

周倩,姜磊,程旅航,等.情感倾向驱动的旅游酒店大数据网络舆情[J].湖南科技大学学报(自然科学版),2020,35(4):67-73. doi:10.13582/j.cnki.1672-9102.2020.04.010

Zhou Q, Jiang L, Cheng L H, et al. A Sentiment Analysis-driven Public Opinion Study for Travel Hotel Reviews [J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2020,35(4):67-73. doi:10.13582/j.cnki.1672-9102.2020.04.010

# 情感倾向驱动的旅游酒店大数据网络舆情

周倩,姜磊\*,程旅航,陈焱,陈艳红

(湖南科技大学 知识处理与网络化制造实验室,湖南湘潭 411201)

**摘要:**从网络舆情中获取评价以支持决策是旅游管理部门和企业关注的重要问题.为此,本文提出情感倾向计算的新方法以便于网络舆情的理解与分析.首先构建反映酒店评论语义特征的情感词库,然后从句法特征的角度对情感词进行扩充,最后采用深度学习得到情感倾向的分类结果.通过湘潭市旅游酒店评论的舆情分析实验,发现该方法能够得到更为真实的消费者对旅游酒店的情感倾向.从3个竞争酒店评论的舆情分析应用实例上可以得出:情感倾向能够挖掘出酒店的各自特色与管理上的短板,从而进行决策支持.

**关键词:**网络舆情;酒店评论;情感倾向计算模型;实例分析与应用

**中图分类号:**C912.63; G350

**文献标志码:**A

**文章编号:**1672-9102(2020)04-0067-07

## A Sentiment Analysis-driven Public Opinion Study for Travel Hotel Reviews

Zhou Qian, Jiang Lei, Cheng Lyuhang, Chen Yan, Chen Yanhong

(Key Laboratory of Knowledge Processing and Networked Manufacture, Hunan University of Science and Technology, Xiangtan 411201, China)

**Abstract:** Obtaining evaluations from online public opinion to support decision-making is an important issue for government and enterprises. For this purpose, a new method of sentiment analysis to analysis online public opinion was proposed. The emotional space model of hotel reviews for extracting features from semantic perspective, was constructed, and then the syntactic elements and network new words in the hotel domain were used as supplementary lexicons. Furthermore, deep learning was used to perform emotional classification. Through sentiment calculations of hotel reviews in Xiangtan, it could be found that this method we used could get more real view of consumer. From the application examples of the three competing hotel reviews, it could be concluded that the sentiment analysis method found useful information to support decision making.

**Keywords:** public opinion; hotel review; sentiment analysis model; case study

人们对具体事情或话题的情感倾向能够帮助政府部门处理突发舆情事件<sup>[1-2]</sup>,也能够辅助企业进行商务决策<sup>[3-4]</sup>,甚至可以用于进行股市走势的判断<sup>[5-6]</sup>,因此,网络舆情的情感倾向分析研究越来越受到研究者的关注.酒店作为旅游景区的重要基础设施,它的运营状况、服务水平对景区的口碑有着非常重要的

收稿日期:2019-01-03

基金项目:教育部人文社科规划基金资助项目(17YJAZH032);湖南省教育厅重点资助项目(15A064);湖南省自然科学基金资助项目(2016JJ20156)

\*通信作者,E-mail: jleihn@126.com

影响。2018年11月爆出的20多家知名五星级酒店的卫生问题,说明酒店在管理的过程中,针对员工工作质量的日常检查,很难做到全方位、全天候的覆盖。而作为酒店服务对象——顾客,他们的体验与评论可以帮助管理者全面迅捷地得到酒店服务的真实质量。采用情感倾向计算方法对酒店评论进行自动理解是得到酒店舆情的有效方法。当前的情感倾向计算主要有3种:基于语义规则、基于机器学习以及基于话题模型的情感分析方法。基于语义规则的情感分析方法在情感词典<sup>[7-9]</sup>的基础上,通过语义规则计算评价词与情感词的距离来进行情感分类。但是,单纯从情感词典出发,对于一些特殊句式可能会造成误判,所以许多研究者将句法<sup>[10-12]</sup>也融入到情感倾向性分析中,也有部分研究者通过构建领域情感词典<sup>[13-15]</sup>来进行解决。基于机器学习的情感分析模型<sup>[16-18]</sup>主要利用机器学习中的各种分类方法,通过事先搭好标签的训练集进行训练得到分类模型,然后对新的网络文本进行情感分析。基于话题模型的方法,如使用LDA<sup>[19]</sup>主题模型,依据指定的情感单元搭配模式,提取情感信息,融入到情感向量空间,从而识别出所谈论的话题以及情感倾向<sup>[20-21]</sup>。

上述3种方法的研究都取得了很好的成果,但是在网络文本通常存在大量错别字的情况下以及面对特定领域的情形下,上述方法表现一般。因此,设计一个融合上述方法优点并对错别字具有容错能力的方法应该具有相当的价值。基于这个想法,本文利用网络爬虫获取旅游景区周边酒店的评论,构建了酒店评论的语义特征情感词库,然后采用深度学习进行情感倾向计算。通过对湘潭地区酒店评论的情感倾向计算实验,发现我们的方法得到的结果更准确。在3个具有竞争关系的酒店舆情分析应用实例中,发现情感倾向驱动的酒店评论舆情分析能很好地评价管理水平,并能指导管理者找出工作的改进方向。

## 1 旅游酒店的情感倾向计算模型

旅游酒店的网络舆情主要来自于各旅游网站上以文本形式存在的酒店评论,它们的情感倾向计算主要分为以下2个步骤:(1)语义特征情感词典的构建;(2)情感的分类。

### 1.1 构建语义特征情感词典

理解酒店评论所体现的舆情可以通过计算其在各个可能涉及的语义维度的情感倾向得到。因此,我们需要将酒店评论所能涉及各个语义维度提取出来作为情感特征,然后在每个特征中构建相应的情感词库以方便情感倾向的分类计算。

#### 1) 酒店评论语义特征构成

为了表征酒店评论这一特定话题,我们通过爬虫从目前国内TOP9的旅游网站上抓取了85万条酒店评论,再应用LDA主题模型进行了话题的分类,然后将酒店评论进一步凝练为10个语义特征<sup>[22]</sup>:整体情感、情绪、服务、卫生、环境、饭菜、交通、价格、硬件和房间整体。即通过我们的研究,认为一条有效的酒店评论只能是就这10个语义的全部或某些部分进行阐述。

#### 2) 酒店评论各语义维度的情感词抽取

对于语义的理解,首先可以想到对于某一特征如“卫生”,如果顾客对它的情感倾向是正面的,显然可以理解为顾客对卫生状况是满意的。情感倾向越正面,意味着卫生状况越好。但是只用如“快乐、兴奋、沮丧、伤心、愤怒”等情感词在酒店评论的语境中是远远不够的,因为有些词虽然没有明显的喜怒哀乐倾向,但是仍然可以影响情感分析结果。例如,“这个酒店的床翻个身都能听到咯吱咯吱响。”这句话就具有负面情绪。在这里“咯吱咯吱响”并不是情感词也不是负面词,但在这个语境中就体现出了情感特征。针对酒店评论而言,需要将这些非情感词纳入到相应的语义特征中。同时,网络中写作比较随意,存在大量错别字与不规范用语,而且写作的人数众多,如“狠差”“棒呆”等,在计算情感倾向时,也必须采用,因此我们也将它们纳入到情感词库中,这样共得到酒店领域情感词汇4 031个(表1)。

#### 3) 反映句法特征的酒店评论情感词

采用情感词对文本进行计算时,对于一些复杂语句难以准确判定其情感倾向,所以一般会采用句法语义来提高准确率。但是在网络情境下,许多网络语言不一定符合语法。这样采用句法语义需要对原有的句法进行容错性扩展,相当麻烦。考虑到酒店评论基本是就服务、交通、环境等方面展开评论,那些包含情感

倾向的语句并不特别复杂,主要为程度不等的否定词.因此本文将酒店评论中那些单纯采用情感词会引起歧义的1362个短语放入相应情感特征补充词库(表2).

表1 酒店评论语义特征情感词典(部分)

特征	极性	情感词(部分)					总数
整体情感	正面	表扬	表扬	棒呆	不虚此行	力赞	140
	负面	人间地狱	很差	很差	虚假	没有下次	241
情绪	正面	大爱哟	顶呱呱	过瘾	杠杠	蛮给力	220
	负面	感到遗憾	烦透了	烦躁	晦气	哭笑不得	571
服务	正面	很灵范	很豪	好暖心	兢兢业业	面带笑容	166
	负面	不予理睬	粗心	恶狠狠	敷衍	黑脸	572
卫生	正面	一尘不染	整洁	干干净净	窗明几净	清香	28
	负面	斑斑点点	残留	臭虫	地不扫	瓜子皮	330
环境	正面	静悄悄	高贵典雅	古香古色	美呆	浪漫气息	108
	负面	暗和闷	憋闷	潮味	吵吵闹闹	车多太吵	260
饭菜	正面	吃饱喝足	美味可口	味道鲜美	撑死	一饱口福	38
	负面	难吃	臭肉	味道不好	变质	倒胃口	46
交通	正面	畅通无阻	超近	专车接送	便利	方便快捷	26
	负面	交通拥堵	老远	后不着店	太绕	偏僻	51
价格	正面	便宜	价格合理	物超所值	价格公道	超所值	47
	负面	冤枉钱	吃大亏	抢钱	不实惠	微贵	112
硬件	正面	雪白	网好	功能齐全	大窗	错落有致	93
	负面	床不大	坑坑洼洼	简陋	灯不亮	开裂	489
房间整体	正面	独一无二	耳目一新	风景如画	独具匠心	博雅	288
	负面	打差评	荡然无存	大不如前	乏善可陈	鸡肋	205

表2 酒店评论领域反映句法特征元素补充情感特征词库(部分)

整体情感	服务	交通	环境	硬件
谁还会再来	欠规范	不会再选择	相差特别远	不太匹配
没有下次	没有特色	没有任何优势	睡不成	不是很齐全
不是特别推荐	无人应答	不是非常便利	懒得再住	未按规定

## 1.2 酒店评论的情感倾向分类模型

文本情感倾向计算一般考虑如下问题:词汇评价的极性,在文本所涉及的话题中有哪些特征帮助情感的判断,这些判断的程度如何?针对上述问题再综合酒店评论这一特定领域中的话题,我们建立情感倾向的分类计算模型  $S_{EPD}$ , 即

$$S_{EPD} = \{E, P, D, C, G\}.$$

式中:  $E$  为情感倾向集合;  $P$  为语义情感特征集合;  $D$  为情感特征程度得分集合;  $C: D \rightarrow P$  为一个满射函数;  $G: C \rightarrow E$  为一个满射函数.

$E$  为酒店评论的情感倾向集合, 设情感共分  $m$  类, 则

$$E = \{e_1, e_2, \dots, e_m\}. \quad (1)$$

$P$  为语义情感特征集合, 设共分  $n$  个特征, 则

$$P = \{p_1, p_2, \dots, p_n\}. \quad (2)$$

$D$  是一条评论在每一个情感特征上程度得分的集合, 即

$$D = \{d_1, d_2, \dots, d_n\}. \quad (3)$$

当一条评论中特征  $p_i$  出现  $k$  个程度词  $d_w$  时, 则

$$d_i = \prod_{j=1}^k d_w. \quad (4)$$

$C: D \rightarrow P$  是一个满射函数连接情感程度到特征, 即

$$C(\mathbf{P}) = d_i p_i. \quad (5)$$

$G: C \rightarrow E$  是一个满射函数连接程度修饰后情感到情感倾向. 其中  $G$  可以用深度学习来进行情感倾向的计算, 其形式化公式如式(6).

$$G(\mathbf{E}) = \beta_0 + \beta_k^T \left[ \prod_{i=1}^n \sigma_i(\alpha_{i,0} + \alpha_{i,m}^T C_i) \right]. \quad (6)$$

式中:  $\alpha$  与  $\beta$  为权值参数;  $\sigma$  为激活函数, 即

$$\sigma = \frac{1}{1 + e^v}. \quad (7)$$

最终情感倾向的类别由 Softmax 函数确定:

$$f_j(\mathbf{E}) = \frac{e^{E_j}}{\sum_{i=1}^J e^{E_i}} \mathbf{E}. \quad (8)$$

## 2 湘潭市酒店评论情感倾向计算实验

### 2.1 湘潭地区酒店评论的获取与处理

我们编写 9 个网络爬虫分别从携程、同程、途牛、蚂蜂窝、去哪儿、艺龙、飞猪、穷游和驴妈妈等网站抓取湘潭地区旅游景区酒店的评论. 抓取的内容如表 3. 然后对评论采用机械压缩算法进行数据清理, 对于那些如“不错不错不错”缩成“不错”以及单纯复制他人的评论予以删除, 得到最终的网络评论, 然后对评论进行分词.

表 3 旅游景区酒店评论

字段名	类型	描述
comment_id	Varchar(50)	评论编号(主键)
comment_time	data	评论时间
business_id	Varchar(50)	酒店编号
Comment_content	txt	评论内容
Comment_score	Float	评分
Grab_time	data	抓取时间
Comment_supportNum	int	评论支持人数

### 2.2 湘潭地区酒店评论情感倾向实验结果

我们将 26 000 条湘潭地区酒店评论作为实验用数据集, 经人工打上标签: 非常好、好、差和非常差等 4 个等级. 这里, 我们使用 20 000 条数据作为训练集, 6 000 条作为测试集. 采用 Keras 深度学习中的 Sequential model 按我们提出的方法进行情感倾向计算, 并将其与采用情感词典计算的结果进行对比. 模型的设计如图 1 所示, 该模型由 4 层组成, 输入层 20 个节点(对应语义情感特征  $\mathbf{P}$  及其正负两方面), 2 个隐层(分别为 15 个节点和 10 个节点)与 1 个输出层(4 个节点, 对应情感分类  $\mathbf{E}$ ). 其中隐层的激活函数采用式(7), 输出层类别由式(8)确定, 计算的结果如表 4.

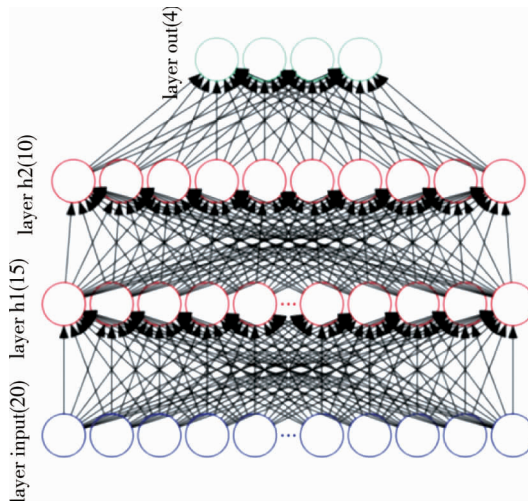


图 1 酒店评论深度学习网络结构

从实验结果可以看出, 采用我们提出的酒店评论情感倾向计算方法, 比单纯使用情感词典极性计算得到的结果要准确, 这主要有 3 个原因.

1) 在酒店这一特定的场景下, 一些中性的词汇或语句表现出了情感意义, 在情感词典方法中, 这一因素没有体现, 而我们提取的特征词汇包括在旅游酒店这个领域中具有情感的非情感词.

表 4 计算结果对比

方法	训练集准确率	训练集召回率	测试集准确率	测试集召回率
Ours	86.64	88.36	84.47	85.72
Ours(加句法特征分词)	90.59	92.83	86.76	86.52
经典的情感词典方法	68.34	66.93	64.48	65.11

2) 消费者进行网络评论时有时候会出现不少错词, 错词不包含在情感词典中, 但确实表达了情感, 如

表1中的“表扬”.而我们提取了旅游酒店中经常出现的错别字作为特征,放到深度学习模型中进行了训练.

3)我们引入句法特征的短句做分词,将那些具有否定句式、双重否定等影响情感词判断的句法特征放入到我们的分词库中,将这些特殊句式整体作为一个词汇来确定其情感极性,从而得到正确的情感极性.而单纯依靠情感词典查询的方法很难对这些语句进行识别.

同时,采用深度学习经过特征集的训练后,更能准确地得到旅游酒店领域中的情感模式.通过训练后得到的旅游酒店深度学习情感模型对于新评论比采用情感词典得到的要更加准确.

### 3 情感倾向驱动的网络舆情应用实例

我们选取湘潭3家具有竞争关系的五星级酒店——湖南华宇国际大酒店、湘潭华天大酒店以及湘潭盘龙山庄大酒店做舆情分析.分别从酒店语义情感特征的10个维度上对它们的评论进行情感倾向的计算得出其口碑,然后进行特色与管理中的短板的挖掘分析,用以决策支持.

#### 3.1 竞争酒店特色分析

竞争酒店特色的语义情感倾向雷达图分析如图2所示.从图2可以看出:3家酒店在各方面表现的都非常好,其中湖南华宇国际大酒店的特色在于卫生最突出,湘潭华天大酒店则在服务与硬件设施上略有胜出,湘潭盘龙山庄大酒店在环境与饭菜上表现突出.从各维度的竞争上各酒店可以看出相较于竞争对手自己的不足,然后确定下步工作的重心.

#### 3.2 竞争酒店的好感度变化趋势及原因分析

竞争酒店2018年1月~8月酒店评论情感好感度变化趋势如图3所示.从图3可以看出:湖南华宇国际大酒店从2018年1月开始消费者的好感度就逐步下降,而湘潭华天大酒店在1月~4月波动后就逐步稳定下来,湘潭盘龙山庄大酒店则呈逐步上升趋势.

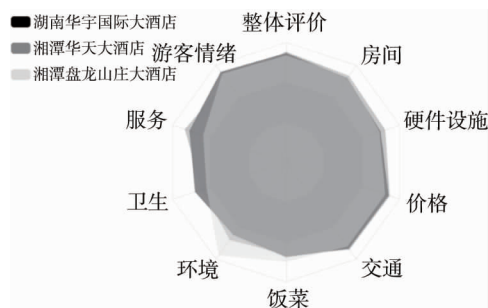


图2 竞争酒店特色的语义情感倾向雷达图分析

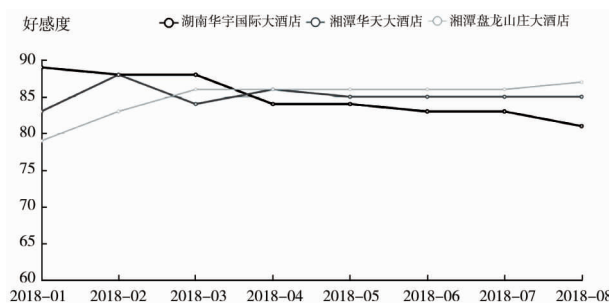


图3 竞争酒店2018年1月~8月酒店评论情感好感度变化趋势

竞争酒店评论的10维情感倾向月度对比如图4所示.从图4可以看出:

1)消费者对湖南华宇国际大酒店的卫生与服务两方面的好感度在3月下降后保持稳定,而对饭菜的好感度一直在下降,其他维度的情感倾向略有波动.

2)湘潭华天大酒店的卫生与服务两方面的消费者好感度在2月上升后保持稳定,对饭菜与价格两方面的好感度在1月~4月波动较大,5月后保持稳定,其他方面波动较小.

3)消费者对湘潭盘龙山庄大酒店卫生的好感度一直略有下降,对饭菜的好感度在2月大幅评价上升后保持稳定,对服务的好感度在2月大幅上升后一直略有上升,对价格的好感度在2月下降后一直在上升.

综上可以得出:湖南华宇国际大酒店的管理水平有所下降,它的饭菜的舆情评价一直下降,并且在3家竞争酒店中长时间最低,特别需要加强膳食部门的管理;湘潭华天大酒店在1月~3月经历春节后,管理稳定;而湘潭盘龙山庄大酒店的提升主要依赖于饭菜的逐步提升,它的卫生在3家酒店中舆情最差,需要进一步提升.

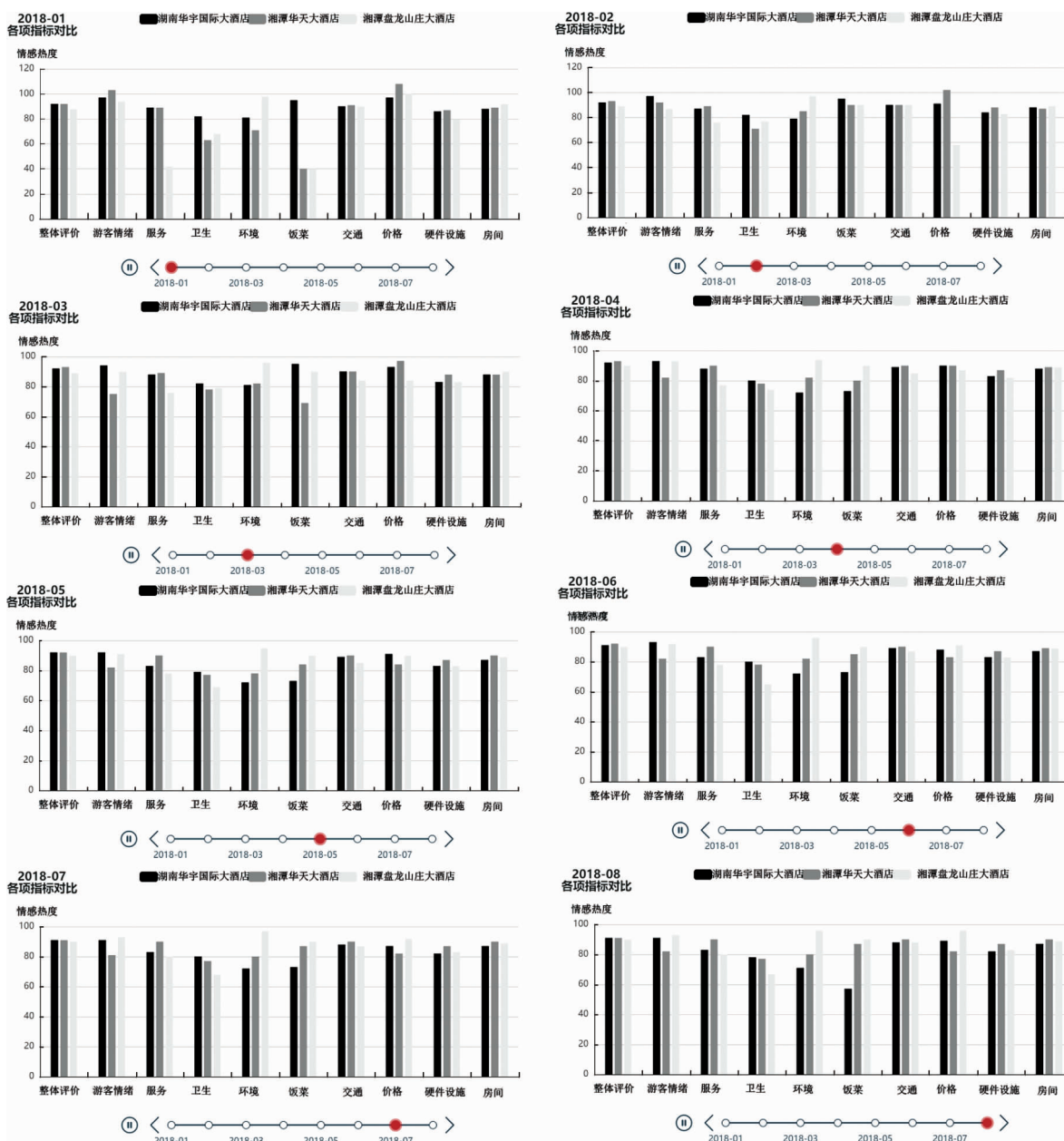


图4 竞争酒店评论的10维情感倾向月度对比

### 4 结论

- 1) 本文构建的情感倾向的分类计算模型能够很好地描述酒店领域中评论文本的情感倾向。
- 2) 建立的酒店语义特征情感词典能够对网络文本的错别字进行容错。
- 3) 实验表明从评论中通过情感挖掘能得到消费者对酒店服务与特色的真实态度。

### 参考文献:

[1] 刘斌.融合社交情感分析的股市预测探究[J].计算机系统应用,2018,27(2):250-256.

[2] 李勇,蔡梦思,邹凯,等.社交网络用户线上线下情感传播差异及影响因素分析——以“成都女司机被打”事件为例[J].情报杂志,2016,35(6):80-85.

[3] 张鲁民,贾焰,周斌,等.一种基于情感符号的在线突发事件检测方法[J].计算机学报,2013,36(8):1659-1667.

[4] 张艳丰,李贺,彭丽徽.基于模糊情感计算的商品在线评论用户品牌转换意向研究[J].现代图书情报技术,2016,32(5):

64-71.

- [5] 郭崇慧,张倚天.一种基于网络评论的商品特征挖掘方法[J].情报学报,2016,35(1):77-83.
- [6] Pai P F, Liu C H. Predicting Vehicle Sales by Sentiment Analysis of Twitter Data and Stock Market Values[J]. IEEE Access, 2018,6:57655-57662.
- [7] 郭宇,王晰巍,李师萌,等.基于情感分析的社会网络用户影响力模型研究[J].情报学报,2017,36(11):1139-1147.
- [8] 桂斌,杨小平,张中夏,等.基于微博表情符号的情感词典构建研究[J].北京理工大学学报,2014(5):537-541.
- [9] Agarwal B, Poria S, Mittal N, et al. Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach[J]. Cognitive Computation, 2015, 7(4): 487-499.
- [10] 唐晓波,刘一平.基于依存句法的跨语言细粒度情感分析[J].情报理论与实践,2018(6):124-129.
- [11] 王洪伟,郑丽娟,尹裴,等.基于句子级情感的中文网络评论的情感极性分类[J].管理科学学报,2013,16(9):64-74.
- [12] 万常选,江腾蛟,钟敏娟,等.基于词性标注和依存句法的 Web 金融信息情感计算[J].计算机研究与发展,2013,50(12):76-91.
- [13] 姜杰,夏睿.机器学习与语义规则融合的微博情感分类方法[J].北京大学学报(自然科学版),2017,53(2):247-254.
- [14] 孙晓,彭晓琪,胡敏,等.基于多维扩展特征与深度学习的微博短文本情感分析[J].电子与信息学报,2017,39(9):2048-2055.
- [15] Deng S, Sinha A P, Zhao H. Adapting sentiment lexicons to domain-specific social media texts[J]. Decision Support Systems, 2017, 94:65-76.
- [16] 王伟,周咏梅,阳爱民,等.一种基于 LDA 主题模型的评论文本情感分类方法[J].数据采集与处理,2017,32(3):629-635.
- [17] 傅向华,刘国,郭岩岩,等.中文博客多方面话题情感分析研究[J].中文信息学报,2013,27(1):47-56.
- [18] Tripathy A, Agrawal A, Rath S K. Classification of Sentiment Reviews using N-gram Machine Learning Approach[J]. Expert Systems with Applications, 2016, 57:117-126.
- [19] 王振飞,刘凯莉,郑志蕴,等.面向微博话题的增强型舆情分析方法[J].小型微型计算机系统,2017,38(4):678-682.
- [20] 郭顺利,张向先.面向中文图书评论的情感词典构建方法研究[J].现代图书情报技术,2016,32(2):67-74.
- [21] 刘鑫磊,张备,沈建京,等.用于涉军网络舆情情感分析的情感词典构建[J].信息系统工程,2016(3):32-34.
- [22] 姜芳,李国和,岳翔.基于语义的文档特征提取研究方法[J].计算机科学,2016(2):254-258.