

# Research on Multi-Dimensional Features of Budget Hotels Based on Online Reviews

Xue Liu, Ning Zhang\*

Business School of Qingdao University, Qingdao Shandong  
Email: \*shiny548984@163.com

Received: Jan. 30<sup>th</sup>, 2020; accepted: Feb. 13<sup>th</sup>, 2020; published: Feb. 20<sup>th</sup>, 2020

---

## Abstract

Online customer reviews (OCRs) are the emotional expressions of customers after product experience. We can use the Importance-Performance Analysis (IPA) to study the multidimensional features of hotels in OCRs. In this paper, Qingdao budget hotels are taken as an example. The methods of natural language processing technology and K-means are used to extract hotel features that customers pay attention to in OCRs, and IPA is used to determine the priority of multidimensional feature improvement. The conclusions show that: 1) service, location, and price are the main advantages of budget hotels. Managers maintain a competitive advantage and ensure the supply of resources. 2) Catering and room facilities are the main disadvantages of budget hotels. Managers should improve these two characteristics as soon as possible to improve customer satisfaction as soon as possible. This research uses OCRs to automatically obtain the features of customer attention, and uses the IPA model to determine the advantages and disadvantages of economic hotel development, which provides a more accurate reference for hotel managers in making decisions.

## Keywords

Budget Hotel, Online Customer Review, Revised IPA

---

# 基于在线用户评论的经济型酒店多维特征研究

刘雪, 张宁\*

青岛大学商学院, 山东 青岛  
Email: \*shiny548984@163.com

---

\*通讯作者。

## 摘要

在线用户评论(online customer reviews, OCRs)是顾客对于产品体验过后的情感表达, 可以借鉴重要性-绩效分析法(Importance-performance analysis, IPA)对OCRs中的酒店多维特征进行研究。本文以青岛市经济型酒店为例, 利用自然语言处理技术、K-means等方法提取OCRs中顾客关注的酒店特征, 结合修正IPA确定多维特征改进的优先次序。结论表明: 1) 服务、位置、价格是经济型酒店的主要优势, 管理者需保持竞争优势, 保证资源的供给。2) 餐饮、客房设施是经济型酒店的主要劣势, 管理者应尽快对这两类特征进行改进, 以尽快提高顾客满意度。该研究通过OCRs自动获取顾客关注的特征, 结合IPA模型确定经济型酒店发展的优劣势, 为酒店管理者进行决策提供更加准确的借鉴。

## 关键词

经济型酒店, 在线用户评论, 修正IPA

Copyright © 2020 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

随着 Web2.0 时代的迅速发展以及普及, 人们更加倾向于通过互联网在社交平台表达自己的想法并进行互动, 已有研究表明在线顾客评论(online customer reviews, OCRs)是一种重要的社会互动形式[1], 因此 OCRs 成为消费者和行业管理者主要的信息来源。与商家发布的产品客观描述相比较, OCRs 是来源于消费者对于产品真实体验过后的情感表达, 因此更具有信任性和说服力, 特别是对于体验型产品而言, 酒店类产品作为最典型的体验型产品之一, OCRs 会对潜在顾客的消费行为产生重大的影响。在酒店行业的发展过程中, 由于旅游业在线销售的持续增长, 国内外游客不断进入国内市场, 直接刺激了对于酒店业的需求, 青岛作为山东重要城市, 正在努力向建设国家中心城市发展, 旅游业是青岛的支柱性产业。青岛作为全国旅游业发展的重要城市, 吸引了大批经济型酒店的入住。

市场营销学以及经济学理论[2]认为产品以及服务都具有多维特征, 消费者对于各特征的偏重点有所不同, 因此对于 OCRs 中展示的各特征的情感表达的关注程度就会不同。识别产品的多维度特征的重要程度, 可以更加精准的根据消费者的需求进行改进以促进更高的销量。总结已有研究发现, 目前业界对于酒店满意度的研究已经比较深入, 但是大多采用调查问卷的形式或者直接采用网站数值评分, 很少对在线顾客评论的文本内容进行分析。

重要性-绩效性分析法(Importance-performance analysis, IPA)是测量顾客满意度的重要模型[3], 其应用领域广泛, 但是也存在一定的缺陷, 其对于模型的使用有严格的假设条件, 实际数据并不能满足模型的假设要求。基于上述问题, 国外很多学者对 IPA 分析法进行了修正, Deng 通过总结其他学者的研究成果, 对 IPA 分析法在统计学方法上做了一定的优化[4], 取得了良好的效果。

本文应用美团网站评论内容进行研究, 根据自主发布的评论探索消费者对于酒店特征的偏好以及关注程度。运用机器学习、情感分析等技术对数据进行处理, 并通过 IPA 模型对酒店的多维特征进行分析,

从而给酒店行业提供更加精准合理的建议。

## 2. 相关研究

### 2.1. 经济型酒店多维特征研究

酒店的特征分为很多类,例如酒店形象、价值等无形特征和酒店位置、网络等有形特征;床上用品、洗漱用品等实用性特征和娱乐设施、用餐服务等体验性特征。对于酒店特征的感知是由消费者在入住酒店过程所感受到的对于与酒店相关联的一切特征的重要性以及满意度的评价。Callan 和 Bowman [5]通过对酒店的 38 项特征进行调查发现,消费者认为重要的特征包括员工服务态度和服务效率、卫生环境、基础设施等,同时有很多被调查者表示体验性特征是相对重要的。Knutson [6]的调查中显示消费者满意度的重要组成部分包括:服务人员的态度、酒店的位置、环境的安全等有形和无形的属性特征。消费者对于酒店特征的重要性与满意度的感知是不确定的,有些特征可能会有很高的满意度但是对于消费者来说可能不是很重要,有的特征很重要但是它的存在和优化并不能很好的提高满意度,三因素理论[7],即必备因素、绩效因素和兴奋因素对以上特征进行解释。梳理和总结现有文献发现国内对于旅游市场的研究较多,而对于经济型酒店特征的研究较少,虽然有文献对国内的消费者在酒店的需求以及满意度进行调查,或者针对特定特征进行研究。高宝俊等[8]根据已有研究发现酒店位置特征是消费者进行决策时会特别关注的一个特征,单晓红等[9]对携程酒店刻画在线评论的用户画像时对涉及到的酒店的类型、房间大小等 4 个属性进行研究。总的来看,对于经济型酒店的所有特征的分析还缺乏系统的深入的研究,因此需要在国内酒店行业研究中,需要进一步对消费者对于经济型酒店属性的感知进行探讨。

### 2.2. 传统 IPA 以及 IPA 的修正

IPA 分析法是将所有酒店特征的满意度和重要性取均值作为各因子的组合评价,依据矩阵中的象限分布,对各因子进行分析从而找出需要急需改进的地方。IPA 分析法在国内外应用相对较多,巫景飞[10]研究国内典型的经济型酒店,发现消费者最为重视的依旧是酒店的选址问题。Kuo 等[11]探讨影响香港消费者选择酒店住宿的因素,运用 IPA 分析法对 26 项酒店服务属性进行研究,发现酒店的地理位置、住宿的环境、餐饮等方面迫切的需要进行改进和优化。

IPA 分析法的使用有两个必要的前提条件,首先,要求矩阵图中两坐标轴是相互独立的;其次,各测量因素的满意度与总体满意度之间的关系必须是线性相关且对称。现实情况下的数据并不能满足使用假设,许多学者对 IPA 分析法进行了修正和拓展。Deng 对以往研究进行总结,针对传统 IPA 中存在的问题进行了修正,提出以引申重要性代替自述重要性,即将各因素满意度评价取对数进行分析,之后将总体满意度与各维度因素的满意度进行偏相关分析,得出的偏相关系数作为引申重要性数值[4]。将各满意度进行对数取值可以增加对于变量的敏感性,并且用偏相关系数更能反映各特征与总满意度之间的真实情况。

### 2.3. OCRs 的情感分析

情感分析也称为意见挖掘或者评论抽取,是对非结构化的文本数据进行主客观分析,同时对提取的主观句进行情感极性分类,即将积极或者消极的情感倾向进行识别,从而获取对于某商品的褒贬意见,为做决策提供依据[12]。情感分析分为粗粒度情感分析和细粒度情感分析,粗粒度情感分析包括篇章级和句子级情感分析,细粒度情感分析是在产品的各个特征层面进行分析,Medhat [13]认为对产品 OCRs 进行细粒度分析时主要的步骤为:情感识别、特征提取、情感分类和情感极性识别。赵学峰[14]通过基于 DBSCAN 的文本聚类过程,依据在线评论和语料库对酒店的信誉维度进行挖掘,结果表明消费者比较关

注的属性包括: 硬件、服务、环境、饮食以及性价比。本文借鉴前人所采用的基于情感词典的方法, 通过情感极性和强度计算情感分数得到满意度与重要性得分, 从而进一步进行修正 IPA 的分析。

## 2.4. K-means 聚类

K-means 聚类是一种无监督的机器学习算法, 由于其原理简单、实现容易、效果比较好, 因此是比较常用的聚类方法。其通过计算对象与对象之间的距离, 判断彼此之间的相似度, 进而进行分类。聚类的主要目的是将数据集分为 K 类, 使得类内的数据距离最近, 类间聚类最大。

K-means 聚类的主要过程包括: 1) 确定聚类数目 K, 之后随机选择 K 个初始点最为聚类中心。2) 数据集中的每一个点与 K 个初始点计算距离, 将点分配给距离最小的聚类中心。3) 将数据集中所有点分好类别后, 重新计算聚类中心。4) 重复步骤(2)和(3), 直到聚类中心不在发生变化, 则认为已经达到最优聚类。

## 3. 研究设计

本文的总体研究框架主要包括文本预处理、Word2Vec 词向量的转换、K-means 聚类、细粒度情感分析以及偏相关分析。

### 3.1. 数据来源

以青岛市为例, 选择我国经济型酒店的代表品牌如家、七天、汉庭等 21 家酒店的评论进行研究。通过编写网页爬虫的 Python 程序共获取 35,398 条 OCRs, 获取的内容包括: 文本评论、产品得分等。对所有数据进行清洗, 剔除无效和低于两个字的评论数据, 最终得到 30,877 条评论数据。

### 3.2. 基于 K-means 的酒店特征的聚类

1. 首先对获取的评论数据进行分词, 形成 Word2Vector 语料, 然后使用 gensim 库中的 Word2Vec 模型对分词后的语料进行训练(window = 10, vector = 300), 将词映射到 K 维向量空间形成相应的词向量。

2. 使用 K-means 聚类算法对词典中的名词进行聚类。K-means 选择最优聚类数的方法两种: 手肘法和轮廓法。手肘法的核心思想是随着聚类数的增大, 样本的划分会更加精确, 那么误差平方和就会逐渐变小。当到达真实聚类数时, 再增加聚类个数所得到的聚合程度回报会迅速变小, 所以误差平方和的下降幅度会骤减, 然后随着聚类数的增大而趋于平缓, 即肘部对应的就是数据的真实聚类数。轮廓系数法的核心指标是轮廓系数(Silhouette Coefficient), 平均轮廓系数越大, 聚类效果越好, 所以平均轮廓系数最大的值便是最佳聚类数。

由于在轮廓系数法确定的值不一定是最优, 有时需要借助 SSE 进行判断, 因此本文选择手肘法来确定最优聚类数。将聚类数设置为 1 到 15 进行反复预测, 最终结果显示聚类数为 13 时最优, 因此本文分为 13 类特征进行研究。

K-means 特征聚类的过程为: 1) 对评论进行分词等处理后, 选出评论中的名词作为特征聚类的候选词, 并保留不低于出现频次为 11 的名词形成名词词典。2) 词典中 Word2Vec 训练后的转化为词向量作为聚类的语料库。3) 调用 sklearn 库中的 K-means 算法对词向量进行聚类, 根据最优聚类数, 输入聚类数 13, 依据距离公式进行判断, 同时借鉴以往相关研究, 借助人工进行分类, 最终的酒店特征分类: 餐饮、床上用品、服务、隔音、环境、价格、基础设施、客房设施、卫生、品牌、网络、卫浴、位置。

### 3.3. 基于词典的情感分析

通常消费者在发表评论时会针对特定特征发表意见, 因此本文进行特征级别的细粒度情感分析。为

为了使分类结果更为准确, 本文加入了常见的与酒店相关的词汇, 以构成酒店进行情感分析的专用词典。具体构建原则如下: 将 HowNet2007 版[15]词典作为基础词汇建立词典, 同时将所有有效评论进行分词等处理, 选取形容词形成酒店的积极和消极词典。将基础词典与酒店词典进行合并, 形成本文酒店的专用情感词典。

除了名词和形容词, 消费者在发表评论时的情感强度有所不同, 因此本文依据 HowNet2007 版, 确定了情感极性的量化标准, 如表 1 所示。特征情感极性分数 = 程度副词极性值 \* 情感词的极性值, 如果没有程度副词极性值则为 1 或者-1。

**Table 1. Sentiment polarity quantification criteria**  
**表 1. 情感极性量化标准**

程度副词	极性值
非常, 绝对……	2
特别, 实在……	1.5
愈, 愈加……	1
稍许, 有点……	0.5
无程度副词的正向情感词	1
无程度副词的负向情感词	-1

### 3.4. 修正 IPA

通过以上实验得到 13 类酒店特征, 情感分数即代表各特征的满意度得分, 对提取出的 13 类特征进行相关性检验, 结果显示 78 个相关系数中仅有 9 个存在不相关, 包括: 餐饮 - 基础设施、餐饮 - 客房设施、餐饮 - 网络、隔音 - 网络、环境 - 网络、品牌 - 卫浴、品牌 - 位置和位置 - 网络。

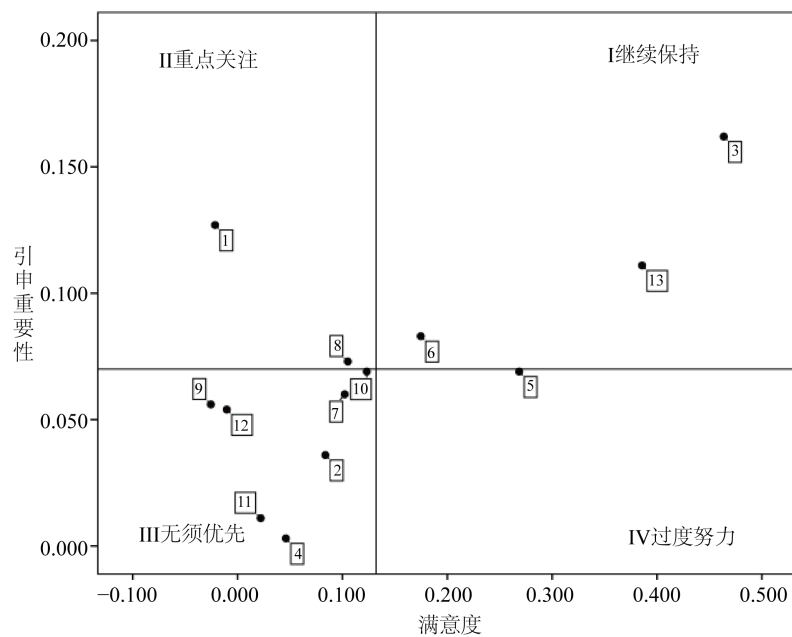
为了解决各特征之间的相互影响, 根据 Deng 的转换方法对数据进行转换。总满意度得分是由爬虫获取的消费者对于酒店的星级评分, 分值范围为 1 到 5 的整数, 其中 1 表示非常不满意, 5 表示非常满意, 基于情感词典的特征满意度与总满意度得分进行偏相关分析, 酒店特征的引申重要性与满意度分值如表 2。

以引申重要性为纵坐标, 满意度为横坐标, 计算所有特征的引申重要性的均值和满意度的均值, 两均值作为中心点将矩阵分为四个象限。将 13 个特征依据表 2 各特征的分值绘制在矩阵象限图中, 结果如图 1 所示。

根据 IPA 分析图发现, 第一象限即满意度和重要性都较高, 包括三个特征: 服务(3), 位置(13), 价格(6)。其中, 服务的满意度与重要性最高, 反映了顾客对于酒店服务人员及服务态度的重视, 随着社会的发展和文明素质的提高, 顾客更加重视对于无形特征的要求, 更加强调对于消费事物的体验性, 而酒店作为典型的体验性产品更应该注重服务的提升。其次为酒店位置, 酒店位置一直以来都被作为重要的参考特征, 以往对于酒店特征的研究中也会单独拿出“酒店位置”进行研究, 因此酒店的选址至关重要, 顾客在进行消费时更加倾向于选择交通方便、景区周围或者周围设施比较齐全的酒店。最后为价格, 经济型酒店在近几年能够快速发展得益于其价格定位, 相比于其他类型的酒店, 例如三星、四星及高级酒店, 经济型酒店在价格方面的优势比较明显, 同时由于大部分消费者的经济状况都是较为普通, 对于价格问题还是比较敏感, 因此经济型酒店应该继续发挥其在价格方面的优势, 在把握好已经占有的市场份额的基础上努力优化以图增加占有率。由于第一象限的特征代表的是企业的竞争优势, 因此采取的策略为“继续保持”。

**Table 2.** Implicitly derived importance and satisfaction score  
**表 2.** 引申重要性和满意度分值

代码	特征	引申重要性	满意度
1	餐饮	0.127	-0.022
2	床上用品	0.036	0.084
3	服务	0.162	0.464
4	隔音	0.003	0.046
5	环境	0.069	0.269
6	价格	0.083	0.175
7	基础设施	0.060	0.102
8	客房设施	0.073	0.105
9	卫生	0.056	-0.026
10	品牌	0.069	0.123
11	网络	0.011	0.022
12	卫浴	0.054	-0.010
13	位置	0.111	0.386



**Figure 1.** IPA analysis of implicitly derived importance and satisfaction  
**图 1.** 引申重要性与满意度 IPA 分析图

第二象限即满意度低重要性高, 包括两个特征: 餐饮(1), 客房设施(8)。饮食是人们出行所必不可少的, 而经济型酒店最初的服务模式为住宿和早餐, 随着现代人们对于健康生活的追求, 早餐成为一天生活的标配, 因此对于酒店提供免费饮食顾客日益关注。由于第二象限的特征代表企业迫切需要改进的方面, 如果忽视将对企业的发展带来严重的威胁, 而人们对于酒店饮食的需求并不苛刻, 并没有要求饮食种类的多样性与丰富性, 因此对于酒店管理人员来说提供相对简单的饮食是必须改进和提高的。对于客

房设施, 例如酒店的空调、水壶等硬件设施和房间大小等基本房间类型, 是顾客进行消费体验的基础, 也是酒店最基本的服务, 直接影响消费者对于酒店的总体感知和满意度, 因此, 此象限采取的策略为“重点关注”。

第三象限即满意度和重要性都比较低, 包括七个特征: 品牌(10), 基础设施(7), 卫生(9), 卫浴(12), 床上用品(2), 网络(11), 隔音(4)。基础设施例如停车场、直梯等, 卫浴例如沐浴设施等, 床上用品例如被子、枕头等, 以及房间的隔音效果等一系列的酒店特征是酒店的基本配备, 因此并不会引起多大关注, 顾客对于这些必备特征总是习以为常, 因此重要性偏低同时对于总体满意度影响也较低。对于品牌特征, 本文发现一定量的 OCRs 中会提及酒店名称, 例如“汉庭”、“如家”等, 因此顾客在进行消费选择时会一定程度上受到品牌的影响, 但是在考虑品牌因素之前大多数顾客会优先考虑别的因素。而卫生特征也是一个重要的基本条件, 在其他基本特征具备的条件下对于所住宿酒店的卫生应该达到一定的标准。网络作为现代化城市必备的因素, 尤其是旅游类城市, 已经可以达到较高的覆盖率, 不再是顾客所额外关注的特征。因此对于象限三的特征采取的策略为“无须优先”, 在酒店资源充分的条件下也可进行改进。

第四象限即满意度高重要性低, 包含一个特征: 环境(5), 在交通比较便利的地方或者景区周围可能环境有一定的嘈杂, 而在相对偏远交通不太便捷的地方环境相对安静, 由于位置是顾客在消费时十分关注的特征, 而位置的便利与环境的安静在一定程度上相矛盾, 因此一般顾客会优先考虑酒店地址而忽略环境安静与否, 因此对于第四象限的特征本文认为是“过度关注”, 在酒店资源有限的情况下优先满足其他象限的特征, 将过度的资源转移到其他方面来增加酒店的总体满意度。

#### 4. 结论

经济型酒店作为典型的体验性产品, 能够获取顾客在各类社交媒体发表的 OCRs, 因此从 ORCs 中自动获取顾客关注的特征尤为重要。本文的重点是实证应用了 Bi [16]将 OCRs 引入 IPA 的方法, OCRs 数据更加真实更易于获取, 基于大量数据的前提下可以得到更加客观精准的结果。因此, OCRs 与 IPA 模型结合的方法可以更加真实有效的了解目前青岛市经济型酒店的发展状况, 在更具有信任的结论基础上提出企业的改进策略尽量满足顾客需求以得到更高的满意度, 为青岛市构建国家中心城市提供酒店方面改进的参考。

本文研究也存在一定的局限。第一, 选择的酒店样本只针对青岛市经济型, 不能代表青岛市酒店行业的整体发展状况, 后续会针对不同类型的酒店进行研究分析。第二, 对于本文所借鉴的修正 IPA, 虽然已经包括了关于不对称的三因素理论, 但是并不能精确地体现各特征的满意度与重要性之间的不对称性, 因此后续将在对 IPA 的修正方法方面进行更深入的研究。

#### 基金项目

本文受山东省社科规划项目(18CHLJ22)和国家民委民族研究项目(2018-GMB-022)支持。

#### 参考文献

- [1] Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., *et al.* (2005) The Firm's Management of Social Interactions. *Marketing Letters*, **16**, 415-428. <https://doi.org/10.1007/s11002-005-5902-4>
- [2] Jerdee, T.H. and Rosen, B. (1974) Effects of Opportunity to Communicate and Visibility of Individual Decisions on Behavior in the Common Interest. *Journal of Applied Psychology*, **59**, 712-716. <https://doi.org/10.1037/h0037450>
- [3] Martilla, J.A. and James, J.C. (1977) Importance-Performance Analysis. *Journal of Marketing*, **41**, 77-79. <https://doi.org/10.1177/002224297704100112>

- 
- [4] Deng, W.J. (2007) Using a Revised Importance-Performance Analysis Approach: The Case of Taiwanese Hot Springs Tourism. *Tourism Management*, **28**, 1274-1284. <https://doi.org/10.1016/j.tourman.2006.07.010>
- [5] Callan, R.J. and Bowman, L. (2000) Selecting a Hotel and Determining Salient Quality Attributes: A Preliminary Study of Mature British Travelers. *International Journal of Tourism Research*, **2**, 97-118. [https://doi.org/10.1002/\(SICI\)1522-1970\(200003/04\)2:2<97::AID-JTR190>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1522-1970(200003/04)2:2<97::AID-JTR190>3.0.CO;2-1)
- [6] Knutson, B.J. (1988) Frequent Travelers: Making Them Happy and Bringing Them Back. *Cornell Hotel and Restaurant Administration Quarterly*, **29**, 82-87. <https://doi.org/10.1177/001088048802900121>
- [7] Anderson, E.W. and Mittal, V. (2000) Strengthening the Satisfaction-Profit Chain. *Journal of Service Research*, **3**, 107-120. <https://doi.org/10.1177/109467050032001>
- [8] 高宝俊, 孙含琳, 王寒凝. 在线评论对酒店订满率的影响研究[J]. 旅游学刊, 2016, 31(4): 109-117.
- [9] 单晓红, 张晓月, 刘晓燕. 基于在线评论的用户画像研究——以携程酒店为例[J]. 情报理论与实践, 2018, 41(4): 99-149.
- [10] 巫景飞. 我国经济型酒店服务质量的 IPA 分析——以如家快捷与锦江之星为调查对象[J]. 华东经济管理, 2007, 21(11): 95-98.
- [11] Kuo, C.M., Chen, L.C. and Lin, S.Y. (2010) Exploring the Relationship between Hotel-Based Service Attribute Importance and Customer Satisfaction at International Conferences in Taiwan. *Journal of Convention & Event Tourism*, **11**, 293-313. <https://doi.org/10.1080/15470148.2010.533073>
- [12] 唐晓波, 刘广超. 细粒度情感分析研究综述[J]. 图书情报工作, 2017, 61(5): 132-140.
- [13] Medhat, W., Hassan, A. and Korashy, H. (2014) Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, **5**, 1093-1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- [14] 赵学锋, 汤庆, 张睿, 等. 基于客户评论和语料库的在线酒店信誉维度挖掘[J]. 图书情报工作, 2012, 56(12): 124-129.
- [15] 董振东. 知网 HowNet [CP/DK]. [http://www.keenage.com/html/e\\_index.html](http://www.keenage.com/html/e_index.html), 2014.
- [16] Bi, J.W., Liu, Y., Fan, Z.P. and Zhang, J. (2019) Wisdom of Crowds: Conducting Importance-Performance Analysis (IPA) Through Online Reviews. *Tourism Management*, **70**, 460-478. <https://doi.org/10.1016/j.tourman.2018.09.010>