基于深度学习与聚类分析的桂林龙脊梯田景区游客服务偏好与满意度研究

刘星雨[1]，姚晓轩[2]

（1. 澳门大学工商管理学院，澳门 000011；2. 广西师范大学数学与统计学院，广西 桂林 541006）

摘要：旅游业正在经历深刻的转型，游客的需求与偏好呈现多元化趋势，精准识别游客偏好及满意度状况至关重要。以桂林龙脊梯田景区为案例，采用深度学习技术和聚类分析方法，从网络评论中提炼游客的服务偏好主题，识别具有相似偏好的游客群体，结合问卷调查数据量化评估服务满意度。研究发现，游客服务偏好涵盖住宿、餐饮、交通、信息等服务，信息服务和住宿服务的改进空间最大，交通服务和餐饮服务亦存在提升需求，据此提出相应建议。

关键词：龙脊梯田景区；服务偏好；深度学习；游客聚类；市场竞争力

中图分类号：F592.7 **文献标志码：**A **文章编号：**1671－1807(2025)12－0062－09

在全球化与数字化的大背景下，旅游业正经历一场深刻的转型，游客的需求与偏好也随之呈现出多元化的趋势。旅游目的地若要脱颖而出，必须不断创新，提供高度个性化且质量卓越的服务，以满足游客日益复杂的需求，从而巩固自身的市场地位。位于中国广西壮族自治区桂林市的龙胜龙脊梯田景区，凭借其独一无二的自然景观与深厚的文化底蕴，成为无数旅游者心中的向往之地。然而，面对游客群体不断变化的期待，如何精准捕捉并满足其服务偏好，成为景区管理者亟须解决的难题。

《乡村振兴战略规划（2018—2022年）》强调乡村传统村落不仅是中华文化的重要载体，发展乡村旅游也是实现扶贫、产业振兴及农村经济多样化的重要战略[1]。龙脊梯田为“全球重要农业文化遗产”，龙脊梯田景区的大寨村入选联合国世界旅游组织“最佳旅游乡村”[2]，其品牌价值与日俱增，为桂林世界级旅游城市的建设注入了新的活力。在这一背景下，如何进一步增强龙脊梯田的品牌效应，提升其在国际旅游市场中的影响力，成为社会热议的话题。为此，探讨如何提升龙脊梯田景区的旅游品牌，显得尤为重要。

近年来，随着大数据分析与人工智能技术的应用，旅游研究领域也迎来数字化转型[3]。数字化转型具体是指利用现代信息技术使传统产业实现全面升级。近年来数字经济的快速增长，促进了数字化与乡村景区更深层次的融合。数字乡村旅游建设通过应用新一代数字技术，推动农业升级、农村进步和农民发展，是实现农业农村现代化和乡村振兴的重要路径[4]。比如应用深度学习与聚类分析等前沿算法，使得从庞杂的数据中抽丝剥茧，提炼出游客对服务的真实需求与满意度成为可能[5-6]。这不仅有助于旅游企业洞察市场趋势，也为个性化服务的定制提供了坚实的数据基础。

然而，当前针对数字乡村旅游的研究主要从应用视角探讨数字乡村的概念、意义、挑战及实施路径，重点关注国内外实践案例，但对典型个案的作用机制，尤其是深入挖掘游客的个性化需求特征，优化游客服务的研究较少[7]。本文致力于运用深度学习技术，从网络评论中挖掘龙脊梯田景区的服务偏好主题，通过聚类分析识别具有相似偏好的游客群体，进而辅以问卷调查数据，实现对服务满意度的量化评估，旨在为龙脊梯田景区提供一套基于数据的决策支持体系，助力其优化服务流程，提升游客体验，进而增强其市场竞争力与品牌影响力。通过研究，不仅能为旅游业的个性化服务与市场细分策略带来新视角，同时也将丰富服务管理理论在旅游实践中的应用内涵。

收稿日期：2025-01-17

基金项目：广西创新驱动发展专项（AA20302011）

作者简介：刘星雨（2001—），女，广西桂林人，硕士研究生，研究方向为数字经济与乡村产业经济；姚晓轩（1999—），男，山西运城人，硕士研究生，研究方向为数统分析。

1 理论基础

1.1 服务质量理论与游客行为理论

服务质量理论是研究客户对服务感知质量的理论框架。Parasuraman 等[8]提出的 SERVQUAL 模型是服务质量理论的经典模型,包含有形性、可靠性、响应性、保证性和移情性五个维度。钟飞燕等[9]以广州从化区乡村民宿为例,通过对 SERVQUAL 模型和重要性-绩效分析(importance performance analysis,IPA)结果的解读,明确了普通民宿和省级民宿在服务质量方面的优劣势,并确定了需要重点改进的服务指标。此外,学者们应用 SERVQUAL 模型探究餐饮服务[10]、交通服务[11]、信息服务[12]等的监管效果与服务满意度。游客行为理论主要研究影响游客行为决策的因素。Ajzen[13]的计划行为理论认为,行为意图是行为的直接前因,而行为意图受态度、主观规范和知觉行为控制的影响。厉新建等[14]基于计划行为理论探讨虚拟旅游体验对实地旅游意向的影响机制,发现虚拟旅游体验、知觉行为控制和主观效应能够显著影响虚拟旅游者的态度,同时虚拟旅游体验和虚拟旅游者的态度可以直接正向影响实地旅游意向,该研究对于深入理解虚拟旅游体验与实地旅游意向之间的关系具有参考价值。

1.2 新质生产力与高质量发展

在追求经济高质量发展的大背景下,新质生产力被视为推动产业升级与转型的关键力量,承载着提升服务质量、优化游客体验和促进可持续发展的重任。已有学者探讨旅游新质生产力的内涵、新质生产力与旅游业高质量发展的关系[15],分析新质生产力对文化旅游业的影响,提出其促进文化旅游业高质量发展的三个方面的理论逻辑,即加速生产方式变革、重构生产关系、提高经济发展质量,并总结了新质生产力在提升文化旅游业效率与质量、丰富模式创新、提升整体竞争力、增强人力资本价值等方面的作用[16]。学者们发现龙脊梯田景区在资源优化配置、人文景观挖掘、人口素质和管理机制等方面,尚存在明显短板[17],认为新技术、新要素等新质生产力是赋能旅游业高质量发展的动力[18],应用现代信息技术,尤其是网络爬虫、自然语言处理、数据挖掘和深度学习等技术,为旅游研究开辟了新的路径[19-21]。

综合上述文献,不难发现,旅游业的高质量发展离不开对游客需求的精准洞察和服务质量的持续优化。利用大数据分析和人工智能技术,深入挖掘游客的偏好和痛点,对于提升旅游目的地的吸引力和竞争力至关重要。

2 数据来源与研究方法

2.1 研究区域概况

龙脊梯田,作为一项跨越千年的农业奇迹,不仅是人类智慧与自然和谐共生的典范,更是当代旅游业中的一颗璀璨明珠。这片历史悠久的梯田,自秦朝始,经由无数代农民辛勤耕耘,至今仍生机勃勃,每年吸引着来自世界各地的游客驻足观赏。然而,尽管拥有得天独厚的自然资源和文化积淀,龙脊梯田景区在产业要素发展上却面临着诸多挑战。受人力资源、资金投入、管理机制等因素制约,在商业资本投入、旅游大力开发等冲击下,梯田生态、农耕文化、民族风俗、村寨风貌、景观意象等保护面临巨大挑战,亟须加强科学维护与管理[22]。尤其在游客体验层面,娱乐项目的稀缺和参与性的不足,导致游客停留时间短,旅游收入增长受限,景区仍处于“门票经济”阶段。

2.2 数据采集与处理

为了深入理解游客对龙脊梯田景区的服务偏好,采用了多源数据采集策略,包括但不限于从携程、马蜂窝、穷游网、小红书和抖音等旅游和社交平台抓取的用户评论、问答和攻略分享。数据采集工作借助 Python 编程语言,利用 BeautifulSoup 和 Scrapy 等库自动化执行。为了确保数据质量,去除无效、重复和无关的信息,以及含有广告或非法字符的评论。此外,中文文本经过 jieba 分词工具处理,结合多份停用词表进行筛选,剔除非实质性词汇,如“的”“在”等,以聚焦于反映服务偏好的关键词。对评论文本进行标准化处理,如转换为小写、去除标点符号等,便于后续分析。

研究应用 TF-IDF 算法对文本数据进行处理,该算法能够评估词语在文档集中的重要程度,帮助识别游客在评价中频繁提及的服务偏好相关关键词。兰晓芳等[23]提出基于 TF-IDF 和 TextRank 的中文文本关键词提取方法。TF-IDF 算法通过计算词频(TF)和逆文档频率(IDF)的乘积,为每个词语赋予权重,从而过滤掉常见的词语,保留对文档集而言具有更高区分度的关键词汇。基于关键词,对每条评论进行情感评分,区分正面、负面及中立情感。此过程通过自然语言处理技术实现,精准捕捉消费者情绪波动。将每条评论按其主题归类,并计算各主题下情感评分的分布情况,揭示不同服务方面的情感倾向差异。

问卷调查部分设计包含了游客的人口统计特征(如性别、年龄、职业、收入水平)以及对住宿、餐饮、交通、信息服务的满意度评价。问卷通过线上渠道发放,使用方便抽样和滚雪球抽样方法,2024年5—8月,共收集到540份有效问卷,其中273名受访者曾亲身体验过龙脊梯田景区旅游。问卷数据经过编码、标准化和缺失值处理后,用于量化评估各项服务的满意度和改进空间。

2.3　TF-IDF算法

使用基于TF-IDF(term frequency-inverse document frequency)的方法来进行主题分类和情感分析。TF-IDF是一种用于信息检索和文本挖掘的常用加权技术,它能够评估一个词语在文档集中的重要程度。

词频(term frequency,TF)表示词条在文本中出现的频率,这个数字通常会被归一化(一般是词频除以文章总词数),以防止它偏向长的文件(同一个词语在长文件里可能会比短文件有更高的词频,而不管该词语重要与否)。TF的计算公式为

$$\mathrm{TF}_{ij}=\frac{n_{ij}}{\sum_{|d_j|} n_{kj}} \tag{1}$$

式中:n_{ij}为词条t_i在文档d_j中出现的次数;TF_{ij}为词条t_i在文档d_j中出现的频率。

逆文件频率(inverse document frequency,IDF)表示关键词的普遍程度。如果包含词条i的文档越少,IDF越大,则说明该词条具有很好的类别区分能力。某一特定词语的IDF,可以由总文件数目除以包含该词语之文件的数目,再将得到的商取对数得到:

$$\mathrm{IDF}_i=\ln\frac{|D|}{1+|j:t_i\in d_j|} \tag{2}$$

式中:$|D|$为所有文档的数量;$|j:t_i\in d_j|$为包含词条t_i的文档数量,加1防止包含词条的数量为0从而导致运算出错的现象发生。

某一特定文件内的高词语频率,以及该词语在整个文件集合中的低文件频率,可以产生出高权重的TF-IDF。因此,TF-IDF倾向于过滤掉常见的词语,保留重要的词语,表达式为

$$\mathrm{TF-IDF}=\mathrm{TF}\cdot\mathrm{IDF} \tag{3}$$

2.4　聚类分析

采用K-means聚类分析算法对游客进行分群,以识别具有共同偏好的游客群体。与其他聚类算法相比,K-means算法是将n个对象分成K个簇,保证簇内指标具有较高的相似度,簇间指标具有较低的相似度,使得每个对象到其所属聚类中心的距离之和最小。算法具有计算简单、可拓展性强等优点。通过计算数据点与聚类中心之间的欧氏距离,将每个数据点分配给最近的聚类,随后更新聚类中心,直至聚类结果收敛。

假设有n个数据点$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$,每个数据点$\boldsymbol{x}_i$是一个$d$维向量,即$\boldsymbol{x}_i\in\mathbf{R}^d$。需要将这些数据点分成$\boldsymbol{k}$个簇$\boldsymbol{C}=\{\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_k\}$,每个簇有一个质心$\boldsymbol{\mu}_k$。

(1)初始化质心:随机选择k个数据点作为初始质心$\{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\cdots,\boldsymbol{\mu}_k\}$。

(2)分配数据点:对于每个数据点$\boldsymbol{x}_i$,计算它与每个质心的距离,并将其分配到距离最近的质心所在的簇。距离通常使用欧氏距离D来衡量。

$$D(\boldsymbol{x}_i,\boldsymbol{\mu}_k)=\|\boldsymbol{x}_i-\boldsymbol{\mu}_k\| \tag{4}$$

数据点$\boldsymbol{x}_i$分配到簇C_j的条件为

$$j=\arg\min_k\|\boldsymbol{x}_i-\boldsymbol{\mu}_k\| \tag{5}$$

(3)更新质心:重新计算每个簇的质心。

$$\boldsymbol{\mu}_k=\frac{1}{|\boldsymbol{C}_k|}\sum_{\boldsymbol{x}_i\in\boldsymbol{C}_k}\boldsymbol{x}_i \tag{6}$$

式中:$|C_k|$为簇C_k中的数据点数量。

(4)迭代:重复步骤(2)和步骤(3),直到质心不再变化或达到最大迭代次数。

通过上述公式和步骤,K-means算法能够有效地将数据集划分为k个簇,每个簇内的数据点具有相似的特征。

2.5　TextCNN模型

TextCNN是一种将卷积神经网络(CNN)应用于文本分类任务的模型。它通过使用多个不同大小的卷积核来提取句子中的关键信息,并通过池化层减少参数数量,最后通过全连接层进行分类。

模型结构:①输入层:输入是一个固定长度的词向量序列;②卷积层:使用多个不同大小的卷积核对输入序列进行卷积操作;③池化层:对每个卷积核的结果进行最大池化操作;④全连接层:将池化后的特征向量拼接起来,通过全连接层进行分类。

具体计算:假设输入的句子由n个词组成,每个词用一个d维词向量表示,那么输入可以表示为一个$n\times d$的矩阵$\boldsymbol{X}$。

(1)输入层:

$$\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n] \tag{7}$$

(2)卷积层:

使用多个不同大小的卷积核对输入进行卷积操作。假设有k个卷积核,每个卷积核的大小为

$h\times d$,步长为 1。

对于第 i 个卷积核 $\boldsymbol{W}_i$和偏置 $\boldsymbol{b}_i$,卷积操作可以表示为

$$\boldsymbol{c}_i(j)=f(\boldsymbol{W}_i x_{j:j+h-1}+\boldsymbol{b}_i) \tag{8}$$

式中:f 为激活函数(如 ReLU),$x_{j:j+h-1}$为从位置 j 开始的长度为 h 的子序列。

卷积结果 $\boldsymbol{c}_i$是一个 $(n-h+1)$ 维的向量:

$$\boldsymbol{c}_i=[\boldsymbol{c}_i(1),\boldsymbol{c}_i(2),\cdots,\boldsymbol{c}_i(n-h+1)] \tag{9}$$

(3)池化层:

对每个卷积核的结果进行最大池化操作,得到每个卷积核的最大值:

$$\hat{\boldsymbol{c}}_i=\max\boldsymbol{c}_i \tag{10}$$

(4)全连接层:

将所有卷积核的最大池化结果拼接成一个特征向量 $\boldsymbol{z}$:

$$\boldsymbol{z}=[\hat{\boldsymbol{c}}_1,\hat{\boldsymbol{c}}_2,\cdots,\hat{\boldsymbol{c}}_k] \tag{11}$$

通过全连接层进行分类,假设输出类别数为 C:

$$y=\mathrm{softmax}(\boldsymbol{W}_f\boldsymbol{z}+\boldsymbol{b}_f) \tag{12}$$

式中:$\boldsymbol{W}_f$为全连接层的权重矩阵;$\boldsymbol{b}_f$为偏置向量;softmax 函数用于将输出转换为概率分布。

3 结果分析

3.1 文本数据主题

应用 TF-IDF 算法对从各大平台收集的游客评论数据进行了关键词提取。通过计算每个词语的重要性,识别出游客在评论中频繁提到的服务相关主题。基于这些主题词,归纳出以下四大服务偏好主题,如表 1 所示。

住宿服务偏好是关于游客对住宿服务的偏好,主要集中在酒店和民宿的设施完备性、卫生情况、地理位置、价格合理性及服务态度等方面。高质量的住宿服务不仅能提供舒适的居住体验,还应当具备便利的预订流程和友好的服务态度。主题词包括酒店、民宿、设施、位置、价格、卫生、舒适、服务态度、房间、预订等。

餐饮服务偏好主要反映游客对食物质量、多样性和服务态度的关注。游客希望在龙脊梯田景区不仅能品尝到美味的食物,还能享受到干净卫生的用餐环境和热情周到的服务。此外,价格的合理性和餐厅的特色菜也是影响游客餐饮体验的重要因素。主题词包括食物质量、多样性、服务态度、价格、环境、特色菜、卫生、口味、餐厅、用餐体验。

交通服务偏好是游客旅游体验的重要组成部分。游客偏好便捷的交通方式、合理的费用和清晰的导航指引。良好的公共交通设施、便利的租车服务和足够的停车场地可以极大地提升游客的旅游便利性。此外,班车服务和良好的道路状况也被认为是优质交通服务的重要指标。主题词包括便捷性、交通方式、费用、地理位置、公共交通、租车、导航、停车场、班车、道路状况。

信息服务偏好主题涵盖游客对旅游咨询、路线规划和信息准确度的需求。游客希望能够方便地获取详细的景点介绍、旅行攻略和电子导览信息,帮助他们更好地规划行程。同时,优质的客服服务和有效的投诉处理机制也是游客期望的重要内容。主题词包括旅游咨询、路线规划、预订便利性、信息准确度、导游服务、电子导览、景点介绍、客服、旅行攻略、投诉处理。

3.2 主题情感

利用 litNlp 库来进行评论情感极性分析,litNlp 是兼容最新版 Tensorflow 2.0 实现的一个轻量级的深度情感极性推理模型,使用字符级代替词语级进一步提升训练和推理速度,可以实现细粒度的多级别情感极性训练和预测。

情感分析模型训练使用的数据,正面评论为 7 000 条,负面评论为 3 000 条,每条评论有一个标签,负面评价标签为 0,正面评价标签为 1。

对前述四大服务偏好主题进行情感分析,如表 1 所示。其中,住宿服务的消极情感比例最高,积极情感比例最低,表明顾客对住宿体验的满意度相对较低;餐饮服务方面,积极情感比例最高,消极情感最少,显示顾客对此服务的满意度较高;交通服务中,中性情感占比最高,表明顾客对此服务的满意度居中,消极情感比例较积极情感比例略高;信息服务领域,中性和消极情感比例接近,积极情感比例最低,显示顾客在获取信息方面的体验相对较差。

表 1 情感分析表

类别	住宿服务	餐饮服务	交通服务	信息服务
积极情感比例/%	26.36	38.46	27.94	24.65
中性情感比例/%	34.77	34.39	41.58	37.60
消极情感比例/%	38.88	27.15	30.48	37.75
积极情感数量/条	141	85	463	508
中性情感数量/条	186	76	689	775
消极情感数量/条	208	60	505	778

3.3 服务满意度

问卷中收集了游客对住宿、餐饮、交通、信息服务的满意度和重要度得分,引入一个新的变量——改进空间得分,每项服务的改进空间得分等于每个游客的重要性得分除以满意度得分的加权平均。

服务越重要、满意度越低，改进空间得分就越高，说明该项服务优化的优先级越高，各项得分如表 2 所示，改进得分分布如图 1 所示。

游客对各项服务的满意度均在中等水平（3.5～3.8 分），其中信息服务满意度最低（约为 3.55 分），而其重要性得分却普遍较高（4.2～4.3 分），显示了游客对这些服务有较高的期待，但实际体验与期望存在差距，尤其是信息服务方面，改进空间得分高达 1.253 7，表明这是急需改进的领域。住宿服务和餐饮服务的满意度虽然略高，但它们的改进空间也较大，尤其是住宿服务的改进空间得分达到 1.217 6，表明提高住宿质量对提升整体旅游体验至关重要。

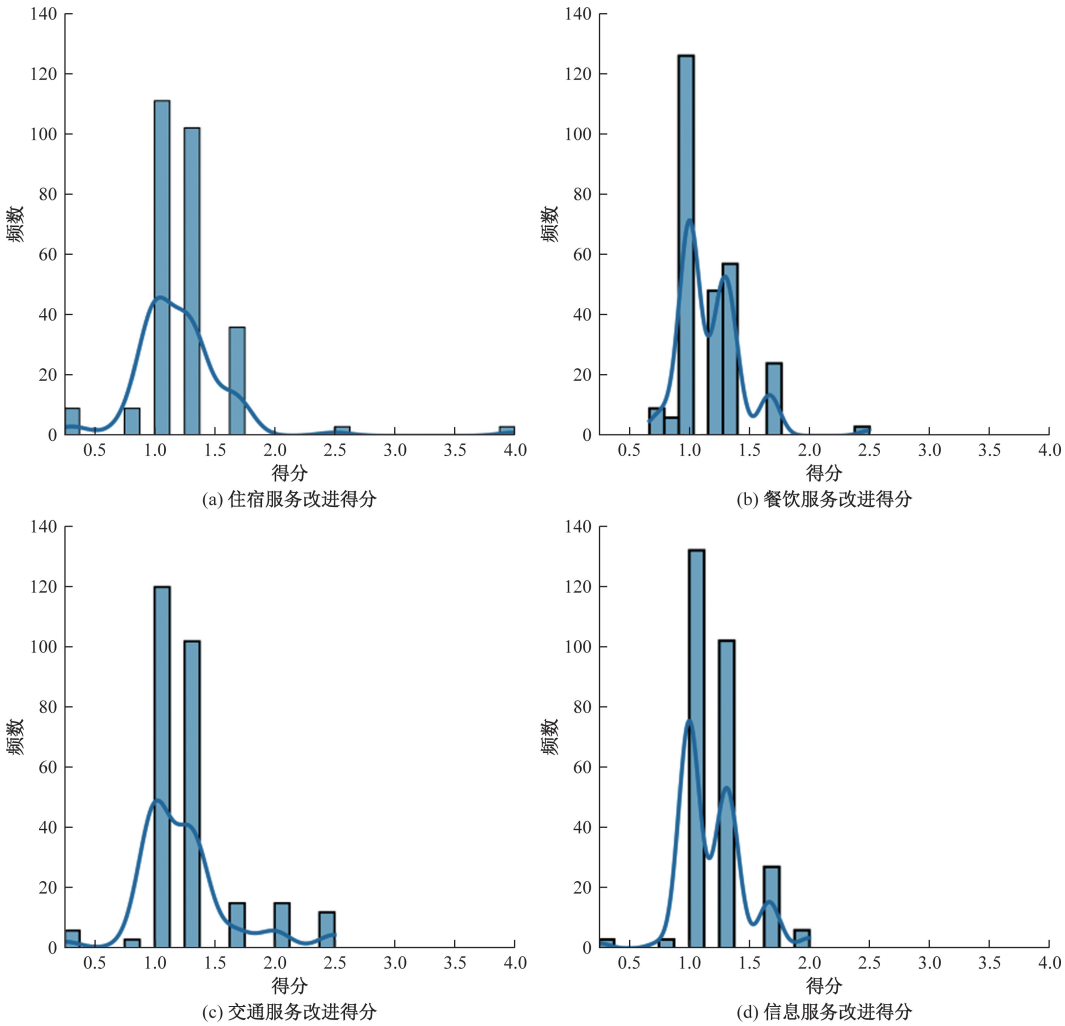
表 2　满意度

类型	满意度得分	重要性得分	改进空间得分
住宿服务	3.681 3	4.274 7	1.217 6
餐饮服务	3.769 2	4.307 7	1.175 1
信息服务	3.549 5	4.197 8	1.253 7
交通服务	3.637 4	4.197 8	1.192 3

通过改进空间得分的分布图分析可以看出，住宿服务和餐饮服务的改进空间得分较为集中，表明虽然这两项服务的满意度中等，但仍有一定的改进空间。其中，住宿服务的改进空间略大于餐饮服务，建议对住宿服务进行优先改进。交通服务和信息服务的改进空间得分较为分散，尤其是信息服务的改进空间得分最高，显示在信息服务领域的改进需求最为迫切。

3.4　服务偏好

房间清洁度是游客最重视的，其次是住宿地点的安全性和周围景观。虽然房间设施和服务人员态度也被认为重要，但关注度略低。价格在住宿选择中并非首要考虑因素，大多数游客偏好经济型住宿（100～300 元），中档型（300～600 元）也有一定市场，而对高端住宿需求较少。游客倾向于住宿带有阳台或露台的房间，对户外空间有所偏好，同时对商务级 WiFi、隔音等设施有一定程度的需求，但私人影院等高端设施需求较低。



(a) 住宿服务改进得分　(b) 餐饮服务改进得分

(c) 交通服务改进得分　(d) 信息服务改进得分

图 1　服务得分分布

游客最关注餐饮的口味,其次是用餐环境和菜品种类,对服务人员态度和价格敏感度较低。偏好尝试当地特色美食,而西式餐点需求最低,显示游客对体验本土文化的倾向。希望获得地方特色体验和互动式餐饮,对定制化菜单需求不大。

交通便捷性是首要考虑,安全性和舒适度次之,价格关注度较低。景区班车和缆车最受欢迎,共享单车/电动车次之,自驾车和租车需求较少。大部分游客对交通费用的接受范围在中低价位(50～100元)。

游客最重视信息的准确性和获取的便捷性,对信息的全面性也有一定需求,而多语言支持需求较低。旅游官网和 App 是主要信息来源,社交媒体也是一个重要渠道,相比之下,旅游咨询中心和导游服务需求较少。

综合来看,游客在旅游服务中,更重视住宿的清洁度、餐饮的地道体验、交通的便捷与经济性,以及信息的准确性和即时获取。此外,他们对价格的敏感度在不同服务领域表现不一,但总体上更倾向于性价比高的选项,并且对体验当地特色有较高的兴趣。

3.5 相关性

通过相关性分析,发现各项服务满意度之间存在一定的正相关关系,各个服务相关性图如图 2 所示。

住宿满意度与餐饮满意度(相关系数为 0.63)、交通满意度(0.46)和信息服务满意度(0.50)均呈正相关,表明住宿体验的好坏会直接影响到其他服务的感知。

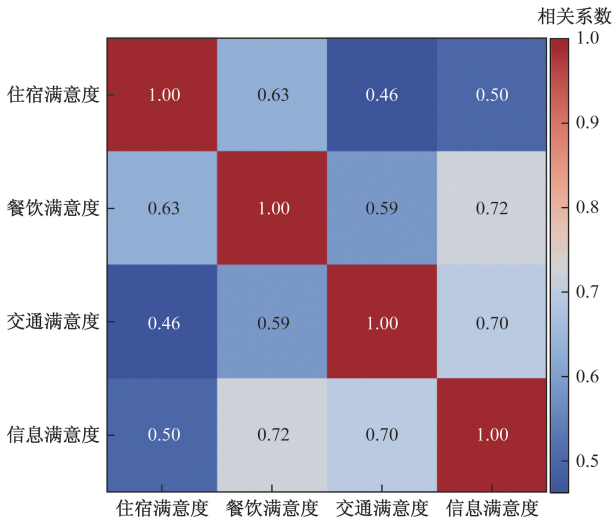


图 2 服务相关性

餐饮满意度与交通满意度(0.59)和信息服务满意度(0.72)的强相关性,揭示了提升餐饮服务质量可能同步提高交通和信息服务的满意度。

交通满意度与信息服务满意度(0.70)的紧密联系,意味着交通便利性与信息的易获取性是游客整体满意度的重要组成部分。

相关性矩阵中的数据表明各项服务之间存在一定的互补性,提升某一项服务的满意度可能会带动其他服务满意度的提高。因此,在制定改进计划时,可以采取综合措施,考虑到这些服务之间的联系。

聚焦优先改进领域,基于各项服务之间的相关性和先前的改进空间分析,信息服务应当成为优先改进的领域。信息服务的满意度提升将有助于整体服务体验的改善。

建议综合提升信息服务,考虑到信息服务对其他满意度的高影响力,建议优先改进信息服务。这包括提升信息的准确性、便捷性和全面性。联动改进住宿和餐饮服务,由于住宿和餐饮满意度之间存在较高的相关性,可以通过提供更好的住宿体验(如加强房间的日常清洁与维护、优化服务人员的工作态度与专业能力等)来间接提升餐饮服务的满意度。注重整体服务提升,由于各服务满意度之间的相互影响,应采取综合性的提升措施,而不是孤立地改进某一单项服务。例如,可以通过加强员工培训、改善基础设施,以及建立有效的顾客反馈机制等方式,实现整体服务质量的提升,从而提高游客的满意度。

3.6 聚类分析结果

通过肘部法则,聚类个数为 4 或 5 为明显转折点,轮廓系数图中,聚类个数大于 2 时趋于稳定,最终确定最佳聚类数为 4,肘部图如图 3 所示,轮廓系数图如图 4 和图 5 所示。

主成分分析(principal component analysis,PCA)降维结果如图 6 和图 7 所示,降维后的主成分能够较好地分离四个群体,观察各群体在降维后的分布情况,确认群体之间的差异性。

通过 *K*-means 算法和 PCA 降维,游客主要被分为四类群体(其他较为零散),每一类具有不同的特征组合,如表 3 所示。

第一类:25～34 岁男性,事业单位工作人员,年收入中等,满意度中等,尤其对交通服务的满意度较低。这一类人群对服务的期望较高。

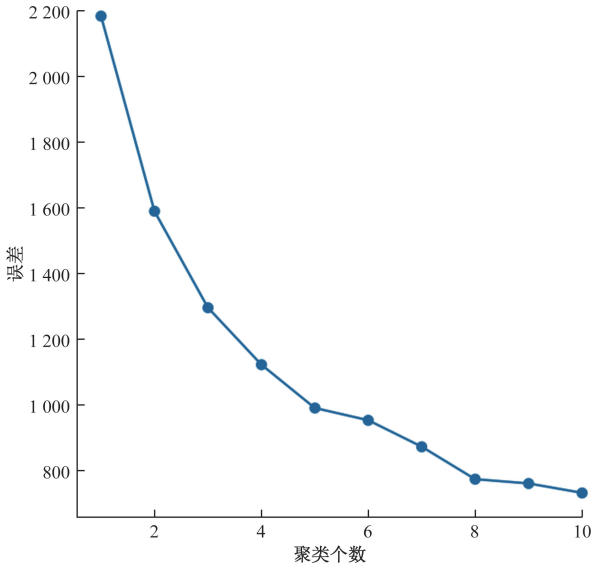


图 3　肘部

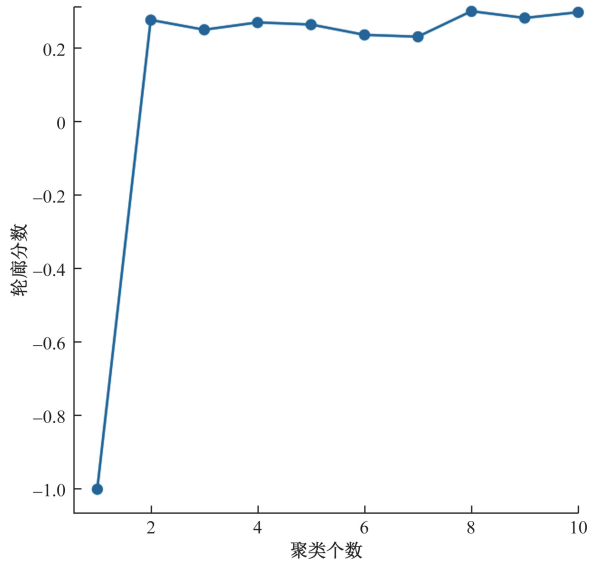


图 4　轮廓系数

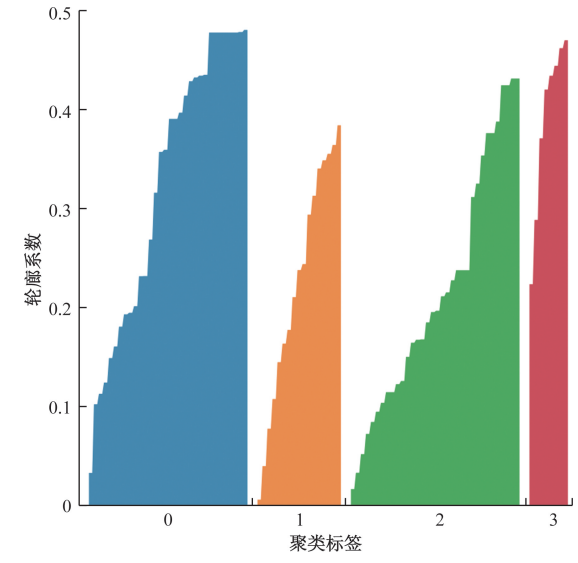


图 5　轮廓系数可视化

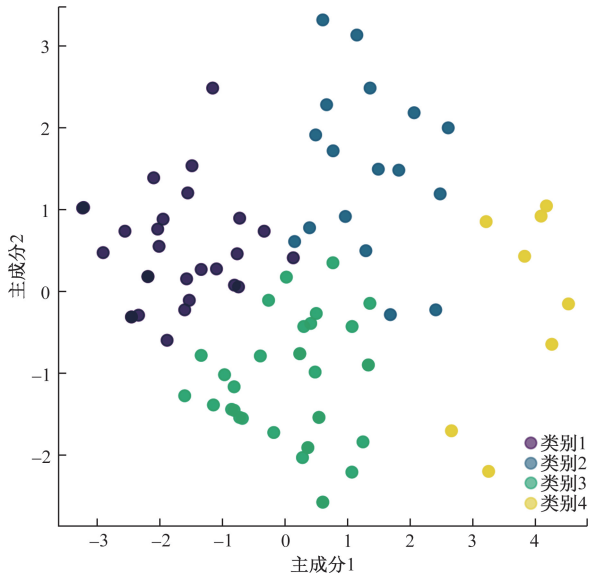


图 6　聚类结果二维可视化

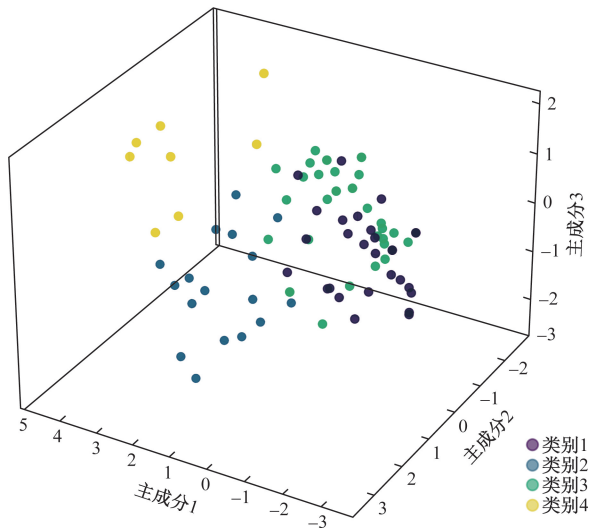


图 7　聚类结果三维可视化

第二类：18～24 岁男性学生，收入低，对所有服务有较高满意度，这一类人群对服务的整体满意度较高，尤其对餐饮服务的满意度最高。

第三类：25～34 岁女性职场人士，收入中等，对所有服务有较高满意度。

第四类：18～24 岁收入较低的工薪族和农民，对所有服务的满意度均非常高，尤其对住宿和餐饮服务的满意度更高。与此同时，他们偏好性价比高的服务。

4　结论与建议

4.1　结论

结合深度学习识别的服务偏好主题和问卷调查的量化数据发现：①游客服务偏好主要涵盖游客对住宿服务、餐饮服务、交通服务、信息服务等的偏

表 3　聚类信息

类别	性别	年龄/岁	职业	年收入/元	住宿	餐饮	交通	信息
第一类	男	25～34	事业单位	5 万～10 万	3.2	3.1	2.8	3.1
第二类	男	18～24	学生	1 万～2 万	3.8	4.1	3.7	3.8
第三类	女	25～34	企业、事业单位	5 万～10 万	3.8	4.0	3.8	3.7
第四类	男、女	18～24	企业、农民	5 万以下	4.7	4.9	4.3	4.5

好。住宿服务偏好体现在设施完备性、卫生情况、地理位置、价格合理性及从业人员服务态度等方面,餐饮服务偏好反映游客对餐饮质量、餐饮多样性与特色、价格的合理性和从业人员服务态度的关注,交通服务偏好主要体现在交通方式、导航指引、公共交通设施、租车服务、班车服务等方面,信息服务偏好反映游客对旅游咨询、路线规划和信息准确度的需求。②信息服务和住宿服务的改进空间最大,尤其是在信息准确度、获取便捷性和住宿卫生、设施完备性方面。交通服务和餐饮服务也存在提升需求。③游客在旅游服务中,更关注住宿的清洁卫生、餐饮的本土体验、交通的便捷与经济性,以及信息的准确和便捷。此外,男性工作族对交通服务的满意度较低。

4.2　优化建议

(1)优化信息服务。建立完善的旅游咨询中心,提供详细的旅游信息和路线规划服务;开发景区官方 App 和小程序,整合景点信息、导航、导览等功能,提供实时景点介绍、旅行攻略和电子导览服务。在景区内设立多个旅游咨询点,并配备专业的服务人员,方便游客获取信息。提升信息准确度定期更新景区信息,确保信息的准确性和时效性;设立专门的客服团队,高效解答游客疑问,及时处理投诉和反馈。建立 24 h 在线客服服务,快速响应游客的咨询和投诉。提供多语言导览服务,满足不同国籍游客的需求;引入智能导览设备,如电子导览器和增强现实(augmented reality,AR)导览服务,并在主要景点设置租赁点。提升游客的导览体验。招募多语言导游,并提供相关培训。

(2)优化住宿服务。提升设施和卫生标准,加强卫生检查与监督,增加清洁人员数量与卫生检查频次,提高清洁服务质量,确保房间的整洁和卫生。优化预订流程,提供多样化的预订渠道,如官网、App,积极与知名第三方平台建立合作关系,扩大预订渠道,方便游客预订;优化在线预订系统,增加智能推荐功能,确保预订流程的便捷性和安全性。同时丰富支付方式,提供预订确认和提醒服务。提供更多价格区间选择,根据不同游客需求,新增不同档次的住宿选择,提供从经济型到高档型的多种住宿,如经济型民宿和高端民宿。定期推出优惠活动,如提前预订优惠和长住折扣,通过优惠活动和打折促销,吸引更多游客。

(3)优化交通服务。提升景区内的班车服务频次,增加班车运行频次,减少游客等待时间;增加班车数量,调整运行时间,确保高峰期有充足的班车服务。规划多种班车线路,覆盖更多周边景点及旅游热点区域。优化交通费用结构,根据游客需求,对现有交通费用进行分析,调整交通费用及价格结构。提供更多优惠措施,如日票、周票和月票,满足不同游客的出行需求。推出多种交通票务选择,并在售票点和在线平台进行宣传。增加停车场地,在景区周边规划并建设新的停车场地,增加停车位数量,提供充足的停车位;设立停车引导系统,并在景区内设置清晰的指示牌,方便游客找到合适的停车位置。

(4)优化餐饮服务。提升食物质量和多样性,定期更新菜单,推出季节性和地方特色菜品;严格把控食材质量,确保食物的新鲜和卫生。加强环境卫生管理,制定餐厅清洁计划,定期进行清洁和消毒,保持用餐环境的干净整洁。提升服务人员素质,定期举办服务培训课程,提高服务人员的专业素质,提高服务质量。实施服务质量评估机制,定期进行服务质量评估,并根据评估结果及时改进服务中的不足。

(5)加强数字化转型。构建智能化旅游服务平台,如建立一个集成的智能化旅游服务平台,该平台能够提供一站式服务,包括在线预订、电子票务、智能导览、实时信息更新等;利用大数据分析和机器学习技术,为游客提供个性化的旅游推荐,包括住宿、餐饮、活动等;技术人员可开发景区专属的移动应用程序,提供导航、导览、预订、社交分享等功能,增强游客的互动体验。加强数字化营销,景区利用社交媒体平台进行精准营销,通过分析用户行为和偏好,推送定制化的内容,提高用户参与度和品牌影响力;运用数字广告技术,如搜索引擎营销(SEM)和程序化购买,以提高广告投放的效率和效果;积极管理在线评价和口碑,通过及时响应和解决问题,提升游客满意度和忠诚度。提升数字化运营效率,部署智能监控系统,实时监控游客流量和

行为模式，以优化资源分配和提高安全管理；建立数据分析平台，通过收集和分析游客数据，为管理层提供决策支持，优化运营策略；对员工进行数字化技能培训，提高他们对新技术的适应能力和服务效率。

参考文献

[1] 李鹜扬，李甜甜，王钺伟，等. 增权视角下的少数民族村落乡村旅游产业发展[J]. 科技与产业，2024，24(15)：31-36.

[2] 陈伍香，刘珊删，朱亚慧，等. 乡村旅游质量游客感知、地方依恋对游客环境责任行为的影响——以广西龙脊梯田景区为例(英文)[J]. Journal of Resources and Ecology，2024，15(3)：673-682.

[3] 侯慧娟，谷慧玲. 数字化转型背景下文旅企业投资现状及对策研究[J]. 科技和产业，2024，24(24)：197-202.

[4] 汪菁. 数字乡村建设的发展现状、行动困境和优化路径——基于浙江省德清县数字乡村建设的实践[J]. 科技与产业，2023，23(4)：128-132.

[5] 王晓慧，覃京燕. 基于深度学习的情感化设计[J]. 包装工程，2017，38(6)：12-16.

[6] 杨家文，陈若宇，段阳. 珠三角公共交通需求-供给关系类型与影响因素分析[J]. 地理研究，2024，43(12)：3215-3227.

[7] 闫岳飞. 乡村旅游景区游客付费意愿的影响因素——以石家庄市井陉村群为例[J]. 科技与产业，2024，24(12)：143-148.

[8] PARASURAMAN A P，ZEITHAML V，BERRY L. SERVQUAL：a multiple-item scale for measuring consumer perceptions of service quality[J]. Journal of Retailing，1988，64(1)：12-40.

[9] 钟飞燕，姚凯宜，何慧. 基于SERVQUAL-IPA模型的乡村民宿服务质量评价与提升路径研究[J]. 科技和产业，2024，24(2)：124-131.

[10] 吕维霞，张鸿颖，肖志璇. 消费者感知视角下网络餐饮服务监管效果评价指标及其影响因素[J]. 东北大学学报(社会科学版)，2023，25(6)：36-44.

[11] 李世伟，辛晓敏，潘福全. 基于模糊综合评价方法的青岛地铁乘客满意度分析[J]. 城市轨道交通研究，2020，23(8)：117-119.

[12] 沈旺，王俊尧，李昕娱，等. 突发事件影响下考虑公平与透明对信息服务满意度的影响研究[J]. 图书情报工作，2024，68(12)：31-42.

[13] AJZEN I. The theory of planned behavior[J]. Organizational Behavior and Human Decision Processes，1991，50：179-211.

[14] 厉新建，李兆睿，宋昌耀，等. 基于计划行为理论的虚拟旅游行为影响机制研究[J]. 旅游学刊，2021，36(8)：15-26.

[15] 厉新建，曾博伟，张辉，等. 新质生产力与旅游业高质量发展[J]. 旅游学刊，2024，39(5)：15-29.

[16] 叶紫青，张颖熙. 新质生产力赋能文化旅游业高质量发展的理论逻辑与发展路径[J]. 经济问题，2024(9)：16-23.

[17] 成官文，王敦球，秦立功，等. 龙脊梯田景区生态旅游开发的生态环境保护研究[J]. 经济地理，2001(S1)：320-322.

[18] 冯学钢，李志远. 新质生产力与旅游业高质量发展：动力机制、关键问题及创新实践[J]. 上海经济研究，2024(9)：38-48.

[19] 魏伟，丁双莹，刘方方，等. 旅游信息科学：多学科交叉融合背景下的理论与实践探究[J]. 地球信息科学学报，2023，25(9)：1747-1764.

[20] 卢欢，杨森甜，张妍妍，等. 基于街景照片的城市旅游情感空间探测研究[J]. 人文地理，2023，38(3)：164-172.

[21] 贺文华，李纯青. 游客体验旅程设计有效性对主题公园重游意愿的影响机理——基于双因素理论[J]. 旅游学刊，2024，39(8)：98-111.

[22] 樊亚明，孙正阳，张晓莎，等. 基于UGC数据的农业文化遗产地景观意象感知研究——以龙胜龙脊梯田为例[J]. 桂林理工大学学报，2024，44(2)：357-365.

[23] 兰晓芳，刘卓，许志豪，等. 基于TF-IDF和TextRank结合的中文文本关键词提取方法——以体育新闻为例[J]. 软件工程，2023，26(8)：6-10.

Research on Tourist Service Preference and Satisfaction of Guilin Longji Terraced Fields Scenic Area Based on Deep Learning and Cluster Analysis

LIU Xingyu[1]，YAO Xiaoxuan[2]

(1. Faculty of Business Administration，University of Macau，Macau 000011，China；
2. School of Mathematics and Statistics，Guangxi Normal University，Guilin 541006，Guangxi，China)

Abstract：The tourism industry is undergoing a profound transformation，with tourists' demands and preferences showing a diversified trend. Accurately identifying tourists' preferences and satisfaction levels is of crucial importance. Taking the Longji Terraced Fields Scenic Area in Guilin as a case study，deep learning technology and cluster analysis methods were employed to extract the themes of tourists' service preferences from online reviews and identify groups of tourists with similar preferences. Combined with questionnaire survey data，the service satisfaction was quantitatively evaluated. The research finds that tourists' service preferences cover accommodation，catering，transportation and information services. The improvement space for information services and accommodation services is the largest，while transportation services and catering services also have room for improvement. Based on this，corresponding suggestions are put forward.

Keywords：Longji Terrace Scenic Area；service preference；deep learning；tourist clustering；market competitiveness