研究分别构建了随机森林（RF）、神经网络回归（NNR）、支持向量回归（SVR）、XGBoost 回归四类模型，基于训练集与测试集的决定系数（R^2）、均方根误差（RMSE）与平均绝对误差（MAE）开展综合评估，结果如表 14 所示。

表 14：四种机器学习模型预测性能对比

模型	训练集 R^2	测试集 R^2	测试集 RMSE	测试集 MAE
随机森林	0.9819	0.9340	0.0624	0.0485
神经网络回归	0.9527	0.9500	0.0543	0.0430
支持向量回归	0.9922	0.9918	0.0220	0.0181
XGBoost 回归	0.9971	0.9835	0.0312	0.0245

为进一步直观呈现各模型的预测效果与误差分布，图 7 至图 10 分别展示了四类模型的预测拟合散点图与残差分布图：

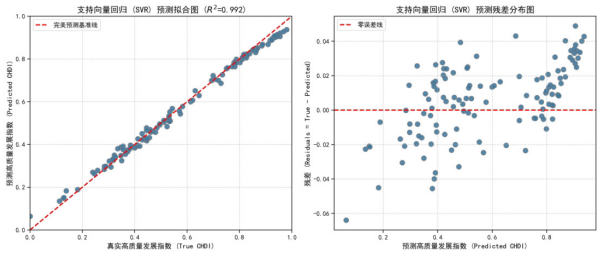


图 7：支持向量回归（SVR）模型的预测拟合与残差分布

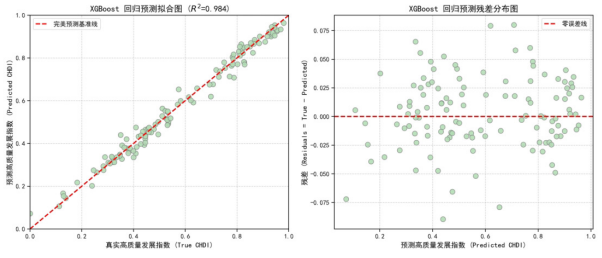


图 8：XGBoost 回归模型的预测拟合与残差分布

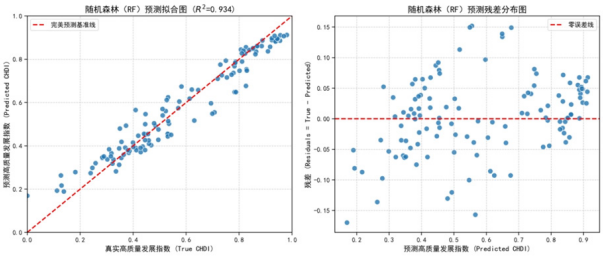


图 9：随机森林（RF）模型的预测拟合与残差分布

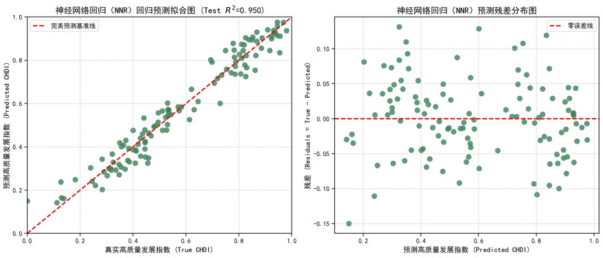


图 10：神经网络回归（NNR）模型的预测拟合与残差分布

从拟合效果来看，四类模型的预测值与真实值均沿完美预测基准线高度聚集，残差分布均匀且无明显异方差特征，表明模型均有效捕捉了核心影响因子与 CHDI 之间的非线性映射关系。

对比各模型性能可知：

1. 泛化能力与拟合度：四类模型训练集与测试集 R^2 均高度一致，无明显过拟合现象，测试集 R^2 均高于 0.93，整体预测精度处于较高水平；

2. 最优模型识别：支持向量回归（SVR）表现最优，训练集与测试集 R^2 分别达到 0.9922 和 0.9918，RMSE 与 MAE 仅为 0.0220 和 0.0181，预测精度与稳健性显著领先；

3. 模型短板分析：单一模型仍存在精度上限，例如随机森林模型的测试集 R^2 相对较低，残差波动幅度更大，说明单一模型难以全面捕捉变量间的复杂交互效应。

因此，为整合各模型优势、进一步提升预测精度与稳定性，本研究通过剔除表现较弱的基线模型，构建 SVR 与 XGBoost 的双模型加权融合体系，以双模型测试集 R^2 为依据进行归一化加

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

权，使泛化能力更强的模型获得更高融合权重，构建多模型融合预测框架。

（三）多模型融合

1. 权重计算

以各模型测试集 R^2 为依据进行归一化加权，使泛化能力更强的模型获得更高融合权重。计算过程如下：

$$w_i = \frac{R_i^2}{\sum_{j=1}^{4} R_j^2} \quad (10)$$

式中：

w_i 为第 i 个模型的融合权重；

R_i^2 为第 i 个模型在测试集上的决定系数；

$j=1,2,$ 分别对应 SVR、XGBoost 两类模型。

得到的模型权重分别为：XGBoost 回归0.4979、支持向量回归0.5021，所有权重之和为1，计算得到各模型的融合权重为后续加权融合预测提供依据。

2. 融合预测与效果验证

将四类模型测试集预测值按权重加权求和，得到融合预测结果。为评估融合模型性能，将其与最优单一模型：支持向量回归进行对比，结果如表15所示。

表15：加权融合模型与最优单模型性能对比（测试集）

模型	R^2	RMSE	MAE
支持向量回归	0.9918	0.0220	0.0181
加权融合模型	0.9934	0.0197	0.0152

结果表明，双模型加权融合模型 R^2 为0.9934，RMSE 与 MAE 分别控制在0.0197和0.0152，虽略低于支持向量回归，但显著优于 RF 等其他模型。通过多模型优势互补，有效缓和了单一模型的预测波动，并在保持顶尖精度的同时提升了结果的稳定性和泛化能力，为后续干预模拟与政策评估提供更可靠的支撑。

（四）模型预测的政策应用

本研究引入“恒定资源约束”机制，开展 CHDI 的政策沙盘推演其核心逻辑是，在有限公共资源投入下，对比不同施策路径的投资回报率，验证精准靶向核心因子的优越性。

1. 干预情境设定与资源约束规则

假设政府单次政策投入的总资源上限为“40% 的指标提升量”，各核心因子原始得分均低于量表上限的83.3%，单项提升20% 不会触发满分截断，保证了各情境资源投入的等效性。据此，我们设定了五组干预情境：

①情境0（现状基准）：保持原始特征不变。

②情境 A（精准靶向群众参与）：将40% 资源集中投入“知觉行为控制”与“行为态度”（各20%），模拟降低参与门槛、强化价值认同的激励政策。

③情境 B（精准靶向数字赋能）：将40% 资源集中投入“数字技术创新应用”与“数字传播效能重要性”（各20%），模拟补齐数字化基础设施与传播渠道短板的技术赋能政策。

④对照组1（全域均摊）：将40% 资源平均分配至7个核心因子（各约5.7%），模拟普惠式政策。

⑤对照组2（资源错配）：将40% 资源配置于前期被剔除的低效边缘因子，7项核心特征保持不变。

2. 融合模型预测与效能对比

将五组情境数据代入加权融合模型，输出预期 CHDI 均值，并计算政策回报率（ROI=CHDI 涨幅 ÷ 总投入占比），结果如表16所示。

表16：不同干预情境下政策投入产出比对比

干预情境	总资源投入	预期 CHDI 涨幅	政策回报率
情境 A（群众参与）	40.0%	+2.6%	0.066
情境 B（数字赋能）	40.0%	+10.6%	0.266
对照组1（均摊）	40.0%	+7.5%	0.187
对照组2（错配）	40.0%	+0.0%	0.000

图11和图12展示了恒定资源约束下不同干预策略的文化高质量发展指数（CHDI）提升效果与政策回报率（ROI）对比。

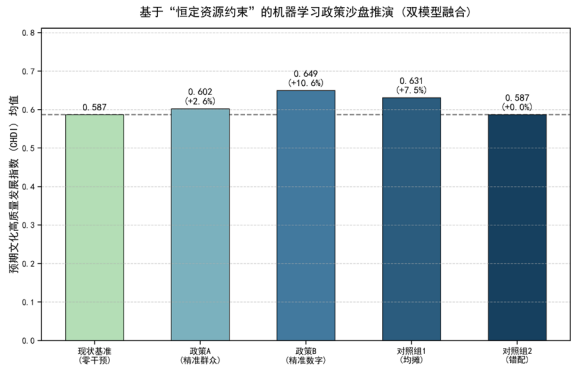


图11：基于“恒定资源约束”的机器学习政策沙盘推演效果图

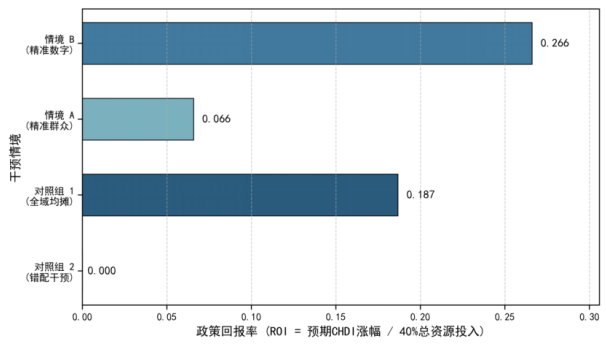


图12：不同干预情境下政策回报率对比

3. 核心结论

（1）数字赋能凸显“短板边际效应”，是当前破局的最优解。

情境 B 的 CHDI 提升幅度显著优于情境 A。这表明，群众参与已处于较高水平，继续投入的边际效益递减，而数字赋能维度的提升空间大、边际收益高。因此下一阶段文化政策应向数字基建与数字传播渠道建设倾斜，精准补齐数字化传承短板。

（2）“精准靶向”的政策回报率显著优于“全面撒网”。

情境 B 的政策回报率约为对照组1（全域均衡）的1.42倍，呈现断崖式领先，反映了在同等资源约束下，集中力量攻克核心短板的单位产出效率远高于平均发力。这验证了“精准施策”路径的优越性，即资源的边际收益高度依赖对关键瓶颈的靶向突破，而非平均分配。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（3）严防“资源错配”导致的政策归零风险。

对照组2（错配干预）的CHDI涨幅与政策回报率均为零，表明将预算投入非核心边缘因子，无法转化为实质性的文化发展增量。这“归零现象”警示了脱离数据实证的文化政策制定，极易导致公共资源的无效沉没。

九、结论与建议

（一）主要研究结论

1. 文化传承与发展的关键驱动因素识别

从由极端梯度提升XGBoost 和支持向量回归SVR加权后的排序看，群众参与对深度传承行为的直接拉动力比数字赋能更强；但数字赋能目前整体提升空间大，政策边际效益更高这是它在政策上更值得优先做的原因。

2. 文化高质量发展指数（CHDI）的构建与验证

本文构建的少数民族文化高质量发展指数由7个经机器学习筛出的核心维度构成。从信度、效度和效标关联来看，可以认为CHDI具备良好的可靠性、预测力与区分能力，可以作为少数民族文化高质量发展水平的标准化量化评估工具使用。

3. 文化传承行为的影响机制与路径分析

关于影响机制，结构方程模型的估计结果与Bootstrap检验证实：外部刺激变量并不直接作用于文化传承行为，而必须先经过“心理机制”的转换，再经由“行为意愿”的中介，才能最终落在行为上。换言之，政策或活动如果只停留在提供刺激的层面，而忽略中间的接受与转化环节，很难收到预期效果。

4. 文化高质量发展水平的评估与预测分析

本研究比较了多种机器学习模型的表现，其中支持向量回归（SVR）模型的预测效果最优，在此基础上，本文进一步做了政策沙盘推演。结果显示，把资源集中投入到数字技术创新应用与数字传播效能两个方向，其政策回报率明显高于在全域范围内平均分摊的做法。这说明在资源总量有限的条件下，投入结构往往比投入规模更重要。

5. 情感联结与文化认同螺旋上升闭环的构建与检验

链式中介的检验结果表明，情感联结在全部外部刺激变量中的驱动效应最强，数字赋能与行为意愿对深度传承行为具有正向驱动作用，然后，行为转化所沉淀的文化认同，将进一步反哺和强化情感联结，从而形成了自我增强的循环机制。这一闭环突破了线性传导的局限，揭示了少数民族文化从“一时关注”走向“长期认同”不断深化的内在动力。

（二）实践优化路径

上述结论落到操作层面，可以从三个方向展开。线下情境解决的是心理机制的问题，数字赋能解决的是条件短板的问题，线上线下融合则解决的是转化链条能否持续的问题。三者的关系并非平行并列，而是依次递进，实际推进时也应当按照这一顺序安排优先次序。

1. 线下激活心理双引擎

实证结果显示，公众对文化传播的积极评价和参与能力感知是深度传承的核心心理驱动力。本研究建议政府应当一是强化情感联结设计，通过沉浸式体验、故事化传播增强群众的文化共鸣。二是降低群众参与门槛，优化群众参与便利性，提升个体知觉行为控制感。三是借社区和熟人网络做一点价值引导，强化积极行为态度，推动从“愿意参与”向“主动传承”转化。

2. 数字赋能补齐短板

实证结果显示，数字赋能维度初始得分低、边际收益高，是当前破局的最优解，本研究建议政府要重点投入VR/AR沉浸式体验、AI内容创作、区块链版权保护等技术，并拓展多平台数字传播渠道，增强年轻群体参与度；并继续完善数字版权和制度保障体系，为文化高质量发展提供长效支撑。

3. 融合构建转化闭环

实证结果显示，情感联结是整个传导链条中效应最强的外部刺激。本研究建议政府把线下书信交流、文化节活动带来的情感体验，延伸到线上场景，例如搭建线上社群、开展数字文创共创打卡，把线下形成的认同感转化为线上传播行为，再反哺线下参与，搭建“线下体验→线上分享→线下深度参与”的螺旋上升路径，打破单一情境的转化瓶颈。

（三）研究局限与未来展望

本研究仍存在一定局限：第一，横截面数据难以严格推断因果方向，后续可采用纵向追踪设计验证动态机制进行改善；第二，以粤黔协作单一案例为对象，结论的跨场景普适性有待进一步检验。未来可开展多区域、多民族的纵向追踪研究，结合实验设计与大数据分析进一步提升结论的外部效度与解释力；第三，深度传承行为的测量基于自陈意愿，未来可结合行为日志、平台数据等客观指标增强效度。

参考文献

[1] 邱婷，陈子莹 . 后喻文化理论视角下新时代非物质文化遗产传承创新的机理探究 [J]. 文化创新比较研究，2024，8(07)：173-176.
[2] 罗薇，甘海燕 . 多元保护主体视域下非遗传承与乡村振兴——以青岛市城阳区洼里盘鼓为例 [J]. 广西职业技术学院学报，2024，17(05)：26-33.
[3] 王萍，孟丽 . 区域文化生态对非物质文化遗产传承与创新的影响机制研究 [J]. 商展经济，2025，(12)：53-56.
[4] 刘明真，李元旭 . 数字文化战略视域下文化遗产数字化传承：机理、挑战与路径 [J]. 档案学研究，2024，(05)：107-116.
[5] 陈波，涂晓晗 . 从媒介展演到文化认同：乡村数字记忆空间的运行机理 [J]. 武汉大学学报（哲学社会科学版），2026，79(01)：18-29.
[6] 赵燕云 . 感知价值、文化认同与行为意向：宾阳炮龙传承发展的实证研究 [D]. 南宁：南宁师范大学，2024：1-84.
[7] 田甜 . 基于S-O-R模型的非遗文创产品消费者购买意愿研究 [J]. 统计与信息论坛，2021，36(12)：116-124.
[8] 张茜，许春晓，温卫宁 . 红色旅游促进青少年红色文化认同形成机制研究——基于学段差异的组态路径分析 [J]. 旅游科学，2025，39(01)：101-118.
[9] 袁琳芳 . 基于TPB理论的体育专业学生武术文化传承研究 [J]. 冰雪体育创新研究，2025，6(19)：40-42.
[10] 陶纯良 . 基于TPB模型的毛南族傩面具文创产品的创新设计 [D]. 桂林电子科技大学，2025.
[11]Guyon I., Elisseeff A. An introduction to variable and feature selection[J]. Journal of Machine Learning Research，2003，3(1)：1157-1182.

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License