

# 中西方游客对民宿满意度的情感特征研究

**摘要：** 本文旨在探讨中西方游客对香港地区民宿满意度的情感特征在哪些方面有着显著差异。首先，采用NLTK工具包对中西方游客的整体满意度进行度量；其次，利用Fightin's Words算法模型得到他们对于民宿属性的不同程度强调；再次，基于神经网络模型对民宿属性的感知进行细粒度情感分析。研究结果表明：(1) 中国游客对民宿整体满意度高于西方游客，且在撰写评论时更加客观；(2) 中国游客更加看重民宿房东的友好程度、房间通风情况、房源附近是否有便利店以及房源价格等方面，而西方游客更多强调与民宿房东的回复速度、房间大小、房源附近是否有酒吧和夜市等方面；(3) 相比于西方游客，中国游客对于房东的满意度较高，但对民宿整体内外部环境和价格方面更为挑剔。

**关键词：** 文化差异；情感分析；在线评论；强调差异；Fightin's Words算法模型；感知差异；神经网络模型

**中图分类号：** C93

## 引言

随着消费水平以及交通便利程度提高，全球旅游业呈现出一片欣欣向荣的景象。2018年全球旅游总人次达到121亿人次，预计2019增速将达百分之五点五，2018全球旅游总收入达5.34万亿美元，预计2019年增速将达到百分之三点七。<sup>[1]</sup>旅游对目的地国和旅游者母国都产生了巨大的经济效益，尤其是发展中国家大多致力于发展旅游业，期望从预期的经济改善中获益，<sup>[2]</sup>而旅游目的地经济发展和旅游业提升可以通过住宿业来实现。<sup>[3]</sup>在当前国际旅游热潮的引领下，如何更精确地解析全球各地游客对住宿的评价、偏好和兴趣等特征，日益成为学术界、业界和政府关注的焦点之一。

共享经济在过去几年得到了普及，民宿已成为了除传统酒店业以外游客旅行住宿的另一主要选择。相比于传统酒店，民宿一方面具有价格优势，有研究表明民宿网站所提供房源的平均价格比绝大多数城市的酒店低；<sup>[4]</sup>另一方面，民宿可以给游客带来不同于标准化酒店的新体验，满足游客选择住宿环境的个性化需求，同时也能使游客更深刻地感受当地的文化生活。上述两方面原因使得民宿备受游客青睐，对传统酒店业产生了巨大冲击。<sup>[5]</sup>随着人们生活水平的提高、购买力的增强和房产投资的兴起，闲置房屋数量也开始上升，将其作为租赁资源的趋势为民宿在旅游市场广受游客欢迎奠定了基础。自2015年后，中央政府、各部委及部分地方政府先后发布涉及民宿行业的指导意见与相关条例（如《国务院办公厅关于加快发展生活性服务业促进消费结构升级的指导意见》、《北京市旅游条例》等）以鼓励民宿行业的发展。

游客住宿体验是其整体旅行体验的重要组成部分，<sup>[6]</sup>其在鼓励游客回访、给予正面评价、向他人推荐等方面起着根本性的作用，进而影响了游客的行为忠诚。<sup>[7-9]</sup>游客满意度及其前导变量的研究中面临一个重大挑战，那就是游客的多样性，需要通过细分市场来加以考虑。<sup>[10]</sup>由于服务具有无形性、异质性以及与顾客同步性等特点，<sup>[11]</sup>因此细分市场和理解游客差异在民宿行业中是至关重要的。以往对传统酒店研究的发现，游客在期望、需求和服务等方面存在不同感知。<sup>[12]</sup>有学者将游客的差异归因为一个主要的变

量,即文化差异,<sup>[13,14]</sup>它包含价值观、信仰和规范等方面,共同区分了特定文化群体。<sup>[15]</sup>而来自不同地方的游客从小就被上述方面以微妙的方式根植于其个体之中,他们依赖于自身的文化背景,<sup>[16]</sup>无论是在国内还是出国旅行时都表现得很明显。<sup>[17]</sup>游客自身的文化背景在认知、信息处理和购买决策中扮演着重要的角色。<sup>[18]</sup>因此,如何更为全面地了解来自不同文化背景游客的整体住宿偏好和情感特征,是当前国际旅游市场研究亟须解决的重要议题之一。

民宿业面对的是一个全球性市场,主要有两种代表性观点:一方面,全球所有游客都拥有一套相近的价值观,进而产生大体相同的旅游偏好和消费行为;<sup>[19]</sup>另一方面,与前述观点相反,全球旅行者由于其所处的社会背景、受教育的方式、文化价值观等并不完全一致,因而他们的旅游偏好和消费习惯也就各有千秋。<sup>[20]</sup>然而尽管研究游客的多样性具有公认的重要性,目前的研究尚不足以判断并解释不同文化背景游客的偏好和情感特征,同时关于是否应该考虑或是在多大程度上考虑不同文化之间差异的争论仍然没有得到解决。<sup>[21-24]</sup>因此本研究试图通过探讨文化差异对民宿在线评论的影响来加入到这场争论中。尽管目前已有学者在旅游业和酒店领域对在线评论有所研究,但针对共享经济这一新商业模式背景下对在线评论的研究还尚不多见,同时已有研究表明共享经济平台与传统电商平台存在诸多不同。<sup>[5]</sup>本研究试图利用文本挖掘对游客的情感展开分析,并通过对大数据的处理尝试构建民宿情感评价的关键词库,以不同文化背景的游客为例,以在线评论为素材,探究游客对民宿的情感倾向,揭示他们对各个属性的情感特征。具体来讲,主要关注以下三个问题:(1)来自不同文化背景的游客对民宿整体满意度上有什么差异?(2)他们分别强调了民宿哪些不同的属性?(3)他们对民宿每一个特定属性又有怎样的感知?本研究从网络大数据中解读出不同文化背景的游客对民宿的情感差异,这是一种基于旅游大数据探究游客情感倾向的尝试,有助于更好地指导民宿针对不同文化背景游客改进其营销策略和方式,具有较高的应用价值。

## 一、理论基础和研究假设

### 1. 在线评论

网络口碑已经成为消费者决策的重要影响因素。<sup>[25]</sup>目前国内外已有众多文献发现口碑的重要性,在电子商务中口碑指的就是消费者的在线评论,<sup>[5]</sup>它是消费者发布的其关于此次消费的体验,对其潜在消费者的购买决策有重要影响,因此也受到越来越多学者和相关从业人员的关注。在线评论是没有任何标准格式的非正式的帖子,与专家意见和面对面的口碑传播(WOM)相比,它们并没有明确的权威或可信度,然而其确实提供了更广泛的产品覆盖面和更多样化的意见。<sup>[26]</sup>比起数值评分,在线评论更能体现真实的情感,从而被潜在消费者所参考和信赖,因此在线评论变得越来越重要,大部分消费者相信在线评论是最容易获得、最值得信赖和使用的信息来源之一。<sup>[14]</sup>民宿的在线评论主要是指游客自由地写下有关住宿的体验,由于游客很难在住宿之前很难感知到民宿的服务,在线评论作为吸引潜在消费者的工具,此时就显得尤为重要。<sup>[8]</sup>与具体的产品相比,由于民宿提供的服务是无形的,那些由营销人员提供的信息如公司网站、广告等常常被认为是具有偏见的,这使得游客做购买决策的不确定性和风险更大。<sup>[24]</sup>因此,在线评论是在对一个相对未知的服务实体做出反应之前使用的重要信息源。所以通过在线评论可以更好地了解游客,这对民宿从业人员快速回应游客需求以及策划营销方案具有重要的意义。目前国内外对民宿在线评论的相关研究较少,这一方面可能是数据的来源不足,对民宿在线评论的探究还处于起步阶段;另一方面可能是对共享经济中消费者评论的影响机制和测量方法尚未有一个权威的标准和规范。<sup>[5]</sup>而本研究对这两方面做了努力和尝试,从海量的在线评论中挖掘出有用的信息,并得到了一些有趣的结果。与其他信息线索相比,在线评论有几个具有吸引力的优势:第一,它能直接捕捉到游客对所提供的服务的看法;第二,它能提供接触大量游客的途径,避免研究者花费大量的努力和成本通过调查来收集;第三,它能允许游客提供定量和定性信息,对特定方面进行评价,并描述对民宿的总体体验。因此,本研究旨在从在线评论的文本内容中提取见解。

### 2. 文化差异

越来越多游客在做出最终决定之前都会参考网上发表的意见,而文化等内在因素可能会影响个体形成期望和评价的标准和权重。<sup>[27]</sup>文化是区分人类群体的规范、价值观和制度特征的同质性,<sup>[28]</sup>了解不同

文化背景下游客行为的差异非常重要。Hofstede 认为文化差异只有通过比较才存在,并提出衡量跨文化差异的五个维度:风险维度、关系维度、权力距维度、性别维度以及时间维度。<sup>[29]</sup>不同的文化背景会导致人们不同认知过程和行为。<sup>[30]</sup>文化维度强调了影响游客欲望和体验的跨文化差异。<sup>[31]</sup>然而,口碑也是文化背景在国际营销研究中的一个突出建构,<sup>[32]</sup>有关在线评论文献的研究表明,在线文化能够代表消费者文化的特定领域,<sup>[33]</sup>这种效应在不同地区或国家之间也可能存在显著差异。

在过去对传统酒店的研究中,Schuckert 等学者将游客文化差异的相关研究进行了理论总结,他们将其分为四个方面:对酒店属性或服务方面的不同感知程度即感知差异(Perception Difference),对酒店属性和服务质量的不同期望程度即期望差异(Expectation Difference),对每个酒店属性或服务方面的不同强调程度即强调差异(Emphasis Difference),以及在投诉中表现出的不同行为即投诉差异(Complaint Difference)。<sup>[34]</sup>本研究参照 Francesco 等人的做法从民宿的属性感知差异和属性强调差异这两个维度出发,<sup>[12]</sup>对民宿在线评论中不同文化背景的游客进行探索以了解哪些属性对游客最重要,并发现哪些属性会影响其决策,以此来接近潜在顾客。<sup>[35]</sup>本研究将从以下三个方面对在线评论中所体现的文化差异进行实证研究。

### 2.1 对民宿整体满意度差异

过去研究使用了 Hofstede 框架中的个人主义与集体主义维度,发现了两种维度下的群体对待产品或服务有不同的评价标准,同时研究表明个人主义国家的消费者比集体主义国家的消费者更愿意在网上分享他们对产品或服务的看法和体验。<sup>[29]</sup>Makoto 等学者的研究得出了相似的结论:西方文化背景下的群体可能更直接地表达自我的感受;另有学者研究表明西方消费者在线评论中使用词汇的范围也要远大于东方消费者,<sup>[14]</sup>Lai 等人的研究同样显示美国消费者在线评论中表达意见的频率高于中国消费者。<sup>[36]</sup>显著的文化差异包括对服务的期望程度和对优质或劣质服务的反应程度。个人主义导向的西方文化比集体主义导向的东方文化更倾向于期望更高的服务质量,<sup>[37]</sup>如果服务不那么理想的话,西方国家比东方国家更有可能抱怨。在 Laroche 等学者的研究中发现相比与西方客人,对劣质服务的宽容程度高于西方客户;<sup>[38]</sup>Makoto 等学者的研究表明西方消费者无论在正面还是负面评论的情绪都比以日本为代表的东方消费者表达得更直接。<sup>[14]</sup>

本研究中选择中国游客和西方游客作为研究对象,基于以上分析,提出假设 1:

H1: 在民宿网站的在线评论中,中国游客和西方游客表现出的整体满意度存在差异,且中国游客的整体满意度要高于西方游客。

### 2.2 对民宿属性的强调差异

在线评论涉及不同属性,<sup>[39]</sup>浏览在线评论有助于帮助游客了解有关民宿的更多信息。Schuckert 等学者强调了游客对各种酒店属性的强调程度存在差异。<sup>[34]</sup>在以往对酒店管理的研究中,有一个研究视角是了解不同的文化背景是否决定了游客对酒店不同属性的强调程度存在差异。虽然已有学者对来自不同文化背景游客的强调差异进行了研究,<sup>[12]</sup>但学者们得到的结论往往是相反的。早期一些学者发现来自东方和西方的游客对各种酒店属性强调程度差异不大,认为没有侧重点上的差异,<sup>[40]</sup>而后来有学者对此提出了相反观点,如 Poon 等学者认为不同文化的游客之间存在显著的强调差异,<sup>[41]</sup>具体而言,他们认为亚洲游客更注重价格和员工的热情好客等因素,而西方游客更注重酒店的外观;Makoto 在对日本和西方消费者的研究中也发现,前者更强调价格环境等因素,而后者更强调服务质量;<sup>[14]</sup>Tsang 的研究指出亚洲游客强调酒店员工的能力,而西方游客更强调员工的态度。<sup>[42]</sup>但相反的是,Kuo 等人的研究中发现美国游客强调员工解决问题的能力,而中国游客则更看重礼仪;<sup>[43]</sup>也有学者发现,中国游客更强调酒店的卫生程度、舒适程度,以及入住和退房的便利性,而西方游客则看重酒店的运动设施是否齐全、住宿是否安静、总体满意度和员工能力以及餐厅和房间升级的便利性。<sup>[12]</sup>基于以上分析,本研究提出假设 2:

H2: 中国游客和西方游客对于民宿属性强调程度存在差异。

### 2.3 对民宿属性的感知差异

文化差异反映在对特定事物的偏好超过其他事物、对特定过程的持续偏好超过其他过程,以及对环境线索的选择性注意以及对其作出的不同反应。<sup>[31]</sup>当评价某件事物时,人们对同一产品或服务的看法可能不同。Weiermair 认为不同文化背景的游客对酒店方面存在独特感知,因此他们对所接受的相同服务质量可能会有显著不同的感知;<sup>[44]</sup>Hsieh 和 Tsai 也在研究中表明文化差异可以产生不同的服务质量感知;



[45]多位学者认为与西方游客相比,亚洲游客面对同样的服务表现,对服务的评价却不那么好。具体来讲,有的学者从集体主义和个人主义的维度出发认为亚洲文化是集体主义的,与来自个人主义社会的西方游客相比,亚洲游客更可能期望得到更多的礼貌和体谅;而有的学者从权力距离的维度出发,认为由于亚洲文化与西方文化之间权力差距较大,亚洲游客很可能认为自己比服务提供者更强大; [42, 43, 46, 47]也有学者在研究中得出了相反结论,如 Zhang 等学者强调个人主义社会的个人,如美国游客和欧洲游客,对于服务方面有更高的期望,因此,与亚洲游客相比,对于相同的服务他们通常不太满意,在服务评估方面更加严格。[48]先前对感知差异的研究大部分集中于对跨文化群体的服务质量感知上,而在最近的研究中发现西方消费者面对同样的产品、服务以及周边的环境都表现出与亚洲消费者不同的感知, [14]Galati 等学者在研究中证明了中国游客和西方游客对于酒店的属性感知上存在着差异。[12]根据目前的研究推测,一个游客的文化背景可能在塑造其对民宿属性的感知上发挥重要作用,因此提出假设三:

H3: 中国游客和西方游客对于民宿属性感知程度存在差异。

## 二、研究方法

### 1. 数据来源

为了对民宿评论进行抽样,本研究选择 Airbnb 网站上的香港地区的所有房源作为本研究的数据源,原因有以下几个。首先,香港是中国最具特色的城市之一,香港旅游发展局公布的数据显示,2018 年访港旅客约 6515 万人次,较 2017 年增长 11.4%,创历史新高。早在 2016 年,香港就超越了伦敦成为游客最多的国际旅游城市, [2]有大量的中国本土游客和西方游客,是一个多文化的旅游城市,这为本文的跨文化研究提供了基础。其次, Airbnb 是最受欢迎也是最大的民宿网站之一, Airbnb 已经在 190 个国家获得了超过 200 万个房源。 [2]同时,在 Airbnb 网站(www.airbnb.com)上提供了中文、英文等语言评论,相同并采用相同的评论格式和功能,这允许本文比较两种语言的评论。因此,本研究通过 Python 爬取了截至 2019 年 6 月 12 日 Airbnb 网站上香港地区共 10970 个房源的所有 240484 条评论,在筛选了除中英文语言以外的其他语言以及其无实际意义的评论以后,共得到 133419 篇英文有效评论和 62285 篇中文有效评论。

本研究参照 Markus 等学者用语言对样本进行分类的相关研究,将在线评论分为两组:英语评论和中文评论。 [34]有研究表明英语是世界上使用最广泛的语言,因此说英语的人可能有相似的文化和习俗,他们与不说英语的人非常不同,会说英文的游客中有较高比例的西方游客,如美国人和欧洲人,而不会说英语的游客中有较高比例的亚洲游客,如中国人。 [34]由于地理边界不一定与语言边界完全一致,本研究发现这种方法非常重要。游客的在线评论可能会分享他们最初在本国文化中获得的规范,而这些规范基于其民族文化,由国家、种族、宗教和其他边界条件所定义。 [49]过去使用同一种语言的人尽管来自不同的国家,但他们可能拥有一些相同的文化,例如,一些日本人住在伦敦,如果他们倾向于使用英语,他们可能更受西方文化的影响,而不是东方文化。 [34]因此,语言的使用反映了其背后的文化。

### 2. 分析方法

本研究主要采用的研究方法是情感分析。传统研究主要通过满意度指标来获取游客对住宿的整体评价,但是实际上仅关注满意与否并不能准确地刻画游客对住宿情感的正负面评价,以及影响这些评价的要素。近几年来,运用情感分析方法对在线评论进行研究成为了一种新趋势。情感是一种由情感和情感分析引发的态度、思想或判断,也称为意见挖掘,主要研究人们对某些事物的情感。 [50]情感分析是利用自然语言处理和数据挖掘算法等对文本内容进行情感的判断,从而把握内容中的观点和态度的计算研究。 [51]传统问卷虽然可以预设游客满意度评价指标和影响因素并进行测度,但是这种做法受限于研究者的知识水平,容易出现偏见和错误的预设。 [52]与传统的市场研究方法(如问卷调查、访谈等)相比,情感分析的优势在于其成本效率和时间效率更高。 [2]同时,情感分析能提取消费者的实时意见和情绪,以避免造成回忆偏差。 [53]

### 3. 分析步骤及过程

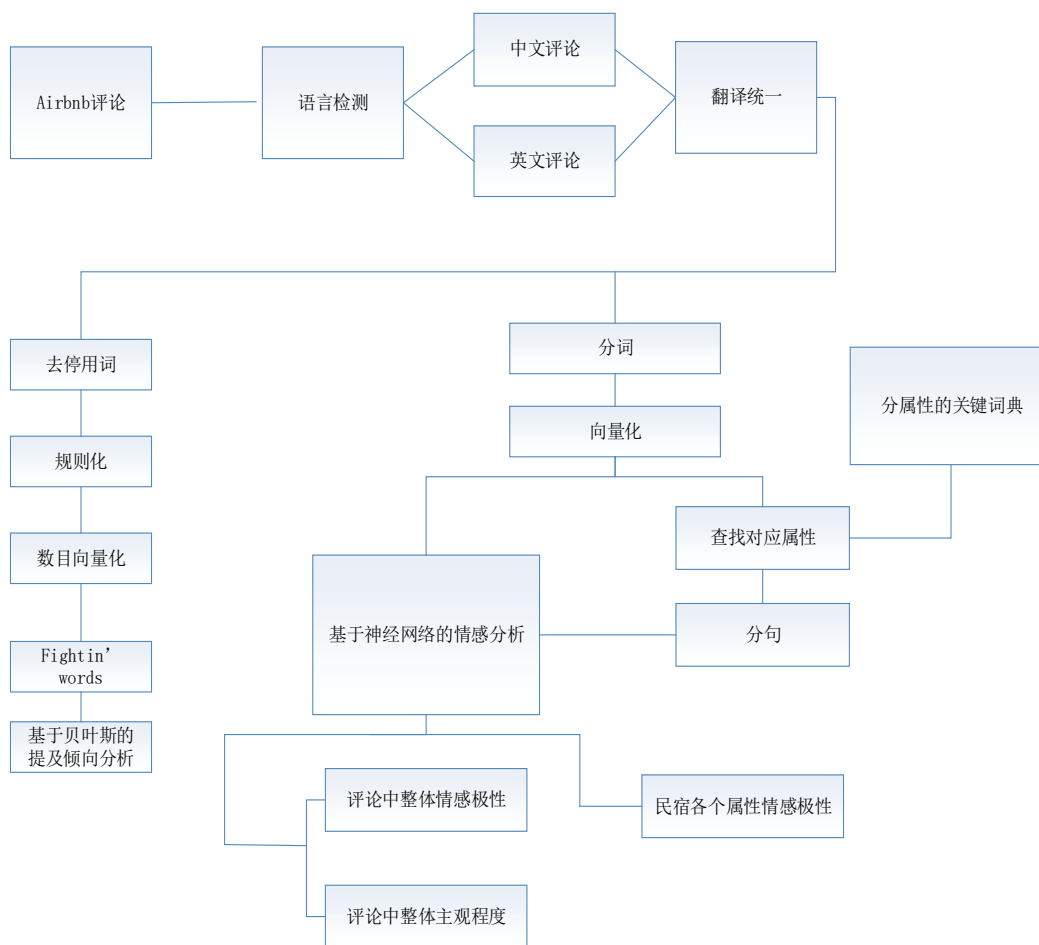


图 1 分析步骤流程图

本研究将所爬取的中文评论通过谷歌进行翻译，以便对其与英文评论统一进行分析，其中主要的分析步骤分为三步：首先对 Airbnb 网站上的中英文评论做整体的情感分析，其次挖掘出中英文评论中对民宿哪些属性有不同程度的强调，最后再在此基础上，对民宿的每一个属性进行单独的情感分析，得到中西方游客的感知差异。分析步骤流程如图所示。

### 3.1 整体满意度的差异分析

本研究在进入细粒度分属性进行情感分析之前，先对所爬取的近 20 万篇中英文评论的整体满意度差异做一个全面化的度量。对于中国游客和西方游客的评论，本研究先分析所有评论的情感极性，再从统计上得到中文评论和英文评论的整体特征并进行比较。通过这种整体的比较方式，可以了解到不同文化群体对民宿整体满意度的差异。同时，为了进一步发现两者之间的差异，本节对中英文评论整体的主观性也进行了一个度量，以期得到一个有趣的结果。

具体来讲，本研究采用 NLTK 工具包（Natural Language Toolkit 自然语言处理工具包）基于模型对所有评论情感极性进行度量，之后对评论主观性（Subjectivity）进行分析。<sup>[54]</sup>模型经过大量数据训练的神经网络，可以将文本转化为对应的情感极性分数与主观性分数。情感极性分数范围在-1 到 1 之间，数值越负表示情感越消极，数值越正表示情感越积极，即越满意。如“这是我所住过的最差的民宿”就是典型的表现消极情感的评论。若评论中性，如仅在描述事物特性，则情感极性为零。主观性分数范围在 0 到 1 之间，越大表示文本越主观。如“在我看来”、“我认为”都表示文本从主观出发进行判断。研究通过模型得到情感极性与主观性分数，并计算中文所有评论、英文所有评论在这两项的平均值、极值以及方差，进行比较分析。这里的平均值表示所有中英文评论中情感积极程度和评论主观程度的平均水平，极值表示所有中英文评论中情感积极程度和评论主观程度的最大最小水平，方差值则表示中西方游客在撰写评论时各自表达差异的大小。

### 3.2 对民宿属性强调差异的分析

当面对大量文字信息的时候,如何提取总结其中重要的文字或词组,以求通过它们来划分不同来源的文本是一个重要的问题。本文希望通过评论文字找到能够凸显出中国与西方游客在选择民宿时关注属性的差异,即强调差异,而这种强调差异可以通过体现该方面的文字受提及的程度来展现。如当厨房相关的事物经常在西方游客评论中出现而很少出现在中国游客评论中,可以说西方游客对民宿的厨房这个属性更加关注。这种基于文字频率的分析常常出现在对于强调差异的分析中,比较常见的方法有:直接的频率对比、<sup>[14]</sup>词袋对比、<sup>[55]</sup>TF-IDF 对比。<sup>[56]</sup>直接的频率对比包括频率差的对比  $f_1-f_2$ , 频率比例的对比  $f_1/f_2$ 。然而这样的频率对比方式有一些缺陷:有的词组可能在中国游客评论中被提到几次,而在西方游客评论中被提及频数为 0,这样会过度加重这一词组的权重;同样,有的词组在两种评论中都非常罕见,这样的词组对区分两类文本并无显著性,因此,本研究需要为有意义词组设置一个门槛,即只考虑  $f$  大于某个值的词组。再者,本文需要纳入考量的是在样本中有显著性的词组有可能并不在整体上有显著性,故需要对频率进行正则化,加入一些先验的概率分布。综合以上问题,本研究引入 Monroe 等人提出的 Fightin's Words 算法模型。<sup>[57]</sup>该模型假设各词分布符合先验分布和多项分布,给出在  $k$  种文本类型中每个词分别的分数,分数越高表示该词越能代表某一种类型。该分数表示在多大程度上该词代表了中国游客或西方游客的评论,在本实验设置下,正分数越正代表该词相对而言更能表示是中国游客的评论,负分数越负代表该词相对而言更能表示这是西方游客的评论。

### 3.3 对民宿属性感知差异的分析

在得到了中西方游客对民宿不同属性的强调以后,本节关注的是对于民宿的同一属性中西方游客的感知差异。在传统的情感分析任务中,研究者往往通过软件对整篇文章的情感进行分析,而大多数情感分析软件采用基于情感词的情感分析法(Lexicon-based approach),这种分析方法通过识别文中可以表现情感的词汇来分析情感强度,如出现“这次旅行让人郁闷”时,“郁闷”一词表现了一种负面的情绪。但这一方法有两方面局限性:首先,当文本较短时,可能文本中并无或少有情感词库中的词汇,导致识别不精确或不成功;其次,当出现复杂的语意,如多重否定、暗示等时,单纯通过识别情感词可能会导致错误。因此,本研究在情感分析时引入基于神经网络的模型分析法(Model-based approach)。神经网络通过大量标注文本的预训练,能够很好的识别复杂的情感表达与句意模式,可以得到一个平滑的情感极性分数以用于之后的分析。在现实中,民宿在线评论这种文本形式可能具有两个特点:

第一,不同句间往往明确在讲不同的评论方面,且各方面之前关联程度较小。如一篇评论:“这间民宿的地理位置相当不错。但是因为楼层较高,厕所味道较重。”此评论中两句话分别在描述不同方面,且有不同的情感倾向。

第二,民宿评论中有反复出现的关键词,如房东、地理位置等信息几乎被绝大多数评论提及,且有一套关键词来提示这些属性,有某种约定俗成的意味。故在这部分,本研究采取句子级别的细粒情感分析。先在上一节利用 Fightin's Words 算法模型得到的一套与民宿属性的关键词典的基础上,来判断这个句子提到具体哪个属性,再对于每一篇评论中的每个句子再进行情感分析,进而得到中西方游客对于各个属性的不同情感。

## 三、研究结果

### 1 整体满意度差异的研究结果

本研究采用 NLTK 工具包的基于模型对情感极性进行度量,以期望得到中英文评论中对民宿整体满意度的差异。实验得到了 20 万评论样本的整体情感极性的分析结果,如表 1 所示,情感极性分数范围在 -1 到 1 之间,数值越负表示情感越消极,数值越正表示情感越积极,即对民宿越满意。从表 1 中的结果可以看出,中国游客和西方游客表现出的整体满意度存在差异,且中国游客的整体满意度要高于西方游客。其中,西方游客最不满意的评论的消极程度比中国游客最不满意的评论还消极 25%,这表明西方游客在给差评的时候更直率。综上, H1 假设成立。

表 1 中英文评论的积极性与消极性

中文评论	英文评论
------	------



Mean	0.26837	Mean	0.2571
Max	1.00000	Max	1.0000
Min	-0.8000	Min	-1.000
Var	0.09422	Var	0.1007

整体来看，英文评论情感极性的方差比中文评论大 10%，这说明西方游客的评论比中国游客的差异要更大。这是否是由西方游客更为直接地表达自己的观点所造成的呢？为了揭开这个问题的答案，本研究进一步对评论的主观性进行了一个度量，具体结果见下表 2，主观性分数范围在 0 到 1 之间，越大表示评论文本越主观。从表中可以看出，平均来看，西方游客评论比中国游客更为主观，因此总体来说西方游客在撰写评论时更加主观地表达自己的观点。

**表 2 中英文评论的主观性与客观性**

中文评论		英文评论	
Mean	0.4612	Mean	0.4923
Max	1.0000	Max	1.0000
Min	0.0000	Min	0.0000
Var	0.9583	Var	0.8889

## 2. 对民宿属性强调差异的研究结果

在比较粗粒度对所有评论整体进行情感分析以后，本文通过词云图初步的了解中西方游客对民宿强调的强调差异，以便细化接下来的研究。词云图结果如图 2 所示。



**图 2 中西方游客民宿评论词云图**

接着，运用 Fightin’s Words 算法模型分别对中英文评论中所提及的民宿不同属性进行分析，得到了中西方游客对民宿属性的强调差异（见附录 5），该分数表示在多大程度上该词代表了中国游客或西方游客的评论，简单地说分数越小越能体现西方游客强调的属性，分数越大越能体现中国人强调的属性。在去掉了无意义、翻译造成的用法不同和同义但表达习惯不同的词组后，经过进一步对初始结果的处理，本研究提取了几个有意思的属性，筛选了部分强调差异较为明显的研究结果（如图 3 所示）：在房东方面，西方游客强调房东回复速度，而中国游客强调的是房东的友好程度；在房间方面，中国游客更加强调的是屋内的空气即通风效果和床的大小及舒适度，而西方游客则更多强调房间空间的大小、装修的新旧或好坏以及洗澡水；在房源周边环境方面，中国游客强调房源附近是否有便利店，西方游客则看重附近是否有咖啡铺、酒吧餐吧和夜市；在房源与游玩地点的距离方面，中国游客强调房源距离维港广场、油麻地、庙街的距离，而西方游客则看重房源是否位于港岛上和离兰桂坊的距离；在出行方式上，中国游客强调房源附近是否有地铁站，而西方游客则看重步行的距离和离巴士车站的距离；其他方面，中国游客更强调房源是否适合家庭出行和价格的高低，而西方游客主要看重是否适合独自住宿。上述六个方面仅筛选了部分强调差异较为明显的结果，剩下的还有诸多属性也存在强调差异，将在后续将形成与民宿有关属性的关键词典用于下一节的研究。

综上，中西方游客对民宿属性的强调存在明显差异，假设 H2 成立。

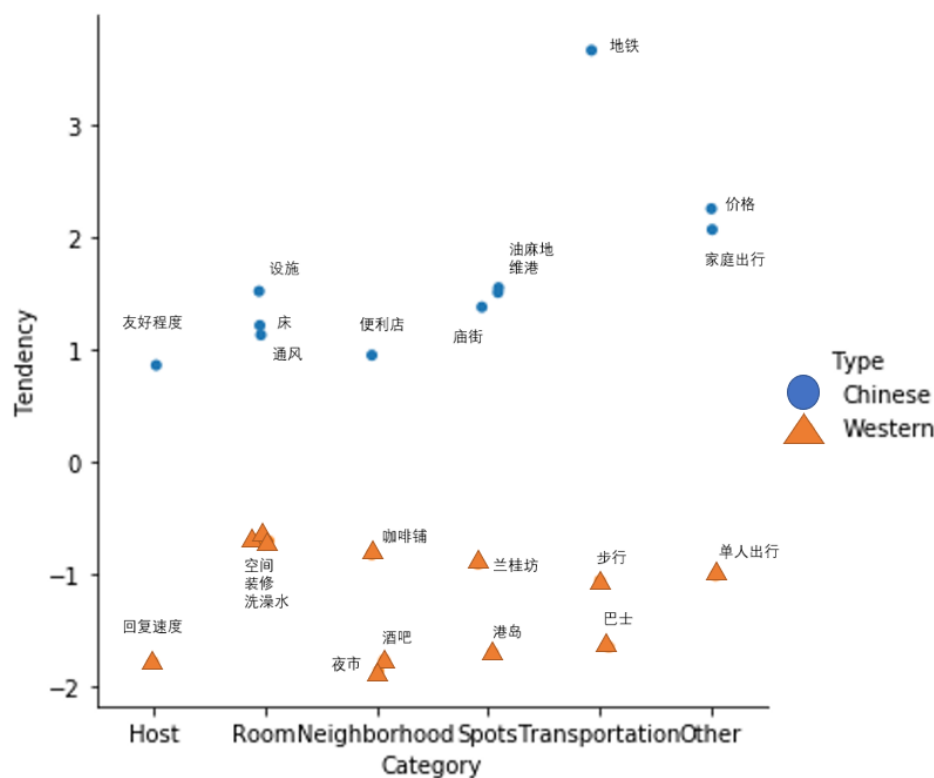


图 3 中西方游客民宿方面强调差异图

### 3 对民宿属性的感知差异研究结果

在得到了中西方游客强调的不同属性以后，为了能够细粒度分析中西方游客对民宿各个属性的感知程度，本研究采用神经网络模型分析法进行句子级别的细粒情感分析，得到了中西方游客对于同一属性的不同情感程度。此实验是建立在 4.2 节 Fightin’s Words 模型得出的结果的基础上，同时参照由丽萍等学者的做法，<sup>[58]</sup>将民宿各项属性进行了归类（总归为四大类：民宿内部和外部环境、房东、价格），形成了一套与民宿属性相关的关键词典。并对民宿内部和外部环境、房东、价格的情感极性分别进行了测量（如表 3 所示），得到每个属性的情感分数（详见附录 5）。

表 3 民宿属性感知测量的关键词典

民 宿 内 部 环境	总体房源	新旧程度	装修	和图片相似程度		和描述相似程度			
		是否适合某种出行意图：单人出行		夫妻出行（情侣出行）；		家庭出行；	商务出行；		
	总体房间	层高		面积	空气	卫生			
	设施	家具	窗户	Wifi		毛巾	投影仪		
		床单	空调		窗帘	冰箱	床		
	卫浴间	马桶	浴缸	洗漱台			洗发露		
		沐浴露		卫生情况			干湿分离情况		
	厨房	做饭便利程度		锅	炉		烤箱	微波炉	
民 宿 外 部 环境	整体交通	出行便利程度		机场	码头		火车	当地交通	
		Mini 巴士	巴士	地铁		步行	骑车		
	整体周边	周边景点	油麻地	旺角	庙街	尖沙咀	体育中心		
		维多利亚港		西贡	其他自然景观		岛屿	山	



		街区便利程度	咖啡铺	商场	购物中心
	位置	便利店	酒吧	药店	夜市
房东	房东整体	房东回复		态度	友好程度
		易沟通程度		耐心程度	
价格	高（贵）		低（便宜）		性价比（值）

由于本节目的是研究中西方游客对各个属性的感知，在得到以上属性的所有情感分数结果以后，本文利用 Seaborn 和 Pandas 通过代码将得到的结果可视化（见附录 4），箱线图（Box 图）中间的横线代表着均值，可理解为对该属性的满意度，横线位置越高代表对该属性越满意，越低代表对这个属性更为挑剔；而箱子的大小代表着方差的大小，箱子越大可以理解为对此属性更为敏感。研究结果显示（图 4）：对于房东方面，中国游客更为满意，而在民宿整体的内部环境、外部环境和价格上又比西方游客显得更为挑剔，尤其是在价格方面。同时西方游客在这四个属性均比中国游客更为敏感。在四个方面中，无论是中国游客还是西方游客都对房东这个属性满意度最高。

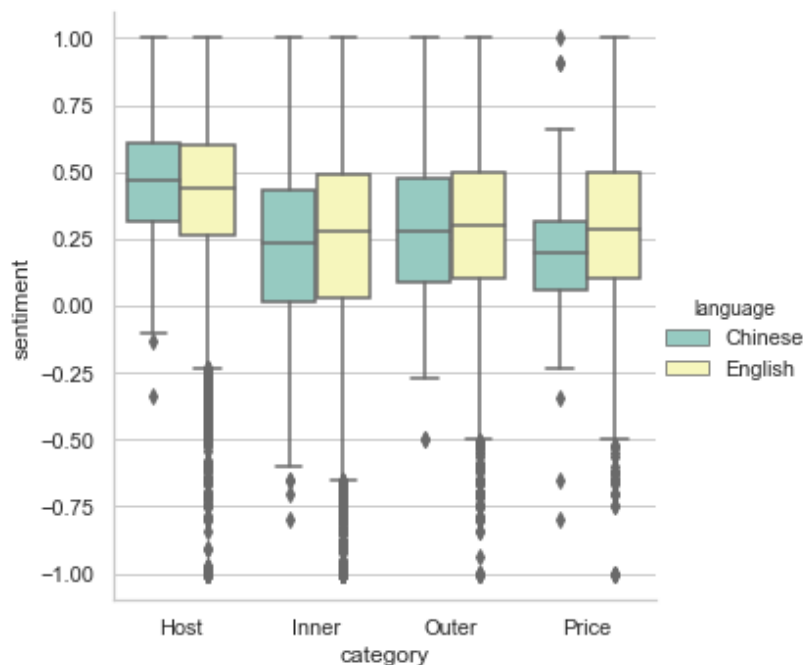


图 4 中西方游客四大属性的感知差异 BOX 图

### 3.1 对民宿房东属性的分析

接着本节对这四个属性进行深入分析，挖掘到更细粒度的领域。在房东方面，本研究对房东的回复速度、态度、友好程度、易沟通程度和耐心程度这六个属性分别做更细粒度的情感分析，研究结果显示（见图 5）：西方游客对房东的回复速度、态度、友好程度、易沟通程度、耐心程度这这六个属性均比中国游客更为挑剔；同时，西方游客在房东回复速度、友好程度、易沟通程度上更为敏感，尤其是对房东回复速度和友好程度这两个属性。

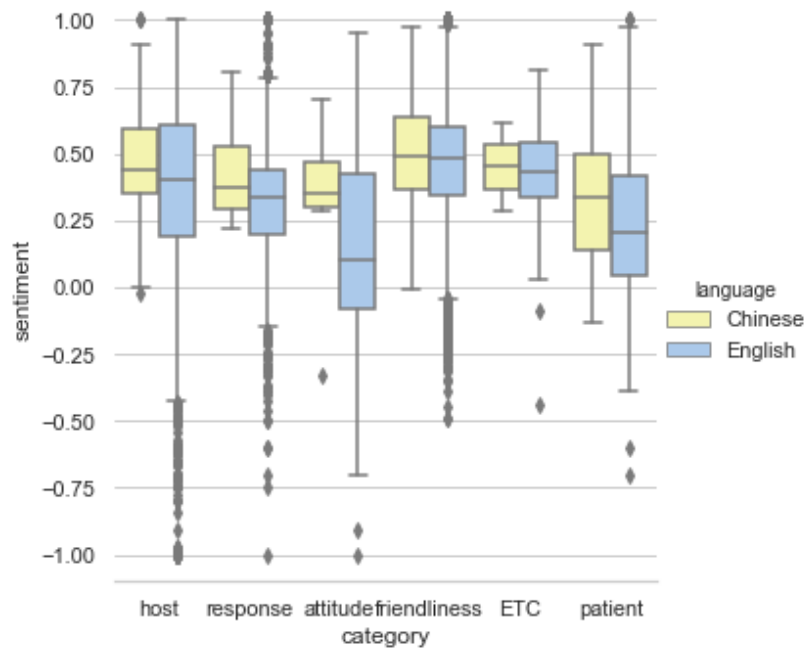


图 5 中西方游客对房东方面六个属性的感知差异 BOX 图

### 3.2 对民宿内部环境属性的分析

对于民宿的内部环境方面，本研究对整体房源、整体房间、设施、卫浴间和厨房分别做了细粒度的情感分析，研究结果显示（见图 6）：中国游客对房源的总体满意度基本与西方游客一致，而在房源的设施、厨房上比西方游客更为挑剔，西方游客则对房间总体、卫浴间更挑剔；西方游客对整体房源、设施、卫浴间和厨房均比中国游客更为敏感，尤其是对房源设施和厨房上。

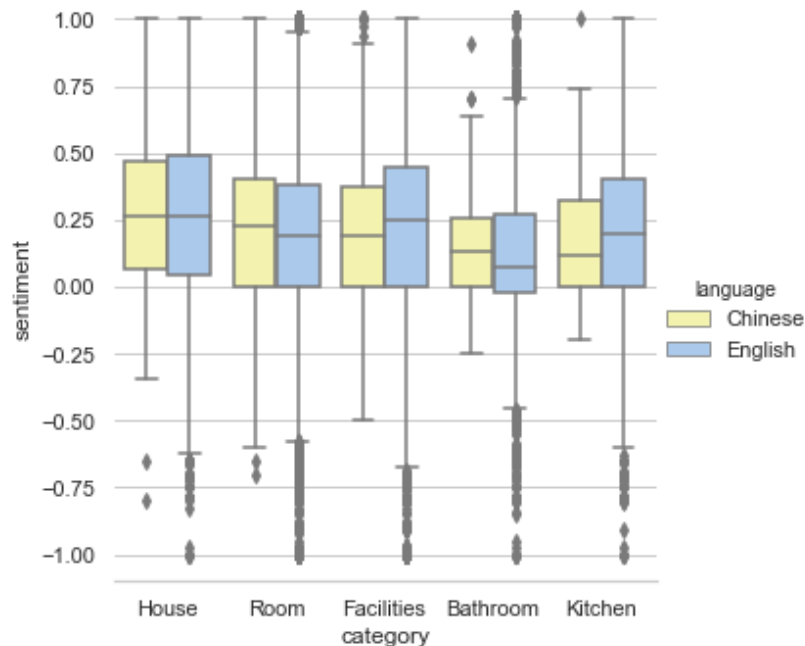


图 6 中西方游客对民宿内部五个属性的感知差异 BOX 图

为了更进一步地理解中西方游客的差异，本文进一步挖掘整体房源、整体房间、设施、卫浴间和厨房，将它们细化到更细的属性（见表三），研究结果显示：在整体房源方面，中国游客对房源与描述相似程度、是否适合家庭或商务出行住宿满意度较高，而对房源的新旧、装修、与图片相似程度、是否适合

单人或情侣出行住宿比起西方游客更为挑剔；在房间整体方面，西方游客对房间的卫生情况较为满意，但相比中国游客，他们却对房间的面积、层高、以及空气情况更为挑剔；在房间设施方面，相比西方游客，中国游客只在床和窗帘这两个属性较为满意，却对家具、窗户、wifi 无线网络、投影仪、毛巾、床单、冰箱和空调都很挑剔；在卫浴间方面，相比中国游客，西方游客对马桶、浴缸、洗漱台、干湿分离情况以及沐浴露都较为满意，而对卫浴间整体的卫生情况和洗发露较为挑剔；在厨房方面，相比于西方游客，中国游客对所有属性包括做饭便利程度、炉、锅、烤箱和微波炉都要更挑剔。

3.3 对民宿外部环境属性的分析

在于民宿外部环境方面，本研究对整体交通、周边景点、整体位置分别做了细粒度的情感分析，研究结果显示（见图 7）：相比于西方游客，中国游客在对房源周边的景点和房源的整体位置这两个属性更挑剔，而对房源附近整体交通情况比较满意；而西方游客对整体位置情况和周边景点的情况更为敏感。

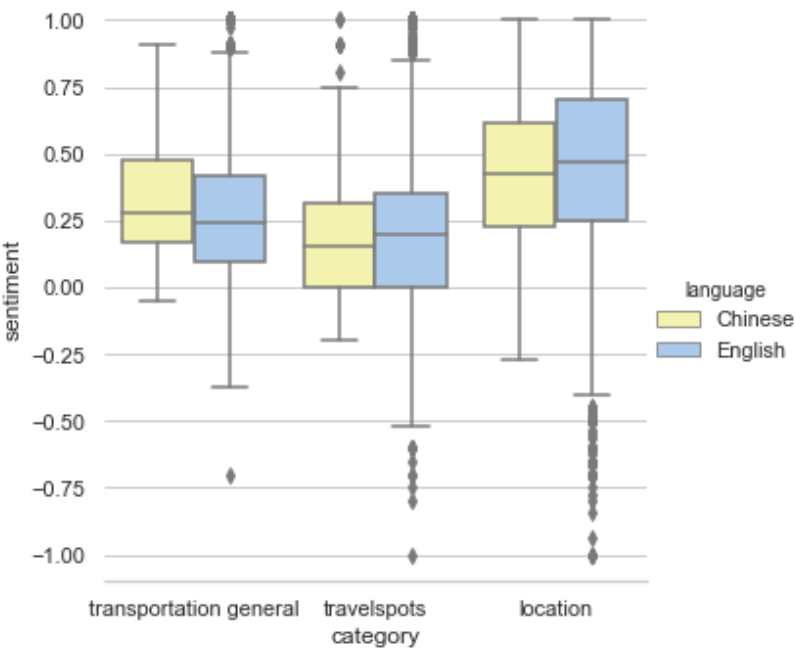


图 7 中西方游客对民宿外部三个属性的感知差异 BOX 图

为了更进一步地理解中西方游客的差异，本文对交通、周边、位置三个方面做了更细粒度的情感分析，分别细化到更小的方面，研究结果显示：在交通方面，相比于西方游客，中国游客对民宿所在地区交通情况和其与机场距离、码头距离、乘大巴距离、乘 mini 巴士距离、乘地铁距离、步行距离更为挑剔；在周边情况方面，相比于西方游客，中国游客对民宿与景点的距离、尖沙咀的距离、铜锣湾的距离、体育中心的距离、维多利亚港的距离、西贡的距离、庙街的距离、自然景观的距离、山的距离更为挑剔；在整体位置方面，相比于中国游客、西方游客中国游客对附近街区便利程度、有无购物中心、咖啡馆、药店、酒吧和夜市更挑剔。

3.4 对民宿价格属性的分析

在于价格方面，本文对整体房源的价格、价格高（贵）、价格低（便宜）、性价比分别作了更细粒度的情感分析，研究结果显示（见图 8），对于高价格的房源，中国游客和西方游客都不太满意，中国游客对于整体房源的价格、低价格的房源和具有性价比的房源都比西方游客更挑剔，而西方游客对整体房源价格和高价格的房源较为敏感。



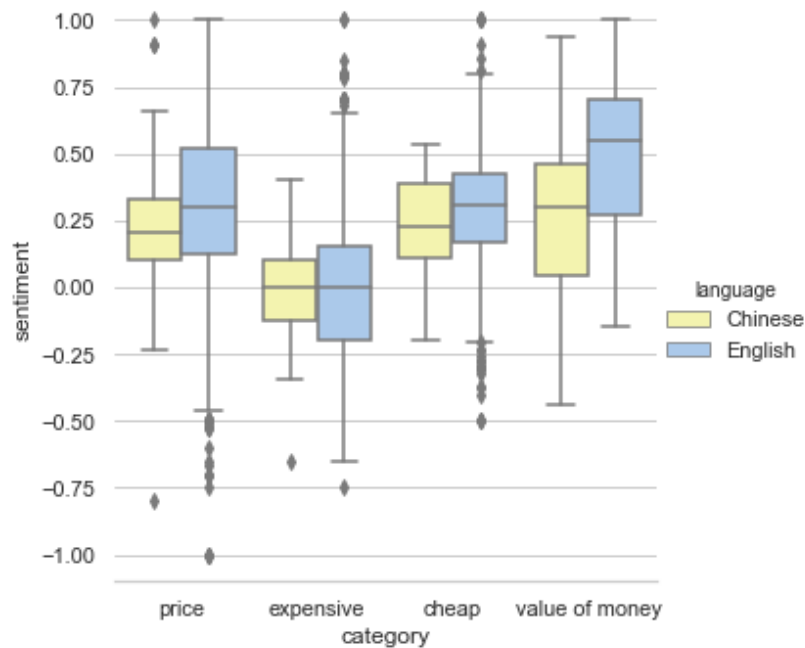


图 8 中西方游客对价格方面四个属性的感知差异 BOX 图

基于以上四个大方面以及所涉及七十多个属性的研究结果，可以得出中西方游客在对民宿属性的感知存在明显差异，故 H3 假设成立。

综上，假设 H1、H2、H3 均成立。

## 四、结论

### 1. 理论贡献

首先，本研究是把 Airbnb 平台的民宿房源作为研究对象，过去的学者研究了影响 Airbnb 房源价格的关键因素，<sup>[59]</sup>也有的对比两地研究了影响民宿口碑的因素<sup>[60]</sup>，也有的学者研究了民宿商家属性对消费者口碑行为的影响。<sup>[61]</sup>之前国内大多研究民宿网络口碑的动机、参与者、影响机制和不同情境的要素，对民宿在线口碑满意度的研究却并不多见。而目前对在线口碑满意度的研究集中于酒店业，然而共享经济平台的民宿评论与传统电商平台的酒店评论又有很大不同，<sup>[5]</sup>随着共享经济的高速发展，这一点很重要，因此民宿的在线评论必须纳入在线口碑满意度的讨论之中，本文丰富了对民宿在线口碑满意度的研究，把对传统酒店业的在线评论的研究拓展到了民宿业。

其次本文通过测量中西方游客异质性对民宿不同属性的强调和各个属性的感知的影响进行了探索，为跨文化的研究提供了新的视角和方向。研究结果显示，不同文化背景的游客无论是在对民宿整体的满意度（H1）、不同属性的强调（H2）还是各个属性的感知（H3）上都存在着差异。具体来讲，中国游客比西方游客在撰写评论时更加积极和客观；在对民宿不同属性的强调上，中国游客与西方游客强调程度差异最为突出的方面体现在对房东、房间、游玩地点、周围环境、出行方式上。在此基础上，本研究形成了一套用于测量民宿各个属性的关键词典，可供未来做相关研究的学者们参考；在对民宿各个属性感知上，中国游客和西方游客对房东、价格、房源内部环境和房源外部环境四个大方面以及其细化的七十多个属性上都有不同程度的感知。目前，国内外鲜有对民宿进行跨文化差异的研究，因此本文的研究成果为跨文化研究文献也积累了一定的贡献。

最后，本文在实验中使用的方法展示了将传统或现代的自然语言处理（NLP）算法引入营销领域的潜力。在最近的尝试中，Makoto 和 Yun<sup>[14]</sup>使用商用分析软件（WCA）和基本的频率比较等进行强调分析和感知分析。而本文使用的方法与传统软件分析的方法相比有两点的优势。一方面，本研究引入了 Fightin's Words 算法，该算法通过对词频分布的最大似然估计得出了比普通频率分析在各方面更准确稳定的结果。该方法在基于数据分析的政治学中早已获得广泛运用，而现在本文将引入营销领域，可以取得较好的效果。另一方面，通过代码实现和基于深度学习的情感分析模型的调用，我们可以更细节化更定制化地分析文本的情感。通过代码实现可以自定与民宿有关的各个属性及相应关键词，形成关键词典，对句子级别进行属性判定和情感分析。深度学习模型近年来在各个自然语言处理的任务上都展现出了卓越的泛用性。在情感分析上，该方法能够识别出更细节更准确的情感表现。通过两种新方法的引入，可以得到基于文本的更深刻见解。当遇到需

要分属性的情感分析与更复杂情况下的强调比较时，未来的研究者可以采用相似的方法进行实验。若与作者联系，作者愿意提供模型与代码。

## 2. 管理启示

本研究从网络大数据中解读出不同文化背景的游客对民宿的情感差异，是一种基于旅游大数据探究游客情感倾向的尝试，有助于更好地指导民宿针对不同文化背景游客改进营销策略和方式，具有较高的应用价值。

对于旅游和民宿从业者来说，研究结果显示了中西方游客之间的关键文化差异。本研究的结果，藉由找出细分游客的重要因素以及游客对这些因素的感知，可以帮助相关从业者在游客满意标准下，决定稀缺财务资源的最佳分配。在对不同文化背景游客的有效细分，以及对民宿属性重点和感知的分析中可能会发现一些有关游客满意度的问题，如果不加以解决，就会损害房源的声誉。一方面，它能帮助理解特定文化背景的游客是否对民宿特定的属性有着特别的关注，例如，的研究结果发现中国游客对附近是否有便利店更为关心，而西方游客则更为关心附近是否有咖啡铺和酒吧，因此针对不同的房源，民宿从业人员可以对不同文化群体的游客投放广告，突出房源的特定属性以吸引相关文化群体的游客；另一方面，通过得到不同文化群体对某方面的感知，能够有针对性地对某方面进行改善，从而提高满意度，例如西方游客对于房东服务质量更为挑剔，而中国游客则体现在房源价格上。那么对于西方游客，民宿从业人员在交流过程中需要更注重自身回复速度、态度、友好程度和耐心程度，以保证带给西方游客高的服务质量感知。而对于中国游客，营销人员在设定房源价格以前，需要更注重市场调研，尽可能使得房源符合性价比的要求。对于相关从业人员在设计基于特定民宿方面的有说服力的营销活动、促销活动和沟通时，确定详细的文化背景和在强调和感知民宿方面的差异可能是至关重要的。本文的研究结果表明，民宿从业人员应该考虑个人的不同需求和期望，以及他们的看法，在设计一个特定的营销活动之前，应该识别并了解他们的特定需求以及他们的文化特性，因为个人认为重要的酒店方面会因他们文化背景的不同而有所不同。

## 3. 研究不足及未来方向

首先，本研究只选择了香港地区 Airbnb 房源作为数据来源，从单个国家的地区提取数据时可能会产生偏见，因此可进一步将分析扩展到其他国际旅游目的地，如新加坡、巴黎或纽约，未来对文化差异的研究可以集中在这些城市。

其次，香港对于中国游客和西方游客这两个文化背景下的群体而言，目的地一个是国内，一个是国外，可能如前人研究所述，<sup>[62-64]</sup>游客对旅行体验的评价可能会因此而有所不同，从而影响到本研究的分析。未来可以对这个问题进行深入研究，探索原籍国的民宿是否会影响在线口碑。

再次，尽管我们引入了一种新的分类方法来处理大数据集，与传统的文化研究相比，产生了稳定的实证结果，但是语言分类的有效性还有待于进一步的检验。本研究把用英语写的评论归为一个组，把中文写的评论归为另一个组，原因是说英语的人即使他们来自不同的国家，也可能有共同的西方文化特征。<sup>[34]</sup>但是关于语言分类的问题还有待进一步的研究，因为一个外国人的国籍和文化背景不能仅仅根据 Airbnb 上的信息来确定。

最后，未来我们还可以在Schucker学者研究的基础上选择剩下两个维度中的期望差异（expectation difference）进行研究，<sup>[34]</sup>本研究考虑跨文化差异时，假设了需求和期望方面的差异，但没有明确。这项措施可以提供额外的关键信息，帮助解释个人之间存在的强调和感知差异。

## 参考文献

- [1] Liu, S. Q., Mattila, A. S.. Airbnb: Online targeted advertising, sense of power, and consumer decisions. *International Journal of Hospitality Management*, 2017,(60):33-41.
- [2] Geetha, M., Singha, P., Sinha, S.. Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis. *Tourism Management*, 2017,(61):43-54.
- [3] Jones, P. H. D. C. D.. Sustainability in the global hotel industry. *International Journal of Contemporary Hospitality Management*, 2014,26(1):5-17.
- [4] Fang, B., Ye, Q., Law, R.. Effect of sharing economy on tourism industry employment. *Annals of Tourism Research*, 2016,(57):264-267.
- [5] 张延宇. 共享经济背景下在线民宿预订评价影响因素分析. 哈尔滨工业大学, 2017.
- [6] Kau, A. K., Lim, P. S.. Clustering of Chinese tourists to Singapore: an analysis of their motivations, values and satisfaction. *International Journal of Tourism Research*, 2005,7(4-5):231-248.
- [7] Gavilan, D., Avello, M., Martinez, N. G.. The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 2018,(66):53-61.
- [8] Radojevic, T., Stanistic, N., Stanic, N., et al. The effects of traveling for business on customer satisfaction with hotel services. *Tourism Management*, 2018,(67):326-341.
- [9] Tan, H., Lv, X., Liu, X., et al. Evaluation nudge: Effect of evaluation mode of online customer reviews on consumers' preferences. *Tourism Management*, 2018,(65):29-40.
- [10] Bodet, G., Anaba, V., Bouchet, P.. Hotel Attributes and Consumer Satisfaction: A Cross-Country and Cross-Hotel Study. *Journal of Travel & Tourism Marketing*, 2016,34(1):52-69.
- [11] 唐建生, 贾慧敏, 王彦彦. 基于归因理论的消费者对第三方服务失败的反应研究. *工业工程与管理*, 2017,22(03):172-177.
- [12] Francesco, G., Roberta, G.. Cross-country analysis of perception and emphasis of hotel attributes[J]. *Tourism Management*, 2019,(74):24-42.
- [13] Forgas, C. S., Palau, S. R., Sánchez, G. J., et al. Urban destination loyalty drivers and cross-national moderator effects: The case of Barcelona. *Tourism Management*, 2012,33(6):1309-1320.
- [14] Nakayama, M., Wan, Y.. Is culture of origin associated with more expressions? An analysis of Yelp reviews on Japanese restaurants. *Tourism Management*, 2018,(66):329-338.
- [15] Pizam, A., Pine, R., Mok, C., et al. Nationality vs industry cultures: which has a greater effect on managerial behavior?. *International Journal of Hospitality Management*, 1997,16(2):127-145.
- [16] Hofstede, G.. *Culture and Organisations: Software of the Mind*. New York: McGraw-Hill.: 1991.
- [17] Crotts, J. C., Erdmann, R.. Does national culture influence consumers' evaluation of travel services? A test of Hofstede's model of cross-cultural differences. *Managing Service Quality: An International Journal*, 2000,10(6):410-419.
- [18] Kim, S., Thorndike, P. D.. Predicting purchase intentions for uni-national and bi-national products. *International Journal of Retail & Distribution Management*, 2000,28(6):280-291.
- [19] Nowak, J., Kochkova, O.. Income, Culture, and Household Consumption Expenditure Patterns in the European Union: Convergence or Divergence?. *Journal of International Consumer Marketing*, 2011,23(3):260-275.
- [20] Mooij, D. M., Beniflah, J.. Measuring Cross-Cultural Differences of Ethnic Groups within Nations: Convergence or Divergence of Cultural Values? The Case of the United States. *Journal of International Consumer Marketing*, 2017,29(1):2-10.
- [21] Gao, B., Li, X., Liu, S., et al. How power distance affects online hotel ratings: The positive moderating roles of hotel chain and reviewers' travel experience. *Tourism Management*, 2018,(65):176-186.
- [22] Huang, S. S., Crotts, J.. Relationships between Hofstede's cultural dimensions and tourist satisfaction: A cross-country cross-sample examination. *Tourism Management*, 2019,(72):232-241.
- [23] Liu, Y., Teichert, T., Rossi, M., et al. Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, 2017,(59):554-563.
- [24] Nath, P., Devlin, J., Reid, V.. The effects of online reviews on service expectations: Do cultural value orientations matter?. *Journal of Business Research*, 2018,(90):123-133.



- [25] Cheung, C. M. K., Thadani, D. R.. The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, 2012,54(1):461-470.
- [26] Chik, A., Vásquez, C.. A comparative multimodal analysis of restaurant reviews from two geographical contexts. *Visual Communication*, 2017,16(1):3-26.
- [27] Stamolampros, P., Korfiatis, N., Kourouthanassis, P., et al. Flying to Quality: Cultural Influences on Online Reviews. *Journal of Travel Research*, 2018,58(3):496-511.
- [28] Eisingerich, A. B., Rubera, G.. Drivers of Brand Commitment: A Cross-National Investigation. *Journal of International Marketing*, 2010,18(2):64-79.
- [29] Hofstede, G.. *Culture's consequences: Comparing values*. Thousand Oaks, CA: Sage Publications, 2001.
- [30] Lalwani, A. K., Shavitt, S.. You Get What You Pay For? Self-Construal Influences Price-Quality Judgments. *Journal of Consumer Research*, 2013,40(2):255-267.
- [31] Kumar, V., Pansari, A.. National Culture, Economy, and Customer Lifetime Value: Assessing the Relative Impact of the Drivers of Customer Lifetime Value for a Global Retailer. *Journal of International Marketing*, 2016,24(1):1-21.
- [32] Akdeniz, M. B., Talay, M. B.. Cultural variations in the use of marketing signals: a multilevel analysis of the motion picture industry. *Journal of the Academy of Marketing Science*, 2013,41(5):601-624.
- [33] Alden, D. L., Merz, M. A., He, Y.. A categorization approach to analyzing the global consumer culture debate. *International Marketing Review*, 2008,25(2):166-182.
- [34] Schuckert, M., Liu, X., Law, R.. A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *International Journal of Hospitality Management*, 2015,(48):143-149.
- [35] Pizam, A.. Hotel guest surveys: Uses and misuses. *International Journal of Hospitality Management*, 2017,(65):125-126.
- [36] Lai, J., He, P., Chou, H., et al. Impact of National Culture on Online Consumer Review Behavior. *Global Journal of Business Research*, 2013,7(1):109-115.
- [37] Ha, J., Jang, S. S.. Effects of service quality and food quality: The moderating role of atmospherics in an ethnic restaurant segment. *International Journal of Hospitality Management*, 2010,29(3):520-529.
- [38] Laroche, M., Ueltschy, L. C., Abe, S., et al. Service Quality Perceptions and Customer Satisfaction: Evaluating the Role of Culture. *Journal of International Marketing*, 2018,12(3):58-85.
- [39] Bagheri, A., Saraee, M., Jong, F.. *An unsupervised aspect detection model for sentiment analysis of reviews*. Natural Language Processing and Information Systems. Springer Berlin Heidelberg, 2013.
- [40] McCleary, K. W., Choi, B. M., Weaver P A. A Comparison of Hotel Selection Criteria Between U.S. and Korean Business Travelers. *Journal of Hospitality & Tourism Research*, 1998,22(1):25-38.
- [41] Poon, W. C., Lock, T. L. K.. Are travellers satisfied with Malaysian hotels?. *International Journal of Contemporary Hospitality Management*, 2005,17(3):217-227.
- [42] Tsang, N. K., Ap, J.. Tourists' Perceptions of Relational Quality Service Attributes: A Cross-Cultural Study. *Journal of Travel Research*, 2007,45(3):355-363.
- [43] Kuo, C.. The Importance of Hotel Employee Service Attitude and the Satisfaction of International Tourists. *The Service Industries Journal*, 2007,27(8):1073-1085.
- [44] Weiermair, K.. Tourists' perceptions towards and satisfaction with service quality in the cross-cultural service encounter: implications for hospitality and tourism management. *Managing Service Quality: An International Journal*, 2000,10(6):397-409.
- [45] Hsieh, A. T., Tsai, C. W.. Does national culture really matter? Hotel service perceptions by Taiwan and American tourists. *International Journal of Culture, Tourism and Hospitality Research*, 2009,3(1):54-69.
- [46] Hsu, C. H. C., Kang, S. K.. Profiling Asian and Western family independent travelers (FITS): An exploratory study. *Asia Pacific Journal of Tourism Research*, 2003,8(1):58-71.
- [47] Manrai, L., Manrai, A.. Hofstede's Cultural Dimensions and Tourist Behaviors: A Review and Conceptual Framework. *Journal of Economics, Finance and Administrative*, 2011,16:23-47.
- [48] Zhang, J., Beatty, S. E., Walsh, G.. Review and future directions of cross-cultural consumer services research. *Journal of Business Research*, 2008,61(3):211-224.

- [49] Hannerz, U.. Transnational connections. Culture, people, places. London: Routledge, 1996.
- [50] Fang, X., Zhan, J.. Sentiment analysis using product review data. *Journal of Big Data*, 2015,2(1).
- [51] 刘玉林, 菅利荣. 基于文本情感分析的电商在线评论数据挖掘. *统计与信息论坛*, 2018,33(12):119-124.
- [52] 刘逸, 保继刚, 陈凯琪. 中国赴澳大利亚游客的情感特征研究——基于大数据的文本分析. *旅游学刊*, 2017,32(05):46-58.
- [53] Rylander, R. G., Propst, D. B., Mcmurtry, T. R.. Nonresponse and Recall Biases in a Survey of Traveler Spending. *Journal of Travel Research*, 2016,33(4):39-45.
- [54] Wiebe, J.. Tracking Point of View in Narrative. *Computational Linguistics*, 1994,(20):233-287.
- [55] Joachims, T., Ls, I., Str, B.. Text Categorization with Support Vector Machines: Learning with Many Relevant Features: In *Proceedings of the 10th European Conference on Machine Learning*, Chemnitz, Germany, 1999.
- [56] Ramos, J.. Using TF-IDF to Determine Word Relevance in Document Queries: *Proceedings of the first instructional conference on machine learning*, Piscataway, NJ, 2003
- [57] Monroe, L. B., Colaresi, P. M., Quinn, K.. Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis*, 2009,16(4):372-403.
- [58] 由丽萍, 白旭云. 基于在线评论情感语义分析和TOPSIS法的酒店服务质量测量. *情报科学*, 2017,35(10):13-17.
- [59] 吴晓隽, 裘佳璐. Airbnb房源价格影响因素研究——基于中国36个城市的数据. *旅游学刊*, 2019,34(04):13-28.
- [60] 李燕琴, 于文浩, 柏雨帆. 基于Airbnb网站评价信息的京台民宿对比研究. *管理学报*, 2017,14(01):122-128.
- [61] 梁晓蓓, 徐真, 李晶晶. 共享短租平台商家属性对消费者网络口碑的影响研究. *数据分析与知识发现*, 2018,2(11):46-53.
- [62] Jeuring, J. H. G., Haartsen, T.. The challenge of proximity: the (un)attractiveness of near-home tourism destinations. *Tourism Geographies*, 2017,19(1):118-141.
- [63] Wildish, B., Kearns, R., Collins, D.. At home away from home: visitor accommodation and place attachment. *Annals of Leisure Research*, 2016,19(1):117-133.
- [64] Pearce, P. L.. The experience of visiting home and familiar places. *Annals of Tourism Research*, 2012,39(2):1024-1047.

# **Study on the Sentimental Features of Chinese and Western Tourists' Satisfaction with the Homestay**

**Abstract:** The research crawled 133,419 English reviews and 62,285 Chinese reviews on Airbnb website in Hong Kong through Python and then discussed the significant differences in the sentiment characteristics of Chinese and western tourists' satisfaction with home stay facilities in Hong Kong. First, NLTK Language Toolkit was used to measure the overall satisfaction of Chinese and western tourists from two dimensions: positive and negative, subjective and objective. Second, we used the Fightin's Words model to get the different attributes of dwellings on which Chinese and western tourists emphasize and constructed a set of key dictionary of dwellings attributes. Additionally, based on the neural network Model perception of each attribute of homestays for fine-grained sentiment analysis.

The results show: (1) Chinese tourists have higher overall satisfaction than western tourists and are more objective in writing their comments; (2) There are emphasis differences between Chinese tourists and western tourists. Chinese tourists emphasize the friendliness of hosts while western tourists focus more on the speed of the reply of hosts. As for the room, the ventilation condition and the overall facilities of the room are important for Chinese, especially the bed, while western tourists value the space size, decoration. For the surrounding property, Chinese tourists value convenience stores nearby, while western tourists care more about the nearby coffee shops, bars and night markets. As for the travel mode, Chinese tourists attach great importance to the convenience of taking the subway, which is the most important attribute for Chinese tourists, while western tourists attach great importance to the convenience of taking the bus. In other aspects, Chinese tourists are more concerned about the price. As for many other attributes that differ from emphasize differences, this study has screened them into a set of key dictionaries related to the attributes of homestays; (3) Compared with western tourists, Chinese tourists are more satisfied with the host of the homestay, but more picky about the overall internal and external environment and the price of the homestay.

**Keywords:** Cultural Difference; Homestay; Online Review; Fightin's Words; Emphasis Difference; Model-Based Approach; Perception Difference



## 附录 1 NLTK 工具包示例图

### NLTK 3.4.4 documentation

[NEXT](#) | [MODULES](#) | [INDEX](#)

## Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to [over 50 corpora and lexical resources](#) such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active [discussion forum](#).

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

NLTK has been called "a wonderful tool for teaching, and working in, computational linguistics using Python," and "an amazing library to play with natural language."

[Natural Language Processing with Python](#) provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analyzing linguistic structure, and more. The online version of the book has been updated for Python 3 and NLTK 3. (The original Python 2 version is still available at [http://nltk.org/book\\_1ed](http://nltk.org/book_1ed).)

### Some simple things you can do with NLTK

Tokenize and tag some text:

```
>>> import nltk
>>> sentence = """At eight o'clock on Thursday morning
... Arthur didn't feel very good."""
>>> tokens = nltk.word_tokenize(sentence)
>>> tokens
['At', 'eight', 'o'clock', 'on', 'Thursday', 'morning',
 'Arthur', 'didn't', 'n't', 'feel', 'very', 'good', '.']
>>> tagged = nltk.pos_tag(tokens)
>>> tagged[0:6]
[('At', 'IN'), ('eight', 'CD'), ('o'clock', 'JJ'), ('on', 'IN'),
 ('Thursday', 'NNP'), ('morning', 'NN')]
```

Identify named entities:

```
>>> entities = nltk.chunk.ne_chunk(tagged)
>>> entities
Tree('S', [(('At', 'IN'), ('eight', 'CD'), ('o'clock', 'JJ'),
 ('on', 'IN'), ('Thursday', 'NNP'), ('morning', 'NN')),
 Tree('PERSON', [(('Arthur', 'NNP')],
 ('didn't', 'VBD'), ('n't', 'RB'), ('feel', 'VB'),
 ('very', 'RB'), ('good', 'JJ'), ('.', '.')])])
```

Display a parse tree:

```
>>> from nltk.corpus import treebank
>>> t = treebank.parsed_sents('wsj_0001.mrg')[0]
>>> t.draw()
```

### TABLE OF CONTENTS

- NLTK News
- Installing NLTK
- Installing NLTK Data
- Contribute to NLTK
- FAQ
- Wiki
- API
- HOWTO

### SEARCH

图1 NTLK工具包界面

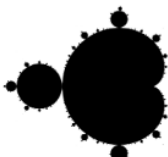
## 附录2 Fightin words算法的实现示例图

```
fighting_words.py
1 import numpy as np
2 from sklearn.feature_extraction.text import CountVectorizer as CV
3 import string
4 exclude = set(string.punctuation)
5
6 def basic_sanitizel(in_string):
7     """Returns a very roughly sanitized version of the input string."""
8     return_string = ' '.join(in_string.encode('ascii', 'ignore').strip().split())
9     return_string = ''.join(ch for ch in return_string if ch not in exclude)
10    return_string = return_string.lower()
11    return_string = ' '.join(return_string.split())
12    return return_string
13
14 def bayes_compare_language(l1, l2, ngram = 1, prior=.01, cv = None):
15     """
16     Arguments:
17     - l1, l2; a list of strings from each language sample
18     - ngram; an int describing up to what n gram you want to consider (1 is unigrams,
19       2 is bigrams + unigrams, etc). Ignored if a custom CountVectorizer is passed.
20     - prior; either a float describing a uniform prior, or a vector describing a prior
21       over vocabulary items. If you're using a predefined vocabulary, make sure to specify that
22       when you make your CountVectorizer object.
23     - cv; a sklearn.feature_extraction.text.CountVectorizer object, if desired.
24
25     Returns:
26     - A list of length |Vocab| where each entry is a (n-gram, zscore) tuple."""
27     if cv is None and type(prior) is not float:
28         print "If using a non-uniform prior:"
29         print "Please also pass a count vectorizer with the vocabulary parameter set."
30         quit()
31     l1 = [basic_sanitizel(l) for l in l1]
32     l2 = [basic_sanitizel(l) for l in l2]
33     if cv is None:
34         cv = CV(decode_error = 'ignore', min_df = 10, max_df = .5, ngram_range=(1,ngram),
35                binary = False,
36                max_features = 15000)
37     # Now sum over Languages...
38     counts_mat = cv.fit_transform(l1+l2).toarray()
39     vocab_size = len(cv.vocabulary_)
40     print "Vocab size is {}".format(vocab_size)
41     if type(prior) is float:
42         priors = np.array([prior for i in range(vocab_size)])
43     else:
44         priors = prior
45     z_scores = np.empty(priors.shape[0])
46     count_matrix = np.empty([2, vocab_size], dtype=np.float32)
47     count_matrix[0, :] = np.sum(counts_mat[:len(l1), :], axis = 0)
48     count_matrix[1, :] = np.sum(counts_mat[len(l1):, :], axis = 0)
49     a0 = np.sum(priors)
50     n1 = 1.*np.sum(count_matrix[0,:])
51     n2 = 1.*np.sum(count_matrix[1,:])
52     print "Comparing language..."
53     for i in range(vocab_size):
54         #compute delta
55         term1 = np.log((count_matrix[0,i] + priors[i])/(n1 + a0 - count_matrix[0,i] - priors[i]))
56         term2 = np.log((count_matrix[1,i] + priors[i])/(n2 + a0 - count_matrix[1,i] - priors[i]))
57         delta = term1 - term2
58         #compute variance on delta
59         var = 1./((count_matrix[0,i] + priors[i]) + 1./((count_matrix[1,i] + priors[i]))
60         #store final score
61         z_scores[i] = delta/np.sqrt(var)
62     index_to_term = {v:k for k,v in cv.vocabulary_.iteritems()}
63     sorted_indices = np.argsort(z_scores)
64     return_list = []
65     for i in sorted_indices:
66         return_list.append((index_to_term[i], z_scores[i]))
67     return return_list
68
```

C:\Users\vxhao\AppData\Local\Temp\Temp1\_FightingWords-master.zip\FightingWords-master\fighting\_words.py\* LF UTF-8 Python GitHub Git (0)

图2 Fightin words算法代码图

## 附录3 Textblob情感分析示例图



### TextBlob

Star 6,368

TextBlob is a Python (2 and 3) library for processing textual data. It provides a consistent API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more.

#### Useful Links

[TextBlob @ PyPI](#)  
[TextBlob @ GitHub](#)  
[Issue Tracker](#)

#### Stay Informed

[Follow @storia](#)

#### Donate

If you find TextBlob useful, please consider supporting its author:

[Flattr](#)

Your donation helps move TextBlob forward.

## TextBlob: Simplified Text Processing

Release v0.15.2. ([Changelog](#))

*TextBlob* is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

```
from textblob import TextBlob

text = '''
The titular threat of The Blob has always struck me as the ultimate movie
monster: an insatiably hungry, amoeba-like mass able to penetrate
virtually any safeguard, capable of--as a doomed doctor chillingly
describes it--"assimilating flesh on contact.
Snide comparisons to gelatin be damned, it's a concept with the most
devastating of potential consequences, not unlike the grey goo scenario
proposed by technological theorists fearful of
artificial intelligence run rampant.
'''

blob = TextBlob(text)
blob.tags          # [('The', 'DT'), ('titular', 'JJ'),
                    # ('threat', 'NN'), ('of', 'IN'), ...]

blob.noun_phrases  # WordList(['titular threat', 'blob',
                    #           'ultimate movie monster',
                    #           'amoeba-like mass', ...])

for sentence in blob.sentences:
    print(sentence.sentiment.polarity)
# 0.060
# -0.341

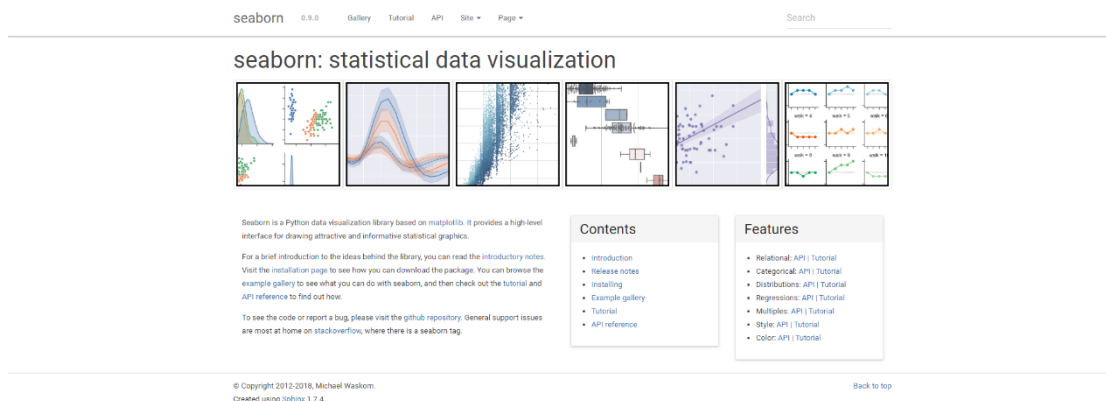
blob.translate(to="es") # 'La amenaza titular de The Blob...'
```

TextBlob stands on the giant shoulders of [NLTK](#) and [pattern](#), and plays nicely with both.

图3 Textblob界面

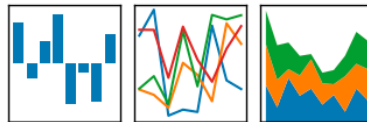


## 附录4 Seaborn和Pandas文献可视化示例图



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



[home](#) // [about](#) // [get pandas](#) // [documentation](#) // [community](#) // [talks](#) // [donate](#)

## Python Data Analysis Library

*pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.

*pandas* is a [NumFOCUS](#) sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to [donate](#) to the project.

A Fiscally Sponsored Project of

**NUMFOCUS**  
OPEN CODE = BETTER SCIENCE

### v0.25.1 Final (August 22, 2019)

This is a minor bug-fix release in the 0.25.x series and includes some regression fixes and bug fixes. We recommend that all users upgrade to this version.

See the [full \*whatsnew\*](#) for a list of all the changes.

The release can be installed with conda from the defaults and conda-forge channels:

```
conda install pandas
```

Or via PyPI:

```
python -m pip install --upgrade pandas
```

### VERSIONS

#### Release

0.25.1 - August 2019

[download](#) // [docs](#) // [pdf](#)

#### Development

0.26.0 - September 2019

[github](#) // [docs](#)

#### Previous Releases

0.25.0 - [download](#) // [docs](#) // [pdf](#)

0.24.2 - [download](#) // [docs](#) // [pdf](#)

0.24.1 - [download](#) // [docs](#) // [pdf](#)

0.24.0 - [download](#) // [docs](#) // [pdf](#)

0.23.4 - [download](#) // [docs](#) // [pdf](#)

0.23.3 - [download](#) // [docs](#) // [pdf](#)

0.23.2 - [download](#) // [docs](#) // [pdf](#)

0.23.1 - [download](#) // [docs](#) // [pdf](#)

0.23.0 - [download](#) // [docs](#) // [pdf](#)

0.22.0 - [download](#) // [docs](#) // [pdf](#)

0.21.1 - [download](#) // [docs](#) // [pdf](#)

0.21.0 - [download](#) // [docs](#) // [pdf](#)

0.20.3 - [download](#) // [docs](#) // [pdf](#)

0.19.2 - [download](#) // [docs](#) // [pdf](#)

0.18.1 - [download](#) // [docs](#) // [pdf](#)

0.17.1 - [download](#) // [docs](#) // [pdf](#)

0.16.2 - [download](#) // [docs](#) // [pdf](#)

0.15.2 - [download](#) // [docs](#) // [pdf](#)

0.14.1 - [download](#) // [docs](#) // [pdf](#)

0.13.1 - [download](#) // [docs](#) // [pdf](#)

0.12.0 - [download](#) // [docs](#) // [pdf](#)

```

In [88]: df = pd.DataFrame()

list_of_tuples = []

i = 0

for cat in aspect_dic["inner"]["host"]["general"]:
    # ["host", "response", "attitude", "nice", "friendly", "easy to communicate", "patient"]
    cn_host_temp = cn_sent_results["inner host general " + cat]
    en_host_temp = en_sent_results["inner host general " + cat]

    if cat in ["nice", "friendly"]:
        cat = "friendliness"

    if cat == "easy to communicate":
        cat = "ETC"

    for data in cn_host_temp:
        list_of_tuples.append(tuple([i, cat, data, "Chinese"]))

    for data in en_host_temp:
        list_of_tuples.append(tuple([i, cat, data, "English"]))

df = pd.DataFrame(list_of_tuples, columns = ["index", "category", "sentiment", "language"])

sns.set(style="whitegrid", palette="pastel", color_codes=True)
# tips = sns.load_dataset("tips")
# ax = sns.boxplot(x=tips["total_bill"])

# ax = sns.boxplot(x="category", y="sentiment", hue="language", data=df, palette="Set3")

# ax = sns.catplot(x="category", y="sentiment", hue="language",
#                 split=True, inner="quart",
#                 legend_out = True,
#                 palette={"Chinese": "y", "English": "b"},
#                 data=df, kind="violin")

ax = sns.catplot(x="category", y="sentiment", hue="language",
                legend_out = True,
                palette={"Chinese": "y", "English": "b"},
                data=df, kind="box")

sns.despine(left=True)

plt.show()

```

图4 代码操作可视化界面

## 附录五 中西方游客强调差异的结果和感知差异的结果（浓缩图）

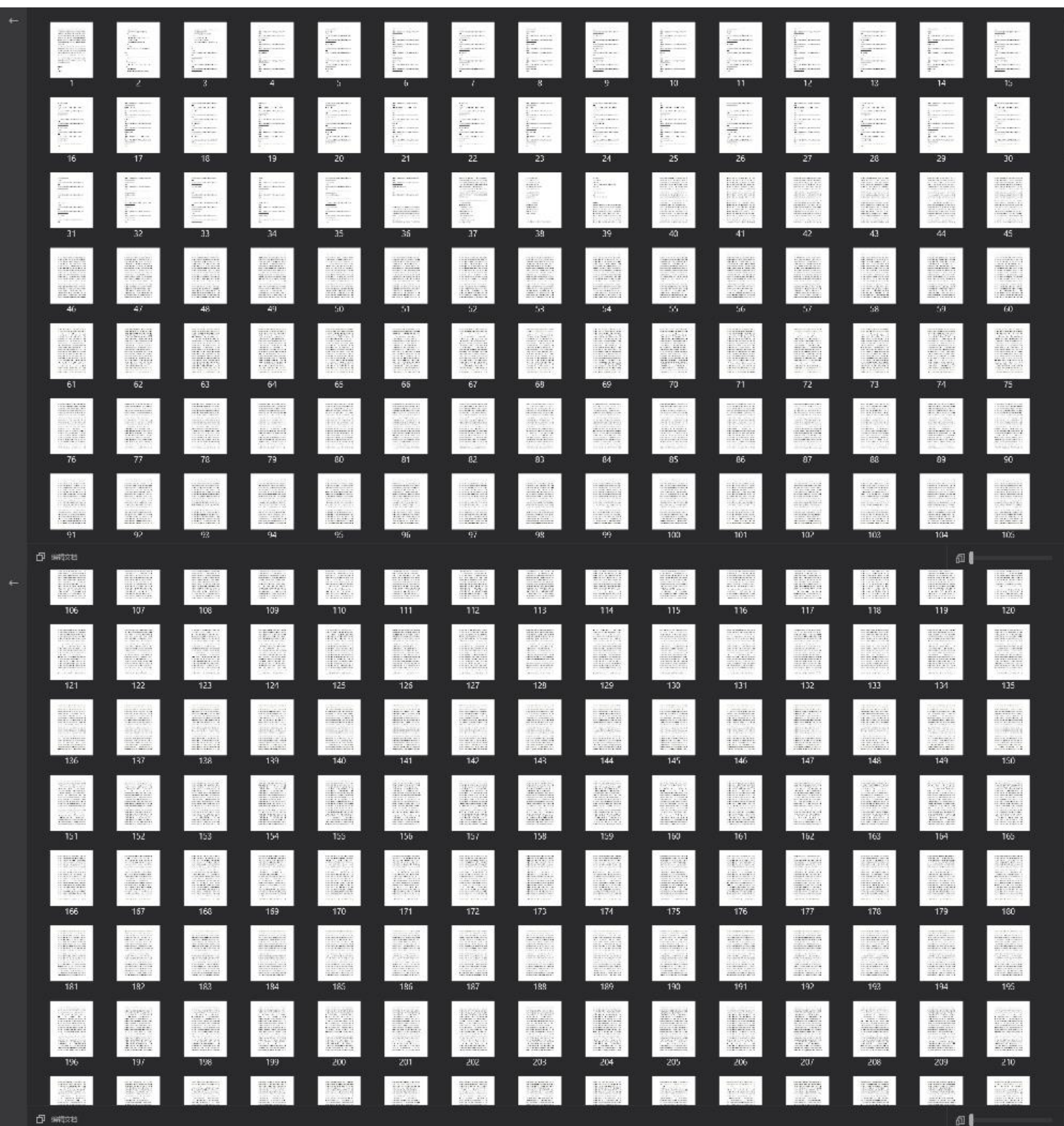


图5 强调差异和感知差异结果浓缩图

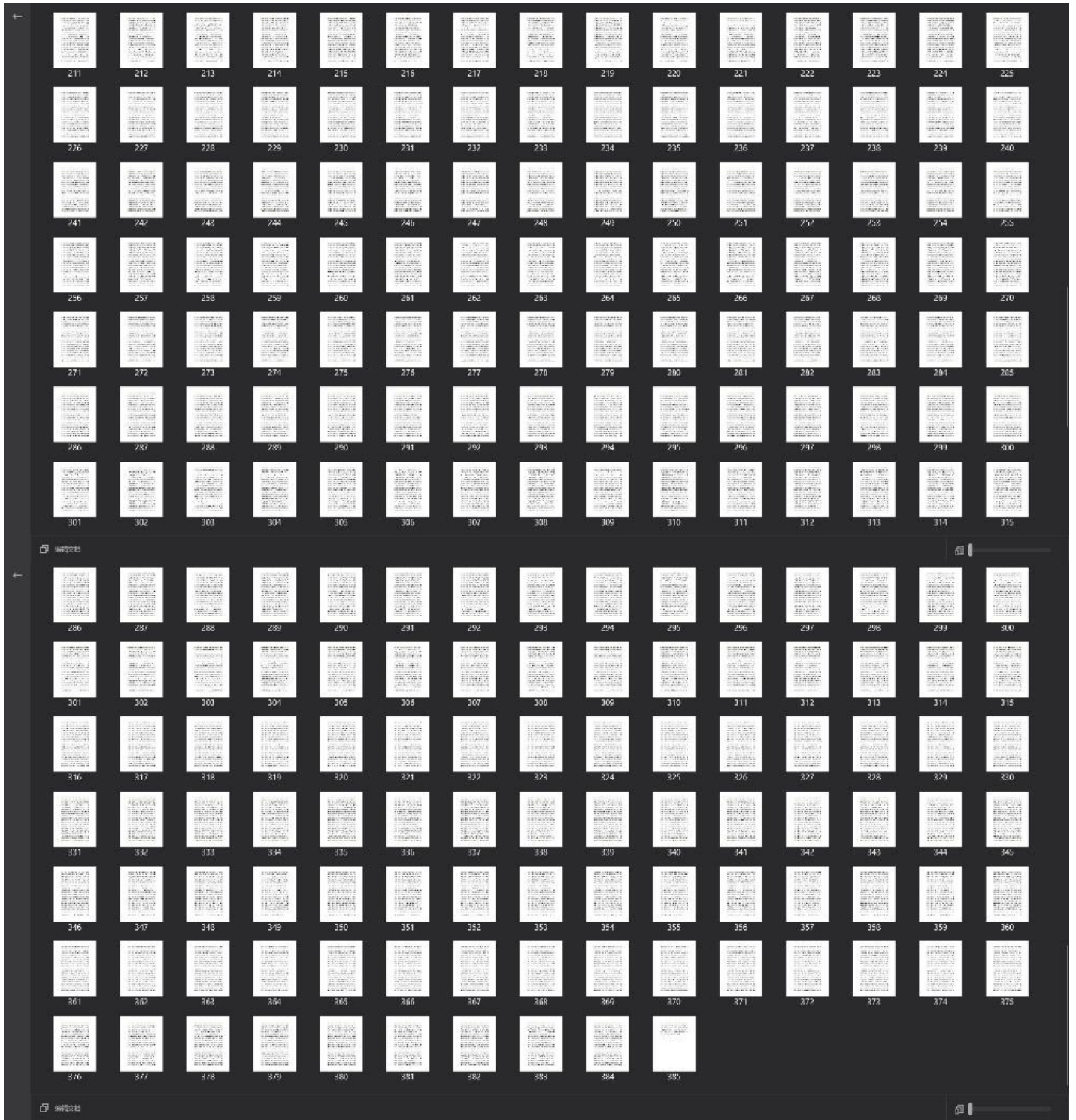


图6 强调差异和感知差异结果浓缩图