

# 融合内容语义与空间分布的在线旅游评论可视化方法

张书美 刘畅 钱乃畅 安洁玉\*  
赣东学院

DOI:10.32629/acair.v4i1.19347

**[摘要]** 在线旅游评论蕴含着极具价值的用户反馈信息,但其海量、非结构化的特征导致传统人工分析难以有效挖掘深层规律。为降低人工分析的低效率与主观偏差,本文设计并实现了一套面向景区口碑的文本挖掘与可视化分析系统。相较于仅依赖词频统计的做法,本研究强调“语义内容—地理空间”的联合分析:通过清洗与特征提取,将评论文本转化为可计算的向量表示;随后以词云展示游客关注重点,并结合地理热力图刻画评论的空间分布。该系统可为游客出行决策、景区管理及行业研究提供数据支持,对文旅行业的数字化与智能化发展具有一定参考意义。

**[关键词]** 在线旅游评论; 文本挖掘; 可视化技术; 热力图; 词云

**中图分类号:** F590 **文献标识码:** A

## A Visualization Method for Online Travel Reviews Fusing Content Semantics and Spatial Distribution

Shumei Zhang Chang Liu Naichang Qian Jieyu An\*  
Gandong University

**[Abstract]** Online travel reviews contain valuable user feedback, yet their massive volume and unstructured nature make it difficult for traditional manual analysis to uncover deeper patterns effectively. To reduce the inefficiency and subjective bias of manual analysis, this paper designs and implements a text-mining and visual analytics system for scenic spot reputation. Compared with approaches relying solely on word-frequency statistics, this study emphasizes a joint analysis of semantic content and geospatial information: through data cleaning and feature extraction, review texts are transformed into computable vector representations; a word cloud is then used to highlight tourists' key concerns, and a geographic heat map is employed to depict the spatial distribution of reviews. The proposed system can support tourists' travel decisions, scenic spot management, and related academic research, and offers practical reference for the digital and intelligent development of the cultural and tourism industry.

**[Key words]** Online tourism reviews; Text mining; Visualization technology; Heat map; Word cloud

### 引言

在数字技术深度融入日常生活的当下,在线旅游行业已全面进入数据驱动的发展新阶段。旅游垂直平台、社交平台、在线旅行社等多元渠道持续积累的海量用户评论,既是涵盖游客食、住、行、游、购、娱全链条的多维度真实体验“原始数据资源”,更是基于文本挖掘的在线旅游评论可视化方法的核心分析对象。挖掘这些数据的价值,对旅游生态的各参与方意义深远<sup>[1]</sup>。

本文聚焦基于文本挖掘的在线旅游评论可视化方法,通过深度融合文本挖掘和文本可视化两大技术实现评论的高效分析与直观展示。在数字化浪潮席卷各行各业的背景下,该方法不仅是激活旅游评论信息资产价值的关键工具,更是突破行业数据

处理瓶颈、释放数据潜能的核心路径,将为旅游行业的数字化、智能化转型注入强劲动力。

### 1 现有文本挖掘技术分析

#### 1.1 预处理技术

预处理是文本挖掘的基础环节,旨在净化数据、降低噪声,为后续分析提供可靠的数据基础。主要包括三步:一是数据清洗,去除评论中可能存在的HTML标签、特殊符号(如“@”“#”)以及重复数字、乱码等无意义内容,减少无关信息对关键词提取和情感分析的干扰;二是词形归一(词干提取/词形还原),将同一词语的不同表达统一为基本形式,如“满意的”“满意地”统一为“满意”,“便捷性”与“便捷”统一为“便捷”,以提升统计与分析的准确性;三是停用词去除,过滤“的”“是”“在”

等高频但语义贡献较低的词语,从而降低数据规模,提高后续处理效率。

### 1.2 特征提取技术

特征提取的核心是将非结构化文本转化为机器学习算法可处理的数值或向量形式,当前主流技术主要分为两类:一是词袋模型<sup>[2]</sup>,将评论表示为“词汇—出现次数”的计数向量,例如“环境优美,服务贴心”可表示为“环境:1,优美:1,服务:1,贴心:1”。该方法实现简单、计算效率高,但不考虑词序与语法结构,语义表达能力有限,适用于以词频统计为主的分析场景;二是TF-IDF<sup>[3]</sup>,通过综合词语在单条评论中的出现频率与在整体语料中的稀缺程度计算权重,词语在单条评论中出现越频繁、在全体评论中越少见,其权重越高。因此,TF-IDF更有利于突出评论中的关键信息(如“排队时间长”“停车难”等),在旅游评论主题挖掘中应用较为广泛。

### 1.3 模型构建技术

模型构建需结合具体任务选择算法,比如情感分析常采用朴素贝叶斯、支持向量机,主题识别则多运用LDA(潜在狄利克雷分配)模型;对评论进行主题提取,并通过余弦相似度对主题间的继承、消失、新生等演化路径进行分析。不过,旅游评论存在口语化、歧义性的特点,像“性价比高”这类表述可能因用户认知差异产生不同解读,导致部分模型的预测精度仍需结合实际场景进一步优化。

## 2 文本可视化相关技术

基于文本挖掘的可视化属于典型的前沿跨学科领域,涉及文本分析、数据挖掘、计算机图形学、人机交互以及认知科学等方面的理论和技术。它能够将文本中的信息先通过文本挖掘的手段处理成文本数据,再经过可视化转换成图形或者图像,让人们以视觉的方式交互式地探索数据背后的信息<sup>[4]</sup>。

### 2.1 词云可视化技术

词云图通过视觉映射机制,将词汇的统计特征转化为视觉属性。以词汇出现频率为核心依据,频率越高的词汇在词云中字体越大;同时采用颜色编码技术,通过颜色区分情感倾向(如绿色表示正面评价、红色表示负面评价)。这种形式能让用户快速捕捉评论中的热点话题(如“景色”、“值得”、“历史”),准确把握游客对景点的整体评价倾向。

### 2.2 热力图可视化技术

热力图可视化基于地理信息系统(GIS),将游客评论与对应的地理位置进行关联,并以颜色梯度表达评论在空间上的分布强度:颜色越深表示评论越密集(如深红色代表高密度区域,浅黄色代表低密度区域)。借助热力图,用户能够直观识别热点与冷点区域,为旅游决策与景区管理提供空间维度参考<sup>[5]</sup>。

## 3 案例验证与系统实现

### 3.1 案例选取与数据构建

本研究以具有典型文化地标属性的滕王阁景区作为实证对象,采集网络平台上的游客评论数据,内容覆盖景观体验、服务

质量、交通出行与配套设施等多个方面。为实现对在线评论的综合呈现与分析,本文构建了一套融合文本挖掘与地理信息系统的可视化分析系统,从“内容—空间”两个维度展示评论特征,使用户能够准确把握评论的关键信息及其对应的语境。

### 3.2 语义特征的可视化呈现

图1为滕王阁评论可视化系统界面。系统中的交互式词云展示了从游客评论中提取的高频关键词,如“南昌”“景点”“赣江”“历史”“服务”等,便于用户快速把握评论关注点与整体评价倾向。词云以字体大小体现关键词的出现频率与相对重要性,并通过颜色区分正面、负面等情感属性。用户点击某一关键词后,系统将实时联动显示相关原始评论,帮助进一步了解具体语境与评价细节,从而更直观地掌握景区体验信息,为出行决策与行程规划提供参考。



图1 在线评论的交互式词云可视化分析视图

### 3.3 评论数据的空间分布特征分析

图2展示了滕王阁及其周边区域游客评论在地理空间上的聚集特征。本文将评论数据匹配到对应地理坐标,并采用核密度估计生成热力图,通过颜色由浅至深的梯度直观反映评论点位的密度差异。

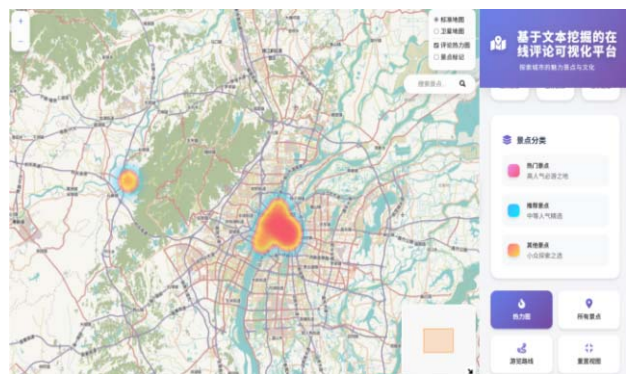


图2 景区评论密度的地理空间分布热力图

从空间格局看,评论分布呈现出以滕王阁为主的核心集聚特征:核心景区附近出现高强度热区,表明该区域是游客活动与

评价信息最为集中的位置;外围区域整体密度较低,仅在个别位置形成零星的次级聚集点,体现出一定的中心向外递减趋势。结合交互式缩放观察可见,评论热点与景区核心游览空间高度重合,说明游客评价更集中于标志性景观与高频停留区域。该结果提示管理者需关注核心区的承载与疏导,同时可考虑通过线路组织与服务节点布局,引导游客关注并带动周边区域的游览活力。

### 3.4 基于多维特征融合的游览路线推荐

针对外地游客的“深度游”需求,系统突破单一热度指标的局限,构建了“空间热度+语义评价”的双因子游览路线推荐模型。该模型的核心逻辑为:一是融合热力图分析得出的高频访问空间节点,确保路线覆盖景区核心游览区域;二是叠加文本挖掘提取的高情感评分标签(如“景色宜人”“文化氛围浓”“交通便利”“服务贴心”等),兼顾游览价值与体验舒适度。通过对两类因子的综合加权计算,系统生成优化后的游览路线,可为希望全方位领略城市风貌与景区特色的游客提供科学、高效的行程规划参考。

### 3.5 案例应用价值与管理启示

通过对滕王阁评论可视化系统的实证应用,本研究验证了基于文本挖掘的可视化方法在旅游管理中的核心价值。该系统打破了传统数据分析的黑盒模式,通过整合预处理、特征提取及多维可视化技术,建立了一套从“数据输入”到“决策输出”的完整闭环。

词云与热力图的有机结合,赋予了景区管理者“上帝视角”般的洞察力:一方面通过系统可以精准捕捉游客的核心诉求,另一方面能基于数据快速定位服务短板(如通过关键词关联发现的“排队拥堵”、“停车困难”等高频痛点)。这使得管理方能够从被动应对转向主动优化,例如针对性地改善交通微循环或优化票务流程。该方法不仅为景区的运营提供了数字化工具,更为同类文化遗产景点的口碑监测与数据资产活化提供了可复制、可推广的范式,具有广阔的行业应用前景。

## 4 结语

本文构建了融合文本挖掘与可视化技术的在线评论综合分析框架,实现了对非结构化评论信息的结构化表达与直观呈现。该系统既可以帮助游客快速把握景区评价重点,为出行决策提供参考,也能为景区管理者识别热点与服务问题、优化运营管理提供数据支撑。与此同时,本文提出的“语义提取-空间映射”技术路径具有一定通用性与可扩展性,可为电商、餐饮等泛服务行业的口碑监测与数字化分析提供方法参考。

### [基金项目]

江西省大学生创新创业训练计划项目(项目编号:S202513432024)。

### [参考文献]

- [1] 缪迅杰,徐鑫茹,徐泽水.在线评论在决策管理中的应用研究进展及展望[J].管理学报,2025,22(06):1179-1188.
- [2] 魏崑,丁香香,郭梦星,等.文本相似度计算方法综述[J].计算机工程,2024,50(09):18-32.
- [3] 胡荣笙,车文刚,张龙,等.融合TF-IDF算法和预训练模型的文本数据增强[J].计算机仿真,2024,41(05):495-500.
- [4] 王曼瑾,何媛.基于文本挖掘的大唐不夜城景区用户评论分析[J].现代计算机,2024,30(13):51-56.
- [5] 孙焕良,马晓慧,王亚星,等.支持热力图等级偏好的路径规划算法[J].计算机与数字工程,2024,52(01):9-17.

### 作者简介:

张书美(2005--),女,汉族,陕西省安康市人,本科生,赣东学院,研究方向:计算机科学与技术基础理论与应用研究。

刘畅(2005--),女,汉族,辽宁省阜新市人,本科生,赣东学院,研究方向:计算机科学与技术基础理论与应用研究。

钱乃畅(2006--),女,汉族,辽宁省铁岭市人,本科生,赣东学院,研究方向:计算机科学与技术基础理论与应用研究。

安洁玉(1982--),男,汉族,山东省日照市人,博士,赣东学院,研究方向:人工智能、可视化。