

基于文本挖掘与关联规则分析的非遗旅游体验优化研究

赵芳, 赵威, 刘一一, 朱娇蕾

西南交通大学希望学院基础部, 四川 成都

收稿日期: 2026年3月3日; 录用日期: 2026年4月22日; 发布日期: 2026年4月30日

摘要

在响应“十四五”文化数字化战略, 研究聚焦川黔布依族非遗保护与旅游体验优化。研究以四川拉落村为核心案例并结合贵州景区数据, 通过Python采集大众点评评论($n = 1305$)与小程序行为记录($n = 50,814$), 运用语义挖掘、情感分析及关联规则算法进行多维分析。研究发现: 景观与服饰是游客满意度的核心驱动力, 积极情绪占比达70.4%, 但服务耗时与质量仍是主要痛点。在活动偏好上, 美食与银饰制作支持度最高, 但项目间关联强度较弱, 反映出景区协同服务设计不足。研究结论指出, 应依托数字化手段强化非遗项目的协同联动, 为布依族非遗的数字化保护与景区服务优化提供实证支撑。

关键词

布依族非遗传承, 舆情分析, 关联分析, 文本挖掘, 旅游体验优化

Optimizing Intangible Cultural Heritage Tourism Experiences Through Text Mining and Association Rule Analysis

Fang Zhao, Wei Zhao, Yiyi Liu, Jiaolei Zhu

Department of Basic Studies, Southwest Jiaotong University Hope College, Chengdu Sichuan

Received: March 3, 2026; accepted: April 22, 2026; published: April 30, 2026

Abstract

Aligning with the “14th Five-Year” Cultural Digitization Strategy, this research focuses on the protection of Buyi intangible cultural heritage (ICH) and the optimization of tourism experiences in the Sichuan-Guizhou region. Taking Laluo Village in Sichuan as the core case study and integrating data

from Guizhou scenic areas, the study utilized Python to collect 1,305 Dianping reviews and 50,814 mini-program behavior records. Methods such as semantic mining, sentiment analysis, and association rule algorithms were applied for multi-dimensional analysis. The findings indicate that natural scenery and ethnic attire are the core drivers of tourist satisfaction, with positive sentiment accounting for 70.4%, while service duration and quality remain primary pain points. Regarding activity preferences, traditional food and silver jewelry making received the highest support; however, the weak association between various projects suggests a deficiency in integrated service design within the scenic areas. The study concludes that digital empowerment should be leveraged to strengthen project synergy, providing empirical support for the digital preservation of Buyi ICH and the enhancement of scenic area services.

Keywords

Bouyei Ethnic Intangible Cultural Heritage Inheritance, Public Opinion Analysis, Correlation Analysis, Text Mining, Tourism Experience Optimization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

以四川宁南拉落村与贵州布依族聚居区为案例,本研究探讨了依托传统节庆、手工艺及音乐舞蹈形成的“非遗+旅游”特色发展模式。拉落村通过参与式体验与产业融合,在显著带动地方经济(旅游收入占比达18%)的同时,也面临着传承群体老龄化、青年参与不足及文化真实性与商业化张力等结构性挑战。在文旅融合持续深化的背景下,大数据、人工智能等数字技术为非遗的个性化传播与延续性实践提供了新路径,如何科学评估传播影响因素并预测其发展趋势,已成为破解保护难题、推动非遗在现代社会广泛传播的关键课题。

2. 文献综述

2.1. 非遗与旅游融合研究

“非遗+旅游”融合是近年研究热点。鲁炜中等(2024)以四川布拖县为例,从乡村振兴视角分析彝族“非遗+旅游”产业发展,提出品牌塑造、体验升级及市场与技术驱动协同的路径[1]。鄢继尧等(2023)基于多源数据采用熵值法与K-means聚类对全国进行分区识别,指出中国非遗资源旅游利用潜力总体偏低且区域差异显著,并提出分类型融合发展策略[2]。在宏观层面,郭延龙、郑晶晶(2022)通过CiteSpace分析2000~2021年113篇文献,发现研究从政策导向向体验、产业化、传播技术等细分路径延伸[3]。张广海等(2023)基于网络文本与核密度分析,研究全国国家级与省级非遗资源的分布及景区响应机制,发现旅游总收入、经济发展水平、科技水平等是驱动因素[4]。贵州作为布依族主要聚居区,其非遗旅游发展同样受到学界关注;四川宁南拉落村则是四川布依族非遗旅游的典型代表,两地共同构成布依族非遗旅游研究的重要场域。

2.2. 数字化保护与传播路径

数字化技术在非遗保护与传播中的作用日益突出。王亮、杜森林(2025)以甘孜藏族自治州为例,分析非遗数字化传播现状与困境,如素材采集广泛但形式单一、互动性与参与感不足[5]。鄢继尧等(2023)基于

多源数据评估中国非遗资源的旅游利用潜力并进行分区,为资源配置提供依据[6]。

2.3. 文本挖掘与情感分析在旅游研究中的应用

文本挖掘与情感分析在旅游研究中应用广泛。郭佳怡等(2023)运用文本挖掘与情感分析方法评估智慧旅游背景下的服务质量,通过网评识别痛点与情感诉求,并建立评价体系[7]。在文化遗产旅游形象感知方面,王皓、李志刚(2023)以平遥古城为例,运用高频词分析、语义网络分析、情感分析及扎根理论编码,识别文化内涵、品牌建设与体验为正向驱动力,治理与服务管理问题会负面影响形象感知[8]。厉建梅等(2023)以洛阳龙门石窟为例,利用网络游记和在线评论进行文本分析,通过词频分析、网络语义分析及文本编码,归纳出游客感知价值包含服务功能价值、情感价值、文化认知价值、成本价值与社会价值五个维度[9]。在文化感知差异研究方面,郑淞尹等(2022)基于文本挖掘分析中外游客在世界遗产地的文化感知差异,通过情感分类、词频与语义网络分析,揭示中外游客在遗产地体验中对文化表达、价值感知等方面的差异[10]。

2.4. 关联规则挖掘在旅游活动分析中的应用

关联规则挖掘在旅游行为分析中应用增多。曹青青等(2024)运用 Apriori 算法分析大学生选择景点特征组合,发现“民俗”“物价”与“文化底蕴”等因素在学生群体中尤为重要[11]。涂晨等(2023)结合 LDA 主题模型与 Apriori 算法对 OTA/UGC 数据进行挖掘,旨在发掘旅游目的地中游客最关注的主题,并探讨主题间隐藏的活动关联[12]。在文创产品设计方面,吴定丙(2024)通过改进的 Apriori 算法挖掘文创产品中的关联特征,用于提高设计效率与产品匹配度,支持度和置信度指标能帮助识别不同文化元素之间的组合规律[13]。唐仲华(2008)对 Apriori 与 FP-Growth 等算法进行对比,并提出改进算法,用于分类与关联规则挖掘场景[14]。

2.5. 游客感知价值与文化认同研究

在游客感知价值研究方面,王佳果等(2022)研究旅游者感知文化距离对民族文化遗产旅游体验意愿的影响,构建“认知-情感-意动”理论框架,选取广西龙脊梯田和贵州肇兴侗寨为案例,通过结构方程模型验证感知文化距离对体验意愿的影响路径[15]。田美玲等(2024)在文化 IP 解码视角下研究旅游者文化感知对文化认同的作用机理,以长江国家文化公园为例,基于 2019~2023 年 258,865 条景区点评数据,通过文本挖掘识别游客文化感知的维度,并结合问卷与回归模型探讨这些维度如何推动文化认同[16]。在景点间语义交互研究方面,陈宇等(2022)提出从游客需求出发,以文本数据挖掘景点间语义交互的方法,用以优化旅游格局与旅游路径设计,为理解景区内活动或设施间的协同效应提供了方法基础[17]。

3. 数据来源与预处理

3.1. 数据来源

本研究利用 Python 爬取大众点评平台上布依族景区游客评价数据,以“布依族景区”等为关键词检索在线评论,共获取有效评论 1323 条;经去重与清洗后用于建模分析的有效评论为 1305 条。评论内容涵盖服饰文化、住宿环境及旅游体验等维度,能够反映游客在景区情境下的主观感知,适用于词频、语义与情感挖掘。

此外,利用在布依族活动小程序平台获取以宁南县各镇为关键词的游客在线非遗活动评分系统记录的信息,获取 50814 条样本数据。该小程序由宁南县文化广电旅游局联合本地文旅企业开发,是宁南县布依族非遗文化传承的官方数字化载体与线下运营体系,是本地游客与居民参与布依族非遗活动的核心入口,数据具有较高的代表性和权威性。两类数据在概念上为互补关系:前者偏“目的地整体体验与

情”，后者偏“具体活动参与记录”；本文在结果呈现中分开展示，并在讨论中对照两类证据所指向的体验触点差异。

3.2. 数据预处理

1. 布依族景区游客评价数据预处理

由于原始爬取的数据具有较高的复杂性，因此需要对其进行数据预处理。评论文本采用 jieba 精确模式分词；停用词在通用停用词表基础上增加景区领域词与无意义符号过滤；Word2Vec 采用 Skip-gram 结构；语义相近词通过余弦相似度排序输出。情感分析采用基于词典与规则/分类模型相结合的打分方法(情绪值 0~1)，并以 0.3、0.7 为中性区间边界。TF-IDF 在单条评论为文档、全量评论为语料库的设定下计算词权重，用于提取情感判别特征词。具体流程为首先对数据进行去重，并对数据格式进行转换，将爬取的数据统一转换为字符串格式，清除 NAN 值以及非文本数据；其次，利用 jieba 库对文本数据进行分析；最后在分词之后的数据上基于 HIT 停用词表并结合自定义景区停用词进行过滤，将高频虚词以及标点符号进行过滤，最终获取有效数据评论 1305 条。

2. 小程序平台数据预处理

(1) 数据预处理

本研究采用分组中位数填充法处理数据缺失值，该方法能有效保持数据分布特征并适配分析需求，通过引入缺失指示变量与控制变量进一步提升结果可靠性；异常值检测采用箱线图法，结果显示满意度、互动性评分及独特性评分均无异常，仅游客推荐意愿存在异常值，鉴于样本数据量充足，最终采用删除法处理该类异常值。

(2) 连续变量分段与独热编码

针对爬取的年龄、参与时长及消费金额三个连续型特征，本研究采用独热编码分段处理。考虑到消费金额与参与时长数值范围差异显著，直接建模易导致特征权重分配偏差。通过将连续变量离散化为区间段，不仅能有效降低数据噪声干扰，还可提升分类模型对特征模式的捕捉能力。经预处理后，最终数据集包含以下特征有游客 ID、年龄段、参与时长段、消费金额段、满意度(1~5 分)、推荐意愿(1~5 分)、互动性评分(1~5 分)、独特性评分(1~5 分)、参与活动项目数量以及参与活动项目。

(3) 特征描述性统计

数据分析结果显示，受访者满意度均值为 3.44 (标准差 0.827)，推荐意愿均值为 3.42 (标准差 1.124)，二者整体处于中等偏上水平；相较之下，推荐意愿的标准差更大，离散程度更高。互动性评分均值最高(3.75)，但标准差亦最大(1.543)，表明游客对互动体验的评价存在较为明显的两极分化；独特性评分均值最低(3.35，标准差 1.43)，反映出景区体验同质化现象较为普遍。就旅游行为特征而言，消费金额段均值 3.90 (标准差 1.976)、参与时长段均值 3.44(标准差 1.537)，二者的变异系数分别约为 0.508 与 0.448 (由标准差与均值之比计算)，显示游客在消费水平与停留时长上个体差异显著；游玩项目数量均值为 3.36 (标准差 1.429)，其离散程度相对更小、分布更为集中。年龄段均值为 3.55(标准差 1.011)。综合来看，用户体验整体处于中等水平，但各维度均呈现不同程度的异质性，其中互动性体验与消费相关行为的差异尤为突出。

4. 基于文本挖掘的布依族文化旅游游客体验分析

4.1. 游客高频词关联与语义网络分析

1. 词频分析与语义关联

Word2Vec 是一种用于将词语转换为向量的技术，属于自然语言处理(NLP)中的词嵌入(Word

Embedding)方法。它能够通过深度学习模型将每个词语映射到一个多维的向量空间中,使得词语之间的语义关系可以通过向量的数学运算来捕捉。

本研究选取“瀑布”“景区”“陡坡”“景点”等景区相关高频词作为中心词,利用 Word2Vec 的 most_similar ()函数计算各中心词与语料库中其他词语的余弦相似度,并按相似度排序,筛选出与每个中心词最相近的若干词语及其相似度数值,从而形成表 1 中的“景区相关预测中心词”及对应相似度。

Table 1. Core predictive terms for tourist attractions

表 1. 景区相关预测中心词

瀑布	景区	陡坡	景点
观光车 0.59	文化 0.67	扶梯 0.60	买 0.72
玩 0.53	比较 0.64	很大 0.47	没有 0.61
寨子 0.54	吃 0.58	位置 0.46	吃 0.60
工作人员 0.53	寨子 0.57	设施 0.45	水雾 0.59
游客 0.52	扶梯 0.56	停车费 0.46	攻略 0.59

数据表明游客对景区的关注和评价主要围绕景观、消费与体验三个维度。从相似度来看,“景点”与“买”、“吃”的相似度较高,说明游客对景点的评价与消费、餐饮体验紧密相关;“景区”与“文化”、“寨子”、“扶梯”的关联较强,反映出游客将景区与民族文化、传统村落及配套设施联系在一起;“瀑布”与“观光车”、“玩”、“寨子”、“工作人员”的相似度较高,说明自然景观与交通、游玩、村落及服务人员共同构成游客体验;“陡坡”与“扶梯”、“很大”、“设施”、“停车费”的关联,则体现了游客对地形、设施及费用的关注。整体而言,表明游客在景区体验中“景观-消费-服务”的语义关联结构,为理解游客感知和优化景区服务提供了语义层面的依据。

2. 网络图可视化与解读

基于高频词共现网络分析,本研究选取前 35 个核心词汇构建共现关联模型,并对边权重进行量化。结果显示,“景区-陡坡-瀑布-值得”形成明显的共现簇,表明在地形存在一定挑战的情况下,游客仍对核心景观给予积极评价,可视化结果见图 1。

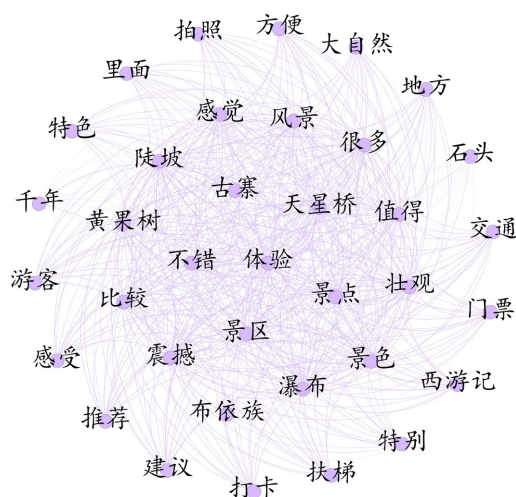


Figure 1. Network association analysis diagram

图 1. 网络关联分析图

4.2. 基于 TF-IDF 与情感分析的游客评价研究

1. 游客情感倾向分布特征

情感分析结果显示，游客评价整体呈正向倾向。从表 2 可见：积极情绪(情绪值 0.7~1)占比 70.50%，其中 0.7~0.8 区间(一般积极)占 63.45%，0.8~1 区间(中高度积极)合计占 7.05%，说明多数游客对核心体验持认可态度，且积极评价以中等偏上为主；中性情绪(0.3~0.7)合计约 9.1%，表明仍有部分评价未形成明确情感倾向；消极情绪(0~0.3)占比 20.31%，其中情绪值 0~0.1 的极端负面评价占 16.09%，在消极情绪中占主导，提示景区在服务与管理等方面存在突出短板，需重点改进。综上，情感分布呈现“以积极为主、消极集中于极端低分”的特征：约七成评价为积极，整体吸引力较强；约两成为消极，且逾八成消极评价集中在 0~0.1 区间，凸显服务质量与基础设施的改进紧迫性；约一成中性评价则显示体验仍有提升空间。该分布为景区实施分层、精准优化策略提供了量化依据。

Table 2. Statistical table of emotion types by segment

表 2. 情绪类型与分段统计表

情绪大类(总分区间)	细分强度分布(数量、占比)	小计(数量、占比)
积极情绪 (0.7~1.0)	一般 828 (63.45%); 中度 51 (3.91%); 高度 41 (3.14%)	920 (70.50%)
中性情绪 (0.3~0.7)	一般 30 (2.29%); 中度 22 (1.68%); 中高度 28 (2.14%); 高度 40 (3.06%)	120 (9.19%)
消极情绪 (0.0~0.3)	一般 24 (1.84%); 中度 31 (2.38%); 高度 210 (16.09%)	265 (20.31%)

2. 影响游客情感的特征值分析

(1) TF-IDF 理论基础

TF(词频)衡量词语在单篇文档中的出现频率，IDF(逆文档频率)衡量词语在所有文档中的稀有程度——词语越常见，IDF 值越低。TF-IDF 为二者之积，用于刻画词语在文档中的相对重要性，能够突出“在特定文档中重要、在其他文档中较不常见”的词语，因而常用于文本分类、关键词提取与信息检索。

(2) 影响游客情感的特征值分析

Table 3. Top 10 feature terms affecting tourist sentiment and Their TF-IDF weights

表 3. 影响游客情感的特征值及 TF-IDF 权重(前 10)

情绪类型	Top10 特征词(TF-IDF)
正面情绪	瀑布(0.681521)、景区(0.408788)、黄果树(0.343257)、壮观(0.166660)、值得(0.151748)、 不错(0.145608)、体验(0.137714)、风景(0.134205)、景点(0.131686)、古寨(0.130696)
负面情绪	瀑布(0.740155)、景区(0.417559)、黄果树(0.385439)、门票(0.157021)、景点(0.120101)、 扶梯(0.086361)、路线(0.043181)、商业街(0.013739)、购票(0.013739)、恶心(0.009814)

在 TF-IDF 理论上，对游客评论文本进行特征提取与权重计算：以词频(TF)表示词语在单条评论中的出现频率，以逆文档频率(IDF)表示词语在全部评论中的稀有程度，二者相乘得到各词的 TF-IDF 权重。据此筛选出文档重要性高、且在正/负情感类别间区分度较大的特征词，用于识别影响游客情感的关键词汇。表 3 给出了正面情绪与负面情绪下 TF-IDF 权重最高的前 10 个特征词及对应权重。

(3) 景观词汇情感双极性挖掘

由表 6 可知, 正面情绪的高权重特征词以景观与体验类为主: 瀑布(0.682)、景区(0.409)、黄果树(0.343)居前三位, 壮观、值得、不错、体验、风景、景点、古寨等紧随其后, 说明自然景观与整体体验是驱动游客满意度的核心因素。负面情绪的高权重词中, 瀑布、景区、黄果树同样位居前三且权重更高(如瀑布 0.740), 同时出现门票、扶梯、路线、商业街、购票等与游览流程与设施相关的词, 以及强烈负面词恶心(0.010), 反映出负面评价集中在“核心景观体验”与“购票、路线、设施”等运营服务环节。

情感强度分析进一步显示, “瀑布”“景区”“黄果树”在正、负两类情绪中均具有较高 TF-IDF 权重, 呈现明显的情感双极性: 同一景观符号既可能承载“壮观/值得”的积极评价, 也可能在排队、购票与动线组织不畅时被再次激活为负面情绪载体。如正面评论中有“瀑布很壮观, 风景值得专程来一趟, 古寨也有味道”; 负面评论中有“瀑布还可以, 但排队和购票太折腾, 整体体验被拉低”。正面列表中景观与体验类词汇占主导, 负面列表中服务与设施类词汇(门票、扶梯、路线、商业街、购票)与极端负面词形成补充, 共同指向景区需在高关注度景点的游览动线、票务与配套设施上重点优化, 以降低负面情绪、提升整体满意度。

5. 基于 APRIORI 算法布依族非遗活动的关联规则挖掘

为揭示游客在布依族非遗活动参与中的组合选择机制, 本文在利用 TF-IDF 识别“耗时”等关键体验痛点的基础上, 引入 Apriori 算法对活动共选关系进行关联规则挖掘, 将“体验摩擦”(如时间成本、等待与衔接不畅)与“活动组合结构”(共现频率与带动效应)进行联动分析, 为景区活动协同供给与服务流程优化提供实证依据。规则解读时, 将提升度划分为三类讨论: 提升度明显大于 1(强正向关联, 本表未出现)、提升度接近 1(弱或近似独立共现)、提升度显著小于 1(负相关, 需谨慎解释样本结构)。结合景区空间布局与活动时间的典型情境: 当手作体验与表演类项目分处不同片区且场次间隔紧凑时, 即便存在共同偏好, 也可能因动线与时间窗不匹配而表现为低提升度; 这提示管理优化应同时调整空间相邻性与排期衔接, 而非仅增加活动供给数量。Apriori 算法对游客活动选择记录(交易数据)进行扫描, 提取满足最小支持度阈值的频繁项集并生成关联规则, 核心指标包括: 支持度(Support), 表征活动组合在全部选择中的出现频率; 置信度(Confidence), 表示在已选前项条件下选后项的条件概率; 提升度(Lift), 表示前项对后项选择概率的相对增益。上述指标为识别高偏好活动与评估活动间带动性提供了量化工具。

5.1. 活动支持度特征

Table 4. Ranking of support for Bouyei ethnic group activities

表 4. 布依族活动支持度排名

排名	活动名称	支持度	排名	活动名称	支持度
1	美食制作	0.1994	6	蜡染工作	0.1984
2	银饰制作	0.1992	7	竹编工艺	0.1983
3	农耕文化	0.1990	8	服饰试穿	0.1983
4	枫香染 DIY	0.1989	9	古歌传习	0.1985
5	民族婚俗	0.1984	10	铜鼓制作	0.1979

由表 4 可知, 支持度排名前 10 的活动其支持度均处于 0.1979~0.1994 之间, 分布集中, 反映出游客偏好呈“多点高频”特征。其中, “美食制作”(0.1994)与“银饰制作”(0.1992)居前两位, 可视为第一梯队, 体现游客对饮食文化与金属工艺类体验的稳定偏好。从类型构成看, 前 10 名中手工艺体验类(银饰

制作、枫香染 DIY、蜡染工作、竹编工艺、铜鼓制作等)占比较高,整体平均支持度约 0.198,说明“可参与、可沉浸、可产出”的体验形态更易形成选择优势,在一定程度上支持了“生产性保护”导向的产品配置合理性。

5.2. 关联规则与活动组合结构

如表 5 显示,各规则支持度集中在 0.036~0.038,置信度约 0.18~0.19,说明相关活动对在游客选择中具有一定共现基础。从提升度分层看,Lift 的范围为 0.92~0.97 均低于 1,属于“弱关联/接近独立共现”区间:宜解释为偏好相近人群在同一行程内的并行选择,而非后项被前项显著带动的互补结构。若结合景区导览图与活动时段安排进行情境化解释:当“枫香染 DIY”与“八音坐唱”在物理距离或场次上不易顺接时,统计上的弱提升度并不否定文化主题的相近性,而可能反映组织层面的可达性约束。然而,各规则提升度(Lift)普遍在 0.92~0.97 之间且多略低于 1,表明前项活动对后项活动选择概率的增益有限,活动间“带动性”总体偏弱。因此,尽管部分活动在同一游客群体中存在并行偏好,这种共现更可能源于共同兴趣或主题吸引,而非明确的因果联动或强互补关系。具体而言:“枫香染 DIY - 八音坐唱”与“八音坐唱 - 枫香染 DIY”双向规则的支持度均为 0.03628,置信度分别为 0.182 与 0.184,共选现象稳定,但提升度均为 0.927,接近且略低于 1,宜解释为共现关系而非显著带动;“农耕文化展 - 八音坐唱”与“八音坐唱 - 农耕文化”双向规则支持度均为 0.03616,置信度约 0.182 与 0.184,提升度均为 0.923,同样未体现明显跨活动转化增益,可能反映游客对传统文化内容的共同偏好;“三月三对歌 - 八音坐唱”的支持度(0.03766)与置信度(0.192)在表内相对最高,提示表演/互动型项目之间更易形成连续参与,但提升度仍为 0.975,接近 1,进一步说明整体联动效应有限。

Table 5. Activity correlation analysis table

表 5. 活动关联分析表

前项活动	后项活动	支持度	置信度	提升度	关联强度
枫香染 DIY	八音坐唱	0.03628	0.18239	0.92713	弱
八音坐唱	枫香染 DIY	0.03628	0.18443	0.92713	弱
农耕文化展	八音坐唱	0.03616	0.18165	0.92341	弱
八音坐唱	农耕文化	0.03616	0.18382	0.92341	弱
三月三对歌	八音坐唱	0.03766	0.19171	0.97455	弱

5.3. 讨论与管理启示

结合 TF-IDF 所识别的“耗时”等体验痛点与 Apriori 所揭示的“活动组合低带动性”,可推断景区活动体系在跨项目衔接与协同供给上仍有改进余地:活动之间虽具备一定共选基础,但在时间成本、等候与动线转换等约束下,尚未形成稳定、可预期的转化路径。为统合文本与行为两类证据,本文提出一条简化的解释链条——情绪倾向(情感分析)刻画游客对服务与服务流程的整体态度;TF-IDF 关键词指向触发该态度的具体议题(景观吸引与流程摩擦的分化);Apriori 提升度则反映活动层面的共选结构及其转化潜力。三者呈递进关系:当高吸引力景观词与负面情绪并存(情感双极性)时,往往提示票务、排队与导引等环节组织不畅;当活动共选相对稳定而提升度持续低于 1 时,则更宜从时间表与空间布局上的弱衔接加以解释,而非简单归因为产品主题失当。据此,优化宜以降低跨活动参与摩擦、提升协同效率为优先,通过排期与时间窗协调、动线与导览优化、联票/预约与推荐触点设计等组合手段,提高活动间的可达性与连续参与概率,进而改善整体体验并提升消费与参与深度。

6. 结论与建议

6.1. 研究结论

综合文本挖掘与关联规则结果, 本节从情绪表达、关键词议题与活动共选结构三方面归纳主要发现如下:

整体体验处于中等偏上, 但异质性显著。满意度与推荐意愿均值分别为 3.44 与 3.42, 处于中等偏上; 其中推荐意愿标准差更高, 表明口碑传播态度分化更明显。互动性均值最高(3.75)但离散度最大, 游客对互动体验呈现“两极分化”; 独特性均值相对较低(3.35)且分歧较大, 提示差异化记忆点仍不足。消费金额与参与时长变异系数分别为 0.508 与 0.448, 反映消费能力与停留行为差异突出, 而游玩项目数量相对更集中。

文本与情绪结果指向“景观强吸引 - 服务/效率短板”并存。高分评论高频词聚焦“布依族服饰”“推荐”“核心景观”等, 说明自然景观与文化元素构成主要吸引力; 低分评论集中于“耗时”“体验差”等, 表明游览效率与服务体验是主要痛点。情感倾向整体正向(正面约 70.4%), 但负面评价占比仍达约 20.3%, 提示在服务管理、流程组织与基础设施方面存在需要优先治理的薄弱环节。

活动偏好“多点高频”, 但组合带动性不足。非遗活动支持度前 10 均高于 0.195, 且以手工艺体验为主, 说明“可参与、可产出”的体验形态更易形成选择优势; 但关联规则的提升度普遍接近 1 且略低于 1, 活动间带动效应不显著, 更多体现为“同一兴趣人群的共选”而非强互补联动。结合“耗时”痛点, 可推断跨活动转化受排队、动线与时间窗不匹配等“体验摩擦”制约。综合前文, 上述结果共同指向“景观与文化展示吸引力较强, 而游览流程组织与活动衔接仍相对薄弱”的结构特征; 这与文本侧的情绪分布(积极为主但极端负向集中)相互印证, 提示治理抓手应从事后舆情回应进一步前移到排队、票务、导引与排期一体化设计。

6.2. 建议

鉴于“枫香染 DIY - 八音坐唱”等活动对呈现一定共现特征而提升度接近 1, 说明其更多反映游客偏好的同步性而非显著的相互带动。建议景区以“主题一致性整合”为导向, 构建非强制性、可组合的产品体系(如“工艺体验 - 展演观赏”主题线路、可选联票/套票), 并在关键节点提供规范化信息供给(场次衔接、预计耗时、空间距离、推荐顺序), 以降低跨活动选择的认知与决策成本, 提升活动组合的自然转化效率。

针对活动间联动不足与时间成本约束并存的情形, 建议将“时间成本治理”作为协同优化的优先方向, 从制度与流程层面提升跨活动参与的可达性与确定性。具体包括: 推进预约、排队与检票规则的一体化管理; 设置与活动节奏相匹配的时间窗与缓冲机制, 实现手作体验与表演类项目的可衔接排程; 同步完善动线指引与实时等待信息发布, 减少排队、绕行与信息不对称导致的机会损失, 从而提升整体体验效率与多项目参与概率。

鉴于“农耕文化展 - 八音坐唱”“三月三对歌 - 八音坐唱”等组合表现为偏好相近下的共现而非显著增益, 建议将关联规则结果用于客群分层与场景化供给, 避免将其简单等同为强关联。可依据游客偏好结构形成差异化产品线与导览方案(如“文化叙事沉浸线”“节庆互动体验线”), 并在空间与时间维度实施相邻化配置(相近区域布设、相邻时段排期), 以增强体验连贯性与主题连续性, 进而提升游客停留时长与多项目参与水平。

参考文献

[1] 鲁炜中, 华袁, 马阿木沙. 乡村振兴战略视野下四川彝族“非遗 + 旅游”文化产业发展研究——基于四川布拖县

- 的调研[J]. 国际经济与发展, 2024, 2(8): 95-97.
- [2] 鄢继尧, 赵媛, 郝丽莎, 郭宇. 基于多源数据的中国非遗资源旅游利用潜力评价与分区研究[J]. 地理研究, 2023, 42(12): 3294-3312.
- [3] 郭延龙, 郑晶晶. 非遗保护视角下文旅融合模式研究——基于 CiteSpace 的可视化分析[J]. 呼伦贝尔学院学报, 2022, 30(1): 132-141.
- [4] 张广海. 非物质文化遗产资源时空分布特征及景区响应机制——基于网络文本信息挖掘与定量测度分析[J]. 地域研究与开发, 2023, 42(1): 102-107.
- [5] 王亮, 杜森林. 新媒体时代非物质文化遗产数字化传播研究——以甘孜藏族自治州为例[J]. 今传媒, 2025, 33(2): 117-120.
- [6] 鄢继尧, 赵媛, 郝丽莎, 郭宇. 基于多源数据的中国非遗资源旅游利用潜力评价与分区研究[J]. 地理研究, 2023, 42(12): 3294-3312.
- [7] 郭佳怡, 方博平, 陆欣怡, 等. 基于文本挖掘和情感分析方法的“智慧旅游”服务质量感知研究[J]. 现代信息科技, 2023, 7(6): 1-5, 12.
- [8] 王皓, 李志刚. 基于文本分析与扎根理论的平遥古城文化遗产旅游形象感知研究[J]. 可持续发展, 2023, 13(5): 1373-1382.
- [9] 厉建梅, 齐佳, 单梦琦. 网络文本分析视角下文化遗产旅游地游客感知价值维度构建——以洛阳龙门石窟为例[J]. 旅游论坛, 2023, 16(4): 89-97.
- [10] 郑淞尹, 孙传明, 谈国新. 基于文本挖掘的中外游客文化感知差异——以世界遗产地为例[J]. 华侨大学学报(哲学社会科学版), 2022(1): 64-77.
- [11] 曹青青, 胡贵育, 霍奥莉, 等. 基于 Apriori 算法的大学生选择旅游景点的研究分析[J]. 现代管理, 2024, 14(2): 253-259.
- [12] 涂晨, 李鑫, 叶程轶. 基于 LDA 主题模型与 Apriori 算法的旅游数据挖掘[J]. 物联网技术, 2023, 13(3): 108-112.
- [13] 吴定丙. 基于改进 Apriori 算法下关联特征挖掘的文创产品开发系统[J]. 河北水利电力学院学报, 2024, 34(4): 56-60.
- [14] 唐仲华. 基于关联规则的分类算法研究及应用[D]: [硕士学位论文]. 广州: 华南理工大学, 2008.
- [15] 王佳果, 杨海鑫, 郑密. 旅游者感知文化距离对民族文化遗产体验意愿影响机制[J]. 热带地理, 2022, 42(1): 55-66.
- [16] 田美玲, 方世明, 寇圆圆, 等. 文化 IP 解码视角下旅游者文化感知对文化认同的作用机理——以长江国家文化公园为例[J]. 中国生态旅游, 2024, 14(5): 980-993.
- [17] 陈宇, 秦昆, 喻雪松, 等. 旅游景点间细粒度语义交互作用挖掘及模式分析: 以云南省为例[J]. 地球信息科学学报, 2022, 24(10): 2021-2032.