

# 基于文本挖掘的酒店满意度分析与优化研究

## ——以途牛网上海市酒店在线评论为例

谢菁菁, 宋瑾钰\*

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: 2024年4月2日; 录用日期: 2024年4月16日; 发布日期: 2024年5月31日

### 摘要

随着酒店业的快速发展, 电商平台的在线评论成为了研究消费者行为和改善酒店服务的重要资源。本研究综合应用了数据爬取技术、LDA主题建模方法以及SnowNLP模型的情感分析, 对途牛网上海市酒店的评论数据进行了全面分析, 深入探讨酒店评论中消费者的满意度及情感分布, 了解消费者对酒店服务的核心需求, 为酒店管理和优化提供了数据支撑。

### 关键词

在线评论, 文本挖掘, LDA主题模型, 情感分析, SnowNLP

# Research on Hotel Satisfaction Analysis and Optimization Based on Text Mining

## —Taking Online Reviews of Shanghai Hotels on Tuniu.com as an Example

Jingjing Xie, Jinyu Song\*

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

Received: Apr. 2<sup>nd</sup>, 2024; accepted: Apr. 16<sup>th</sup>, 2024; published: May 31<sup>st</sup>, 2024

### Abstract

With the rapid development of the hotel industry, online reviews on e-commerce platforms have become an important resource for studying consumer behavior and improving hotel services. This

\*通讯作者。

research comprehensively applies data crawling technology of the Selenium library, LDA topic modeling method, and sentiment analysis of the SnowNLP model to conduct a thorough analysis of the review data of Shanghai hotels on Tuniu.com. It delves into the satisfaction and emotional distribution of consumers in hotel reviews, understanding the core needs of consumers for hotel services, and provides data support for hotel management and service optimization.

## Keywords

Online Reviews, Text Mining, LDA Topic Model, Sentiment Analysis, SnowNLP

Copyright © 2024 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

在当今社会快速发展的大环境下,无论是商务出差、旅游休闲还是探亲访友等多样化场景,人们对于出行的需求显著增加,进而导致了对酒店住宿的需求量持续上升。

传统酒店服务交易通常是通过线下预定或者电话预定实现,但在数字化的今天,电商平台已经成为消费者获得住宿服务的首选途径。与能够线下查看检验商品的传统方式不同,通过电商平台预订服务或购买商品,消费者必须依赖于商品描述和其他消费者提供的在线评价来进行抉择。在这个过程中,用户评价扮演了关键角色,它不仅为消费者提供了宝贵的决策支持,也为企业提供了洞察消费偏好和调整市场策略的关键信息。

现有较多学者将文本挖掘技术运用到消费者满意度研究中,通过具体的技术手段分析消费者对相关产品的满意度影响因素及提升策略。例如,陈思含等人[1]通过构建隐含狄利克雷(LDA)模型对健康医疗可穿戴设备评论数据进行主题识别和对主题下的评论数据进行情感分析,了解用户对商品在不同主题下的满意度及商品之间的差异。陈泽明[2]利用 LDA 主题模型和 Python 的 SnowNLP 模块对样本进行主题分类以及情感分析,基于游客感知视角对影响西塘古镇旅游形象的感知因素进行分析。马梦曦,张骏[3]为了有效地获取弹幕的情感极性,利用 TF-IDF 提取弹幕文本的特征,通过空间向量模型进行特征表示,最后,利用 SVM 对文本的情感极性进行分类。

主题模型(如 LDA)、情感分析[4]、TF-IDF 权重计算以及机器学习算法(如 SVM [5]),这些文本挖掘技术应用于消费者满意度研究,从大规模文本数据中提取有用信息,更深入地理解消费者的需求和偏好,为企业提供改进产品和服务的参考依据。

因此,本文通过爬取途牛网用户在线评论,使用数据挖掘技术,对 4846 条数据进行分词、高频词分析、LDA 主题聚类、TF-IDF 权重计算等操作,通过对评论的量化分析,获取影响消费者满意度的重要因素,然后针对具体内容进行拓展分析,进一步提高企业的服务质量,从而提升消费者消费体验和满意度。

## 2. 数据收集以及研究设计

### 2.1. 数据收集

途牛网(<https://www.tuniu.com>)创建于 2006 年 10 月,以“让旅游更简单”为企业使命。作为中国领先的数字一体化旅游服务商,途牛为线上、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。特别是在酒店预订领域, 通过关键词“上海市”进行搜索, 得到的酒店选项总数达到 45,949 家, 这一数字超过了北京、成都、武汉等其它热门旅游目的地的酒店总数, 显示了其数据具有较高的代表性。因此, 本文选择途牛网中关于上海市酒店的在线消费者评论作为研究对象。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, 最终收集到关于 320 家酒店的 4846 条消费者在线评论。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 拥有众多丰富的第三方库, 例如 NumPy、Matplotlib、Pandas 和 WordCloud 等, 已经成为文本数据分析领域中不可缺少的分析工具[6]。这些库使 Python 在处理复杂的数据分析任务, 特别是在文本挖掘和自然语言处理领域中, 展现出了其独特的优势。

LDA (Latent Dirichlet Allocation)是基于共轭先验原理以及贝叶斯框架的文档主题生成模型, 包括文档、主题及特征三层结构, 所以也被叫做三层贝叶斯概率模型[7]。此模型在处理和大量在线评论文本数据方面表现出了显著的优势, 尤其是在识别文本数据中隐含的主题分布及其相应的特征词分布方面。通过自动地从文本中提炼出主题和特征词, LDA 主题模型为深入理解和分析消费者评论背后的意见、态度和需求提供了强有力的支持, 极大地促进了文本数据分析的深度和广度, 对于文本分析研究具有重要意义。

SnowNLP 算法[8]是基于自然语言处理技术(NLP) [9], 专门为中文文本设计的算法。主要具有中文分词(算法是 Character-Based Generative Model)、词性标注、情感分析、文本分类(原理是朴素贝叶斯)、转换拼音、繁体转简体、提取文本关键词、提取摘要、分割句子、文本相似等功能[10]。

本文主要使用 Python 数据挖掘技术, 结合 ROSTEA 软件, 对 4846 条在线评论进行量化分析, 包含分词、高频词提取、词云生成、LDA 主题挖掘、通过 SnowNLP 模型获取消费者情感特征词及其 TF-IDF 权重。

## 2.3. 研究流程

首先利用爬虫技术收集途牛网上海市酒店的用户评论, 并通过 Python 进行数据预处理, 包括建立自定义词典、剔除停用词和应用 Jieba 进行精确分词。接着, 通过统计词频来识别高频词, 并绘制词云图以直观展示。进一步, 运用 LDA 主题模型对数据进行聚类分析, 以识别出主要主题及其关键词。最后, 通过 ROSTAE 工具统计在线评论的情感分布, 在此基础上, 借助 SnowNLP 模型分析消费者的情感倾向, 并利用 TF-IDF 方法确定情感特征词的权重, 结合 LDA 主题关键词进行综合研究, 深入探讨消费者关注的核心问题及其满意度。流程图见如下图 1。

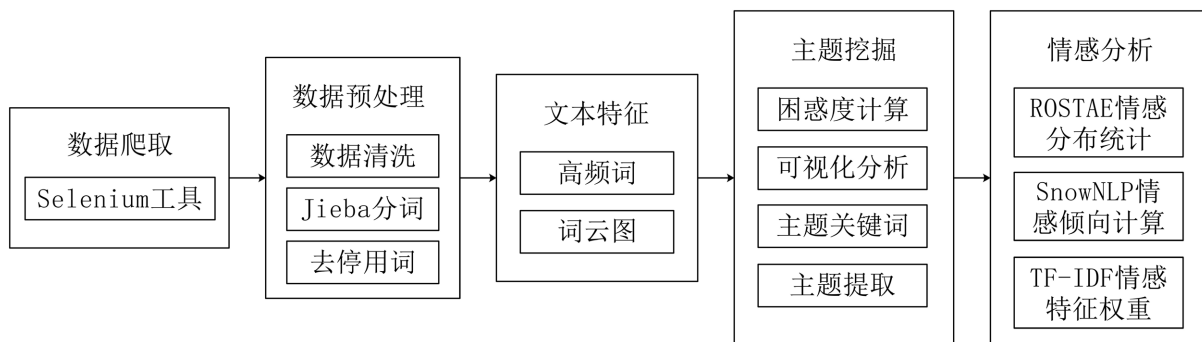


Figure 1. Text mining and analysis flowchart  
图 1. 文本挖掘与分析流程图

### 3. 数据分析与结果

#### 3.1. 词频分析

词频分析是文本分析的重要步骤, 高频词能够体现文本的关键信息。本文运用 Python 中的 Jieba 分词(精确模式), 对采集的评论进行分词处理, 去除副词、无意义词语、合并同类词后, 得到的词频统计如下表 1 所示, 并使用 Python 的 WorldCloud 库绘制了在线评论的词云图如下图 2 所示。

**Table 1.** Word frequency statistics

**表 1.** 词频统计

高频词	频数
酒店	2827
房间	1271
服务	1266
迪士尼	1078
干净	701
早餐	614
位置	572
环境	560
前台	379
班车	366
设施	354
免费	339
接送	337
交通	302
卫生	260
地铁	233
舒适	211
孩子	208
乐园	205



**Figure 2.** Word cloud diagram of consumer online reviews

**图 2.** 消费者在线评论的词云图

### 3.2. 基于 LDA 主题模型的特征分析

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。

在 LDA 模型中, k 是一个超参数,需要在训练模型之前指定。它决定了模型将数据分成多少个主题。正确选择 k 的值对模型的性能和主题的质量有重要影响,过多的主题可能会导致各主题间区分度降低,增加模型的复杂度,而设定的主题数目过少可能导致未能全面映射文档集内的关键主题。

运用 Python 对上海市酒店在线评论进行 LDA 主题挖掘,并通过 pyLDAvis 实现结果可视化。如图 3 所示,圆圈代表 LDA 聚类得出的主题,圆圈的面积反映主题的相对重要性,圆圈之间的距离则体现各个主题之间的差异性。通过多次模拟并对不同 k 值下的 LDA 模型可视化结果,发现当 k 值为 3 时,模型展现出较低的困惑度,并且主题间重叠最少,差异最为显著。这表明在此条件下聚类精确度较高,更适合作为聚类主题数,能够更好地概括和解释数据集。因此,将上海市的酒店在线评论分为三种主题较为合理。

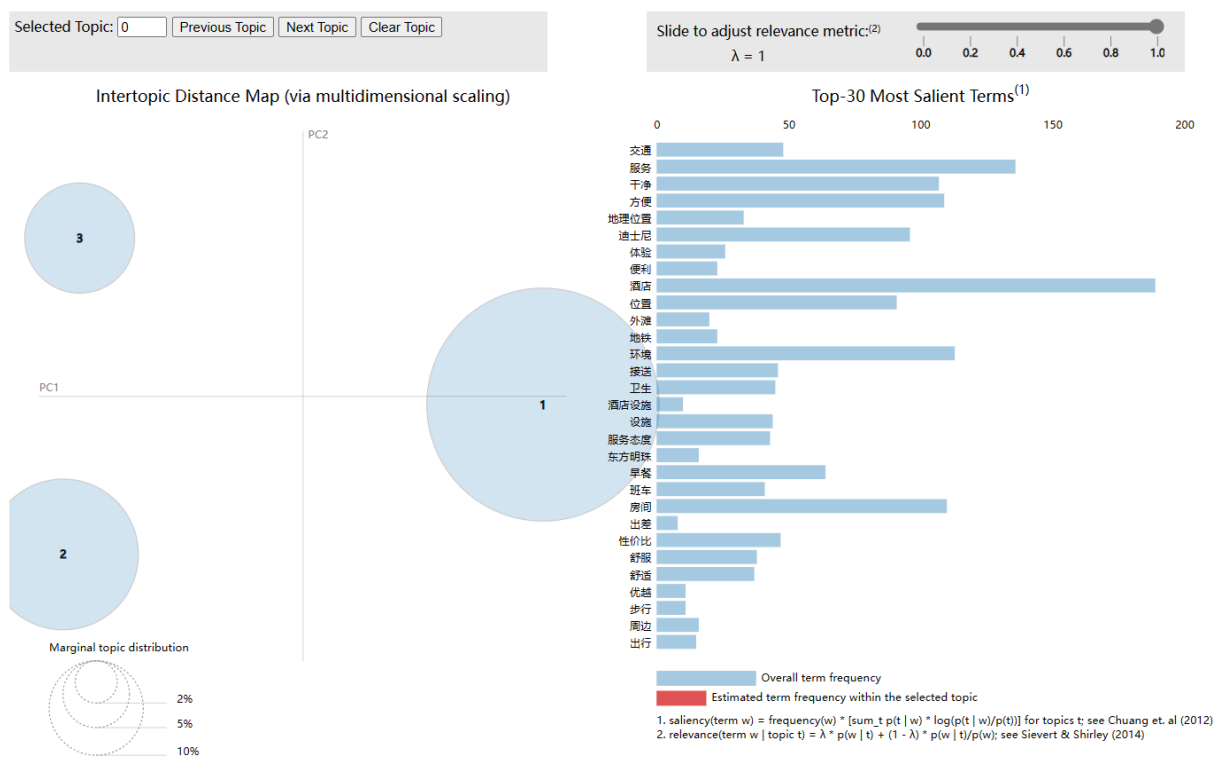


Figure 3. Visualization of the LDA model with three topics  
图 3. 主题为 3 的 LDA 模型可视化图

本文使用 LDA 主题聚类模型得到 3 个主题,并选取了每个主题排名前 16 的特征词,结果如表 2 所示。结合日常生活信息,并根据表内高权重特征词的含义,最终将 3 个主题概括为服务细节、设施环境和位置交通。

第一类主题主要和服务有关,排名前十六的词汇包含许多有关酒店设施、服务的名词,其中高权重词“酒店”“服务”“干净”“环境”“房间”等与基础服务有关,“早餐”“接送”“班车”等则代表了更加全面细致的服务。该主题主要体现了消费者对于酒店服务要求的全面性和细节性。消费者对服

务的评价不仅限于基本的清洁和房间条件,更扩展到了酒店提供的额外服务,如免费早餐、接送服务、班车等便利条件。同时,“迪士尼”这一特征词的出现,意味着某些酒店因靠近迪士尼乐园会受到更多家庭游客的欢迎。这些游客通常会对离迪士尼近且能提供额外服务的酒店有更高的预期。因此,将该主题概括为服务细节类。

第二类主题反映了消费者对于酒店整体环境的关注。该主题排名前十六的词汇中,“体验”“酒店设施”“出差”“市区”等词汇出现频率较高,说明该主题内的消费者对于环境有各自的要求,在住宿后也乐于分享对于不同环境的想法感受。例如,“出差”客户可能特别关心酒店的商务中心、会议室等商务相关设施,而“市区”表明了一个便捷的地理环境对于快速访问城市中心和其他重要地点的重要性。这些词汇共同描绘了一幅客户对酒店环境细节的期待图。由于该主题的数据反映了消费者对酒店内外环境的高度关注,因此概括为设施环境类。

第三类主题聚焦在酒店的地理位置及其交通连通性上,这些因素对于游客和商务旅客的便利性至关重要。在这一主题的特征词中,“方便”一词的权重最高,表明方便性是评价酒店时考虑的首要因素。同时,“交通”与“地铁”等词的出现频率较高,揭示了客户在选择酒店时会考虑到其接近主要交通线路的程度,以确保能够轻松到达城市中其他地区或旅游景点。诸如“出行”和“步行”这样的词汇,也暗示了游客倾向于选择那些可以方便他们步行至附近热门地点或商业区的酒店。此外,“位置”本身的出现也指向了对酒店所处环境的细节评价,包括对周围地区的安全、便利店或餐饮选项的丰富性。在特征词中还出现了如“外滩”“东方明珠”等知名地标名称,这可能反映了一些客户特别是旅游者对于靠近城市著名景点的酒店的偏好。以上可以看出消费者在评论中特别强调了地理位置对于整个酒店体验的影响。他们不仅关心到达酒店的便捷性,也关注酒店本身提供的位置上的优势,如接近商务区或旅游景点等,因此概括为位置交通类。

**Table 2.** LDA topic categories and feature word weights  
**表 2.** LDA 主题类别及特征词权重

服务细节		设施环境		位置交通	
特征词	权重	特征词	权重	特征词	权重
酒店	0.275433756	体验	0.116422667	方便	0.194006671
服务	0.222284215	酒店设施	0.060357473	交通	0.186823354
干净	0.174653528	出差	0.056858308	位置	0.129330693
环境	0.170391158	市区	0.032826735	地理位置	0.126706371
房间	0.162412858	一如既往	0.032288309	便利	0.091169867
迪士尼	0.157256279	环境卫生	0.03129247	酒店	0.080512785
早餐	0.10159407	全季	0.026701963	外滩	0.078008776
方便	0.097741712	完美	0.023520858	地铁	0.072891234
位置	0.095090244	风格	0.021365383	东方明珠	0.057056748
性价比	0.076706401	闹中取静	0.02022417	周边	0.046968543
接送	0.074463551	水族馆	0.01807532	出行	0.045361261
卫生	0.073328522	宾馆	0.017912927	优越	0.042633938
设施	0.072014212	档次	0.017762711	步行	0.042077869
服务态度	0.070330494	老旧	0.016776014	房间	0.041728004
班车	0.067386918	现代	0.014780514	吃饭	0.037775141
前台	0.06534526	装修	0.014738794	周围	0.036272376

### 3.3. 情感分析

#### 3.3.1. 消费者情感分布分析

为进一步探讨消费者对酒店的满意程度, 本文对采集的 4846 条在线评论进行了情感分析, 统计出消费者的情感分布类型与所占比例, 具体消费者情感分析如表 3 所示。

**Table 3.** Consumer sentiment distribution statistics

**表 3.** 消费者情感分布统计结果

	情绪类型	数量(条)	所占百分比(%)
	积极情绪( $5, +\infty$ )	3942	81.35%
	中性情绪 $[5, 5]$	16	0.33%
	消极情绪( $-\infty, 5$ )	888	18.32%
积极情绪分段统计	一般 $(5, 15]$	985	20.33%
	中度 $(15, 25]$	1174	24.23%
	高度 $(25, +\infty)$	1783	36.79%
消极情绪分段统计	一般 $[-15, 5)$	134	2.77%
	中度 $[-25, -15)$	46	0.95%
	高度 $(-\infty, -25)$	26	0.54%

从表 3 的消费情绪分析可见, 积极反馈占据了主导地位, 积极评论总数达到 3942 条, 占比高达 81.35%, 这说明在通过电商平台购买的酒店住宿体验中, 消费者普遍表现出了高度的满意和积极态度。特别地, 积极情绪中“高度”积极的反馈最为显著, 有 1783 条评论归于此类, 占有所有积极反馈的 36.79%, 相关评论为: “环境卫生都很好交通很便利” “很不错, 干净整洁, 带孩子入住放心” 等, 充分体现了消费者对服务的积极评价, 这与酒店本身硬件软件各方面属性有关。而“中度”和“一般”的比例相近, 对于“中度”积极反馈, 共计 1174 条评论, 占积极情绪总数的 24.23%, 相关评论为: “不错, 还送了一次儿童早餐” “房间舒适, 服务也好免费接送” 等。进一步来看, “一般”积极反馈则显示了消费者对酒店基础设施和服务的基本满意。此类评论数量为 985 条, 占总积极反馈的 20.33%, 如“房间不大, 位置好, 方便” 等。虽然这些反馈可能没有涉及到过于具体或深入的细节, 但却稳固地支撑着消费者对电商平台酒店类商品的整体正面看法。

中性情绪的反馈相对较少, 共计 16 条, 占总评论数的 0.33%。这表明大多数消费者在体验完整酒店服务后都有明确的情感倾向, 不过也有一小部分消费者对其体验持中立态度。其中大部分消费者持有“房间不错, 就是停车位少了点” “酒店环境还不错, 早餐品种少了点” 等评价。企业需要将中性情绪顾客转变为积极情绪, 细化服务和提升体验是关键, 比如增加停车位和扩展早餐种类, 可以有效提升顾客满意度, 进而避免消费者转变为消极情绪。

在消极情绪方面, 共有 888 条评论, 占总评论数的 18.32%。这些消极反馈主要集中在酒店服务或设施不满足消费者期望的方面, 评论多为“其他都很好, 就是隔音太差” “酒店很一般, 只不过离迪士尼乐园近” 等。消费者通常会因为一个或几个突出的消极因素影响其对整体住宿体验的评估。在消极情绪的分段中, “一般”消极情绪的评论最多, 共有 134 条, 反映出消费者对酒店某些方面的基本不满, 但可能并不足以影响整体住宿体验。而“中度”和“高度”消极情绪的评论数量较少, 分别为 46 条和 26 条, 消费者的这些评论涉及到了酒店服务或设施的明显缺陷, 如“房间有异味, 影响睡眠” 或“预订的房型与实际不符, 非常失望” 等。虽然消极情感占比比较低, 但仍需保持警惕, 以防“一般”与“中度”

程度的消费者转换为“高度”消极情绪的消费者。

综上, 消费者对于电商平台购买的酒店住宿体验普遍持有积极态度, 消极情绪的表达相对较少, 且大多数消极情绪是适度的。尽管如此, 企业仍需关注那些表达中等和一般消极情感的消费者, 改进酒店硬件软件设施, 提升售前售后服务质量, 满足消费者的感知体验, 同时针对中性反馈进行细化服务提升, 以防止消极情绪的增加。

### 3.3.2. 基于 TF-IDF 的情感特征分析

为进行更加深入细致的数据分析, 通过 SnowLNP 模型得出各条文本数据的情感倾向分值。通常情况下, SnowNLP 的情感得分范围是 0 到 1, 得分越接近 1 表示越积极, 得分越接近 0 表示越消极, 0.5 作为阈值, 大于等于 0.5 的为积极评论, 小于 0.5 的为消极评论。将积极评论和消极评论分别进行词频统计, 以分析不同情感类型的消费者品论动机与心理, 从而发掘酒店业可提升角度。运用 SnowNLP 对于 LDA 主题模型生成的三类主题进行情感分析, 得到的各主体的情感倾向如表 4 所示。三类主题情感得分相差较小, 其中主题二、三得分较高, 说明消费者对酒店内部或周边的硬件配置都体验较好, 但对于酒店服务等软件配置较为一般。

**Table 4.** Sentiment orientation of three topic categories

**表 4.** 三类主题情感倾向

主题	SnowNLP 情感倾向
服务细节	0.8443855683295166
设施环境	0.9970075427766368
位置交通	0.9997276192908005

**Table 5.** Consumer emotional characteristic words and TF-IDF weights (Top 16)

**表 5.** 消费者情绪特征词及 TF-IDF 权重(前 16)

积极情绪		消极情绪	
特征词	TF-IDF 权重	特征词	TF-IDF 权重
酒店	0.713644937	酒店	0.618017778
服务	0.532847708	房间	0.445444268
迪士尼	0.444319731	服务	0.296461145
环境	0.425022715	迪士尼	0.265729431
干净	0.401106101	前台	0.244147757
房间	0.389831851	早餐	0.181685313
位置	0.347910813	干净	0.173106387
早餐	0.26905321	位置	0.165974045
交通	0.216670819	乐园	0.148717421
班车	0.20402382	卫生	0.145015286
性价比	0.197203299	设施	0.134142307
接送	0.1919692	空调	0.115360913
设施	0.184186536	价格	0.109881104
卫生	0.17961216	隔音	0.107310082
舒适	0.15735259	班车	0.106075372
态度	0.1570695	环境	0.103798776



为进一步了解消费者积极、消极评价是如何落地的,如表 5 所示,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 16 个高频词及 TF-IDF 权重。从总体上看,无论是积极评价还是消极评价,由双方共有的特征词“酒店”“服务”“迪士尼”“位置”可见,消费者在选择或评价酒店时,都普遍关注几个核心方面:酒店的整体服务、清洁度、地理位置及其周边的娱乐设施(如迪士尼)。

## 4. 结语

### 4.1. 研究结论

本文采用数据挖掘技术和 ROSTAE 软件对途牛网上海酒店的在线评论进行了综合分析,从而探究消费者对酒店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,分析揭示了顾客普遍关注的服务质量、地理位置和交通便利性等因素。结果表明,顾客特别重视酒店的服务细节、设施环境以及位置交通的便利性,其中位置交通的方便性和设施环境的优质被认为是选择酒店时的主要吸引因素。

情感分析的结果也显示了 18.65% 的评论带有中性或消极情绪,指出了酒店服务和基础设施方面的不足。这一发现突出了服务质量和酒店设施作为顾客决策和评价中的关键要素,同时指明了酒店需要关注并改进的领域,以提升顾客满意度和吸引更多顾客。

### 4.2. 优化策略

#### 4.2.1. 优化服务质量

提升服务质量是提高顾客满意度和积极评价的核心策略之一。酒店需着重培训前台接待和服务人员,确保他们能够以积极、专业的态度迎接每一位顾客。此外,引入客户管理系统以更有效地响应和处理客户的需求和投诉,可以显著提高服务的响应速度和效率。通过定期收集顾客反馈,酒店管理层能够及时调整服务流程,针对客户反映的问题采取具体改进措施,从而持续提升服务质量。

#### 4.2.2. 改善环境卫生

环境卫生对于酒店的整体印象至关重要。酒店应定期对公共区域和客房进行深度清洁,并确保所有设施均保持良好的卫生状态。此外,引入环保和可持续的清洁产品不仅有助于改善酒店的环境卫生,还能提升酒店的品牌形象。通过增加清洁人员的巡检频率,以及在必要时进行即时清洁,可以确保酒店环境始终保持清洁舒适,为顾客提供更加健康和安全的住宿环境。

#### 4.2.3. 提升房间配置和舒适度

房间是顾客体验酒店服务的主要场所,其配置和舒适度直接影响顾客的满意度。酒店应根据顾客反馈和市场趋势,定期更新房间内的设施和装饰,如升级床上用品质量、增加 USB 充电端口、提供高速无线网络服务等。同时,考虑到声音隔离是影响顾客睡眠质量的一个重要因素,改善房间的隔音效果也是提升顾客满意度的关键步骤。

#### 4.2.4. 优化餐饮服务和提供便捷交通服务

餐饮服务的质量和多样性是提升顾客满意度的另一个关键因素。酒店应考虑引入多样化的餐饮选择,满足不同顾客的饮食需求和偏好。定期举办餐饮文化活动或引入特色菜品,也能为顾客提供独特的用餐体验。对于位置较为偏远的酒店,提供便捷的交通服务(如机场接送、市中心班车等)可以显著提升顾客的便利性和满意度。

## 基金项目

浙江省一流课程建设项目。

## 参考文献

- [1] 陈思含, 张云秋, 李博诚. 基于 LDA 模型的健康医疗可穿戴设备评论情感分析[J]. 中华医学图书情报杂志, 2020, 29(12): 41-47. <https://doi.org/10.3969/j.issn.1671-3982.2020.12.007>
- [2] 陈泽明. 基于 LDA 主题模型的景区旅游形象情感分析[J]. 新丝路(中旬), 2019(3): 20-22.
- [3] 马梦曦, 张骏. 基于 TF-IDF 与 SVM 的弹幕文本情感极性分析模型[J]. 现代商贸工业, 2019, 40(8): 205-207. <https://doi.org/10.19311/j.cnki.1672-3198.2019.08.103>
- [4] 崔连超. 互联网评论文本情感分析研究[D]: [硕士学位论文]. 济南: 山东大学, 2016.
- [5] 肖健华, 吴今培, 杨叔子. 基于 SVM 的综合评价方法研究[J]. 计算机工程, 2002(8): 28-30.
- [6] 蔡颖. 基于 Python 的文本数据处理研究[J]. 软件, 2023, 44(5): 179-183.
- [7] 毛晓莉, 施本植. 新能源汽车普通消费者参与的大数据研究——基于文本挖掘和深度学习[J]. 海南大学学报(人文社会科学版), 2022, 40(5): 201-210. <https://doi.org/10.15886/j.cnki.hnus.20211022.001>
- [8] 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. <https://doi.org/10.16607/j.cnki.1674-6708.2018.18.052>
- [9] 林奕欧, 雷航, 李晓瑜, 等. 自然语言处理中的深度学习: 方法及应用[J]. 电子科技大学学报, 2017, 46(6): 913-919.
- [10] 夏玉芹, 单雪微. 基于 Python 的简单文本情感分析[J]. 阴山学刊(自然科学版), 2018, 32(4): 58-62.