

在线评论有用性的深度数据挖掘

——基于 TripAdvisor 的酒店评论数据

○ 史 达 王乐乐 衣博文

摘要 在线评论有用性有助于浏览者快速获取有效信息。本文以双重过程理论的启发式—系统式(HSM)模型为理论基础,基于 TripAdvisor 的酒店评论,采用 Word2Vec、LDA 和机器学习等方法量化变量,通过多元回归对评论有用性的影响因素进行分析。研究发现:(1)基于 HSM 模型建立的有用性影响因素理论模型,将影响因素分为启发式和系统式线索。在启发式线索中,评论长度与有用性呈倒 U 型关系,历史评论数、评论时间、是否有照片存在正面影响,评论者类型不同影响不同;在系统式线索中,所有评论主题、评论负面程度均存在正面影响。(2)采用 Word2Vec 方法计算零有用票数评论的有用性得分,弥补了有用性量化的缺失;采用表现更优的三种机器学习方法(LR、NB、SVM)进行情感分析,并证明所得情感得分优于评论评分对评论情感的衡量。本文以双重过程理论为框架,系统分析了评论有用性的影响因素,更加准确地解释了顾客酒店体验的真实感受,对评论平台和酒店未来需提升的维度和发展方向有一定的借鉴意义。

关键词 评论有用性; HSM; 语义分析; 情感分析; 机器学习

引言

在线评论平台不断发展,越来越多的顾客通过评论

获取商品信息,以减少购买过程中的不确定性和交易成本。^[1]在线评论能够反映商品质量并减少商品不确定性,不但影响顾客对卖家和商品的信任程度、^[2]态度、^[3]质量感知^[4]等,还会影响顾客决策过程中的购买意愿。^[5]在线评论同样影响酒店服务等体验型商品的销售,^[6]如了解在线酒店评论反映的客户投诉内容有助于酒店提高服务质量、客户满意度等。^[7]随着在线评论重要性的增加,评论平台已经成为顾客了解餐馆、酒店、景点等的主要信息来源。然而,在线评论的便利性和信息过载问题并存,^[8]顾客难以判断海量信息的有用性。^[9]因此,事先甄别评论的有用性非常重要。

评论有用性有助于顾客快速筛选商品并进行购买决策,并且评论有用性越高,对顾客决策的影响越大。^[3,10]目前评论有用性的研究趋势,一方面是探索新影响因素或拓展新影响情景,^[11]另一方面是关注如何改进评论有用性的量化方式。^[12]由此,本文首先基于双重过程理论的启发式—系统式模型(HSM),将评论有用性影响因素分为启发式线索和系统式线索,建立影响因素研究模型,探究双重过程对评论有用性的影响;同时,鉴于现有研究在评论者特征方面未曾考虑评论者类型,本文将其作为新影响因素纳入模型中进行检验。

其次,本文侧重量化方法的改进,针对评论有用性的计算方法和情感分析方法进行了算法优化。现有的评

作者简介 史达,东北财经大学萨里国际学院教授、博士生导师,研究方向为旅游体验;王乐乐,东北财经大学国际商学院硕士研究生,研究方向为商业数据分析;衣博文,东北财经大学旅游与酒店管理学院博士研究生,研究方向为旅游大数据分析、旅游共享经济

基金资助 本文受辽宁省教育厅科学研究经费项目基础研究项目(LN2020J26),东北财经大学校级创新团队项目(2020XJTD02),东北财经大学社会服务能力提升项目创新团队专项项目(CXTD2020-4)资助

论有用性量化研究多以有用投票数^[13]或感知有用性的量表评分^[14]来测量评论有用性,进而分析其影响因素。这些研究在评论有用性影响因素分析方面作出了一定贡献,但仍存在以下不足:第一,现有研究对于评论有用性的测量方法较为粗旷,存在遗漏评论中的有用信息、增大估计结果误差等问题;第二,现有评论语义分析方法相对简单,以关键词、主客观词语为参考或根据词典判断情感为主,^[15]未从评论主题角度深度挖掘,存在量化依据不够准确、参照词典相对静态、不能完全涵盖中文表达的丰富含义等问题。基于现有研究存在的不足,本文采用 Word2Vec 算法,根据有用评论计算零有用票数评论的有用性,可弥补零有用票数评论有用性分数的缺失;在语义分析中,针对酒店行业关键词众多的特点,本文运用具有分类、归纳功能的 LDA 模型,基于词频与词汇重要性不一致的考虑将 TF-IDF 方法与 LDA 模型相结合,深入挖掘评论的语义内涵。

综上,现有研究对评论有用性影响因素的归纳和总结较为微观、随性,鲜少将这些因素纳入某个理论框架内进行讨论,缺乏整体性。因此,本文以双重过程理论的 HSM 模型为理论基础,以 TripAdvisor 的酒店评论为例,通过多元回归方法,探究多维因素对评论有用性的影响,不仅拓展了酒店评论有用性的现有理论框架,且在实践中通过准确反映顾客酒店体验的真实感受,对酒店未来需提升的维度和发展方向提供指导作用。

一、文献回顾

1. 评论有用性

(1) 评论有用性及测量

学界对于“评论有用性”的界定较为一致。一般来说,得到消费者认可的有价值的评论为有用评论。^[16]由于有用评论可在一定程度上降低认知成本,^[9]近年来有关评论有用性的研究有所增加。评论有用性的测量是目前学界关注的重点之一。现有的测量方式主要有以下几种:评论的有用投票数、^[13,17]评论有用投票数占总投票数的比例、^[16]采用里克特量表进行打分^[14]等。上述测量方式较为直观易行,但精确度不足。为了更精确地测量评论有用性,邱凌云等采用贝叶斯平均分值修正有用性投票比率,^[18]更准确地还原数据集中每条评论有用性的相对高低;Xiang 等利用 TF-IDF 方法和余弦相似度,根据有用评论计算未被评价有用性评论的有用程度,^[12]但未考虑语境语义的影响。因此,本文拟采用基于神经网络语言模型的 Word2Vec 词向量方法,结合余弦相似度计算评论有用性得分,更加准确地测量评论有用性。

(2) 评论有用性的影响因素

评论有用性的影响因素一直是“评论有用性”的研究重点,如旅游产品在线评论有用性、^[19]搜索型商品评论有用性^[20]的影响因素等。近年来,随着技术的发展和评论信息的多元化,关于评论有用性的影响因素研究也变得更加丰富,如探索个体评论与总体评论一致性对评论有用性的影响、^[18]基于人工智能方法分析照片对评论有用性的影响、^[21]研究写作风格对评论有用性的影响^[22]等。

国内外关于评论有用性影响因素的研究可大致分为评论内容特征和评论者特征。评论内容特征包括评论长度、评分等,评论者特征包括评论者的经验、声誉等。^[11,23]对于评论内容特征,现有研究主要集中于两个方向:一是研究评论内容的外在定量属性对评论有用性的影响,如马超等发现图文结合的多模态评论更具有参考价值等,^[19]Racherla 等认为越长的评论有用性越高;^[24]二是研究评论内容的内在定性属性对评论有用性的影响,如评论语义内涵、^[15]评论情感^[25]等。评论者特征是影响评论有用性的另一个重要因素。浏览者在阅读评论时,还希望了解评论者的背景和信誉,可靠评论者发表的评论被认为更具说服力,^[26]如 Liang 等发现获得点赞数越多的评论者所发布的评论更有用。^[11]

综上,现有关于评论有用性影响因素的研究比较丰富,但各影响因素的呈现方式比较松散,鲜有研究对酒店评论有用性提供一个系统的理论模型。另外,目前关于在线评论有用性量化方式的研究比较有限,所以在算法上仍存在很大的改进空间。

2. 双重过程理论 (Dual Process Theory) 与启发式—系统式模型 (HSM)

双重过程理论用于区分不同类型的信息处理过程,^[27,28]解释了信息的内容及其影响因素对信息有用性的影响。^[29]该理论在说服力和态度改变的研究领域最具影响力,^[30]并且对群体意见中的有效沟通有很大作用。^[31]HSM 模型是双重过程理论的典型代表,^[32]Chaiken 等认为这一模型可用于解释个体信息行为过程。^[33]该模型将个体信息行为过程分为启发式行为和系统式行为。启发式行为基于直觉,通常根据信息的外部线索进行简单判断;而系统式行为基于理性,需要利用足够多的认知资源对信息进行评估。^[34]现有文献鲜有基于 HSM 模型对评论有用性进行研究。由于 Zhang 等认为消费者对在线评论的接受是一个包括启发式和系统式两种行为的双重过程,^[35]故本文以 HSM 模型为基础,对评论有用性进行讨论。

基于双重过程理论的 HSM 模型的特点在于,它并

不是由几个特定变量构成的结构固定的理论模型,而是行为决策研究的一般框架和行为范式。^[36]该模型将影响评论有用性的因素分为启发式线索和系统式线索。在本文中,启发式线索主要是评论者个人的可获得非主观信息,包括评论者特征(评论者个人信息和评论者类型)和评论定量特征(如评论长度、评论时间等)。对于系统式线索,通常认为系统式的信息处理需经过仔细观察和认真思考等高认知努力后才可获得,且偏向用评论内容的质量进行界定。^[37]所以,系统式线索是信息中的主观内容,并以客观方式用于系统信息处理。在本文中,系统式线索为评论定性特征(评论内容的语义分析和情感分析)。

因此,本文以双重过程理论的HSM模型为研究框架,参照Liu等将评论内容特征细分为定量特征和定性特征,^[9]总结归纳以往研究,建立如图1的理论模型图,将评论有用性的影响因素分为启发式线索(包括评论者特征、评论定量特征)和系统式线索(包括评论定性特征),为评论有用性在酒店领域提供一个可供参照的理论框架,并通过优化算法在一定程度上提高了测量评论有用性的准确性。

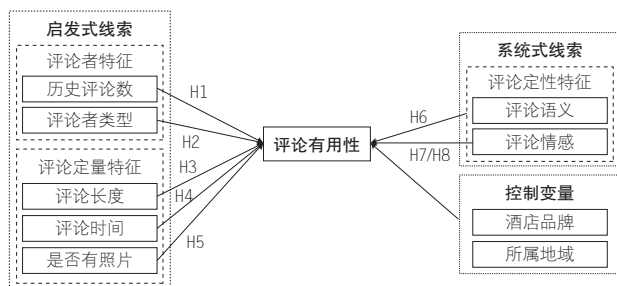


图1 评论有用性影响因素研究模型

二、研究假设

1. 启发式线索

(1) 评论者特征

① 评论者个人信息

TripAdvisor关于评论者的个人信息包括注册时间、所属地域、历史评论数、上传图片总数和各类徽章等。其中,历史评论数指评论者自注册以来发表的评论总数。历史评论数越多,则评论者的等级越高,评论的有用性可能越高。该变量能体现评论者的专业程度且代表性最强,^[11,24,33,38]能够间接影响评论者发表评论的可信度。^[38]因此,本文用历史评论数衡量评论者的专业性。由此,提出以下假设:

H1: 评论者的历史评论数越多,评论有用性越高

② 评论者类型

评论者和浏览者均可在TripAdvisor上选择类型(家庭、情侣、结伴、商务、独自和其他)后进行评论或查看评论。评论者类型对评论模式存在影响,Banerjee等认为商务客人相对其他类型评论者对酒店评分最严格,家庭旅行者偶尔也会有严格评分。^[39]由此,提出以下假设:

H2: 不同评论者类型对评论有用性的影响存在差异

(2) 评论定量特征

① 评论长度

学界关于评论长度对评论有用性的影响持有不同观点。一种观点认为,评论长度对评论有用性有积极影响。Racherla等认为评论长度代表信息量和评论者的严谨性,是影响顾客行为意图的重要因素,对评论有用性有积极影响;^[24]Mudambi等认为长评论可能包含商品的细节信息,如商品如何被购买和使用,有助于提高评论有用性;^[10]Fang等在对景点评论的分析中也获得同样的结论。^[17]但另一种观点认为,评论长度与评论有用性存在非线性关系。尽管丰富的评论信息有助于消费者进行决策,但记忆容量有限,信息处理需求超过上限会导致消费者的认知超载,反而影响决策。^[40]由此,提出以下假设:

H3: 评论长度与评论有用性呈倒U型关系

② 评论时间间隔

评论时间间隔可以用天数、^[41]月数^[42]等来衡量。近期评论包含了最新信息,对浏览者可能更有用。^[42]时间间隔短的评论被认为有较高的时效性,更能反映近期商品的质量。^[20]汪涛等基于归因理论发现,时间间隔较短的体验型商品的评论具有较高的感知有用性。^[41]由于酒店服务属于体验型商品,本文提出以下假设:

H4: 评论时间间隔越长,评论有用性越低

③ 评论是否有图

从评论形式的角度来看,与仅包含文本或仅包含图片的单模态评论相比,图文结合的多模态评论更具有参考价值。^[19]Mudambi等认为,评论深度除了通过评论长度、内容的极端性及表述的感情色彩表现外,还可以借助购得产品的真实图片来强化评论说服力。^[10]Walters等认为,生动的图像刺激对顾客的购买意愿有积极影响。^[43]Xu等认为视觉信息能够吸引注意力,更可信、更具说服力、更有用。^[44]Filieri认为,顾客经常根据视觉信息来判断评论可信度。^[45]由此,提出以下假设:

H5: 含有照片的评论,评论有用性更高

2. 系统式线索

(1) 评论内容的语义分析

评论内容的语义分析方法在不断深化。现有研究主要以评论所含信息量^[15]评论内容深度^[24]等来衡量语义,如潜在语义分析和主成分分析、^[46]基于主题词人工划分主题后建立二元共现语义网络^[47]等。评论的内容和风格是吸引顾客的重要影响因素。由此,提出以下假设:

H6: 评论包含的语义越丰富,评论有用性越高

(2) 评论内容的情感分析

情感分析已成为文本分析的一个重要方面。^[48]现有研究多以评分来量化评论者对商品的情感倾向和体验效果,^[16,20,42]但此类通过单一尺度反映评论者评价的数字评级方法,难以准确衡量评论者的真实情感。

现有研究已转向利用在线软件、基于情感词典或通过机器学习方法进行情感分析。Baek等利用SentiWordNet词典匹配情感词汇,以评论中负面词汇的比例衡量情感,发现负面词汇占比越大,评论有用性越高。^[25]Deng等基于认知评估理论和情绪传染理论建立理论模型,分析了不同情绪对感知评论有用性影响机制的差异。^[22]Xiang等利用朴素贝叶斯方法获得每条评论的情感得分,发现负面评论的有用性更高。^[12]虽然大多数顾客的体验是正面的,但负面的感官体验对顾客评价的影响更大。^[49]除了正负性之外,极端性是对评论情感的另一重要度量。Ghose等以评分衡量情感发现,温和评论(评分为3)的有用性更低。^[13]Forman等也发现,评论情感的极端程度会影响评论有用性。^[38]由此,提出以下假设:

H7: 情感越极端,评论有用性越高

H8: 情感越负面,评论有用性越高

三、研究设计

1. 被解释变量: 评论有用性

如前文所述,直接以有用投票数来衡量有用性并不准确,因为有的评论为浏览者提供了有效信息,却未获得有用投票。本文采用基于Word2Vec的加权相似度算法有效地解决了该问题。具体方法如下:

评论经过数据清洗后,先通过Word2Vec获得每个词语的词向量,再计算每条评论包含词语的平均词向量作为该条评论的评论向量,然后利用评论向量,计算每条评论与所有有用评论的余弦相似度。之所以计算余弦相似度,是因为在本文语料中,少数评论具有较高的有用投票数,多数评论的有用投票数为0。最后,考虑到有用投票数越多的评论对评论相似度的影响应越大,故本文再以有用评论的投票数作为权重,对上一步所计算

的余弦相似度进行加权,计算每条评论与所有有用评论的加权相似度,即每条评论的有用性得分代表评论的有用性。这种算法充分考虑到了浏览者阅读到某条评论的偶然性及个体习惯差异。

Word2Vec包括多种算法。在本文的语料库中,50%以上的词汇为低频词,因此本文选取对低频词更敏感、语义分析效果更好的Skip-Gram算法。

2. 解释变量: 启发式线索、系统式线索

根据前文综述,启发式线索包括评论者特征、评论定量特征,系统式线索包括评论定性特征。为了尽量控制其他因素对评论有用性的影响,本文在模型中增加了酒店品牌、酒店所属地域两个控制变量。

历史评论数(HisRev): 评论者自注册以来发表的评论总数。

评论者类型(Type): 该指标为哑变量,包括情侣游、全家游、结伴旅行、商务行和独自旅行。部分评论者未选择类型,则标记为“其他”。

评论长度(Length): 汉字、字母和数字的总数。

评论时间间隔(Time): 评论发表时间距2020年1月的月数。例如,2020年1月的评论时间间隔均为0,2019年12月的评论时间间隔均为1。

是否有照片(Photo): 该指标为哑变量,若评论含照片取值为1,否则取值为0。

酒店品牌(Hotel): 该指标为哑变量,根据品牌对所有酒店进行分类,^[12]包括万豪、凯悦、希尔顿、洲际、锦江国际、雅高等17个酒店品牌,涵盖66个酒店子品牌。

地域(Area): 该指标为哑变量,根据酒店所属地域分为11类。

评论语义(Semantic): 首先根据困惑度确定所有评论的主题数,再采用基于TF-IDF方法的LDA模型计算每条评论含有每个主题的概率,这些概率即为评论语义的变量值。

其中,(1)确定最佳主题数是本文的关键问题之一。比较粗略的做法是通过主观选择方法即根据经验判断获得。而目前更多采用客观选择方法,即根据若干指标,如困惑度、^[42]对数似然函数、^[20]主题相似度^[43]等来确定。这些指标各有长短: 对数似然函数无法刻画泛化能力,因此会影响预测效果;主题相似度计算方法相对简单,更适合作为约束条件;困惑度适用于无监督学习且应用最广泛。因此,本文参考困惑度指标来确定主题数。(2)高频词通常被认为更重要,但应对常用词和关键词进行区分,例如“的”和“干净”,因此计算TF-IDF得分代替词频。该方法基于词频,以逆文档频率为约束因

子, 计算评论中每个词语的加权重要程度。(3) LDA 模型的分析结果为两个概率矩阵: 评论主题概率矩阵和主题词汇概率矩阵。前者估计了每条评论含有每个主题的概率, 后者估计了每个主题含有每个词汇的概率。

评论情感 (Sentiment): 部分评论同时含有正面情感和负面情感, 因此概率比标签 (0 和 1) 更能体现其中的复杂情感。本文首先人工标注 3075 条评论的情感标签 (0 和 1), 获得 2115 条正面评论和 960 条负面评论, 再采用机器学习方法 (逻辑回归、朴素贝叶斯、随机森林、支持向量机、K 近邻)^① 进行训练, 过程中利用 k 折交叉验证方法, 根据 AUC 值和精确度进行参数调优, 最终获得每条评论含有负面情感的预测概率, 即检验 H8 所用变量值 (Sentiment_neg)。检验 H7 时, 变量值 (Sentiment_bias) 为预测概率与 0.5 差值的绝对值, 绝对值越大则极端程度越大。

对相关算法说明如下:

第一, k 折交叉验证取 k 次训练结果的均值为最终正确率, 可以避免单次训练结果的偶然性, 提升预测模型的泛化能力。

第二, 逻辑回归是一种广义的线性回归模型, 可用于估计样本属于某个类别的概率。

第三, 本文采用连续型和离散型朴素贝叶斯分别进行训练并择优使用。

第四, 随机森林根据信息熵或基尼系数选择出最优特征。其中, 信息熵更适用于连续数据, 基尼系数更适用于离散数据。本文同时采用这两种方法进行训练并择优使用。

第五, 支持向量机是对样本进行分割, 本文选取多种核函数进行训练并择优使用。

第六, K 近邻是一种无监督的分类方法。K 较小时, 容易过拟合; K 较大时, 可降低泛化误差但增加训练误差。本文根据性能优度指标确定 K 值。

综上, 各种变量在 TripAdvisor 上的显示情况, 如图 2 所示。

3. 模型构建

如上所述, 情感极端程度 (Sentiment_bias) 和情感负面程度 (Sentiment_neg) 的定义有重合部分。同时, 这两个变量高度相关。为了避免部分概念重合和高度相关增大估计误差, 本文分别构建模型 (1) 和模型 (2)。

为检验 H7, 建立第一个多元回归方程:

$$\begin{aligned} Usefulness = & \beta_0 + \beta_1 HisRev + \beta_2 Type + \beta_3 Length + \beta_4 Length^2 \\ & + \beta_5 Time + \beta_6 Semantic + \beta_7 Sentiment_bias \\ & + \beta_8 Photo + \beta_9 Hotel + \beta_{10} Area + \epsilon \end{aligned} \quad (1)$$

为检验 H8, 建立第二个多元回归方程:

$$\begin{aligned} Usefulness = & \beta_0 + \beta_1 HisRev + \beta_2 Type + \beta_3 Length + \beta_4 Length^2 \\ & + \beta_5 Time + \beta_6 Semantic + \beta_7 Sentiment_neg \\ & + \beta_8 Photo + \beta_9 Hotel + \beta_{10} Area + \epsilon \end{aligned} \quad (2)$$

此外, 为检验机器学习计算的情感得分与评论评分之间的差异性, 用评论评分替换回归模型中的情感变量, 建立第三个多元回归方程:^②

$$\begin{aligned} Usefulness = & \beta_0 + \beta_1 HisRev + \beta_2 Type + \beta_3 Length + \beta_4 Length^2 \\ & + \beta_5 Time + \beta_6 Semantic + \beta_7 Rating + \beta_8 Photo \\ & + \beta_9 Hotel + \beta_{10} Area + \epsilon \end{aligned} \quad (3)$$



图2 TripAdvisor网站在线酒店评论示意

四、实证分析

1. 数据来源及数据清洗

(1) 数据来源

本文参考 TripAdvisor 发布的中国热门目的地榜单, 以经济中心或旅游胜地为标准, 选取 11 个目标城市; 参考 Xiang 等^[12] 和 Geetha 等^[48] 在研究中选择星级酒店的标准、全球酒店行业权威媒体《HOTELS》杂志 2017 年和 2018 年的全球酒店排行、各酒店品牌在中国的受众度等, 选取 17 个酒店品牌 (涵盖 66 个子品牌)。同时, 考虑到评论时效性, 将时间范围限制于 3 年内。最终获取 TripAdvisor 网站 2017 年 1 月至 2020 年 1 月, 来自 11 个城市、560 个酒店的 67010 条简体中文评论。^③

此外, 本文选取高端酒店作为研究对象。从酒店的角度看, 星级越高服务越全面, 并且更注重提升客户体验, 因此评论内容对酒店提升服务品质非常重要; 从顾客的角度看, 选择星级酒店是为了追求与价格对等的服务和体验, 而评论正是他们获取和反馈信息的最佳途径。

(2) 数据清洗

首先剔除无效评论。原则如下: 评论长度过短或存在凑字数、乱码等问题, 没有分析意义的和非简体中文评论。^④ 其次补充分词词库和停用词文档, 对评论进行

分词、删除停用词。最终获得 63997 条有效评论。相关信息如表 1 所示。

表1 评论的城市分布表

城市	评论	城市	评论	城市	评论	城市	评论
澳门	3725	广州	2938	上海	10018	苏州	4046
北京	14658	杭州	8958	深圳	6033	西安	4759
成都	5640	丽江	132	香港	3090		

2. 描述性统计

(1) 评论者特征

83.2% 的评论者发表过评论，其中 56% 的评论者仅发表过 1 条评论，评论者最多发表过 722 条评论。多数评论者为商务出行（28.8%），其次为结伴旅行（18.4%）、全家游（17.8%），再次为独自旅行（13%）、情侣游（8.2%）。13.8% 的评论者未选择出行类型，考虑到该部分评论占比不低，故未剔除。

(2) 评论定量特征

评分从低至高分别为1、2、3、4、5。96.1%的评论获得4或5, 仅1.9%的评论获得1或2。19%的评论含有照片, 共获得有用投票数2589票, 平均有用投票数为0.21票; 81%的评论不含照片, 共获得有用投票数5803票, 平均有用投票数为0.11票。根据表2所示, 多数评论的长度为50-100个字符(61.2%), 最长评论含733个字符(但该评论并未获得有用投票)。91.3%评论的有用投票数为0, 仅少数评论获得较高票数。由此验证本文的判断: 评论长度与有用投票数不存在正相关。

表2 评论长度和评论有用投票数频数表

评论长度	占比	评论长度	占比	有用投票数	占比	有用投票数	占比
(0,50]	14.0	(150,200]	5.3	0	91.3	5-10	0.2
(50,100]	61.2	(200,250]	2.5	1-2	7.9	10 以上	0.1
(100,150]	13.0	(250, ∞)	4.0	3-4	0.5		

(3) 词云图

根据语料库的词频统计结果,绘制词云图。由图3可见“房间”“入住”“服务”“前台”“热情”“设施”等词频较高,即为客户较关注的方面,如环境、设施、服务、地理位置等。图4为有用评论的词云图,可见有用评论内容涉及酒店各个方面。

3. 计算特征值

(1) 有用性得分

根据模型训练和计算结果,评论有用性的最低得分为 0.36,最高得分为 0.87,平均得分为 0.72。61% 的评论有用性得分介于 0.7-0.8,31.4% 的评论有用性得分介于 0.6-0.7。

(2) 语义得分

首先, 分别利用 LDA 模型和基于 TF-IDF 方法的 LDA 模型训练评论语料, 获得不同主题数下的困惑度, 如图 5 所示。可见当主题数从 5 变为 6 时, 两个模型的困惑度均明显下降, 再考虑实际语境(主题数过多会失去分析意义), 确定主题数为 6。

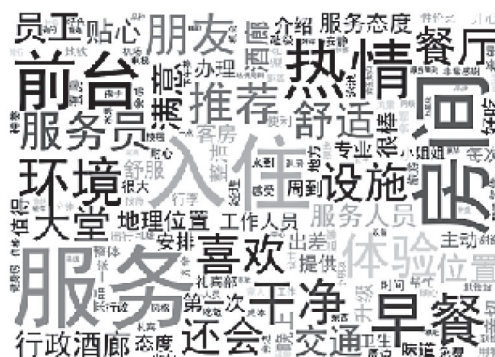


图3 全部评论的词云



图4 有用评论的词云

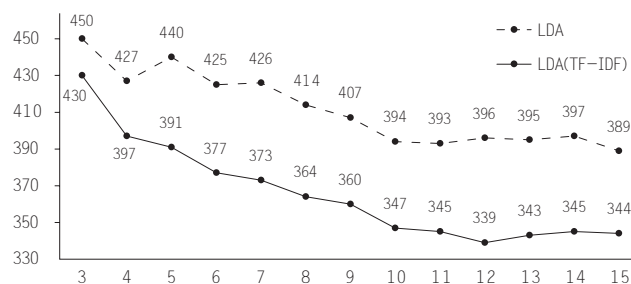


图5 LDA模型和基于TF-IDF方法的LDA模型
在不同主题数下的困惑度对比

其次,按照所确定的主题数训练两个模型,获得每个主题的词汇分布,词汇按出现的后验概率降序排序,如表3所示。与LDA模型比较,基于TF-IDF方法的LDA模型降低了主题间的词汇相似度,主题特色更鲜明。因此,尽管LDA模型的困惑度略优,但本文仍选择后者。

表3 LDA模型和基于TF-IDF方法的LDA模型主题词汇分布结果(斜体为LDA模型)

No	主题1	主题2	主题3	主题4	主题5	主题6
1	房间 0.008	菜品 0.012	健身 0.009	环境优美 0.014	入住 0.012	干净 0.022
	房间 0.043	服务 0.043	入住 0.063	服务 0.044	房间 0.051	房间 0.051
2	套房 0.005	吃 0.012	游泳 0.006	风景 0.011	服务 0.011	早餐 0.017
	入住 0.024	环境 0.012	服务 0.053	环境 0.026	晚上 0.011	干净 0.042
3	入住 0.004	用餐 0.011	舒心 0.006	人性化 0.010	前台 0.010	交通 0.014
	升级 0.015	贴心 0.010	前台 0.040	朋友 0.023	时间 0.009	早餐 0.036
4	升级 0.004	好吃 0.011	健身房 0.006	卫生 0.010	热情 0.010	房间 0.013
	套房 0.013	体验 0.009	热情 0.033	吃 0.022	走 0.007	服务 0.026
5	早餐 0.004	环境 0.011	无比 0.006	舒适 0.010	礼宾部 0.007	整洁 0.013
	早餐 0.013	感受 0.009	房间 0.017	好吃 0.022	想 0.007	交通 0.024
6	会员 0.004	服务员 0.010	客务 0.006	网络 0.010	介绍 0.007	卫生 0.012
	酒廊 0.012	员工 0.008	体验 0.014	餐厅 0.020	孩子 0.007	地理位置 0.020
7	免费 0.003	味道 0.010	泳池 0.006	服务人员 0.010	体验 0.007	便利 0.011
	行政 0.010	喜欢 0.008	办理 0.014	服务员 0.019	做 0.006	热情 0.020
8	晚上 0.003	餐厅 0.009	完美 0.006	乐于 0.009	满意 0.006	环境 0.011
	会员 0.010	行政酒廊 0.008	介绍 0.012	喜欢 0.018	发现 0.006	舒适 0.020
9	很大 0.003	朋友 0.009	布局 0.006	窗外 0.009	第一次 0.006	前台 0.010
	行政酒廊 0.008	一种 0.008	礼宾部 0.011	菜品 0.015	提前 0.005	入住 0.018
10	时间 0.003	喜欢 0.008	顾客 0.006	地理位置 0.009	贴心 0.006	地理位置 0.010
	楼 0.008	专业 0.007	第一次 0.011	热情 0.015	吃 0.005	卫生 0.018

六个主题侧重不同方面,其中主题4和主题6略微重复。主题1包含“升级”“会员”“免费”等词汇,突出会员权益体验效果。不同级别会员可享受差异化权益,如免费升级套房、享受免费早餐或点心等。主题2包含“菜品”“味道”“餐厅”等词汇,突出餐饮水平。酒店行政酒廊、特色餐厅及送餐服务可满足客人不同需求,特色餐饮服务或可成为营销亮点。主题3包含“健身房”“泳池”等词汇,突出设施完善性。差旅、度假等长期住户,尤其关注阅览室、会议室、健身房、游泳池等设施完善程度。主题5包含“服务”“热情”“礼宾部”等词汇,突出人工服务水平。酒店员工的专业素养、仪态谈吐直接影响客人的入住体验,是酒店品牌形象的重要组成部分。主题4和主题6包含“房间”“卫生”“交通”“地理位置”等词汇,突出房间条件和交通便捷性。客人会关注房间软硬件条件,倾向于赞赏外景别致的景观房。同时,地理位置好、交通便捷的酒店更能获得客人青睐。根据六个主题在全部评论中出现的平均概率,评论者最关注酒店人工服务水平(0.21)、房间条件和交通便捷性(0.21),其次是设施完善性(0.15)、会员权益体验效果(0.15),最后是餐饮水平(0.14)。

本文通过相关性分析进一步检验主题相似度。结果显示,主题4和主题6弱相关,其他主题间均不相关,可表明主题相似程度低,有助于避免回归过程中的多重共线性。

(3) 情感得分

机器学习方法的训练结果见表4,本文选取性能较优的三个分类器进行情感分析。

表4 五种机器学习方法的性能排行

分类器	AUC	精确度	排名	分类器	AUC	精确度	排名
逻辑回归	0.995	0.966	1	随机森林	0.970	0.905	4
离散型朴素贝叶斯	0.992	0.956	2	K近邻	0.925	0.836	5
支持向量机	0.989	0.944	3				

根据Bagging思想,应用多种分类器时,可以对各个分类器的结果取均值或投票。因此,本文先取三个分类器的预测概率均值,再以0.5为阈值划分正负性,作为均值结果;先以0.5为阈值,获得三个分类器对每条评论的正负性预测结果,再依照投票思想决定最终正负性,作为投票结果。具体分类情况如图6所示。

图6表明,三个分类器的结果存在较大差异,为减小偶然性、过拟合、欠拟合等问题导致的误差,本文采用均值结果。图7显示,5.1%的评论趋于中性,6.2%的评论偏负面。

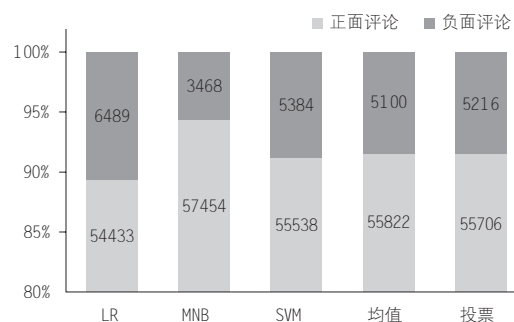


图6 机器学习方法的预测结果

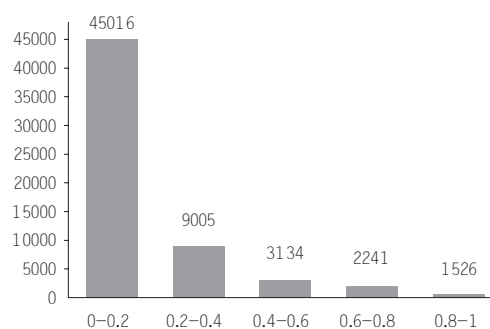


图7 评论负面情感预测概率的频数分布

4. 回归结果

为避免完全多重共线性,剔除与主题4低度相关且平均概率最小的主题6后进行回归。

第一个方程(H7)、第二个方程(H8)的调整R方分别为0.491、0.488。由表5可知,评论者的历史评论数越多,评论越有用,接受H1;不同评论者类型对评论有用性的影响存在差异,接受H2;评论长度与有用性存在倒U型关系,即随着评论长度的增加,有用性先增后减,接受H3;时间间隔越长,评论越有用,不接受H4;含有照片的评论更有用,接受H5;所有主题均对有用性有正面影响,接受H6;情感越极端,评论有用性越低,不接受H8;情感越负面,评论越有用,接受H9。此外,部分酒店品牌和酒店所属地域影响显著。

值得注意的是,评论包含任意主题都能提高有用性,各主题对有用性的影响程度从大到小依次是主题2(餐饮水平)、主题5(人工服务水平)、主题4(房间条件和交通便利性)、主题3(设施完善性)、主题1(会员权益体验效果)。

表5 多元回归方程结果

变量	H7	H8	Rating	变量	H7	H8	Rating
Constant	0.5585***	0.5457***	0.5364***	Topic1	0.0813***	0.0817***	0.0818***
HisRev	0.0000***	0.0000***	0.0000***	Topic2	0.1120***	0.1175***	0.1206***
Length	0.0031***	0.0032***	0.0032***	Topic3	0.0837***	0.0885***	0.0916***
Length_sq	-0.0000***	-0.0000***	-0.0000***	Topic4	0.0912***	0.0924***	0.0927***
Time	0.0004***	0.0004***	0.0004***	Topic5	0.0997***	0.1003***	0.1005***
Type_family	0.0016**	0.0011*	0.0005	Bias	-0.0309***	—	—
Type_business	0.0004***	0.0000	-0.0006	Neg	—	0.0046***	—
Type_lovers	-0.0000	-0.0005	-0.0011***	Rating	—	—	0.0021***
Type_oneself	-0.0009	-0.0015**	-0.0022***	Hotel	***	***	***
Type_friends	-0.0048***	-0.0055***	-0.0062***	Area	***	***	***
Photo	0.0025***	0.0028***	0.0027***				

注:*, **, *** 分别表示在0.1、0.05、0.01水平上显著

为进一步检验H2,本文采用方差分析进行多重比较,结果见表6。可见部分评论者类型之间存在显著不同,如“全家游”与“商务行”、“全家游”与“结伴旅行”。

表6 方差分析的多重比较结果

组1	组2	Reject	组1	组2	Reject	组1	组2	Reject
1 全家游	其他	True	6 其他	商务行	True	11 情侣游	独自	False
2 全家游	商务行	True	7 其他	情侣游	True	12 情侣游	结伴	True
3 全家游	情侣游	False	8 其他	独自	True	13 商务行	独自	False
4 全家游	独自	True	9 其他	结伴	True	14 商务行	结伴	True
5 全家游	结伴	True	10 情侣游	商务行	True	15 独自	结伴	True

第三个方程(Rating)的调整R方为0.488。用评论评分替换情感变量后,估计系数和显著性发生不同程

度变化,且结论与第一、二个方程均矛盾。如图6所示,本文情感分析预测负面评论(5100条)占8.4%,而评分为3以下的评论仅占1.9%,说明情感和评分之间存在一定差异。本文采用的情感分析方法能更准确地识别评论中的负面情绪,衡量评论真实情感。

五、研究结论与建议

1. 研究结论

本文以双重过程理论的HSM模型为研究框架,将评论有用性的影响因素分为启发式线索(包括评论者特征、评论定量特征)和系统式线索(包括评论定性特征),建立了评论有用性在酒店领域一个可供参照的理论框架,探究多维因素对评论有用性的影响,得到如下结论:

(1)在双重过程理论的启发式线索中,评论者的历史评论数对评论有用性存在正面影响,表明浏览者在判断评论可信度时,会参考评论者的专业程度和可信度。不同评论者类型对有用性的影响程度存在差异,如“全家游”与“商务行”对有用性的影响存在显著差异,可能是因为“全家游”与“商务行”的评论者对酒店服务的关注点不同,且评论风格也存在明显差异。需特别指出的影响因素是评论长度。随着评论长度的增加,有用性先增后减。因为人们通常希望快速获取有效信息,但短评所含信息有限,长评则需耗时浏览。该结论扩展了现有研究关于评论长度对有用性影响机理解释。此外,时间间隔越长,评论越有用。因为时间积累有助于有用投票数的积累。含有照片的评论更有用,由于照片可以更直观地传递信息,且含图评论的比重越来越大,未来可以考虑利用图像识别等技术深入探讨图片信息对有用性的影响机理。评论者的历史评论数、评论者类型、评论长度、评论时间间隔和是否有照片等因素经常触发启发式线索,证明了有用性评分中存在双重过程理论中的启发式过程。

(2)在双重过程理论的系统性线索中,情感极端程度存在负面影响,但情感负面程度存在正面影响。因此,评论有用性评分中存在双重过程理论中的系统式过程,且评论情感为系统性线索的触发因素。这是因为人们更倾向在公开平台发表负面评论以揭示商家的不良行为或自己的不幸遭遇,从而提醒他人避免重蹈覆辙。^[50]这在一定程度上验证了现有研究认为负面评论比正面评论更有价值的观点。同时,这也符合现实场景,即在大量正面评论和少量负面评论面前,人们更容易信任后者。另外,本研究利用机器学习方法计算的情感得分与评论评分之间

存在差异,经检验前者能更准确地衡量评论的实际情感。

(3) 回归结果证明,大多数启发式线索对评论有用性有显著影响,浏览者进行评论有用性决策时偏向于启发式过程。同时,系统式线索对于评论有用性的影响也值得重视。因此,构建双重过程的启发式和系统式线索,对评论有用性影响因素的研究非常重要。一方面,启发式线索的构建,包括提供评论者类型标签、建议适当的评论长度等,有助于降低浏览者的研究成本,从而触发启发式过程;另一方面,由于系统式线索可增加浏览者决策的精准性,所以需关注具有负面情感的评论,防止系统式线索的触发,避免带来负面效应。

2. 理论贡献

(1) 本文基于双重过程理论,以HSM模型为研究框架,研究启发式和系统式线索对评论有用性的影响,为酒店评论有用性探索了一个可供参照的理论视角。HSM模型源自双重过程理论,用于解释个体决策行为,而评论有用性问题是该理论在互联网情景中的具体研究切入点。现有研究鲜少对酒店评论有用性建立一个系统的理论模型,本文根据该理论的范式要求,以启发式线索和系统式线索为统筹,以整体的视角关注在线评论有用性的影响因素,不仅丰富了双重过程理论的应用情景,还拓展了酒店评论有用性的现有理论框架。

(2) 本文采用基于神经网络语言模型的Word2Vec的加权相似度算法,根据有用评论计算零有用票数评论的有用得分,优化了现有研究关于评论有用性的计算方法。该算法结合了TF-IDF方法,在考虑语境语义的情况下,训练得到词向量和评论向量。同时,以有用评论的投票数作为权重,对每条评论与有用评论的余弦相似度进行加权,获得加权相似度,即有用性得分。该算法充分考虑到投票数越多的有用评论对相似度的影响越大,并考虑了浏览者阅读到某条评论的偶然性及个体习惯差异,能够有效减小现有研究对有用性的测量误差,优化评论有用性量化方法。

(3) 本文在评论有用性研究中深化了现有研究对在线评论的语义分析和情感分析。基于TF-IDF的LDA主题模型,从评论主题角度深度挖掘评论语义,拓展了酒店评论语义分析的研究视角;采用多种机器学习方法分析评论情感,对比情感分析得分与在线评分两种测量方法的差异,在一定程度上提高了现有研究对酒店评论情感分析的精度,对评论者在酒店评论中情感的真实表达有了更加准确的预测。

3. 管理建议

本文对在线旅游平台和酒店的长远发展有一定的借鉴意义,对其管理建议主要体现在如下两个方面:

(1) 对在线评论平台的建议

首先,优先推荐优质评论。很多在线旅游平台,如TripAdvisor、携程等,大多按照时间顺序,由近及远展示评论。由于大部分浏览者不会看完所有评论,因此将优质评论从大量评论中提取出来优先推送给浏览者,有助于提高网站的信息利用效率,节约浏览者的时间成本。很多在线平台通过赋予评论者荣誉勋章等做法来肯定评论者的专业程度和信誉,并以此激励评论者。因此未来在线评论平台可参考评论者的勋章等级、评论的有用投票数等指标,推送优质评论者的评论或优质评论。

其次,加强信息种类自主选择功能。如今很多在线评论平台为浏览者提供语言分类、评论者类型、评分类型和时节分类的选择,除此之外,还提供评论长度、评论关键字和是否有照片的选择,辅助浏览者进行决策。第一,若浏览者只想阅读长度为150字以内的评论,则经过筛选后只展示符合条件的评论。第二,在线旅游平台除了提供酒店信息之外,还有很多如餐厅、旅游景点等其他浏览者需要的内容,因此可以针对不同的评论对象提取对应关键字,如酒店的关键字可以是“床”“吃”“交通”“泳池”等,餐厅的关键字可以是“味道”“特色菜”“卫生”“快慢”等,关键字还可以包含评论对象的特征或地方特色,让浏览者能够快速筛选出感兴趣的内容。第三,很多在线旅游平台已经通过提取评论中的所有照片形成照片库,以辅助浏览者快速查看评论照片,同时还可以让浏览者选择优先查看含有照片的评论,图文结合更有利于信息传递。

最后,优化移动端页面结构。手机、Pad等的普及使移动端的阅读体验越来越重要。以TripAdvisor平台为例,浏览者可在电脑端看到评论者头像下的历史评论数,而移动端受限于屏幕尺寸无法显示这些数据。此外,页面信息量过多会挤占屏幕空间,使视觉感知较为拥挤,页面信息量少则不利于有需求的浏览者快速获取信息。因此,建议在线评论平台优化移动端页面结构,如根据呈现信息种类的不同而灵活调整,兼顾信息容量和视觉感知等。

(2) 对酒店的建议

首先,关注评论照片,提升细节品质。评论照片不仅能够为浏览者提供图片信息,更能够增加评论内容的真实性。通过评论照片,可直接看到酒店产品和服务中被肯定或否定的方面,进而采取具有针对性的改进措施。

照片所反映的问题可能为细节问题,非亲身入住体验不会发现,因此酒店应重视评论中的图片内容,进而提升酒店产品质量和服务品质。

其次,把控服务质量,重客人之重。本文在语义分析结果中发现,客人通常会关注酒店人员的服务水平、餐食供应、房间条件、交通便利程度等方面,由于各个酒店的具体情况不同,客人所关注的方面也会存在差异,且不同类型的客人侧重点不同。因此,各酒店可以对获得的评论有针对性地分析,寻找客人的关注点,有重点地提升酒店服务质量,找到适合酒店自身的经营策略。

4. 研究不足与展望

虽然本文结合了多种研究方法对评论有用性的影响因素进行分析,并取得一定成果,但仍然存在一些不足,希望能够在未来研究中得到提升:

(1) 词库有待优化。本文对酒店评论进行分析,评论内容涉及酒店行业的专有名词(如行政酒廊、酒店餐厅的名字等)、酒店所在地的专有名词(包括地方特色小吃、风景点名称及地标性建筑的名字等)等词汇。由于Python内含词库的词汇量有限,因此在分词过程中,需根据分词情况补充词汇并加载至分词程序中,从而不断改进分词结果,完善词库词汇。随着学者们对酒店评论的深入研究,未来研究可进一步完善酒店行业词库,优化酒店评论的分词效果。

(2) 语义分析方法有待优化。基于TF-IDF方法的LDA模型解决了词频权重的问题,而在评论分词结果中,仍存在大量近义词。而且汉语语义丰富、表达多样,对相似度高的词语进行简单替换可能会改变语义,且用于替换的相似度临界值与误差大小相关。因此在未来的研究中,可着眼于识别相同或不同的词语在不同语境下的语义相似程度,进一步发展近义词识别方法,从而改进语义模型。

(3) 本文虽然较好地应用了双重过程理论的HSM模型为研究框架,将评论有用性的影响因素分为启发式线索和系统式线索,但是在具体每种线索的特征选择上还没有做到面面俱到,如在系统式线索中,由于技术难度等原因,只选择了评论的文字定性特征。在未来的研究中,可拓展HSM模型启发式线索和系统式线索的具体特征指标,如在系统式线索中增添图片分析等内容,深化双重过程理论在评论有用性等相关研究领域的应用。

参考文献

- [1] Ye, Q., Law, R., Gu, B., Chen, W.. The Influence of User-generated Content on Traveler Behavior: An Empirical Investigation on the Effects of E-word-of-mouth to Hotel Online Bookings. *Computers in Human Behavior*, 2011, 27(2): 634-639.
- [2] Sparks, B. A., Browning, V.. The Impact of Online Reviews on Hotel Booking Intentions and Perception of Trust. *Tourism Management*, 2011, 32(6): 1310-1323.
- [3] Purnawirawan, N., Eisend, M., Pelsmacker, P. D., Dens, N.. A Meta-analytic Investigation of the Role of Valence in Online Reviews. *Journal of Interactive Marketing*, 2015, (31): 17-27.
- [4] Torres, E. N., Singh, D., Robertson-Ring, A.. Consumer Reviews and the Creation of Booking Transaction Value: Lessons from the Hotel Industry. *International Journal of Hospitality Management*, 2015, (50): 77-83.
- [5] Nieto-García, M., Muñoz-Gallego, P. A., Óscar González-Benito. Tourists' Willingness to Pay for an Accommodation: The Effect of EWOM and Internal Reference Price. *International Journal of Hospitality Management*, 2017, (62): 67-77.
- [6] Duverger, P.. Curvilinear Effects of User-generated Content on Hotels' Market Share: A Dynamic Panel-data Analysis. *Journal of Travel Research*, 2013, 52(4): 465-478.
- [7] Hu, N., Zhang, T., Gao, B., Bose, I.. What Do Hotel Customers Complain about? Text Analysis Using Structural Topic Model. *Tourism Management*, 2019, (72): 417-426.
- [8] Frías, D. M., Rodríguez, M. A., Castañeda, J. A.. Internet vs. Travel Agencies on Pre-Visit Destination Image Formation: An Information Processing View. *Tourism Management*, 2008, 29(1): 163-179.
- [9] Liu, Z., Park, S.. What Makes a Useful Online Review? Implication for Travel Product Websites. *Tourism Management*, 2015, (47): 140-151.
- [10] Mudambi, S. M., Schuff, D.. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon. com. *MIS Quarterly*, 2010, 34(1): 185-200.
- [11] Liang, S., Schuckert, M., Law, R.. How to Improve the Stated Helpfulness of Hotel Reviews? A Multilevel Approach. *International Journal of Contemporary Hospitality Management*, 2019, 31(2): 953-977.
- [12] Xiang, Z., Du, Q., Ma, Y.. A Comparative Analysis of Major Online Review Platforms: Implications for Social Media Analytics in Hospitality and Tourism. *Tourism Management*, 2017, (58): 51-65.
- [13] Ghose, A., Ipeirotis, P. G.. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *Social Science Electronic Publishing*, 2011, 23(10): 1498-1512.
- [14] Casaló, L. V., Flavián, C., Guinalíu, M., Ekincic, Y.. Avoiding the Dark Side of Positive Online Consumer Reviews: Enhancing Reviews' Usefulness for High Risk-averse Travelers. *Journal of Business Research*, 2015, 68(9): 1829-1835.
- [15] McKinney, V., Yoon, K., Zahedi, F. M.. The Measurement of Web-customer Satisfaction: An Expectation and Disconfirmation

- tion Approach. *Information Systems Research*, 2002, 13(3): 296-315.
- [16] 王智生, 李慧颖, 孙锐. 在线评论有用性投票的影响因素研究——基于商品类型的调节作用. *管理评论*, 2016, 28(7): 143-153.
- [17] Fang, B., Ye, Q., Kucukusta, D.. Analysis of the Perceived Value of Online Tourism Reviews: Influence of Readability and Reviewer Characteristics. *Tourism Management*, 2016, (52): 498-506.
- [18] 邱凌云, 肖娴, 庞隽. 个体评论与总体评分一致性对评论有用性的影响. *南开管理评论*, 2019, 22(6): 200-210.
- [19] 马超, 李纲, 陈思菁, 毛进, 张霁. 基于多模态数据语义融合的旅游在线评论有用性识别研究. *情报学报*, 2020, (2): 199-207.
- [20] 何有世, 李娜. 搜索型商品评论有用性影响因素研究. *情报杂志*, 2016, 35(12): 202-206, 201.
- [21] Ma, Y., Xiang, Z., Du, Q., Fan, W.. Effects of User-provided Photos on Hotel Review Helpfulness: An Analytical Approach with Deep Learning. *International Journal of Hospitality Management*, 2018, (71): 120-131.
- [22] Deng, C., Ravichandran, T.. Positivity Bias: Effects of Discrete Emotions on Review Helpfulness. *Twenty-fourth Americas Conference on Information Systems*, New Orleans, 2018.
- [23] Park, S., Nicolau, J. L.. Asymmetric Effects of Online Consumer Reviews. *Annals of Tourism Research*, 2015, (50): 67-83.
- [24] Racherla, P., Friske, W.. Perceived "Usefulness" of Online Consumer Reviews: An Exploratory Investigation across Three Services Categories. *Electronic Commerce Research & Applications*, 2012, 11(6): 548-559.
- [25] Baek, H., Ahn, J., Choi, Y.. Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues. *International Journal of Electronic Commerce*, 2012, 17(2): 99-126.
- [26] Sparks, B. A., Perkins, H. E., Buckley, R.. Online Travel Reviews as Persuasive Communication: The Effects of Content Type, Source, and Certification Logos on Consumer Behavior. *Tourism Management*, 2013, (39): 1-9.
- [27] Higgins, E. T.. Promotion and Prevention as a Motivational Duality: Implications for Evaluative Processes. *Dual-process Theories in Social Psychology*. New York: Guilford Press, 1999: 503-525.
- [28] 田虹, 袁海霞. 企业社会责任匹配性何时对消费者品牌态度更重要——影响消费者归因的边界条件研究. *南开管理评论*, 2013, 16(3): 101-108.
- [29] Zhang, W., Watts, S.. Capitalizing on Content: Information Adoption in Two Online Communities. *Journal of the Association for Information Systems*, 2008, 9(2): 73-94.
- [30] Smith, E. R., DeCoster, J.. Dual-process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review*, 2000, 4(2): 108-131.
- [31] Cheung, M. Y., Luo, C., Chen, S. H.. Credibility of Electronic Word-of-mouth: Informational and Normative Determinants of On-line Consumer Recommendations. *International Journal of Electronic Commerce*, 2009, 13(4): 9-38.
- [32] Chu, S. C., Kamal, S.. The Effect of Perceived Blogger Credibility and Argument Quality on Message Elaboration and Brand Attitudes. *Journal of Interactive Advertising*, 2008, 8(2): 26-37.
- [33] Chaiken, S.. Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology*, 1980, 39(5): 752-766.
- [34] Chen, S., Chaiken, S.. The Heuristic-systematic Model in Its Broader Context. *Dual-process Theories in Social Psychology*. New York: Guilford Press, 1999: 73-96.
- [35] Zhang, K. Z. K., Zhao, S. J., Cheung, C. M. K., Lee, M. K. O.. Examining the Influence of Online Reviews on Consumers' Decision-making: A Heuristic-systematic Model. *Decision Support Systems*, 2014, 67(C): 78-89.
- [36] 陈明红, 刘莹, 漆贤军. 学术虚拟社区持续知识共享意愿研究——启发式—系统式模式的视角. *图书馆论坛*, 2015, (11): 83-91.
- [37] Zhang, W., Watts, S. A.. Capitalizing on Content: Information Adoption in Two Online Communities. *Journal of the Association for Information Systems*, 2008, 9(2): 73-94.
- [38] Forman, C., Ghose, A., Wiesenfeld, B.. Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*, 2008, 19(3): 291-313.
- [39] Banerjee, S., Chua, A. Y. K.. In Search of Patterns among Travellers' Hotel Ratings in TripAdvisor. *Tourism Management*, 2016, (53): 125-131.
- [40] 殷国鹏. 消费者认为怎样的在线评论更有用——社会性因素的影响效应. *管理世界*, 2012, (12): 115-124.
- [41] 汪涛, 王魁, 陈厚. 时间间隔何时能够提高在线评论的有用性感知——基于归因理论的视角. *商业经济与管理*, 2015, (2): 46-56.
- [42] 江晓东. 什么样的产品评论最有用——在线评论数量特征和文本特征对其有用性的影响研究. *外国经济与管理*, 2015, 37(4): 41-55.
- [43] Walters, G., Sparks, B., Herington, C.. The Effectiveness of Print Advertising Stimuli in Evoking Elaborate Consumption Visions for Potential Travelers. *Journal of Travel Research*, 2007, 46(1): 24-34.
- [44] Xu, P., Chen, L., Santhanam, R.. Will Video be the Next Generation of E-commerce Product Reviews? Presentation Format and the Role of Product Type. *Decision Support Systems*, 2015, (73): 85-96.
- [45] Filieri, R.. What Makes an Online Consumer Review Trustworthy? *Annals of Tourism Research*, 2016, (58): 46-64.
- [46] 陈江涛, 张金隆, 张亚军. 在线商品评论有用性影响因素研究: 基于文本语义视角. *图书情报工作*, 2012, 56(10): 119-123.
- [47] Hou, Z., Cui, F., Meng, Y., Lian, T., Yu, C.. Opinion Mining from Online Travel Reviews: A Comparative Analysis of Chi-

- nese Major OTAs Using Semantic Association Analysis (Article). *Tourism Management*, 2019, (74): 276-289.
- [48] 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.
- [49] Mehraliyev, F., Kirilenko, A. P., Choi, Y.. From Measurement Scale to Sentiment Scale: Examining the Effect of Sensory Experiences on Online Review Rating Behavior. *Tourism Management*, 2020, (79): 104096.
- [50] Qiang, Y., Zhou, S., Wu, S.. The Influences of Tourists' Emotions on the Selection of Electronic Word of Mouth Platforms. *Tourism Management*, 2018, (66): 348-363.

注释

- ① 在情感分析中, Fang 等采用支持向量机、朴素贝叶斯、随机森林方法,^[17]Geetha 等^[48]和 Xiang 等^[12]采用朴素贝叶斯方法。本文则尝试了更多机器学习方法,并择优列示于文中。
- ② 多数评论的有用投票数为 0, 较适合 Tobit 回归模型,^[44]因此本文没有使用有用投票数替换有用性得分,作为被解释变量进行回归,以检验两者之间的差异性。
- ③ 本文仅参考中文评论。一是因为选取的酒店都在中国境内,入住客人以中国人居多,且中国客人相对具有更多境内酒店的入住体验,因此首选中文评论能够提供更具有参考性的评价。二是因为不同语言存在表达差异。若分析外语原文,仍需分开处理,无法统一归纳;若分析译文,则容易在翻译过程中产生疑义。
- ④ 本文样本含有少量繁体中文评论和其他语言评论,所用停用词文档针对简体中文,因此在数据预处理过程中,剔除非简体中文评论。

Deep Data Mining for Online Reviews Usefulness: Hotel Reviews Data on TripAdvisor

Shi Da^{1,2}, Wang Lele³, Yi Bowen¹

1.School of Tourism and Hospitality Management, Dongbei University of Finance and Economics; 2.Surrey International Institute, Dongbei University of Finance and Economics; 3.International Business College, Dongbei University of Finance and Economics

Abstract The usefulness of online reviews can help customers get effective information quickly and affect their purchase decision. Based on heuristic-systematic model (HSM) of dual process theory, this paper collects Chinese online reviews of hotels from TripAdvisor as raw data and measures the explanatory variables and interpreted variables using tools like word2vec, LDA and machine learning algorithms, which extends the previous quantitative research on the online review usefulness. Meanwhile, it constructs multiple regression models to analyze the different influencing factors of the online review usefulness. The results show that, (1) based on HSM of the dual process theory, the paper establishes a theoretical model of the influencing factors of the review usefulness, and divides the

influencing factors of review usefulness into peripheral cues and systematic cues, which extends the existing theoretical framework of the review usefulness in hospitality. Among peripheral cues, review length, review time, photo and the type of reviewer have significant impacts on review usefulness. Moreover, there is an inverted U-shape relationship between review length and review usefulness; review time and photo also have significantly positive impacts on review usefulness while different types of reviewers have different impacts. In terms of systematic cues, all the review topics and negative reviews have significantly positive impact on review usefulness. (2) The paper uses word2vec method to calculate the usefulness scores of non-vote online reviews based on the reviews with useful votes, which improves the quantification of review usefulness quantification. In addition, it uses three machine learning methods (LR, NB, SVM) with better performance in sentiment analysis, and proves that the machine learning methods perform better to measure the review sentiment compared to use review ratings directly. This paper gives a more accurate explanation for the hotel experience of customers, which has great significance and practical implications for the development of online review platform and hotel in the future.

Key Words Review Usefulness; HSM; Semantics; Sentiment; Machine Learning