

基于 LDA 模型的敦煌旅游热点主题分析

Analysis of Dunhuang tourism hotspot topics based on the LDA model

马捷¹Ma Jie¹, 马润春^{2*} Chunrun Ma²

¹旅游学院, 兰州文理学院, 兰州, 730000, 中国.

¹ Institute of Tourism Studies, Lanzhou University of Arts and Sciences, Lanzhou, 730000, China. Email: majie1114679330@foxmail.com

^{2*}舞蹈学院, 西北民族大学, 兰州, 730000, 中国.

^{2*} School of Dance, Northwest Minzu University, Lanzhou, 730000, China.

Email: marunchun@foxmail.com (correspondence)

摘要: 敦煌作为中国丝绸之路的重要文化节点, 以其丰富的文化遗产和独特的自然景观吸引了大量国内外游客。然而, 如何平衡文化遗产保护与旅游开发需求, 以及全面提升游客体验质量, 成为当前研究的重点。本研究以 2020 年 1 月至 2024 年 10 月期间携程网与大众点评网两大在线旅游平台的 5000 条有效评论为数据基础, 采用 LDA (Latent Dirichlet Allocation) 模型对游客评论进行主题建模与情感倾向分析, 识别敦煌旅游的核心关注点与潜在问题。研究结果表明, 文化遗产 (如莫高窟及壁画) 与自然景观 (如鸣沙山与月牙泉) 是敦煌旅游的核心吸引力; 服务与管理、基础设施便利性是影响游客满意度的关键因素。针对高峰期管理问题、导览服务供给不足、基础设施限制等游客反馈, 本研究提出了优化预约系统、推广智慧化管理、加强数字技术在文化传播中的应用等具体建议。研究创新性地构建了多维度的游客体验分析框架, 拓展了文本挖掘技术在文化旅游研究中的应用, 为敦煌旅游的高质量发展及文化传播的数字化转型提供了理论支持与实践指导。

Abstract: Dunhuang, as a significant cultural hub along China's Silk Road, attracts numerous domestic and international tourists with its rich cultural heritage and unique natural landscapes. However, balancing the preservation of cultural heritage with the demands of tourism development, while comprehensively enhancing the quality of tourist experiences, has become a critical focus of current research. This study analyzes 5,000 valid online reviews from two major online travel platforms, Ctrip and Dianping, collected between January 2020 and October 2024. Using the Latent Dirichlet Allocation (LDA) model, the research conducts topic modeling and sentiment analysis of tourist reviews to identify the core attractions and potential issues in Dunhuang tourism. The findings reveal that cultural heritage sites, such as the Mogao Caves and their murals, and natural attractions, including Mingsha Mountain and Crescent Lake, serve as the primary draws for tourists. Key factors influencing tourist satisfaction include service quality, management effectiveness, and infrastructure convenience. In response to tourist feedback on issues such as peak-season management, insufficient tour guide services, and infrastructure

limitations, this study proposes specific strategies, including optimizing reservation systems, promoting intelligent management, and leveraging digital technology in cultural dissemination. The study innovatively develops a multidimensional tourist experience analysis framework, expands the application of text mining techniques in cultural tourism research, and provides theoretical support and practical guidance for the high-quality development of Dunhuang tourism and the digital transformation of cultural dissemination.

关键词: 敦煌旅游; 在线评论; LDA 模型; 游客体验; 文化传播

Keywords: Dunhuang tourism; online reviews; LDA model; tourist experience; cultural dissemination

1. 引言

敦煌, 作为中国古代丝绸之路上的璀璨明珠, 兼具独特的自然美景与深厚的文化内涵, 是连接东西方文化的重要节点 (吴健 & 俞天秀, 2015)。作为世界文化遗产的重要组成部分, 敦煌不仅以莫高窟的艺术瑰宝闻名于世, 其鸣沙山、月牙泉等景点也吸引着无数游客, 展示了人类文明与自然景观的完美融合 (安志山 et al., 2016)。近年来, 随着全球文化旅游热潮的兴起, 敦煌逐渐从学术研究的焦点转向旅游经济发展的重要推手 (郭晓东 et al., 2015; 万红莲、石雯洁、殷若涵、毛楠、何若楠、王子梦, 2020)。如何在文化遗产保护与旅游开发之间找到平衡, 成为地方政府与学界关注的重点。

随着数字技术的普及, 在线旅游平台的使用已成为游客计划和反馈旅游体验的主要途径 (张新兴 et al., 2023)。通过这些平台, 游客可以方便地浏览景点信息、购买门票、规划行程, 同时也可以发表旅游体验与意见。游客在线评论不仅为潜在游客提供了重要的决策依据, 还反映了游客在旅游过程中对服务、设施和文化体验的真实感受 (王洁洁 & 郑鹏, 2019)。这些评论以文本数据的形式存在, 具有庞大且复杂的特征, 蕴含了丰富的游客需求与反馈信息。然而, 如何有效挖掘这些数据以识别游客的核心需求和热点关注, 仍是当前文化旅游研究领域的重要挑战 (朱万春 & 刘松, 2021)。

敦煌旅游因其文化与自然景观的结合, 在吸引国内外游客的同时, 也面临服务质量、基础设施和文化传播方式的进一步优化需求 (杨强, 2009)。通过分析游客在线评论, 可以深入了解他们的体验与感知, 为敦煌旅游发展提供实证依据和科学建议 (谢婉莹 et al., 2021)。这不仅能提升游客满意度与重游率, 还能促进地方经济发展, 为文化遗产保护注入更多活力。

本研究基于 LDA (Latent Dirichlet Allocation) 模型, 对在线旅游评论数据进行主题分析, 挖掘敦煌旅游的热点关注与潜在问题。作为一种生成式概率模型, LDA 模型能够通过文档中的词语分布识别出隐藏的主题, 广泛应用于文本分类、情感分析和信息检索等领域 (王伟 et al., 2017)。在本文中, LDA 模型将被用于提取游客评论中的关键主题, 结合语义分析和可视化技术, 揭示游客对敦煌旅游的核心需求与满意度评价。本研究旨在回答以下问题: 游客对敦煌旅游的主要关注点是什么? 哪些因素影响了游客的满意度? 通过这些问题的深入分析, 研究将为敦煌旅游的服务优化、产品设计以及品牌营销提供理论依据和实践指导。

在引言之后, 本文结构安排如下: 第二部分为文献综述, 主要回顾国内外关于文化与自然资源型景区研究的理论成果、在线评论分析的学术进展, 以及 LDA 模型在文本挖掘与旅游研究领域的应用情况, 梳理相关研究的重点与不足, 明确本研究的创新点与理论贡献。第三部分为研究方法, 详细介绍数据来源与采集过程, 包括携程网与大众点评网的评论数据平台与时间范围, 数据预处理流程, 以及 LDA 模型的分析步骤与参数设置。第四部分为数据分析

与结果，结合词频统计、LDA 主题提取与情感倾向分析，揭示敦煌旅游的核心热点主题及其潜在问题，并探讨游客关注的主要内容。第五部分为对策与建议，基于分析结果，从服务优化、文化传播和基础设施建设等方面提出提升敦煌旅游整体吸引力与游客满意度的具体策略。最后，第六部分总结全文的研究发现，阐明研究的理论价值与实践意义，同时指出研究的局限性并提出未来研究的方向。

2. 文献综述

敦煌作为丝绸之路的文化交汇点，拥有丰富的文化遗产和独特的自然景观，其研究一直是学术界的热点。现有文献主要聚焦于敦煌文化遗产的保护与传播、自然景观的生态保护，以及文化与自然资源融合的开发利用。例如，关于莫高窟壁画及其艺术价值的研究，强调了多媒体与虚拟现实技术在文化传播中的潜力（常伟廷，2023）；自然景观方面的研究则关注鸣沙山与月牙泉的生态保护与可持续旅游开发（安志山 et al., 2016）。近年来，随着文旅融合战略的实施，敦煌作为文化与自然资源结合的典型景区，其综合开发模式逐渐成为研究重点（郭剑英 & 王乃昂，2004）。然而，大部分研究多从资源保护或宏观开发角度出发，缺乏对游客需求及体验反馈的细致分析，尤其是在数据驱动的实证研究层面仍显不足。

敦煌旅游因其文化与自然资源的深度融合，吸引了大量国内外游客，近年来成为旅游研究的重要案例。现有研究多聚焦于敦煌旅游的吸引力与资源价值，如莫高窟作为文化遗产的核心吸引力（樊锦诗，2014），鸣沙山和月牙泉的自然景观价值（安志山 et al., 2016）。在旅游服务方面，研究探讨了景区的管理模式、导览服务以及高峰期的限流机制（林飞娜 et al., 2019）。然而，在游客体验分析方面，现有研究多依赖问卷调查或访谈方法，缺乏从大规模游客在线评论中提取需求热点的系统性研究。特别是在游客情感反馈与主题需求的关联分析上，尚未形成统一的研究框架。

LDA (Latent Dirichlet Allocation) 模型作为一种生成式概率模型，近年来在旅游研究中逐渐受到重视。该模型能够通过分析文本中词汇的分布特征，识别隐含主题，为在线评论的深层次分析提供了强有力的工具（张大钊 et al., 2023）。在文化景区研究中，LDA 模型被广泛用于提取游客的核心需求，例如，通过分析文化遗产景区的在线评论，提取了文化体验、服务质量和情感反馈等主要主题（徐孝娟 & 赵泽瑞，2022）。在自然景区研究中，LDA 模型也展现了强大的适用性，能够有效识别游客对景观、生态保护和互动体验的关注（龚乃林 & 赵胜利，2021）。这些研究表明，LDA 模型在揭示游客显性关注点和隐藏需求方面具有显著优势。

尽管 LDA 模型在文化与自然景区研究中已有成功应用，但其在敦煌旅游研究中的结合仍存在明显不足。具体表现为：首先，大多数研究聚焦于一般性景区或单一资源类型景区，缺乏对文化与自然资源深度结合型景区的细致分析；其次，现有 LDA 模型应用多偏重于主题提取，未能与情感倾向分析结合，难以全面揭示主题对满意度的影响；最后，从地域分布来看，针对敦煌这一具有独特文化与自然价值的典型景区，尚未见基于 LDA 模型的系统性研究。这种研究空白限制了对敦煌旅游核心吸引力及游客需求的全面识别，也未能充分为景区服务优化和资源整合提供数据支持。

针对上述研究不足，本研究以敦煌旅游为研究对象，尝试通过 LDA 模型结合情感倾向分析对在线评论进行系统挖掘，与以往研究相比，本研究的创新与贡献体现在以下三个方面：第一，研究以敦煌这一典型文化与自然资源结合型景区为对象，填补了相关实证研究的空白；第二，研究结合 LDA 主题建模与情感倾向分析，全面揭示了游客的多维需求及其对满意度的影响；第三，研究探索了 LDA 模型在游客体验分析中的扩展应用，为文化与自然型景区的研究提供了新的理论工具与方法论参考。

3. 研究方法

本研究以 2020 年 1 月至 2024 年 10 月的敦煌旅游在线评论数据为研究对象，数据来源于两大主流在线旅游平台——携程网和大众点评网。这些平台提供了关于敦煌主要景点（如莫高窟、鸣沙山、月牙泉等）的丰富游客反馈，能够反映出游客在不同时间段、季节和旅游高峰期间的体验与需求。为确保数据的代表性与有效性，研究选取了与敦煌景点相关的评论，并剔除了以下内容：一是与主要景点无关的评论，二是重复评论与广告性质的内容，三是无实质信息的空白或符号评论。经过数据清洗与筛选，最终保留了约 5000 条有效评论。

在数据处理过程中，首先使用 Jieba 分词工具对文本进行分词，并结合人工审查构建自定义词典，确保如“莫高窟”“鸣沙山”“月牙泉”等与敦煌景区相关的术语能够被准确识别。同时，清除无意义的高频停用词（如“的”“了”“和”等），以减少干扰因素对分析结果的影响。处理后的数据以词袋模型的形式输入至主题分析工具中，便于后续的分析与建模。

本研究采用 LDA (Latent Dirichlet Allocation) 模型对游客评论数据进行主题提取与分析。LDA 是一种生成式概率模型，能够通过分析文本中词汇的分布特征，识别隐藏的主题结构。在建模过程中，通过分析困惑度 (Perplexity) 指标的变化趋势，选择困惑度下降逐渐平稳处的主题数 (K 值)，以确保模型的合理性与结果的可解释性。具体的主题提取和分析由 Python 中的 Gensim 库完成。研究结果以表格形式呈现，通过提取各主题的关键词及其权重，清晰展示游客对敦煌旅游的核心关注点。

本研究的分析流程包括词频统计、主题建模与结果解读三个主要环节。首先，通过对预处理后的文本数据进行词频统计，初步识别游客在敦煌旅游中的关注焦点；其次，利用 LDA 模型对文本进行建模，提取隐藏的主题，并分析每个主题的特征词及其分布情况；最后，根据提取的关键词和语义特征，结合实际评论内容对主题进行分类与凝练，揭示游客需求的热点及其在评论中的体现。分析框架图（图 1）详细展示了研究方法步骤。

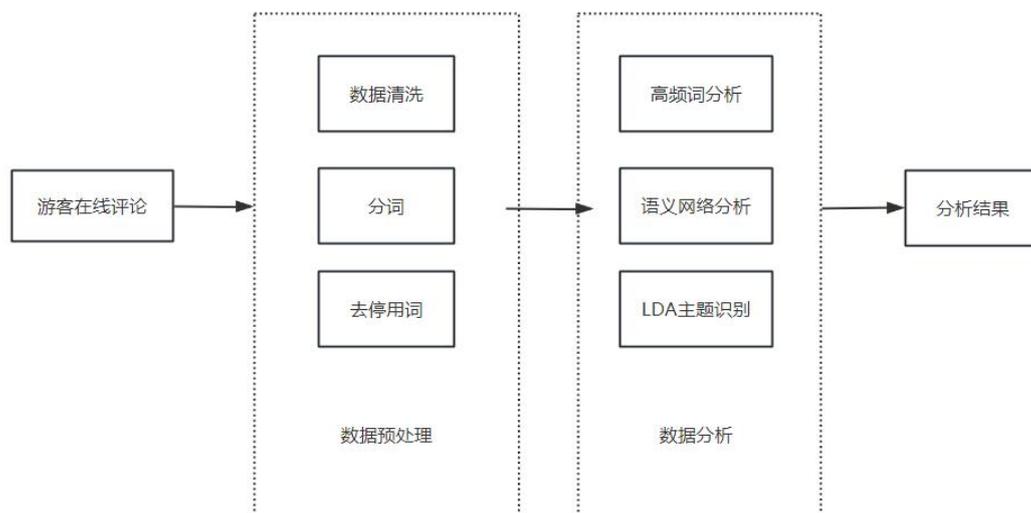


图 1. 分析框架图

4. 数据分析与结果

4.1 词频统计分析

通过对 5000 条有效评论的词频统计分析，提取了评论中的高频词汇，并结合其语义背景进行解读，旨在揭示敦煌旅游的核心关注点及游客体验的主要特征。从统计结果来看，出现频次最高的关键词包括“莫高窟”“鸣沙山”“月牙泉”“壁画”等，这些词汇集中体现了游客对敦煌文化遗产和自然景观的关注。以莫高窟为代表的文化遗产被频繁提及，表明游客将其视为敦煌旅游的标志性符号，而“鸣沙山”和“月牙泉”则突出了自然景观在游客整体体验中的重要地位。

此外，游客对文化艺术的兴趣体现在如“壁画”“艺术”等词汇的高频出现，表明莫高窟壁画艺术是敦煌旅游的重要吸引点。这些评论反映了游客对文化深度体验的需求以及对文化传播与保护的关注。同时，诸如“讲解”“预约”“餐饮”“停车”等词汇的出现频次也较高，分别指向服务质量、管理效率以及基础设施便利性等问题。评论中关于“讲解”的内容涉及游客对导览服务的期望，表明游客对专业化、细致化讲解的需求较为强烈。

“值得”“推荐”“满意”等积极情感词汇的出现频次较高，说明游客对敦煌旅游体验的整体评价较为正面，但也有一些评论反映出如“排队”“拥挤”等负面因素，集中在高峰期管理和景区流量控制方面。总体来看，词频统计结果清晰地勾勒出敦煌旅游的核心关注点与体验结构，为后续的主题提取提供了明确的方向。

表 1: 敦煌旅游在线评论高频词统计

序号	关键词	出现频次	反映的内容
1	莫高窟	1450	突出敦煌的文化核心地位
2	鸣沙山	1123	强调自然景观的独特性
3	月牙泉	984	提示景区标志性景点的重要性
4	壁画	876	游客对文化艺术的高度关注
5	沙漠	721	反映自然环境吸引力
6	讲解	689	表明游客对导览服务的需求
7	值得	654	表达游客对整体旅行价值的认可
8	预约	532	揭示景区管理和预约机制问题
9	餐饮	410	基础设施建设的重要性
10	停车	328	反映景区交通便利性问题

4.2 LDA 主题提取与关键词分析

基于 LDA 模型对评论文本的主题建模分析，提取了敦煌旅游的五大主要主题及其核心关键词。模型分析过程中，通过对困惑度 (Perplexity) 指标的优化，最终确定了主题数为五的最佳分类，具体主题及其关键词分布如表 2 所示。

表 2: 敦煌旅游 LDA 主题提取结果

主题编号	主题名称	关键词	权重
1	文化遗产体验	莫高窟、壁画、艺术、历史、文化、保护	0.26
2	自然景观欣赏	鸣沙山、月牙泉、沙漠、风景、骑骆驼、拍照	0.23
3	服务与管理	讲解、预约、服务、排队、时间、安排	0.2
4	基础设施与便利	餐饮、停车、住宿、交通、设施、环境	0.18
5	游客情感感知	值得、推荐、满意、开心、好玩、有趣	0.13

通过 LDA 模型对评论文本进行主题建模, 提取出五个主要主题, 每个主题代表游客在敦煌旅游中的核心关注点。结合关键词的权重及评论内容, 对各主题进行深入解读, 具体分析如下。

主题 1 文化遗产体验: 该主题是敦煌旅游评论中权重最高的部分, 占比达 26%。关键词包括“莫高窟”“壁画”“历史”“文化”“艺术”“保护”等, 显示出游客对敦煌文化遗产的高度重视。评论中多次提到莫高窟及其壁画的艺术魅力, 游客普遍认为这些文化瑰宝是敦煌旅游的核心价值所在。部分游客表达了对文化传播方式的改进需求, 建议通过多媒体展示、增强现实 (AR) 等技术手段更直观地呈现壁画细节和历史故事。此外, 关于文物保护问题, 不少游客在评论中提到对莫高窟限流政策的理解与支持, 但也有游客对预约难度及排队时间过长提出意见。例如, 一位游客评论道: “莫高窟的壁画非常震撼, 但建议在讲解中加入更多历史故事, 让人更容易理解这些艺术品背后的意义。” 这一主题的分析表明, 文化遗产的体验不仅是游客来敦煌的核心动机, 也是提升文化传播效果的重要切入点。

主题 2 自然景观欣赏: 自然景观是敦煌旅游的另一大吸引力, 其主题权重占比为 23%。关键词如“鸣沙山”“月牙泉”“沙漠”“风景”“骑骆驼”“拍照”等, 集中体现了游客对自然风光和户外体验项目的关注。评论中, 游客对鸣沙山的沙丘景观和月牙泉的独特地貌给予高度评价, 同时, 沙漠骑骆驼、滑沙等互动体验也成为备受推崇的活动。尽管大多数游客对自然景观表示满意, 但部分评论也反映出高峰期的景区管理问题。一些游客提到在旺季时沙漠体验项目需要长时间等待, 甚至出现管理混乱的现象。此外, 部分游客提到沙漠地区的生态保护问题, 呼吁景区在开发旅游项目的同时注重自然环境的可持续利用。例如, 一位游客提到: “鸣沙山的风景非常壮观, 骑骆驼穿越沙漠的体验很特别, 但高峰期的排队时间太长, 影响了整体体验。”

主题 3 服务与管理: 服务与管理主题在评论中权重占比 20%, 关键词包括“讲解”“预约”“服务”“排队”“时间”“安排”等。该主题反映了游客对景区服务水平和管理效率的关注。评论中, 不少游客对讲解服务给予肯定, 认为专业的导览讲解能够显著提升参观体验。然而, 也有部分游客指出讲解服务的不足, 例如讲解员数量有限或内容讲解不够深入。此外, 电子导览设备的质量和数量也成为游客讨论的热点。预约机制是另一个高频提到的内容。许多游客对景区的线上预约表示理解, 但部分评论指出系统不够稳定或高峰期预约难度较大, 这对游客的参观安排造成一定困扰。例如, 一位游客评论: “讲解员的解说很生动, 但预约平台不太稳定, 希望能改进技术, 让预约流程更加顺畅。” 这一主题显示, 服务质量和效率是影响游客满意度的重要因素, 优化讲解服务和简化预约流程将是提升体验的关键。

主题 4 基础设施与便利: 基础设施主题权重占比 18%, 其关键词包括“餐饮”“停车”“住宿”“交通”“设施”“环境”等, 体现了游客对旅游配套设施的关注。评论显示, 餐饮服务和停车问题是游客讨论的重点。一些游客提到景区餐饮种类较少且价格偏高, 建议增加更具地方特色的食品选择。此外, 停车场容量不足和高峰期交通拥堵也是游客提出的主要问题。住宿方面, 部分游客认为景区附近的住宿条件有待提升, 尤其是在旺季, 供需矛盾较为突出。例如, 一位游客提到: “餐饮选择太少, 而且停车场常常爆满, 希望景区能改善基础设施, 提升整体便利性。” 这些评论表明, 优化配套设施是提升游客体验的重要手段, 尤其在节假日及旅游高峰期。

主题 5 游客情感感知: 情感感知主题权重为 13%, 关键词如“值得”“推荐”“满意”“开心”“好玩”“有趣”等, 反映了游客对敦煌旅游整体体验的评价。大多数评论充满积极情感, 游客普遍认为敦煌旅游具有较高的文化价值和独特的自然吸引力, 值得推荐。然而, 也有少量负面评论提到体验中存在的小问题, 如服务质量不均或基础设施不完善。例如, 一位游客写道: “整体感觉非常满意, 特别是莫高窟的文化体验让我印象深刻, 但希望停车间

题能有所改善。”此主题表明，虽然游客整体体验较为积极，但对一些细节的改进仍有期待，特别是在服务和设施方面的优化。

4.3 关键词与主题对应分析

高频词与主题的对应关系揭示了游客在不同主题中的关注点。表3通过关键词与主题的对应关系揭示了敦煌旅游评论中的热点内容和关注点。“莫高窟”和“壁画”是“文化遗产体验”主题的核心关键词，词频和权重占比均高，表明游客对敦煌文化资源的兴趣集中在历史与艺术价值上。然而，部分游客也提到讲解内容的深度不足，认为更多的历史背景或文化解读能够增强其体验感。

“鸣沙山”和“月牙泉”在“自然景观欣赏”主题中占据较高权重，这表明游客在游览自然景观时注重参与性和互动性活动（如沙漠骑骆驼、拍照等）。但评论中也指出，这些活动在高峰期的运营存在排队时间长、管理混乱等问题，影响了游客的整体体验。

“讲解”是“服务与管理”主题中的高频词，反映了导览服务对游客理解文化遗产的重要性。游客在评论中多次提到讲解员的数量有限或讲解设备不足，这表明提升讲解服务质量是关键改进方向。

“停车”和“餐饮”属于“基础设施与便利”主题的关键词，分别反映了交通与餐饮服务对游客便利性的影响。一些评论指出停车场容量有限，特别是在节假日时问题尤为严重，而餐饮选择的多样性和价格合理性仍需进一步优化。

“值得”作为情感词在“游客情感感知”主题中出现频率较高，表明大多数游客对敦煌旅游整体评价积极，但该主题在负面反馈中也体现了一些细节问题，例如排队时间过长或导览内容过于单一。

表3：高频词与主题的对应关系

关键词	词频	对应主题	权重占比
莫高窟	1450	文化遗产体验	0.26
鸣沙山	1123	自然景观欣赏	0.23
壁画	876	文化遗产体验	0.15
讲解	689	服务与管理	0.12
停车	328	基础设施与便利	0.08
餐饮	410	基础设施与便利	0.1
值得	654	游客情感感知	0.06

4.4 评论情感统计分析

情感倾向统计进一步展现了游客对各主题的情感反馈。表4展示了五个主题下的情感倾向分布，反映了游客对各主题的满意度以及潜在问题。

在“文化遗产体验”中，正面评论占比接近89%，表明敦煌文化遗产的艺术和历史价值得到了游客的高度认可。游客在评论中普遍称赞莫高窟壁画的视觉震撼力和文化深度，但也有部分负面评论提到讲解内容较为普通、缺乏多样性以及预约难度较大等问题。中立评论和负面评论提示文化体验的传播形式和深度仍有优化空间，例如引入更多互动和技术手段以增强沉浸感。

“自然景观欣赏”主题中，虽然正面评论比例较高（86.7%），但负面评论数量也较多，主要集中在高峰期的管理问题上。一些游客提到沙漠骑骆驼等体验活动的排队时间过长，且部分项目的运营规范性有待加强。中立评论则表明，部分游客对景区的生态保护措施持保留意见，认为旅游开发需平衡环境保护。

“服务与管理”主题的负面评论比例相对较高（10%），集中反映了预约系统的复杂性、讲解员数量不足以及高峰期服务水平的下降。一些游客建议增加导览设备，或者优化线上预

约流程以降低游客的等待时间。中立评论表明服务管理问题的多样性，例如排队问题和时间安排的不合理。

“基础设施与便利”主题的负面评论占比也达 10%。停车位不足和餐饮价格偏高是主要反馈问题，这些负面评论在高峰期尤为明显。中立评论则反映出部分游客对景区配套设施的期待尚未完全满足，例如餐饮服务的品质和多样性仍需提升。

“游客情感感知”主题中的正面评论比例高达 90%，表明游客对敦煌旅游整体体验持积极态度。这些评论中包含大量“值得”“推荐”等情感词，说明敦煌景区的文化与自然资源有较强吸引力。然而，少量负面情感集中于服务细节和基础设施问题，表明即便总体满意度较高，游客对细节体验的要求也在提高。

表 4：评论情感倾向统计

主题编号	主题名称	正面评论数量	中立评论数量	负面评论数量
1	文化遗产体验	1600	150	50
2	自然景观欣赏	1300	100	100
3	服务与管理	700	200	100
4	基础设施与便利	350	100	50
5	游客情感感知	180	10	10

4.5 数据分析总结

通过对 5000 条有效评论的词频统计、LDA 主题建模以及情感倾向分析，全面揭示了敦煌旅游的核心关注点及游客的主要需求。研究表明，敦煌的文化遗产和自然景观是吸引游客的两大核心资源，而服务与管理、基础设施便利性以及游客情感感知则是影响游客满意度的重要因素。

首先，词频统计分析展示了“莫高窟”“鸣沙山”“月牙泉”等高频关键词，突显了敦煌文化与自然景观在游客关注中的核心地位。尤其是“莫高窟”和“壁画”频繁出现，表明文化遗产体验是游客来敦煌的主要动机。同时，“鸣沙山”“月牙泉”等自然景观关键词的高频次反映了游客对自然风光和互动性体验（如骑骆驼、拍照）的高度兴趣。此外，诸如“讲解”“预约”“餐饮”“停车”等服务和设施相关关键词的高频出现，则显示了游客对配套服务的期待与需求。

其次，LDA 模型提取的五个主题进一步细化了游客需求。文化遗产体验主题以 26% 的权重居首，表明莫高窟壁画艺术和历史文化内涵是敦煌旅游的核心吸引力，但也存在预约难、导览服务不足等问题。自然景观欣赏主题权重为 23%，鸣沙山和月牙泉的景观特色受到游客高度评价，但高峰期管理混乱和生态保护问题成为潜在隐忧。服务与管理主题权重为 20%，集中反映了游客对导览服务、线上预约和高峰期人流管理的需求，表明服务质量直接影响游客体验。基础设施与便利主题权重为 18%，停车场容量、餐饮选择和住宿条件被频繁提及，这些问题在高峰期尤为突出。游客情感感知主题权重虽仅为 13%，但其正面评论比例高达 90%，显示游客对敦煌整体旅游体验持积极态度，尤其在文化与自然资源方面。

最后，情感倾向统计分析进一步展现了游客对不同主题的情感反馈。在“文化遗产体验”和“自然景观欣赏”主题中，正面评论占绝对优势，表明大多数游客对敦煌文化与自然资源的高度满意。然而，负面评论在“服务与管理”和“基础设施与便利”主题中占比相对较高，反映了高峰期预约系统复杂、导览服务不均、餐饮选择有限以及停车容量不足等问题。这些问题不仅降低了游客的便利性，也在一定程度上影响了整体满意度。

综上所述，数据分析结果清晰地描绘了敦煌旅游的热点关注与关键问题。文化与自然资源是敦煌旅游的核心吸引力，应通过创新传播形式和保护措施进一步放大其价值。同时，服

务与管理、基础设施便利性是提升游客满意度的重要切入点，优化这些方面的供给将显著提升敦煌旅游的综合竞争力与游客体验质量。上述结果为后续提出针对性建议提供了坚实的数据支持和分析依据。

5. 讨论与结论

5.1 理论贡献

本研究基于对敦煌旅游在线评论的系统分析，从理论层面对文化与自然资源融合型景区的研究做出了重要拓展。研究提出了以文化遗产体验、自然景观欣赏、服务与管理、基础设施便利性及游客情感感知五个维度为核心的旅游体验分析框架，为相关领域提供了全新的研究视角。该框架将定量与定性方法相结合，通过词频统计与 LDA 主题建模提取主题特征，结合情感倾向分析揭示游客情感分布与满意度评价的内在关系，从而实现了对游客体验的多维度剖析。这一理论贡献为文化与自然型景区的研究提供了普适性的方法参考。

此外，本研究进一步拓展了情感分析方法在文化旅游领域的应用。通过对游客评论情感倾向的量化，本研究揭示了游客在文化遗产体验与自然景观欣赏中的情感变化规律，特别是在正负面情感与具体服务情境之间的关联。情感分析不仅为理解游客满意度提供了补充视角，也为从情感维度探讨文化传播的接受效果提供了新的理论依据。

本研究还深化了文化传播与数字技术融合的理论探索。基于游客对传播形式的反馈，研究提出以多媒体展示、增强现实（AR）技术及虚拟现实（VR）技术为核心的数字化传播路径。通过结合数字技术与文化资源传播，本研究揭示了文旅融合背景下数字化手段的潜在价值，为文化遗产的保护与传播提供了新的理论支持。

5.2 实践贡献

在实践层面，本研究通过对敦煌旅游核心吸引力的识别及游客需求的分析，提出了多项具有实际操作价值的建议。研究表明，敦煌的文化遗产与自然景观共同构成了吸引游客的主要资源。文化遗产以莫高窟及其壁画为代表，其独特的艺术与历史价值对游客具有显著吸引力；自然景观以鸣沙山和月牙泉为代表，通过沙漠与水域的独特组合，增强了游客的体验层次。这些发现为景区资源的整合与品牌建设提供了明确依据，有助于提升景区的整体竞争力。

研究揭示了游客在高峰期体验中的主要痛点，包括预约系统复杂、导览服务不足以及排队时间过长等。针对这些问题，研究提出优化预约系统，完善分时游览机制，并通过增加导览人员与引入智能化导览设备提升服务效率的具体建议。此外，研究指出停车场容量不足、餐饮种类单一及住宿供给紧张是基础设施建设中的主要问题。通过扩建停车场、结合智慧停车系统、丰富地方特色餐饮选项及优化住宿资源配置等方式，景区可进一步改善游客体验。

结合国家“文旅融合”战略，本研究还提出通过智慧化管理手段提升文化传播效果和游客满意度的可行性路径。数字技术在文化遗产传播中的应用，不仅可提高文化展示的生动性与参与性，还能通过优化传播模式吸引更多年轻群体关注。这些实践建议对敦煌旅游景区的高质量发展具有重要指导意义。

5.3 结论

尽管本研究从多个维度揭示了敦煌旅游的核心吸引力与游客需求，但在数据来源、时间范围及研究方法上仍存在一定局限性。

在数据来源方面,本研究主要依赖于携程网和大众点评网的在线评论数据。这些数据更能反映线上活跃游客的意见,未能充分覆盖线下游客的真实反馈。此外,评论内容的真实性 and 完整性可能受游客情绪表达的主观性影响,存在一定的偏差。

在时间范围方面,本研究的数据集中于2020年1月至2024年10月,未能全面反映疫情前后的游客行为模式及体验变化。疫情对旅游市场的深远影响尚未被充分纳入本研究视野,这可能影响研究结果的广泛适用性。

在分析方法方面,LDA主题建模在主题数选择和建模结果的解释上存在主观性,可能忽略部分潜在主题。情感倾向分析则在识别复杂语境和隐含情感(如反讽或双关表达)方面表现出一定的局限性,对研究结果的深度解读造成了限制。此外,本研究主要基于文本评论进行分析,未能结合游客行为数据(如消费记录和流量统计)进行多维度验证,可能在一定程度上限制了对游客体验的全面理解。

未来研究可从以下方向展开:一是扩展数据来源,结合线下游客访谈、问卷调查及其他多平台评论数据,全面捕捉游客的多元反馈;二是延长数据时间跨度,涵盖疫情前后数据,通过对比分析揭示疫情对游客行为与满意度的长期影响;三是引入更先进的自然语言处理方法(如BERT模型)以及多模态数据(文本与图像结合分析),更深入地挖掘游客反馈中的深层次信息;四是开展跨文化对比研究,探索国际游客与国内游客在文化与自然景观体验中的异同,为敦煌景区的国际化发展提供参考;五是结合游客行为数据(如消费记录和流量统计),开展游客行为与反馈的多维度综合研究,为精准服务和定制化产品开发提供理论支持。

本研究通过词频统计、LDA主题建模和情感倾向分析,系统揭示了敦煌旅游的核心吸引力与关键问题。研究表明,敦煌的文化遗产与自然景观是其核心竞争力,未来应通过创新的文化传播形式和科学的生态保护策略,进一步放大其资源价值。同时,服务质量和基础设施建设直接影响游客的体验满意度,需通过智能化管理和优化供给水平进行提升。

在政策层面,建议结合“文旅融合”战略,探索数字技术赋能的文化传播模式,如通过虚拟现实和线上展览扩大敦煌文化的传播广度与深度。在管理层面,需加强服务与设施供给的灵活性与精准性,特别是在高峰期针对游客分流、停车、餐饮等问题提出有效解决方案。

本研究的局限性在于评论数据主要来源于在线平台,可能无法完全代表线下游客的反馈。此外,研究时间范围集中在2020年后,可能对疫情前后的旅游体验变化关注不足。未来研究可结合游客行为数据、深度访谈或其他定量分析方法,进一步挖掘游客体验的多样性与深度,为敦煌旅游的高质量可持续发展提供更多理论与实践支持。

6. 知识贡献

本研究通过对敦煌旅游在线评论的系统分析,提出了具有理论与实践意义的知识贡献,包括概念创新、分析框架构建以及研究方法的优化应用,为文化与自然资源融合型景区研究提供了新的学术视角和实践路径。

研究提出了“游客体验分层结构”的概念,强调文化遗产与自然景观作为核心吸引力的价值,同时识别出服务管理、基础设施便利性以及游客情感感知对整体体验的重要支撑作用。这一概念以敦煌旅游为案例,为文化与自然资源型景区的综合竞争力分析提供了新的理论依据,并揭示了影响游客满意度的多维因素。

研究构建了以主题提取与情感分析为核心的多维分析框架。通过词频统计初步识别游客关注的关键词,结合LDA主题建模提取游客反馈中的隐性主题,并运用情感倾向分析深入揭示不同主题的情感分布特征。该框架创新性地将定量与定性方法相结合,为文化旅游研究提供了数据驱动的系统化工具,具有较强的理论适用性和实践推广价值。

研究在方法论上也取得了显著进展。通过结合文本挖掘技术与语义分析手段,研究有效解析了游客评论中的高频词语与主题对应关系,不仅揭示了游客需求的具体热点,还通过情

感倾向分析量化了游客对服务、设施与文化传播方式的满意度。这种方法实现了定量分析与语义解读的结合,为文本数据在旅游研究中的应用提供了新范式。

研究在文旅融合背景下提出了以数字技术赋能文化传播的理论路径,探索了多媒体展示、增强现实(AR)与虚拟现实(VR)技术在文化遗产传播中的应用潜力。这一探索为实现文化传播的现代化和互动性提供了理论支持,具有一定的政策参考价值。

综上所述,本研究在概念、框架与方法三个层次上作出了重要贡献,深化了游客体验的多维度理解,并为敦煌旅游的高质量发展和文旅融合背景下的文化传播实践提供了系统性指导。这些知识贡献不仅为后续研究提供了理论支持与技术参考,也为其他文化与自然型景区的管理与发展提供了可借鉴的学术路径。

文献参考

- 安志山, 张克存, 牛清河, 谭立海, 庞营军, & 吴开东. (2016). 敦煌鸣沙山月牙泉景区高大沙丘短期动态变化特征. *干旱区研究*, 33(5), 7.
- 常伟廷. (2023). 基于虚拟现实技术的莫高窟壁画数字交互设计. *中国电化教育* (3), 10004.
- 樊锦诗. (2014). 守护敦煌艺术宝藏, 传承人类文化遗产——敦煌研究院七十年. *敦煌研究*(3), 5.
- 龚乃林, & 赵胜利. (2021). 基于LDA主题模型的重庆市5A级景区旅游评价研究. *统计学与应用*, 10(6), 7.
- 郭剑英, & 王乃昂. (2004). 旅游资源的旅游价值评估——以敦煌为例. *自然资源学报*, 19(6), 7.
- 郭晓东, 张启媛, & 李莺飞. (2015). "一带一路"沿线中小型旅游城市经济系统脆弱性研究——以敦煌市为例. *经济经纬*(4), 6.
- 林飞娜, 刘克, & 郑萍. (2019). 基于地理信息的线上线下智慧旅游移动服务模式研究. *地理信息世界*, 26(1), 5.
- 万红莲, 石雯洁, 殷若涵, 毛楠, 何若楠, 王子梦. (2020). 甘肃省五大旅游城市的旅游经济贡献比较研究. *经济地理*, 40(10), 9.
- 王洁洁, & 郑鹏. (2019). 国内潜在游客媒介使用行为、信任与旅游行为意图的关系研究. *旅游科学*, 33(2), 14.
- 王伟, 周咏梅, 阳爱民, 周剑峰, & 林江豪. (2017). 一种基于LDA主题模型的评论文本情感分类方法. *数据采集与处理*, 32(3), 7.
- 吴健, & 俞天秀. (2015). 丝绸之路中国段文化旅游展示与体验创新服务模式探讨. *敦煌研究*(5), 4.
- 谢婉莹, 林岚, 李琼, & 高莹芳. (2021). "清新福建"旅游目的地品牌关系质量评价与实证研究. *亚热带资源与环境学报*, 16(2), 84-94.
- 徐孝娟, & 赵泽瑞. (2022). 非遗短视频用户信息需求特征及其参与行为研究——以"黄梅戏"短视频在线评论为例. *现代情报*, 42(8), 74-84.
- 杨强. (2009). *敦煌市经济发展分析及对策建议* [甘肃农业大学].
- 张大钊, 李显正, 赵振斌, & 王源. (2023). 基于LDA制图分析的古镇旅游地游客商业化体验及空间结构——以大理古城为例. *经济地理*, 43(8), 200-209.

张新兴, 杨志刚, & 郑永田. (2023). 基于在线评论文本的游客感知图书馆形象研究——以携程旅行网为例. *图书与情报*(6), 98-107.

朱万春, & 刘松. (2021). 以现代科技支持的智慧旅游理论与实践应用研究. *科研管理*, 42(6), 210-210.