

基于LDA主题模型的景点在线评论文本挖掘分析

——以携程旅行南浔古镇景点评论为例

张清露¹, 张馨尹¹, 宋瑾钰^{2*}

¹浙江理工大学启新学院, 浙江 杭州

²浙江理工大学计算机科学与技术学院(人工智能学院), 浙江 杭州

收稿日期: 2025年4月24日; 录用日期: 2025年5月16日; 发布日期: 2025年6月19日

摘要

互联网的快速发展给各大景区带来机遇的同时, 也带来了挑战。很多人在互联网上进行景点评论搜索并根据景区评论选择目的地。通过分析游客的评论, 旅游景区可以更好地找到景区存在的问题, 以便提出相应的解决方案, 从而提升游客的满意度和旅游体验。本文利用爬虫技术对携程旅行上南浔古镇景点的用户评论进行爬取, 利用LDA主题模型、SnowNLP情感分析和TF-IDF特征提取技术, 对爬取的评论数据进行了全面分析, 深入探讨景区评论中消费者的满意度及情感分布, 了解消费者对旅游景区的核心需求, 为景区改善服务提供了数据支撑。

关键词

南浔古镇, 在线评论, 情感分析, 行为分析

Text Mining Analysis of Online Reviews of Scenic Spots Based on LDA Topic Model

—Taking Ctrip Nanxun Ancient Town Scenic Spot Review as an Example

Qinglu Zhang¹, Xinyin Zhang¹, Jinyu Song^{2*}

¹Qixin School, Zhejiang Sci-Tech University, Hangzhou Zhejiang

²School of Computer Science and Technology (School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Apr. 24th, 2025; accepted: May 16th, 2025; published: Jun. 19th, 2025

*通讯作者。

文章引用: 张清露, 张馨尹, 宋瑾钰. 基于 LDA 主题模型的景点在线评论文本挖掘分析[J]. 电子商务评论, 2025, 14(6): 1783-1793. DOI: 10.12677/ecl.2025.1461926

Abstract

The rapid development of the Internet has brought both opportunities and challenges to major scenic spots. Many people search scenic spot reviews on the Internet and choose destinations according to the reviews. By analyzing tourists' comments, tourist attractions can better find problems existing in scenic spots, so as to propose corresponding solutions, thus improving tourists' satisfaction and travel experience. This paper uses crawler technology to crawl the user reviews of scenic spots in Nanxun Ancient Town on Ctrip, and uses LDA theme model, SnowNLP sentiment analysis and TF-IDF feature extraction technology to conduct a comprehensive analysis of the crawled review data, deeply explore the satisfaction and emotional distribution of consumers in scenic spot reviews and understand the core needs of consumers for tourist attractions. It provides data support for the service improvement of scenic spots.

Keywords

Nanxun Ancient Town, Online Review, Emotion Analysis, Behavior Analysis

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

近年来，旅游已经成为人们日常生活中不可或缺的娱乐休闲方式。文化和旅游部发布的《2024年国庆节假期文化和旅游市场情况》数据显示，2024年国庆节假期，全国文化和旅游市场平稳有序。经文化和旅游部数据中心测算，国庆节假日7天，全国国内出游7.65亿人次，按可比口径同比增长5.9%，较2019年同期增长10.2%；国内游客出游总花费7008.17亿元，按可比口径同比增长6.3%，较2019年同期增长7.9% [1]。

同时，随着互联网票务服务的普及，旅游评价体系得到了显著发展。在用户认知水平不断提高的背景下，网络评价内容日益客观地呈现了游客的真实体验，逐渐演变为供需双方信息交互的关键渠道。这一现象推动了旅游服务提供商依据用户意见优化服务品质、提升行业竞争力，但是同时也带来了新的挑战：如何从庞大的评价数据中挖掘有效信息，为运营者提供决策支持，从而强化景区特色优势，解决景区现存不足。基于此，本研究聚焦于运用先进的数据处理技术，实现评价文本关键信息的智能提取和分析，为优化景区运营策略、提升对游客的吸引力提供科学依据。

2. 数据准备和研究方法

2.1. 数据来源

携程网(<https://www.ctrip.com>)作为中国领先的数字一体化旅游服务商，为线上、线下消费者提供包含酒店预订、旅游套餐、航班票务在内的产品与服务。本文选择携程网关于南浔古镇的游客评论作为研究对象，通过Python语言与Selenium工具的结合应用，经过对重复及无效空白评论的有效筛选，最终收集到2444条在线评论。对于具有高度相似性并且在短时间内发布的大量评论，或者使用一些特定的、夸张的词汇来吸引注意的评论，有较大的水军评论虚假评论的可能，给予屏蔽删除。因此，收集的数据具有

一定的代表性，适合用于深入的文本分析研究。

2.2. 研究方法

Python 语言凭借其丰富的第三方资源库在数据处理领域占据重要地位。诸如数值计算库 NumPy、数据可视化库 Matplotlib、数据处理库 Pandas 以及词云生成库 WordCloud 等工具，为文本信息处理提供了全面的技术支持[2]。这些工具集显著提升了 Python 在数据解析方面的能力，尤其在文本信息抽取和语言理解等应用场景中展现出卓越的性能。

在文本分析领域，潜在狄利克雷分布(LDA)模型作为一种基于概率图模型的主题建模方法，构建了文档-主题-词汇三层级联关系。该模型采用概率统计方法确定文档的主题归属，并通过概率分布表征主题特征。文档与主题之间呈现多对多映射关系，而每个主题则由一组具有代表性的词汇构成[3]。这一模型在海量网络评论文本的处理与解析方面具有独特优势，能够有效识别文本中潜在的主题结构及其对应的关键词分布。通过自动化提取文本主题特征，LDA 模型为洞察用户评价背后的情感倾向、观点态度和潜在需求提供了可靠的技术支撑，不仅拓展了文本分析的维度，也深化了研究的深度，在文本挖掘领域具有重要的应用价值。

本文首先通过网络爬虫技术自动采集携程网上关于南浔古镇的游客在线评论；其次，结合 Jieba 库进行分词，运用数据预处理技术清洗和整理所收集的原始数据；然后通过统计词频来识别高频词，并绘制词云图以直观展示。再采用 LDA 模型对预处理后的文本数据进行主题挖掘，以识别游客评论中的关键主题。完成后，利用 SnowNLP 库进行情绪分析，将文本评论分为积极评论和消极评论，利用 Jieba 库进行分词，将文本转为 TF-IDF 矩阵，计算文本的情感特征权重，生成消极和积极文本的权重结果文件，结合 LDA 主题关键词进行综合研究，深入探讨游客关注的核心问题及其满意度。

3. 数据收集与分析

3.1. 数据收集

运用 Python 网络爬虫技术，借助 Selenium 浏览器库，以“南浔古镇”、“旅游”作为关键词，自动抓取相关商品的全部评论内容，经过对重复及无效空白评论的有效筛选后保留了 2944 条评论信息。如表 1 所示，这些信息涵盖了用户 ID、评分、评论内容、IP、点赞数等多个维度。

Table 1. Comment data

表 1. 评论数据

用户	评分	评价文本	IP 地址	点赞数
fsx123	5 分	去了那么多古镇，这里是我认为最好的江南古镇，人也不多……	甘肃	2
M42****3333	5 分	第一次来南浔古镇，到了以后已经是 5 点多钟了，吃了一碗双交面……	浙江	1
Chen1299	5 分	南浔古镇本身是免费开放的，其中一些小景点单独收费。建议慢下……	浙江	1

在数据收集过程中，对网络爬虫的运行频率和并发量进行了严格控制，以确保数据的合法获取和有效性。所收集的数据仅供学术研究之用，不作其他用途[4]。

3.2. 分词处理和词频分析

文本挖掘中数据处理最基本的步骤便是分词。分词是将一个单词序列切分成单个单词的过程，是保证后续任务顺利进行的首要工作。本文选用最常用的 Jieba 分词包，它是一个专门的中文分词包，是 Python 内的一个分词开源库[5]。

Table 2. Word segmentation result show**表 2.** 分词结果显示

序号	用户评论
1	古镇 最好江南 古镇 人 景色 美丽 饭菜 好吃 晚上 白天 景色 各有千秋 拍 照片可惜以后 有空去 享受 美景 悠闲 时光
2	第一次 南浔 古镇 以后 吃 双交面 非常 不错 张静江故居 了解 传奇 故事 看 百间楼 夜景 三道 茶 收获 满满
3	南浔 古镇 本身 免费 开放 其中 一些 景点 单独 收费 建议 慢 逛 玩 体验 不同 时间 古镇 魅力 可以 考虑 古镇 看看 清晨 安静 闲适 古镇 美景

通过上表 2 可以看出，“的”、“也”、“了”等词语出现频率很高，此类无实际语义的词汇在文本分析中缺乏实质性研究价值，不仅可能降低数据分析的准确性，还会显著增加数据处理复杂度，因此需构建停用词表进行过滤处理。通过停用词过滤技术，原始文本冗余信息得以有效清除，词汇总量显著降低，具体词频统计结果详见表 3。此外，本文借助 Python 语言的 WordCloud 工具包生成词云可视化图谱，相关结果如图 1 所示。

Table 3. Word frequency statistics**表 3.** 词频统计

高频词	频数
古镇	3485
南浔	2152
江南	708
小莲庄	422
景色	375
百间楼	365
建筑	321
故居	294
历史	250
特色	237
安静	212
文化	204
风景	191
水乡	184
酒店	151
刘氏梯号	102
张石铭	97
嘉业堂	93
便宜	89
坐船	88



Figure 1. Word cloud map of Nannun Ancient Town online review
图 1. 南浔古镇在线评论的词云图

3.3. 基于 LDA 主题模型的特征分析

在 LDA 主题建模过程中，一致性分数(Coherence Score)是评估主题语义合理性的关键指标。其计算原理为：基于滑动窗口统计语料中词汇共现概率，计算点互信息(PMI)与余弦相似度，最终通过归一化处理得到 0~1 区间的评分值。较高的分数表明主题内词汇具有显著的语义关联性，例如在旅游评论分析中，“古巷”、“民宿”、“夜景”等词的高共现频率会提升主题一致性，而混杂“天气”、“晚上”等无关词汇则会降低该指标。

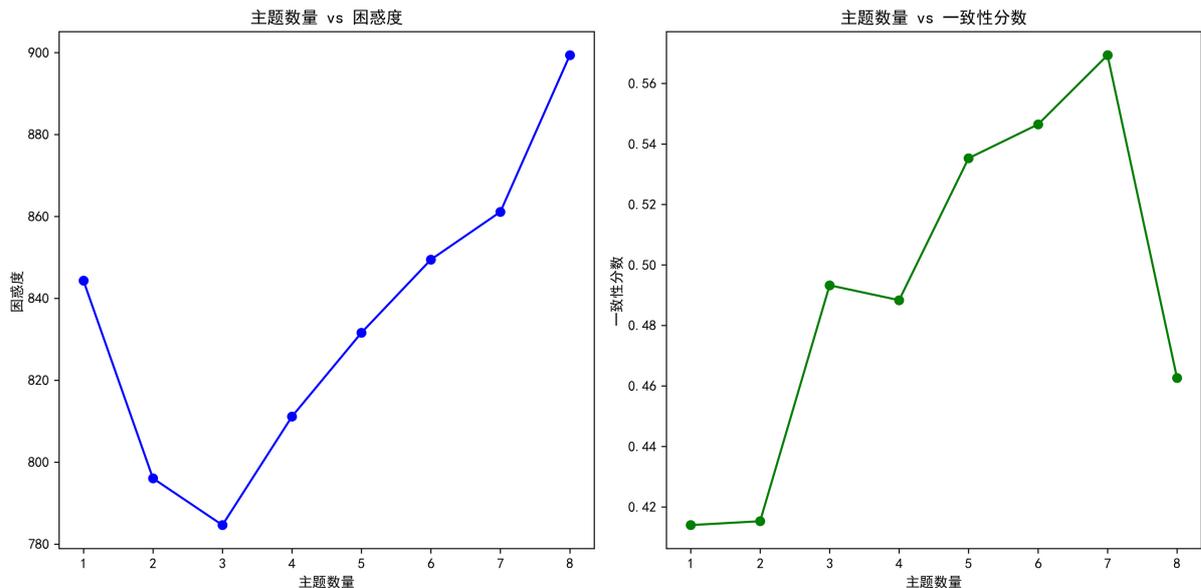


Figure 2. Plot of confusion and consistency scores as the number of topics changes
图 2. 困惑度和一致性分数随着主题数量的变化情况图

困惑度(Perplexity)是衡量语言模型预测文本概率分布能力的一个指标，常用于评价 LDA 模型的性

能。较低的困惑度表明模型对数据有较好的预测能力。随着主题数量的增加，模型的复杂度也会增加。当主题数量较少时，模型可能无法充分捕捉到数据中的所有重要信息，导致困惑度较高。随着主题数量的增加，模型能够更好地描述数据的内在结构，困惑度开始下降。当主题数量过多时，模型可能会开始捕捉到噪声和无关的信息，导致过拟合，从而使困惑度再次上升[6]。

如图 2 所示，困惑度曲线随着主题数量 k 的增加呈现出先下降后上升的趋势，表明当 k 较小时，模型的泛化能力不足，而当 k 过大时，模型可能过度拟合数据[7]。在本研究中，困惑度在 3 时达到最低点，表明该模型在复杂度和泛化能力之间达到了较优的平衡。因此，从困惑度的角度来看，选择 3 作为 LDA 模型的主题数是较为合适的。

另一方面，一致性分数的变化趋势表明，随着 k 的增加，主题间的语义聚合度逐步提升，在 $k=7$ 时达到峰值，而在 $k=8$ 时出现明显下降。尽管一致性分数在 $k=7$ 时最高，但综合考虑困惑度的最低点在 $k=3$ ，且 $k=3$ 时一致性分数也处于较高水平，因此最终选择 $k=3$ 作为 LDA 主题数，以确保模型既能较好地概括数据，又能保持合理的语义聚合度。

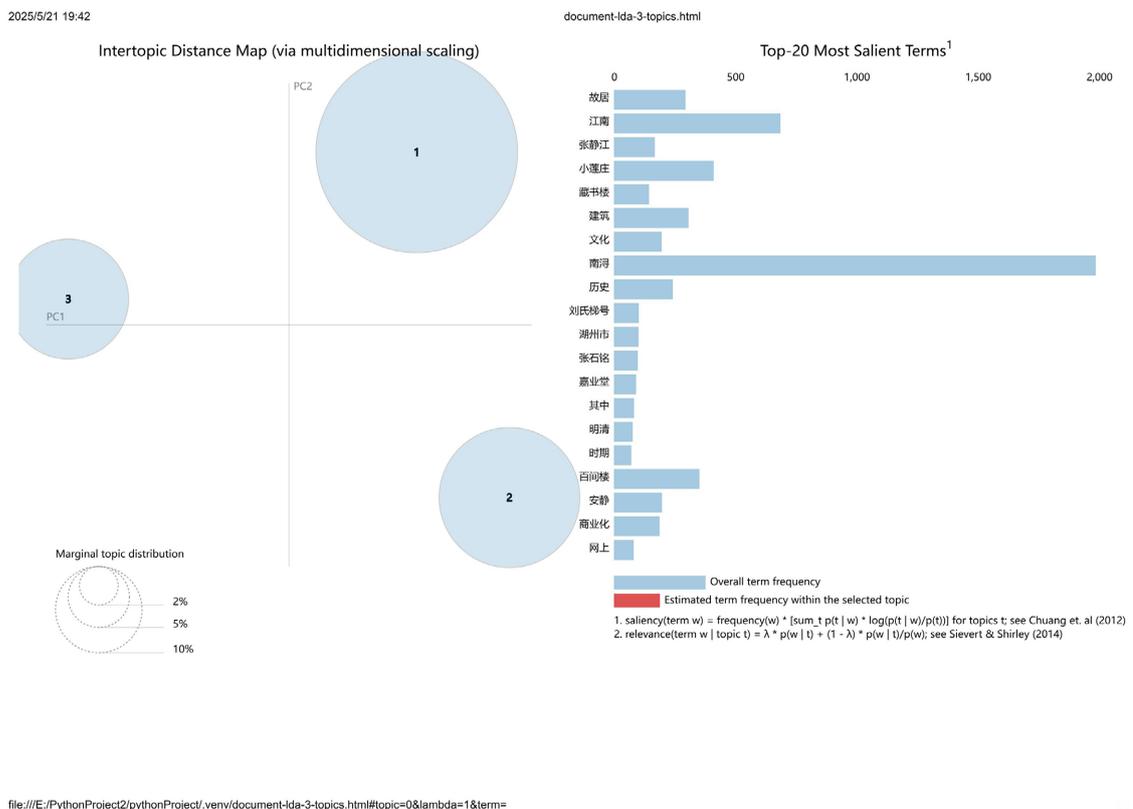


Figure 3. Visualization of the LDA model when the number of topics is 3

图 3. 主题数量为 3 的 LDA 模型可视化图

如图 3 所示，LDA 聚类得到的三个主题在可视化中表现出较好的区分度。圆圈的面积反映了各主题的重要性，而圆圈之间的距离表明了不同主题的语义差异。与其他 k 值的可视化结果相比， $k=3$ 时主题之间的区分度较为清晰，且主题内部的语义聚合程度较高，说明该模型能够有效提取出游客评论中的核心主题[8]。因此，将携程网南浔古镇的评论归纳为三个主题是合理的，这不仅提升了主题分析的精准度，也增强了数据的解释力。

采用 LDA 主题模型对文本数据进行聚类分析，成功提取出 3 个核心主题，并基于词频权重筛选出各

主题前 20 个高频特征词，具体结果详见表 4。通过结合领域背景知识与日常经验，综合语义分析与权重分布，最终凝练出“古镇自然景观特征”、“古镇的旅游体验”和“历史文化遗产”三个主题命名。

Table 4. Top 20 feature words for three topics
表 4. 三个主题排名前 20 的特征词

排名	关键词	权重	排名	关键词	权重	排名	关键词	权重
1	古镇	4.08%	1	古镇	5.95%	1	古镇	4.41%
2	南浔	2.05%	2	南浔	5.56%	2	南浔	4.27%
3	不错	1.69%	3	江南	3.29%	3	故居	3.49%
4	景点	1.41%	4	建筑	1.50%	4	江南	2.62%
5	景区	1.26%	5	故居	1.29%	5	小莲庄	2.60%
6	门票	1.22%	6	小莲庄	1.28%	6	张静江	1.98%
7	没有	1.06%	7	历史	1.23%	7	藏书楼	1.69%
8	值得	1.06%	8	景区	1.09%	8	建筑	1.52%
9	还是	1.03%	9	江南水乡	0.99%	9	百间楼	1.51%
10	很多	0.98%	10	文化	0.99%	10	文化	1.46%
11	感觉	0.82%	11	百间楼	0.99%	11	历史	1.38%
12	景色	0.80%	12	一个	0.95%	12	刘氏梯号	1.22%
13	地方	0.72%	13	小桥流水	0.94%	13	湖州市	1.20%
14	晚上	0.66%	14	张静江	0.73%	14	张石铭	1.18%
15	里面	0.65%	15	景点	0.70%	15	嘉业堂	1.09%
16	一个	0.63%	16	藏书楼	0.70%	16	其中	0.95%
17	不是	0.62%	17	水乡	0.67%	17	明清	0.93%
18	就是	0.61%	18	人家	0.64%	18	时期	0.86%
19	方便	0.60%	19	人文	0.64%	19	南浔区	0.74%
20	乌镇	0.59%	20	名镇	0.55%	20	充足	0.72%

第一类主题主要围绕南浔古镇的自然景观特色展开。高权重关键词包含“古镇”、“南浔”这两个核心主体，同时涵盖“江南”、“景色”、“江南水乡”、“小桥流水”、“建筑”等词汇。其中，“江南水乡”、“小桥流水”直接勾勒出古镇典型的水乡自然风貌，体现游客对水域景观、古朴建筑与自然环境融合的关注；“景色”一词则概括了游客对古镇整体自然景观的直观感受。这些关键词共同反映出游客对南浔古镇自然景观特色，如水乡肌理、建筑与自然交融之美的聚焦与欣赏，因此将此主题命名为“古镇自然景观特征”。

第二类主题主要围绕在南浔旅游的实际体验环节。高权重关键词涵盖“古镇”、“南浔”、“小莲庄”、“网上”、“取票”、“停车”等。“古镇”、“南浔”依然是核心主体，“小莲庄”作为具体景点，体现了对特定景区的关注。“网上”、“取票”、“停车”、“买票”、“购票”、“优惠”等词表明游客对线上线下购票流程、停车便利程度以及价格优惠情况的重视。“景色”、“干净”、“交通”、“房间”、“服务”、“热情”等词汇则反映出游客在游玩过程中对景区景色、卫生状况、交通条件、住宿情况以及服务态度的体验与评价。这一主题充分展现了游客在南浔旅游时从前期准备到游玩全程各方面实际体验的关注，因此将此主题命名为“古镇的旅游体验”。

第三类主题主要聚焦于南浔古镇的历史文化。高权重关键词中“古镇”、“南浔”是基础主体，“故居”、“小莲庄”、“张静江”、“藏书楼”、“刘氏梯号”、“张石铭”、“嘉业堂”等都是南浔古镇极具历史文化价值的遗迹与人物相关内容。“江南”、“历史”、“文化”、“明清”、“时期”等词汇进一步强调了古镇所处的地域文化背景以及悠久的历史发展阶段。整体反映出游客对南浔古镇丰富的历史人文遗迹以及其背后深厚历史文化底蕴的浓厚兴趣，所以将此主题命名为“历史文化遗产”。

3.4. 情感分析

3.4.1. 情感分布分析

本文基于 2944 条网络评论数据样本，采用情感分析技术对南浔古镇游客满意度进行量化评估。通过构建情感分布矩阵并计算各类型占比，系统揭示了游客情感倾向的结构化特征，具体分析结果详见表 5。该分析为旅游目的地服务质量优化提供了实证依据。

Table 5. Statistical results of emotion distribution

表 5. 情感分布统计结果

情绪类型	数量(条)	所占百分比(%)
积极情绪	2655	84.9
消极情绪	472	15.1

从表 5 的情绪分析可见，积极反馈占据了主导地位，积极评论总数达到 2655 条，占比高达 84.9%，这说明游客在南浔古镇之旅中普遍表现出了高度的满意和积极态度。特别地，积极情绪中，相关评论为：“景色美丽饭菜好吃”、“历史悠久文化灿烂风景秀美”等，充分体现了游客对景区的积极评价。

在消极情绪方面，共有 472 条评论，占总评论数的 15.1%。这些消极反馈主要集中在客流量较多时呈现的拥挤、景区购票问题等方面，评论多为“节假日人太多了都挤不动路”、“朱家角居然白天还要收门票”等。游客通常会因为一个或几个突出的消极因素影响其对景点的整体评估。

3.4.2. 基于 TF-IDF 的情感特征分析

为提升数据分析的深度与精度，采用 SnowNLP 情感分析模型对文本数据进行情感极性评分。该模型的情感评分区间设定为[0, 1]，其中数值趋近于 1 表示积极情感，趋近于 0 则反映消极情感。以 0.5 为情感倾向的分界点，评分值 ≥ 0.5 的文本归类为积极评价，评分值 < 0.5 的则判定为消极评价。基于此方法，各主题维度的情感分布情况详见表 6 [9]。三类主题情感得分相差较小，其中主题一得分最高，说明游客对古镇景色十分满意，对于古镇的旅游体验、历史文化的关注度略低但仍然是一个重要因素。

Table 6. Three kinds of thematic emotional tendencies

表 6. 三类主题情感倾向

主题	SnowNLP 情感倾向
古镇的自然景色	0.943
古镇的旅游体验	0.876
古镇的历史文化	0.846

通过 TF-IDF 算法对经过 SnowLNP 模型分类得出的积极评论与消极评论进行文本挖掘，得出前 20 个高频词及 TF-IDF 权重，如表 7 所示。无论是积极评价还是消极评价，由双方共有的特征词“身份证”、“门票”、“取票”、“进入”可见，游客在景区游玩时，都普遍关注的几个核心点有：进景区的方式，

景区是否需要买票, 入口等待的时间。

积极情绪特征词多聚焦于旅游服务流程(如“门票”、“取票”)、基础设施(如“客栈”)、服务便捷性的评价(如“方便”)。

核心节点“门票”、“取票”、“订票”与“身份证”、“网上”、“携程”形成强关联, 指向游客对数字化服务流程(如线上购票、身份验证)的体验; “客栈”、“晚上”、“古镇”等词与“安静”、“直接”等词形成语义簇, 反映游客对住宿环境与景区氛围的积极感知。此外, “提前”、“当天”等时间类词汇与“订票”、“进入”联动, 体现游客对行程规划的重视。整体来看, 积极情绪特征词通过“服务流程-设施体验-时间管理”的语义链条, 系统呈现游客对南浔古镇数字化服务效率、住宿舒适性及行程便利性的认可。

消极情绪特征词集中于购票问题(如“门票”、“二维码”)、客流管理(如“拥挤”、“半天”)、游客在入园、消费等环节的负面体验(如“花钱”、“进入”)。

核心节点“拥挤”与“门票只是”、“半天”等词形成强关联, 反映游客对景区客流量过大、门票政策(如强制购票)的不满; “身份证”、“取票”、“进入”等词与“排队”联动, 凸显入园流程中的低效与混乱。此外, “花钱”、“不买票”等词与“二维码”、“网上”形成语义矛盾, 暗示部分游客对隐性消费或线上服务复杂性的抵触。整体来看, 消极情绪特征词通过“客流压力-购票争议-流程低效”的语义链条, 集中暴露景区在客流管理、票务政策透明性及服务流程便捷性方面的短板。

Table 7. Consumer sentiment feature words and TF-IDF weight (Top 20)

表 7. 消费者情绪特征词及 TF-IDF 权重(前 20)

积极情绪		消极情绪	
特征词	TF-IDF 权重	特征词	TF-IDF 权重
身份证	0.2012	拥挤	0.2678
进去	0.1669	身份证	0.2012
门票	0.1631	进去	0.1669
取票	0.1220	取票	0.1220
收费	0.1180	收费	0.1180
客栈	0.1123	客栈	0.1123
网上	0.0865	门票	0.1631
花钱	0.0777	网上	0.0865
不买票	0.0742	花钱	0.0777
二维码	0.0721	不买票	0.0742
晚上	0.0715	门票只是	0.0927
古镇	0.0715	网上	0.0865
提前	0.0686	不要	0.0815
不需要	0.0673	花钱	0.0772
不收	0.0673	不买票	0.0742
直接	0.0661	不要门票	0.0729
当天	0.0651	二维码	0.0721
订票	0.0644	进入	0.0717
携程	0.0649	半天	0.0706
随便	0.0525	提前	0.0686

综上所述,积极与消极情绪特征词均聚焦于“古镇的旅游体验”主题,但呈现两极分化的评价。积极评价集中于数字化服务(如线上购票、身份验证)的高效性、住宿舒适性及行程规划的便捷性,体现景区在功能服务优化上的成功。消极评价则暴露客流管理失衡、票务政策争议及入园流程低效等问题,亟需通过限流措施、政策透明化及流程简化进行改进。

4. 总结与建议

4.1. 总结

本文采用数据挖掘技术对携程网上南浔古镇的在线评论进行了综合分析,从而探究游客对古镇的主要关注点和满意度。通过识别高频关键词、建立 LDA 主题模型以及评估情绪倾向,揭示了游客普遍关注的自然风景、人文景观和旅游体验等因素[10]。结果表明,游客特别重视景区内的风景景观和旅游体验,其中风景优美、无需门票被认为是选择南浔古镇的主要吸引因素。情感分析的结果也显示了在古镇入口排队时间的评论带有消极情绪,指出了入口秩序方面的不足。这一发现突出了景区服务质量作为游客决策和评价中的关键要素,同时指明了古镇需要关注并改进的领域,以提升游客满意度和吸引更多游客。

4.2. 建议

根据调查结果分析,游客对南浔古镇的文化底蕴和历史氛围方面持有特别关注。建议古镇进一步挖掘自身文化特色,开发更具参与性和互动性的文化体验项目,例如设计以“江南水乡生活”、“南浔商帮文化”等为主题的游览路线,将景点串联起来,并融入角色扮演、情景再现等元素,增强游客代入感。针对部分游客对古镇的服务质量存在一定意见,古镇应加强服务意识培训,提升服务人员的专业素养和服务水平。例如完善景区导览系统,在关键位置设置清晰指示牌,开发智能导览 APP,提供语音讲解、路线规划。优化在线预订平台,简化流程、及时反馈预订信息,提供更便捷、智能的服务。优化客流管理,推行分时段预约制度,缓解节假日拥挤问题;透明票务政策,明确标注免费与收费项目,减少隐性消费争议。简化服务流程,优化线上平台功能(如一键购票),减少操作复杂度;强化基础设施,增设入园通道,提升高峰期通行效率。可以借助 LDA 主题模型分析结果,提炼出游客对南浔古镇的核心关注点和情感倾向,并将其融入到品牌营销策略中,针对不同目标客群,制定差异化的营销策略:针对历史文化爱好者,可以重点宣传古镇的文化底蕴和历史遗迹;针对休闲度假游客,可以突出古镇的江南水乡风情和慢生活体验,推出特色民宿、休闲活动;针对年轻游客,可以利用抖音、小红书等新媒体平台,发布优质旅游内容,吸引更多年轻游客关注。

总而言之,古镇开发者可以充分利用自身优势,不断优化旅游产品和服务,提升游客体验,打造独具特色的江南水乡古镇旅游品牌,实现可持续发展。

参考文献

- [1] 中华人民共和国文化和旅游部. 2024 年国庆节假期文化和旅游市场情况[EB/OL]. (2024-10-08) https://www.gov.cn/lianbo/bumen/202410/content_6978757.htm, 2024-10-24.
- [2] 蔡颖. 基于 Python 的文本数据处理研究[J]. 软件, 2023, 44(5): 179-183.
- [3] 秦春秀, 祝婷, 赵捧未, 张毅. 自然语言语义分析研究进展[J]. 图书情报工作, 2014, 58(22): 130-137.
- [4] 苏宇. 网络爬虫的行政法规制[J]. 政法论坛, 2021, 39(6): 41-53.
- [5] 窦欣怡. 基于 LDA 主题模型的商品在线评论文本挖掘分析[J]. 电子商务评论, 2024, 13(3): 8710-8718.
- [6] 倪超, 吴伟敏. 消费者购买教育机器人满意度影响因素研究——基于 LDA 和 IPA 的在线评论分析[J]. 电子商务评论, 2024, 13(3): 8245-8254.
- [7] 王轶群, 陈成鑫. 基于 LDA 的欧盟信息迷雾政策报告主题挖掘与治理研究[J]. 图书情报工作, 2025, 69(5): 94-106.

-
- [8] 徐远廷, 徐嘉, 沈晓芳. 基于 LDA 主题模型的智慧体育发展现状分析及热点研究[J]. 冰雪体育创新研究, 2025, 6(2): 104-107.
- [9] 姚正旺, 梁利, 史江兰, 等. 西江千户苗寨景点评论分析[J]. 电子商务评论, 2024, 13(3): 8988-8999.
- [10] 朱登赟, 万福成, 国旗, 等. 针对北京旅游景点的评论文本情感标注数据集[J/OL]. 中国科学数据(中英文网络版): 1-10. <https://doi.org/10.11922/11-6035.csd.2022.0026.zh>, 2025-03-07.