

# 基于评论语义挖掘与K-Means聚类的古村落文旅电商体验优化与品牌塑造研究

左秉聪<sup>1</sup>, 陈金菲<sup>2</sup>

<sup>1</sup>山东建筑大学管理工程学院, 山东 济南

<sup>2</sup>山东建筑大学商学院, 山东 济南

收稿日期: 2025年11月28日; 录用日期: 2025年12月11日; 发布日期: 2025年12月31日

## 摘要

古村落作为重要的文化遗产, 其文旅产业的高质量发展对乡村振兴至关重要。随着文旅电商成为游客消费的主要渠道, 在线评论成为优化体验与塑造品牌的关键数据源。本研究以联合国“世界最佳旅游乡村”名录中的中国古村落为案例, 采集“携程网”的9334条游客评论作为数据基础。通过情感分析、LDA主题模型挖掘评论的整体情感倾向与核心关注主题, 并创新性地结合K-Means聚类算法, 构建融合情感得分与主题分布的特征向量对用户进行分群。研究成功识别出六个特征鲜明的游客群体: “环境批判者”、“自然风光爱好者”、“乡村生活向往者”、“历史人文探索者”、“服务体验关注者”和“独特价值探寻者”。基于此立体化用户洞察, 本研究从“产品 - 服务 - 营销 - 品牌”四个维度, 系统提出了针对古村落文旅电商的体验优化与品牌差异化塑造策略。研究验证了文本挖掘与聚类分析在旅游用户细分中的有效性, 为古村落实现精准化运营和可持续发展提供了理论与实践指导。

## 关键词

古村落旅游, 网络评论, 语义分析, K-Means聚类, 用户画像

# Research on Experience Optimization and Brand Building of Ancient Village Cultural Tourism E-Commerce Based on Comment Semantic Mining and K-Means Clustering

Bingcong Zuo<sup>1</sup>, Jinfei Chen<sup>2</sup>

<sup>1</sup>School of Management Engineering, Shandong Jianzhu University, Jinan Shandong

<sup>2</sup>Business School, Shandong Jianzhu University, Jinan Shandong

Received: November 28, 2025; accepted: December 11, 2025; published: December 31, 2025

文章引用: 左秉聪, 陈金菲. 基于评论语义挖掘与 K-Means 聚类的古村落文旅电商体验优化与品牌塑造研究[J]. 电子商务评论, 2025, 14(12): 5637-5651. DOI: 10.12677/ecl.2025.14124534

## Abstract

Ancient villages, as significant cultural heritage sites, see their high-quality cultural tourism development playing a vital role in rural revitalization. With online travel platforms becoming primary channels for tourist consumption, online reviews have emerged as a key data source for optimizing experiences and shaping brands. This study focuses on Chinese ancient villages listed in the UN-WTO's "Best Tourism Villages" initiative, collecting 9334 tourist reviews from platforms like Ctrip. By employing sentiment analysis and LDA topic modeling to uncover overall sentiment tendencies and core discussion themes within the reviews, this research innovatively integrates the K-Means clustering algorithm. It constructs feature vectors combining sentiment scores and topic distributions for user segmentation. The study successfully identified six distinct tourist clusters: "Environmental Critics", "Nature Enthusiasts", "Rural Life Seekers", "History & Culture Explorers", "Service Experience Focusers", and "Unique Value Seekers". Based on this multi-dimensional user insight, the study systematically proposes strategies for experience optimization and brand differentiation for ancient village cultural tourism e-commerce across four dimensions: product, service, marketing, and branding. This research validates the effectiveness of text mining and clustering analysis in tourist segmentation and provides theoretical and practical guidance for the precise operation and sustainable development of ancient villages.

## Keywords

Ancient Village Tourism, Online Reviews, Semantic Analysis, K-Means Clustering, User Profile

---

Copyright © 2025 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

古村落作为重要的文化遗产，其旅游开发对乡村振兴与文化传承意义重大。随着数字经济发展，携程、美团等文旅电商平台已成为游客预订古村落相关服务的主要渠道。然而，当前古村落旅游普遍存在产品同质化、文化体验浅层化、商业氛围过浓等问题，导致游客体验不佳、品牌辨识度弱，制约其可持续发展。在当今数字时代，游客在电商平台发布的在线评论成为反映其真实体验的“数据富矿”。这些非结构化文本客观呈现了游客的情感态度、关注焦点与价值诉求，是诊断文旅痛点、挖掘品牌要素的关键依据。传统问卷调查等方法难以高效处理海量评论文本，而基于自然语言处理技术的评论挖掘已成为旅游管理研究的前沿。本研究聚焦“古村落文旅电商”场景，以语义分析为核心方法，结合 K-Means 聚类算法，旨在实现三方面目标：宏观上，通过情感分析与 LDA 主题模型把握游客评价的整体基调与核心议题；微观上，通过用户分群构建立体画像，识别差异化需求；应用层面，建立“用户画像 - 体验优化 - 品牌塑造”的传导模型，为古村落文旅电商的精准运营与品牌建设提供策略支持。本研究对推动古村落文旅高质量发展具有重要理论与实践价值。

## 2. 文献综述

### 2.1. 古村落旅游与文旅电商研究

国内外关于古村落旅游的研究已形成较为丰富的体系。早期研究多集中于文化遗产保护[1]、社区参

与[2]以及保护性开发规划[3]等传统领域。这些研究为理解古村落旅游的价值与可持续发展奠定了坚实基础。随着电子商务的普及,文旅电商的研究逐渐兴起,但焦点多集中于机票预订[4]、景区门票销售[5]和平台定价[6]等标准化程度较高的标准化、商业化程度较高的旅游场景和交易环节,然而针对古村落这一特殊文化载体的研究相对匮乏,尤其缺乏从游客消费后体验的视角,因此本文尝试将大数据分析方法应用于古村落旅游研究,旨在深化对文旅融合背景下古村落可持续发展的理解。

## 2.2. 网络评论在旅游研究中的应用

在线评论作为游客消费后体验的直接载体,已成为旅游研究的重要数据来源。学界普遍认为,在线评论是衡量目的地形象、游客满意度与忠诚度的关键指标。例如,刘逸等(2017)通过构建融合旅游专属词库、语义逻辑规则与情感乘数的情感评价模型,验证了基于在线评论大数据对旅游目的地进行情感评价的科学性与准确性,为旅游文本大数据的应用提供了关键方法论支持[7]。具体而言,分析手段已从利用ROST-CM等软件进行基础的词频与社会网络分析,发展为采用Python等工具进行情感分析、主题建模等更复杂的语义挖掘。例如,赵杨等(2018)综合运用K-Means聚类与深度学习模型,实现了对用户评论的多维度满意度测评[8]。然而,现有基于网络评论的研究存在明显的场景局限性与分析深度不足的缺陷。一方面,多数研究聚焦于大城市或标准化热门景区,对文化底蕴独特、体验更具敏感性的古村落关注不足。另一方面,尽管分析方法日益先进,但多数研究仍停留在对评论内容的整体描述层面,缺乏对异质性游客群体的深入细分,难以揭示不同游客类型的差异化需求与行为特征。因此,本文将网络评论分析聚焦于古村落这一独特场景,并致力于对游客群体进行精细化划分,旨在深入揭示文化型旅游地的游客体验内在结构差异,为古村落的精准运营与品牌提升提供理论参考。

## 2.3. 游客关注主题的跨目的地比较研究

与滨海旅游、城市旅游、主题公园等成熟业态相比,古村落旅游的游客关注主题具有显著特殊性。现有研究为这种差异提供了证据:滨海旅游吸引力侧重于“水域景观、旅游安全”等自然环境(廖倩倩等,2024)[9];城市文化旅游聚焦于“商业中心、艺术表演”等现代设施(周琰等,2024)[10];主题公园则围绕“活动参与、景区价格”等娱乐价值(林开森等,2019)[11]。这些现有研究为识别不同旅游形态的核心吸引力提供了有价值的参照系。然而,与之相比,针对古村落游客关注主题的系统性研究仍显不足,这恰恰凸显了针对该独特场景进行专门探讨的必要性。因此,本文旨在立足游客视角,通过分析在线评论数据,精准提炼古村落游客的核心关注主题,并进一步据此提出优化古村落文旅电商体验与塑造品牌差异化的策略,以弥补现有研究的不足。

## 2.4. 语义分析与聚类算法相关研究

情感分析、LDA主题模型[12]等技术已相对成熟。情感分析能有效量化文本的情感极性[13];LDA主题模型能无监督地提取文本中的隐含主题。K-Means聚类[14]作为经典的无监督机器学习算法,因其简洁高效被广泛应用于客户分群、推荐系统等领域。在旅游研究中,已有学者尝试运用K-Means对游客进行细分。例如,张瑜等(2023)通过构建包含餐饮、游览等维度的指标体系对游客消费体验进行聚类与评价[15]。但聚焦于古村落这一独特文化场景,利用聚类分析技术对游客进行深入细分的研究比较缺乏。特别是将K-Means聚类与文本语义分析(LDA、情感值)相结合,基于非结构化评论数据构建用户画像并提出针对性策略的研究,尚属探索前沿。

## 3. 研究设计与方法

本研究的技术路线主要包括数据获取、数据处理、语义分析、用户分群与策略推导四个阶段,具体

流程如下：数据采集(爬虫)→数据预处理(清洗、分词)→语义分析(情感分析 + LDA 主题建模)→特征向量构建→K-Means 聚类→用户画像生成→精准策略(体验优化 + 品牌塑造)。

### 3.1. 数据采集

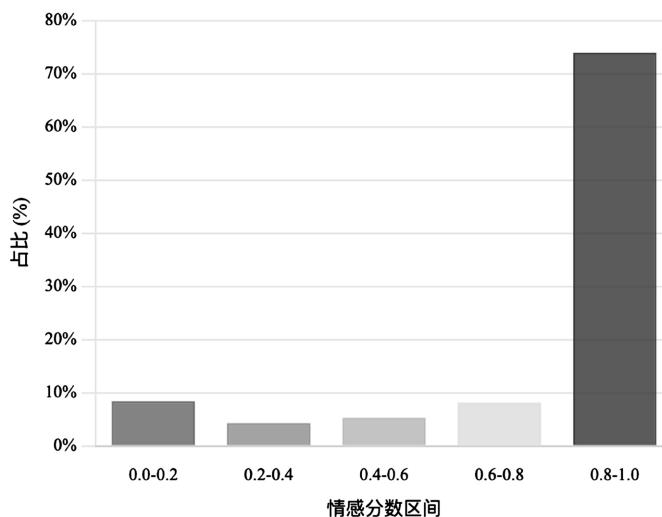
以国内知名古村落(纳入联合国“世界最佳旅游乡村”名录的 19 个村)为研究对象, 利用 Python 编写网络爬虫程序, 抓取“携程网”上的相关景区游客评论数据。采集字段包括评论正文、评分、评论时间、用户昵称等, 共计 9674 条。

### 3.2. 数据预处理

在数据预处理阶段, 为确保研究数据的时效性与有效性, 本文对获取的 9674 条在线评论进行了系统清洗。流程如下: 首先, 进行时间筛选。以数据采集日为基准, 将评论发布时间标准化后, 仅保留 2019 年 1 月 1 日至 2025 年 11 月 11 日期间的记录, 以聚焦近期用户反馈, 避免陈旧数据带来的时序偏差。其次, 执行缺失值处理。在确定时间窗口后, 检查核心字段(如评论内容、评分)的完整性, 对关键信息缺失的记录予以剔除, 确保后续分析的完整性。随后, 开展文本清洗。对评论内容进行去噪处理, 包括移除特殊字符、表情符号、广告文本与超链接, 实施繁简转换, 并基于停用词表过滤无实义词汇, 以提升语料质量。然后, 进行有效性与唯一性校验。核查评分是否处于合理区间(1~5 分), 并依据评论 ID 或内容识别并删除重复条目, 保证数据有效与独立。最后, 使用 Jieba 分词工具[16], 辅以古村落相关自定义词典, 对评论文本进行中文分词。经上述处理, 最终获得 9334 条高质量有效评论, 为后续情感分析与主题建模提供了可靠数据基础。

### 3.3. 语义分析基础模块

#### 3.3.1. 情感分析



**Figure 1.** Bar chart showing sentiment analysis of the review text

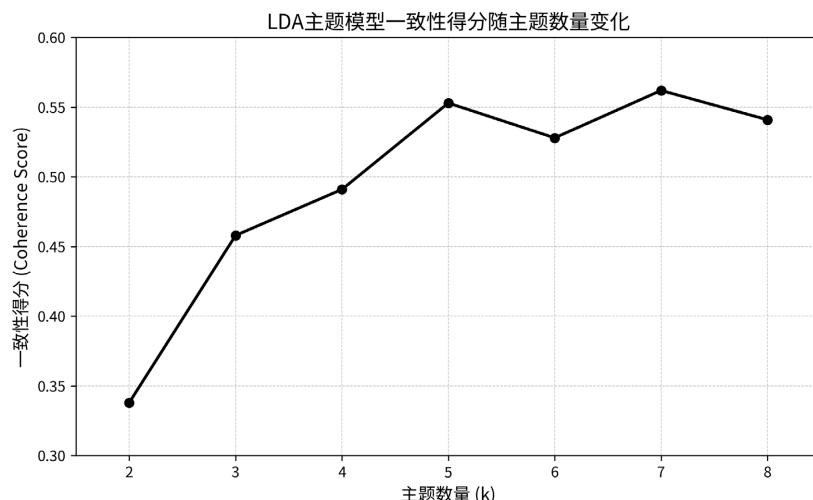
**图 1.** 评论文本情感分析柱状图

为系统量化用户评论文本的情感倾向, 本研究采用基于 SnowNLP 的中文情感分析框架[17]。该框架基于朴素贝叶斯分类算法, 利用大规模中文社交媒体语料训练而成, 能够有效识别文本中的情感特征。具体而言, 首先对预处理后的评论文本进行特征提取, 随后通过 SnowNLP 情感分类器计算每条评论属于正向情感的后验概率, 得到连续型情感得分(score ∈ [0, 1]), 得分越高表明文本情感倾向越积极。最终,

此得分将作为“情感特征”(见3.4.1节)并入特征向量,从而将抽象情感转化为可量化的模型输入变量。情感分析结果如图1所示,用户评论的情感分布呈现显著正向聚集特征:在全部9334条评论中,高达73.96%的评论情感得分集中在0.8~1.0区间(共6904条),而其他分数段合计占比不足26%。这一分布格局清晰表明,绝大多数用户持积极评价态度,但同时也反映出部分用户持消极评价的情况。该结果印证了情感分析在主题挖掘中的关键作用:若仅关注文本主题而忽略情感维度,则难以全面把握用户态度的整体倾向与强度差异。通过将情感特征纳入分析框架,本研究为用户态度理解提供了量化依据,并有助于揭示潜在的评价动机与满意度驱动因素,从而为数据驱动的实证研究奠定了基础。

### 3.3.2. LDA 主题模型分析

本研究运用GenSim库中的LDA主题模型对评论文本进行主题挖掘。首先,通过计算不同主题数( $K$ )下模型的一致性分数来评估模型质量,结果如图2所示。当主题数设定为5时,一致性曲线出现明显拐点,分数达到一个相对较高且开始趋于稳定的水平。因此,本文选择5作为最优主题数,以期在模型的表达力与解释性之间取得最佳平衡。在此基础上,对预处理后的9334条古村落景区评论数据进行LDA模型训练,并提取每个主题中概率最高的特征词,从而识别出游客讨论的核心内容维度。分析结果最终归纳出五个核心主题,其主题词分布如表1所示。



**Figure 2.** Curve of consistency calculation scores  
**图2.** 一致性计算得分曲线

**Table 1.** Analysis of tourist focus dimensions based on LDA topic model ( $K = 5$ ,  $N = 9334$  reviews)  
**表1.** 基于LDA主题模型的游客关注维度分析( $K = 5$ , 数据量: 9334条评论)

| 主题名称   | 占比    | 核心关键词                 | 主题解读                   |
|--------|-------|-----------------------|------------------------|
| 自然生态环境 | 约 25% | 自然、生态、空气、环境、绿色、竹林、荷花  | 反映游客对古村落绿水青山与生态保护的高度重视 |
| 历史文化建筑 | 约 22% | 建筑、历史、文化、特色、古建筑、保护、价值 | 体现游客对传统建筑特色与文化遗产价值的关注  |
| 旅游体验服务 | 约 21% | 服务、设施、管理、旅游、体验、导游、停车  | 凸显游客对景区服务质量与配套设施的需求    |
| 乡村生活文化 | 约 18% | 乡村、生活、文化、民俗、传统、村民、慢生活 | 指向游客对传统生活方式与文化体验的向往    |
| 具体景区特色 | 约 14% | 西递、宏村、徽派、这里、有趣、独特、只有  | 强调游客对景区个性化差异与独特体验的重视   |

该主题结构表明, 游客对古村落景区的关注呈现多元化特征, 其中自然生态环境为议题核心。该结果为古村落景区服务质量优化与精准营销策略制定提供了数据支持, 例如针对生态旅游爱好者突出“绿水青山”理念, 或面向文化探索者强化历史建筑保护与展示。

### 3.4. 特征向量构建

为将非结构化的评论文本转化为可用于聚类分析的数值化数据, 本研究构建了一个多维特征向量以全面表征每条评论的核心信息。特征工程主要包括以下两个步骤: 多维度特征提取与数据标准化。

#### 3.4.1. 多维度特征提取

特征向量由两大类特征构成, 分别捕捉评论的情感属性与语义内容:

① 情感特征。作为衡量游客情感倾向的指标, 本研究直接从情感分析结果中提取 sentiment\_score。该特征是一个连续型数值, 其值域为[0, 1], 数值越高代表评论的情感极性越积极。

② 主题特征。为量化评论的语义内容, 本研究利用已训练好的 LDA 主题模型。对于每条评论  $d$ , 模型会输出一个其在五个主题上的概率分布向量  $\vec{T}_d$ , 表示为:

$$\vec{T}_d = [P(topic_1|d), P(topic_2|d), P(topic_3|d), P(topic_4|d), P(topic_5|d)]$$

其中, 每个概率值  $P(topic\_id)$  表示评论  $d$  归属于第  $i$  个主题的似然度, 且满足  $\sum_{i=1}^5 P(topic_i|d) = 1$ , 该向量有效刻画了评论内容在不同关注维度(如自然生态、历史文化等)上的权重分配。

最终, 每条评论  $d$  被映射为一个六维的特征向量:

$$\begin{aligned}\vec{V}_d &= [sentiment\_score, \vec{T}_d] \\ &= [sentiment\_score, P(topic_1|d), P(topic_2|d), P(topic_3|d), P(topic_4|d), P(topic_5|d)]\end{aligned}$$

至此, 整个数据集(共 9334 条有效评论)被转化为一个规模为  $9334 \times 6$  的特征矩阵, 为后续的聚类分析奠定了基础。

#### 3.4.2. 数据标准化

由于特征向量中各维度的数值范围与分布存在差异, 为避免某些特征因其量纲或数值范围较大而在聚类算法中占据主导地位, 本研究对所有特征进行了标准化预处理。

具体采用 Z-score 标准化方法, 使用 StandardScaler 将每个特征维度的数据转化为均值为 0、标准差为 1 的标准正态分布。对于特征  $j$  的某个原始值  $x$ , 其标准化值  $x'$  的计算公式为:

$$x' = \frac{x - \mu_j}{\sigma_j}$$

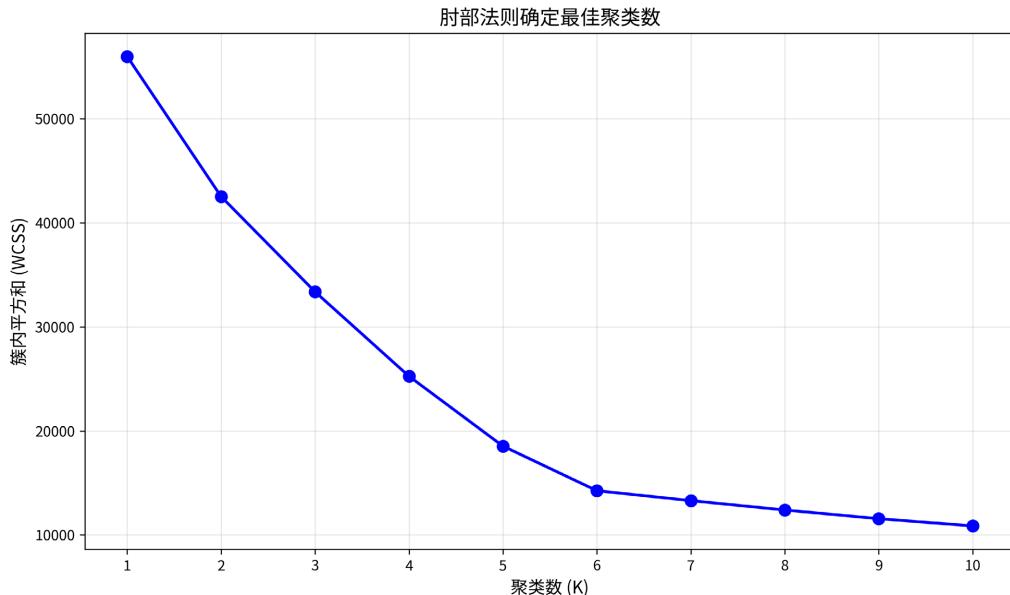
其中,  $\mu_j$  和  $\sigma_j$  分别是特征  $j$  在训练集上的均值和标准差。此处理确保了所有特征在后续的聚类分析中具有可比性, 从而提升模型的性能与稳定性。

### 3.5. 基于 K-Means 聚类的用户分群与画像生成

#### 3.5.1. 聚类数量确定与用户群体划分

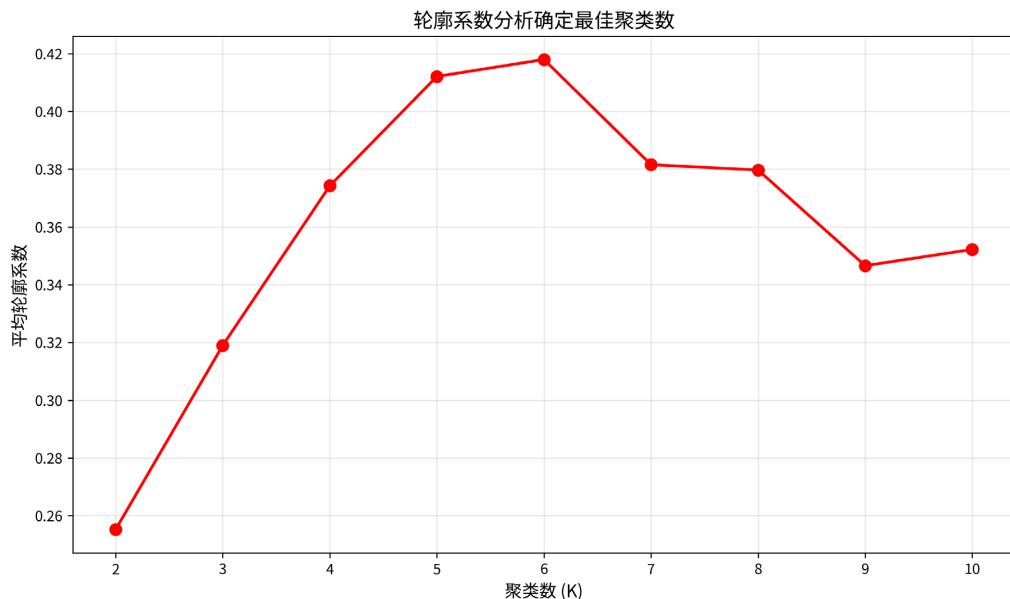
为确定最优的聚类数量( $K$  值), 本研究综合运用肘部法则(Elbow Method)与轮廓系数(Silhouette Coefficient)进行评估。肘部法则通过绘制不同  $K$  值对应的簇内平方和变化曲线, 以曲线的拐点(即弯曲幅度最大的点)作为  $K$  值选取依据; 轮廓系数则量化了样本与自身簇及其他簇的分离程度, 其取值范围为[-1, 1],

数值越大表明聚类效果越好。结果如图 3 与图 4 所示。肘部法则曲线在  $K = 6$  之后弯曲幅度显著减小，趋于平缓，表明继续增加聚类数的收益有限；而轮廓系数在  $K = 6$  时达到峰值 0.418 (虽未达到高度分离水平，但已为当前数据集最优)。综合两项指标的评估结果，并权衡模型简洁性与聚类质量，最终确定本数据集的最优聚类数目为  $K = 6$ 。



**Figure 3.** The elbow method determines the optimal number of clusters

**图 3. 肘部法则确定最佳聚类数**



**Figure 4.** The silhouette coefficient analysis determines the optimal number of clusters

**图 4. 轮廓系数分析确定最佳聚类数**

在确定最优  $K$  值后，执行 K-Means 聚类算法，将全部 9334 条评论数据划分为 6 个互斥的用户群体，

对每个簇进行统计分析, 计算簇内的高频词、平均情感分值、各主题权重分布等, 结合主题标签, 为每个游客群体进行命名和画像描述。聚类结果如表 2 所示, 各簇规模分布相对均衡, 介于 6.0% 至 24.9% 之间。

**Table 2.** Analysis of user group characteristics based on K-Means clustering (K = 6, Average Silhouette Score = 0.418)  
**表 2.** 基于 K-Means 聚类的用户群体特征分析结果(K = 6, 平均轮廓系数 = 0.418)

| 聚类标签 | 集群规模(占比)     | 平均情感得分 | 主导主题(权重)       | 群体特征解读           |
|------|--------------|--------|----------------|------------------|
| 簇 0  | 1425 (15.2%) | 0.1442 | 自然生态环境(0.7250) | 对自然环境不满的用户群体     |
| 簇 1  | 2099 (22.5%) | 0.8989 | 自然生态环境(0.6628) | 对自然环境满意的用户群体     |
| 簇 2  | 1466 (15.7%) | 0.9199 | 乡村生活文化(0.6167) | 对乡村文化体验满意的用户群体   |
| 簇 3  | 2093 (22.4%) | 0.8889 | 历史文化建筑(0.6556) | 对历史文化感兴趣的用户群体    |
| 簇 4  | 1741 (18.6%) | 0.9123 | 旅游体验服务(0.6537) | 对旅游服务体验满意的用户群体   |
| 簇 5  | 507 (5.6%)   | 0.8215 | 具体景区特色(0.5563) | 对景区独特性有特别关注的用户群体 |

综上, 本用户分群模块成功地将非结构化的评论数据转化为具有明确语义标签的离散用户群体。此分群结果不仅揭示了隐藏在评论数据背后的差异化用户模式, 更为本研究后续的用户画像深度生成与精准化策略制定提供了关键依据。

### 3.5.2. 用户分群特征解读

此聚类分群结果不仅揭示了隐藏在评论数据背后的差异化用户模式, 同时显示用户群体不仅因内容偏好(主题)而区隔, 更因情感倾向而细分。这一点在簇 0 与簇 1 的对比中尤为显著, 二者均以“自然生态环境”为核心主题, 但情感得分却呈现出极端的对立(0.1442 vs. 0.8989)。这表明, 传统的主题分析若忽略情感维度, 可能会掩盖重要的用户态度差异, 从而简化了复杂的用户反馈结构。基于分群结果, 六个簇的特征可归纳为表 3 所示:

**Table 3.** Characterization of user segmentation results  
**表 3.** 用户分群结果特征归纳

| 群体分类       | 包含簇号      | 群体规模  | 情感特征            | 核心特征与战略意义   |
|------------|-----------|-------|-----------------|---|
| 高满意度群体     | 簇 1、2、3、4 | 79.2% | 情感得分均超过 0.88    | 构成了评论的主体, 在自然生态、乡村文化、历史建筑和旅游服务等核心维度上获得了良好体验, 证明了景区在这些方面的整体吸引力 |
| 高关注且低满意度群体 | 簇 0       | 15.2% | 情感得分极低 (0.1442) | 极低的情绪分与高主题权重形成强烈反差, 是不容忽视的“预警信号”群体, 需重点关注其反馈内容                |
| 独特价值关注群体   | 簇 5       | 5.6%  | 情感得分中等 (0.8215) | 规模最小, 关注点更为聚焦和独特, 代表对景区个性化价值有特殊需求的细分市场                        |

### 3.5.3. 精细化用户画像构建

基于上述特征, 本研究为每个簇赋予了具象化的用户画像标签与内涵。如下:

① 环境批判者(簇 0)。“环境批判者”是基于聚类分析识别出的关键用户群体。其画像特征表现为一种鲜明的矛盾性: 在主题分布上高度聚焦于“自然生态环境”(主题权重达 0.725), 但平均情感得分却位列最低(仅为 0.1442)。文本分析进一步揭示了其负面态度的焦点: 高频词(图 5)如“商业化”、“嘈杂”、“失望”、“破坏”、“垃圾”等占据主导。典型评论如“商业化太严重了, 失去了古村的宁静”及“河

里能看到垃圾，生态保护令人担忧”。这些内容表明，该群体的负面评价主要源于对观察到的环境破坏、商业过度开发以及现实与预期落差的强烈不满，这反映出他们是一群对原生态与宁静氛围有较高要求的游客。因此，该用户画像的价值在于，其评论数据构成了景区生态管理与商业规划的高敏感性预警指标。针对此群体的反馈进行有效干预，对防范口碑风险、提升整体满意度和实现古村落可持续发展具有关键意义。



**Figure 5.** Online review word cloud of environmental critics  
**图 5. 环境批判者在线评论词云图**



**Figure 6.** Online word cloud of natural scenery enthusiasts' reviews  
**图 6.** 自然风光爱好者在线评论词云图

② 自然风光爱好者(簇 1)。“自然风光爱好者”是基于聚类分析识别出的规模最大的用户群体，其画像特征表现为高度的一致性：在主题分布上显著集中于“自然生态环境”(主题权重达 0.6628)，同时平均情感得分位列前茅(达 0.8989)。文本分析进一步印证了其积极态度的来源：高频词(图 6)如“自然”“景色优美”“绿水青山”“空气清新”等占据主导。典型评论如“清晨的薄雾笼罩古村，山水如画，仿佛置身仙境”及“这里的自然环境保护得非常好，满眼皆绿，呼吸都是一种享受”。这些内容表明，该群体的积极评价主要源于对优美自然景观的深度满足以及对良好生态环境的充分肯定，反映出他们是一群以自然审美与生态体验为核心诉求的游客。因此，该用户画像的价值在于，其评论数据构成了景区自然旅游资源吸引力的核心实证依据。针对此群体的偏好进行产品优化与宣传推广，对巩固景区核心优势、提升游客满意度和实现古村落品牌差异化具有战略意义。

③ 乡村生活向往者(簇 2)。“乡村生活向往者”是基于聚类分析识别出的情感反馈最为积极的用户群体, 其平均情感得分高达 0.9199, 显著优于其他簇群。该群体的核心特征体现为对“乡村生活文化”主题的集中关注(主题权重为 0.6167), 呈现出对乡村生活体验的深度认同与情感共鸣。文本分析进一步揭示了其积极态度的具体内涵: 高频词(图 7)如“慢生活”“淳朴”“民俗”“宁静”“田园”等占据主导。典型评论如“在这里真正感受到了慢节奏的生活, 村民淳朴热情, 仿佛时光都慢了下来”及“体验了磨豆腐、做糍粑等传统民俗, 找回了久违的乡村记忆”。这些内容表明, 该群体的高度满意度主要源于对乡村原生生活方式、传统文化氛围及宁静自然环境的深度沉浸与情感共鸣, 反映出他们是一群追求精神放松、向往质朴生活的深度体验型游客。



**Figure 7.** Online review word cloud of those who aspire to rural life  
**图 7.** 乡村生活向往者在线评论词云图

④ 历史人文探索者(簇 3)。“历史人文探索者”是基于聚类分析识别出的文化导向型用户群体(集群规模占比 22.4%), 其画像特征表现为对历史文化内涵的深度关注与高度认可。该群体在主题分布上显著集中于“历史文化建筑”(主题权重达 0.6556), 且情感得分处于较高水平(0.8889), 反映出其文化体验满意度与情感投入的深度一致性。文本分析进一步揭示了其核心关注点: 高频词([图 8](#))如“古建筑”“历史”“文化”“传统”“保护”等占据主导地位。典型评论如“徽派建筑的雕刻工艺精湛, 每一处细节都承载着深

厚的历史底蕴”及“通过导游的讲解, 深刻感受到这座古村落的文化传承, 仿佛穿越时空与历史对话”。这些内容表明, 该群体的积极评价主要源于对古建筑艺术价值的欣赏、历史文脉的感知以及文化遗产保护效果的认可, 反映出他们是一群注重知识获取、追求精神共鸣的深度文化体验者。因此, 该用户画像的价值在于, 其评论数据不仅验证了景区文化遗产资源的吸引力, 更构成了文旅融合深度发展的关键依据。针对此群体对专业性、知识性文化体验的核心诉求, 优化深度解说系统、开发互动式文化项目及强化遗产保护展示, 对提升文化旅游品质、强化学术影响力及实现古村落可持续发展具有重要战略意义。



Figure 8. Online review word cloud of historical and cultural explorers  
图 8. 历史人文探索者在线评论词云图

⑤ 服务体验关注者(簇 4)。“服务体验关注者”是基于聚类分析识别出的服务导向型用户群体(集群规模占比 18.6%), 其画像特征表现为对旅游服务品质的高度关注与明确认可。该群体在主题分布上集中于“旅游体验服务”(主题权重达 0.6537), 情感得分处于较高水平(0.9123), 体现出其对服务质量的满意度与情感认同的一致性。文本分析进一步揭示了其关注焦点: 高频词(图 9)如“服务”“设施”“便捷”“管理”“导游”等占据主导。典型评论如“景区接驳车班次密集, 工作人员态度友善, 整体体验非常顺畅”。这些内容表明, 该群体的积极评价主要源于对景区基础设施完善度、服务流程效率及人员服务态度的系统性认可, 反映出他们是一群重视旅游便利性、追求舒适体验的务实型游客。因此, 该用户画像的价值在于, 其评论数据直接反映了景区综合服务体系建设的效果, 是衡量旅游接待能力成熟度的重要指标。针对此群体对标准化、高效化服务的核心诉求, 持续优化智慧导览、无障碍设施、服务人员培训等环节, 对提升游客满意度、塑造专业可靠的旅游目的地形象具有关键作用。

⑥ 独特价值探寻者(簇 5)。“独特价值探寻者”是基于聚类分析识别出的差异化需求导向型用户群体(集群规模占比 5.6%), 其画像特征表现为对景区独特价值的专注追求与中等偏上的情感反馈(情感得分 0.8215)。该群体在主题分布上显著集中于“具体景区特色”(主题权重达 0.5563), 体现出其对目的地差异化吸引力的高度敏感性。文本分析进一步揭示了其核心诉求: 高频词(图 10)如“独特”“只有”“特色”“有趣”“发源地”等占据主导。典型评论如“这里是两山理论的发源地, 其生态文明的实践意义是其他古村落无法比拟的”。这些内容表明, 该群体的评价焦点集中于景区独有的文化符号、地标性资

源或创新性体验项目，反映出他们是一群追求稀缺性、注重目的地不可替代性价值的高认知游客。因此，该用户画像的价值在于，其评论数据既是景区差异化竞争优势的验证，也揭示了独特价值挖掘与传播的优化方向。针对此群体对稀缺性体验的核心诉求，进一步深化独特 IP 的体验设计、强化独特价值的叙事传播，对提升景区品牌辨识度、吸引高价值游客群体具有重要战略意义。



**Figure 9.** Service experience engagement: online review word cloud graphic

图9. 服务体验关注者在线评论词云图



**Figure 10.** Online review word cloud of unique value seekers  
**图 10.** 独特价值探寻者在线评论词云图

## 4. 讨论：基于数据驱动的精准优化与品牌塑造策略

基于 K-Means 聚类分析生成的六类立体化用户画像，以及对应的典型评论节选、高频特征词等直接数据证据，本研究的讨论部分将深入探讨如何将这些数据洞察转化为切实可行的管理策略。以下从数据驱动决策的价值、精准体验管理策略、品牌差异化塑造路径三个方面系统阐述研究成果。

### 4.1. 数据驱动决策在古村落旅游管理中的价值

传统旅游管理多依赖于宏观统计数据和经验判断，难以触及游客需求的微观差异。本研究构建的用户画像体系将抽象的用户群体转化为具象、可操作的数据指标，使管理者能够精准识别不同用户群体的独特需求与行为特征。例如，环境批判者(簇 0)的高频词云显示，该群体对“商业化”、“嘈杂”、“失望”等负面体验高度敏感，这与自然风光爱好者(簇 1)关注的“自然”、“景色优美”、“绿水青山”形成鲜明对比。这种基于数据的洞察力，使管理策略从“一刀切”转变为精准靶向，显著提升了决策的科学性与有效性。数据驱动的用户画像不仅提升了决策精度，还为景区资源配置提供了优化依据。通过分析各用户群体的规模、关注焦点及情感倾向，管理者可合理分配管理资源，优先解决负面口碑集中的问题，同时强化正面体验的亮点。例如，针对环境批判者反映强烈的商业化与垃圾问题，景区可调整商业布局，加强环境整治；而对于历史人文探索者(簇 3)关注的古建筑保护与文化传承，则可增加专业讲解服务，深化文化体验。这种基于数据的资源配置方式，能够实现有限资源的最大化利用，提升整体运营效率。

### 4.2. 基于用户画像的精准体验管理策略

#### 4.2.1. 负面体验的预警与补救机制

环境批判者(簇 0)的评论数据表明，该群体对环境破坏和过度商业化问题最为敏感。典型评论如“商业化太严重了，失去了古村的宁静”及“河里能看到垃圾，生态保护令人担忧”，直接指出了景区当前存在的痛点。针对这一问题，景区应建立实时评论监测系统，设置负面关键词预警机制，一旦发现相关评论即刻启动应急处置流程。同时，可设计“生态守护者体验路线”，邀请部分批判者参与环境巡查与保护活动，将其从批评者转化为共建者，有效化解负面情绪，提升口碑评价。此外，景区可建立问题响应闭环机制，对负面评论中反映的问题进行分类、归口、整改和反馈，并通过官方渠道展示改进成果。这种公开、透明的处理方式不仅能够挽回用户信任，还能展现景区负责任的态度，为品牌形象[18]增添正面价值。

#### 4.2.2. 高满意度群体的体验深化策略

针对自然风光爱好者(簇 1)与乡村生活向往者(簇 2)这两个高满意度群体，其评论数据反映了对自然景观与乡村生活的深度认同。前者高频词集中于“自然”、“景色优美”、“空气清新”等，后者则聚焦于“慢生活”、“淳朴”、“民俗”等元素。对此，景区可设计时序主题产品，如“四季摄影之旅”、“农耕体验活动”等，将静态资源转化为动态体验，满足用户对深度旅游的需求。

对于历史人文探索者(簇 3)，其评论中频繁出现的“古建筑”、“历史”、“文化”等词汇，表明该群体对文化深度与专业性的重视。景区可开发 AR/VR 导览系统，还原历史场景，增强文化体验的沉浸感。同时，推出“专家讲解团”、“非遗工坊”等高端产品，满足该群体对知识深度的需求。

服务体验关注者(簇 4)对“服务”、“设施”、“便捷”等因素的关注度最高，景区可通过优化智慧导览系统、完善无障碍设施、强化员工培训等方式，提升服务流程的标准化与人性化，巩固该群体的满意度。

#### 4.2.3. 独特价值群体的产品创新策略

独特价值探寻者(簇 5)作为高价值小众群体，其评论中“独特”、“只有”、“发源地”等词汇的出现频率较高，反映出对稀缺性体验的追求。针对这一需求，景区可挖掘自身独有文化 IP，设计“限量版”高端定制产品，如“古建筑修复观摩”、“文化传承人面对面”等活动，满足该群体的个性化与专属感需求。

此外，景区还可结合该群体的特征，打造主题研学路线，将独特的自然与文化资源转化为具有教育意义和体验深度的旅游产品，进一步提升品牌差异化和市场竞争力。

### 4.3. 品牌差异化塑造路径

用户画像的分析结果为品牌差异化定位提供了直接依据。基于各群体的特征数据，景区可从“单一形象”传播转向“多元子品牌”叙事，实现精准化营销[19]。

针对自然风光爱好者和乡村生活向往者，品牌叙事应突出“心灵栖所，归隐田园”的形象，传递宁静、治愈的情感诉求。内容创作可围绕其评论中的“山水如画”、“慢生活”等关键词展开，通过视觉与文字的组合，强化自然与田园的品牌调性。

对于历史人文探索者，品牌需塑造“时光走廊，文明解码”的专业形象，强调景区的历史深度与文化价值。传播内容可侧重文物解读、历史故事等元素，吸引对该领域感兴趣的知识型游客。

服务体验关注者更注重旅游的便捷与舒适，品牌应主打“智慧古村，无忧之旅”的承诺，突出基础设施、服务流程等方面的优势。而针对独特价值探寻者，则可通过“秘境探寻，专属发现”的叙事，强调景区的独特性和不可替代性。

在营销渠道选择上，景区可根据不同用户群体的媒介使用偏好，实施跨平台精准推送。例如，在抖音等短视频平台面向自然风光爱好者推送视觉冲击力强的内容；在知乎、小红书等知识型平台面向历史人文探索者推送深度文化解读。这种基于用户画像的营销策略，能够有效提升传播效率和投资回报率。

## 5. 结论与展望

### 5.1. 研究结论

本研究以古村落文旅电商场景下的在线评论为数据基础，构建了一套融合情感分析、LDA 主题模型与 K-Means 聚类算法的“数据洞察 - 用户分群 - 策略推导”研究框架。主要结论如下：

#### 5.1.1. 方法有效性

实证表明，将非结构化的评论文本转化为“情感得分 + 主题分布”的特征向量，并应用 K-Means 聚类，能够有效识别出内在特征迥异的游客群体。本研究成功划分出六个具有清晰画像的群体，揭示了传统主题分析难以发现的“主题相同、情感相异”的复杂模式(如簇 0 与簇 1)，验证了该方法在游客细分中的优越性。

#### 5.1.2. 用户洞察深度

研究发现，古村落游客并非同质化的整体，而是依据其关注焦点和情感态度分化为“环境批判者”、“自然风光爱好者”、“乡村生活向往者”、“历史人文探索者”、“服务体验关注者”和“独特价值探寻者”六类典型群体。这一立体洞察精准刻画了不同游客的动机、行为与满意度，为精细化运营提供了靶心。

#### 5.1.3. 策略价值

基于分群结果，本研究从精准体验管理策略、品牌差异化塑造两个层面提出了系统性的策略建议。

策略强调了对负面口碑的预警与转化、对优势体验的深化与传播, 以及基于细分市场的品牌差异化叙事构建, 为古村落文旅电商实现从粗放管理到精准服务、从单一形象到多元品牌的价值升级提供了清晰的路径指引。

## 5.2. 研究局限与展望

本研究仍存在一定局限, 也为未来研究指明了方向。首先, 数据来源较为单一, 未来可纳入社交媒体(如小红书、抖音)的笔记和视频评论, 进行多源数据融合分析, 以获取更全面的用户画像。其次, 本研究主要基于评论文本内容, 未来可结合用户消费频次、浏览行为等结构化数据, 构建更丰富的用户特征向量。此外, 本研究侧重于策略构建, 后续研究可选取典型古村落进行策略落地实践, 并通过对比干预前后的评论数据来量化评估策略效果, 形成闭环验证。最后, 本研究聚焦于古村落, 该方法论有望应用于更广泛的旅游目的地(如历史文化街区、自然保护区等)的智慧管理与品牌建设研究。

## 参考文献

- [1] 冯骥才. 传统村落的困境与出路——兼谈传统村落是另一类文化遗产[J]. 民间文化论坛, 2013(1): 7-12.
- [2] 保继刚, 孙九霞. 雨崩村社区旅游: 社区参与方式及其增权意义[J]. 旅游论坛, 2008(4): 58-65.
- [3] 王小明. 传统村落价值认定与整体性保护的实践和思考[J]. 西南民族大学学报(人文社会科学版), 2013, 34(2): 156-160.
- [4] 程璠. 谁是最好的旅行机票 APP? [J]. 今日工程机械, 2014(6): 120, 122.
- [5] 潘秋玲, 曹三强. 中外旅游景区门票价格的比较研究——兼论门票价格的定价依据[J]. 地域研究与开发, 2008(1): 64-69.
- [6] 程贵孙. 旅游电子商务平台兼并对双边定价策略的影响[J]. 旅游学刊, 2017, 32(3): 20-28.
- [7] 刘逸, 保继刚, 朱毅玲. 基于大数据的旅游目的地情感评价方法探究[J]. 地理研究, 2017, 36(6): 1091-1105.
- [8] 赵杨, 李齐齐, 陈雨涵, 等. 基于在线评论情感分析的海淘 APP 用户满意度研究[J]. 数据分析与知识发现, 2018, 2(11): 19-27.
- [9] 廖倩倩, 程叶青, 费雪艳, 等. 基于游客感知的滨海旅游吸引力研究——以三亚市为例[J]. 旅游科学, 2024, 38(12): 93-110.
- [10] 周琰, 李永强. 城市旅游产品中市民价值、游客价值共创探讨[J]. 西藏发展论坛, 2024(5): 81-87.
- [11] 林开森, 郭伟锋, 鄂俊利. 基于网络文本的影视主题公园游客旅游体验质量研究[J]. 西昌学院学报(自然科学版), 2019, 33(4): 31-36.
- [12] 关鹏, 王曰芬. 科技情报分析中 LDA 主题模型最优主题数确定方法研究[J]. 现代图书情报技术, 2016(9): 42-50.
- [13] 王婷, 杨文忠. 文本情感分析方法研究综述[J]. 计算机工程与应用, 2021, 57(12): 11-24.
- [14] 杨俊闯, 赵超. K-Means 聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14, 63.
- [15] 张瑜, 周丽永. 基于 K-Means 聚类算法的西安市游客消费体验评价[J]. 现代商业, 2023(8): 73-78.
- [16] 石凤贵. 基于 Jieba 中文分词的中文文本语料预处理模块实现[J]. 电脑知识与技术, 2020, 16(14): 248-251, 257.
- [17] 赵志升, 靳晓松, 温童童, 等. 基于 Python-SnowNLP 的新闻评论数据分析[J]. 科技传播, 2018, 10(18): 104-105.
- [18] 李艳, 牛志文. 旅游景区品牌形象塑造研究[J]. 北京第二外国语学院学报, 2001(5): 58-66.
- [19] 李静. 基于大数据精准营销的网络营销策略研究[J]. 商业经济研究, 2017(11): 46-47.