

DOI: 10.11830/ISSN.1000-5013.201709015



# 采用在线评论的景点个性化推荐

王少兵, 吴升

(福州大学 福建省空间信息工程研究中心, 福建 福州 350003)

**摘要:** 通过对旅游网站的景点评论进行情感分析,综合利用自然语言处理技术和领域本体构建技术,准确把握游客对旅游目的地的满意度和需求;将群体智慧和个人偏好有效地结合,为游客出行制定合理的个性化推荐策略.实验结果表明:所提出的推荐策略能够有效地将碎片化的游客评论数据转化为对其他游客出行地选择的辅助信息,提高了游客获取旅游知识的效率,真实地反映游客的旅游感受,为游客景点选择提供参考.

**关键词:** 旅游网站; 在线评论; 情感分析; 个性化推荐

**中图分类号:** TP 391      **文献标志码:** A      **文章编号:** 1000-5013(2018)03-0467-06

## Attractions Personalized Recommendations Using Online Reviews

WANG Shaobing, WU Sheng

(Spatial Information Research Center of Fujian, Fuzhou University, Fuzhou 350003, China)

**Abstract:** Through analyzing attractions online reviews of travel sites based on sentiment analysis, this paper comprehensively utilizes natural language processing technology and domain ontology construction technology, accurately grasps the tourist satisfaction and the demand for tourism destination, effectively combines the wisdom of crowds and personal preference, and sets reasonable personalized recommendation for tourists travel strategy. The experimental results show that the proposed strategy can effectively translate fragmented visitors review data into ancillary information of tourists traveling, improve the efficiency of the tourists to obtain knowledge, truly reflect the tourists feel, and provide the reference for tourist choosing attractions.

**Keywords:** travel sites; online reviews; sentiment analysis; personalized recommendations

旅游互联网应用的快速发展,产生了大量和旅游景点相关的评论信息.这些评论以不同的形式出现在博客、BBS 或论坛网站等各种媒体,成为一种越发重要的经验信息载体,潜在游客会花费很多时间阅读在线评论,协助旅行决策<sup>[1-3]</sup>.然而,游客撰写评论信息具有随意性,评论的质量良莠不齐,大量冗余的信息严重干扰了潜在游客获取和分析有用的信息,造成了信息超载问题,使潜在游客难以做出合适的个性化旅行决策.信息过滤技术和信息定制技术是目前解决信息超载问题的常用技术手段<sup>[4-5]</sup>.相比信息定制技术,信息过滤技术将适合游客的信息经过筛选排序,以推送的方式为用户提供符合其兴趣需求的个性化推荐服务,因此,更加受到游客青睐.但是目前比较流行的信息过滤技术,如个性化推荐系统,普遍仅从单个游客的兴趣偏好、属性,以及游客之间的社交关系等方面挖掘用户的喜好,而忽视群体贡献

**收稿日期:** 2017-09-11

**通信作者:** 吴升(1972-),男,教授,博士,主要从事时空数据分析与可视化、信息共享与智慧政务、应急信息系统的研究. E-mail:ws0110@163.com.

**基金项目:** 国家政务大数据应用工程技术研究中心培育项目(2016L3007);福建省科技创新平台建设项目(2015H2001)

的经验信息. 因此, 本文提出一种基于在线评论的景点个性化推荐策略.

1 相关研究

在线评论不仅是游客表达意见、建议和满意度的常见方式之一, 也进一步形成了景点的网络口碑 (electronic word-of-mouth). 因此, 国内外研究者从多角度展开旅游在线评论的有用性研究, 证实了评论信息对游客的旅行决策产生重要影响, 以评论数据作为研究景点推荐的素材, 具有重要的意义. Gretzel 等<sup>[6]</sup>通过网络调查游客评论, 研究评论信息在旅游规划过程中的作用和影响, 其中, 四分之三的游客都在确定旅游景点之前查看评论. Trusov 等<sup>[7]</sup>将口碑营销与传统营销手段进行比较, 结果表明, 相较于传统营销行为, 口碑推荐具有更大的作用. Zhao 等<sup>[8]</sup>通过回归分析的方法, 验证了在线评论的有用性、时效性、数量及全面性等特征对游客旅游预订意向有显著的正向影响.

旅游地属于体验性商品, 它的产品质量主要通过服务质量体现<sup>[9]</sup>. 和普通产品的评论数据相比, 旅游在线评论反映了游客对旅游服务的感受, 涉及游客旅游活动相关的, 包括吃、住、行、游、购、娱的多个方面. 因此, 只有通过细粒度评论数据挖掘, 才能获取游客个性化需求. 目前, 学者的研究兴趣包括在线评论的主观性内容识别、褒贬义情感倾向分类, 以及旅游推荐等. Yu 等<sup>[10]</sup>通过几种用于区分主观意见和客观事实的模型实验, 发现贝叶斯分类器能够达到非常高的识别精度. Turney<sup>[11]</sup>提出一种无监督分类学习算法, 根据评论语句中形容词或副词短语的平均语义情感倾向, 将评论语句的类型分为正向推荐或负向不推荐. Cao 等<sup>[12]</sup>提出一种综合用户历史评论主题和情感倾向的方法.

根据推荐系统数据来源的不同, 主流推荐系统可分为: 基于内容推荐、协同过滤推荐、基于知识推荐和组合推荐<sup>[13]</sup>. 近年来, 由于社会化网络的成熟及互联网的快速发展, 利用用户社会关系、信任关系提高推荐系统精度的社会化推荐系统成为研究热点. Rendle 等<sup>[14]</sup>通过将用户的年龄、性别、微博数量、标签、关键词, 以及社交网络关注的用户作为分类输入变量, 利用贝叶斯推理和马尔科夫链蒙特卡洛方法, 建立基于社交网络项目推荐模型. Yang 等<sup>[15]</sup>通过给用户朋友圈的朋友赋予不同的权重, 对用户间的信任关系进行建模, 提出一种基于朋友圈的社交网络推荐方法. 李鑫等<sup>[16]</sup>通过考虑不同的兴趣圈上的社会关系, 提出一种在社交网络中基于兴趣圈的社会关系挖掘推荐算法, 更好地解决冷启动问题.

已有的关于在线评论的研究虽取得不错的成果, 但还存在一些问题: 在旅游领域, 鲜有以旅游景点在线评论作为数据来源进行推荐系统的研究; 评论的情感分析主要集中在句子级别, 很难通过细粒度评论数据挖掘、获取游客个性化需求; 现有的社会化推荐方法, 大多把单个游客的社会关系信息作为重要影响因子引入推荐模型, 忽略了其他游客实地旅行之后的旅游体验与感受, 而其他游客的评论信息作为旅游经验信息, 对游客的旅行决策行为产生重要影响.

2 基于在线评论的景点个性化推荐流程

首先, 针对旅游在线评论内容, 利用自然语言处理技术和领域本体构建技术, 一方面, 选取待推荐景点的评论数据, 针对评论对象进行细粒度情感分析; 另一方面, 选取待被推荐景点的游客历史评论数据, 通过词频统计, 抽取其评论对象的关注重点. 然后, 结合个人偏好和群体智慧, 构建景点推荐指数数学模型. 最后, 为游客提供个性化景点推荐. 基于在线评论的景点个性化推荐流程, 如图 1 所示.

2.1 旅游评论数据的获取和预处理

通过网络爬虫程序从携程网、蚂蜂窝等旅游网站的评论专区随机采集景点评论数据和特定游客的评论数据. 每条评论数据记录内容包括游客名 (或游客 ID) 和评论内容. 由于爬取的评论数据有很多没有实际研究意义, 为了避免影响实

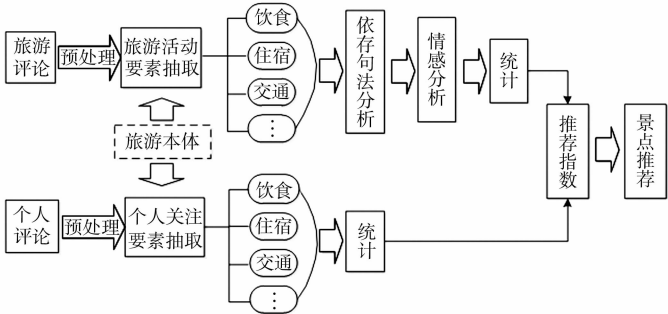


图 1 基于在线评论的景点个性化推荐流程  
Fig. 1 Attractions personalized recommendation process based on online reviews

## 2.2 旅游评论本体的构建

旅游领域中的主要概念及其关系,即景点评论内容中和游客旅游活动有关的,包括吃、住、行、游、购、娱的多个要素的关系.如饮食方面,游客可能从口味、餐饮服务、饮食环境、位置等方面进行评论,购物方面可能涉及商品、购物环境、购物服务、物价等方面的评论.通过咨询领域专家,并查阅相关文献<sup>[17-18]</sup>,结合景点评论内容,归纳旅游领域的概念(或个体)及其关系.利用 Protege 5.1 本体构建工具构建旅游评论本体知识库,作为和景点评论对象匹配的规则信息.旅游评论本体知识库各个概念间的逻辑关系,如图 2 所示.

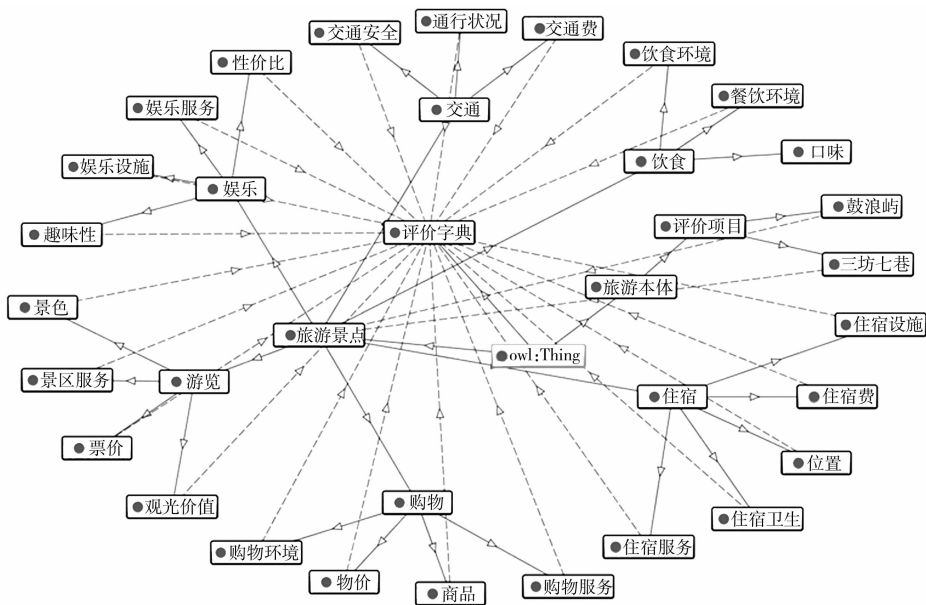


图 2 旅游评论本体概念关系图

Fig. 2 Concept diagram of travel review ontology

- 步骤1 利用自然语言处理技术对旅游评论数据分词、词性标注。
- 步骤2 遍历句法树中所有叶子节点,找到叶子节点为名词短语(NP)或常用名词(NN)的节点,匹配旅游评论本体知识库,识别评论中的评论对象。
- 步骤3 若找到的叶子节点为评论对象,查找其父节点的所有兄弟节点。
- 步骤4 若兄弟节点有谓词短语(ADVP)或形容词(VA),提取情感词并组成评论对象-情感词对。
- 步骤5 若步骤2中没有识别出评论对象,继续向上遍历父节点。递归执行步骤3,4,寻找评论对象的情感词。

情感词进行情感分析属于词级别的情感分析,是为了获取景点评论数据中游客对评论对象的情感倾向,包括正向(褒义)情感和负向(贬义)情感.

由于在情感词的情感极性判断时,情感词前常出现的程度副词及否定词导致对表达的情感有影响.程度副词起到加深或减弱情感倾向的作用,否定词则使情感倾向发生逆转.为了全面了解游客评论对象的整体情感倾向,根据《知网情感分析用词词语集》给出的程度副词不同的修饰等级,将情感词和带有程度副词的情感词的极性进行量化,构建情感极性量化表,如表 1 所示.

对有程度副词修饰的情感词,其情感总极性值( $S$ )=程度副词的极性值 $\times$ 情感词的极性值.例如,“景区极其漂亮”的情感极性值为 $2\times 1=2$ ;“票价很贵”的情感极性值为 $1.8\times (-1)=-1.8$ .对于否定词前置的情感词,如“美”、“不美”,分词时,将否定词与情感词合并为负向情感词.最终,情感总极性值 $S$ 的分类结果据表 1 分为 3 个区间: $S>1$ (褒义); $1\geq S\geq 0$ (中性); $S<0$ (贬义).

2.5 推荐指数模型构建

数据来源于携程网、蚂蜂窝等旅游网站采集的景点评论数据和特定游客的历史评论数据.特定游客的历史评论信息反映其选择旅游目的地时,关注的旅游活动要素.通过统计该游客评论对象的词频数量,可以发现该游客旅游活动关注的主要要素.利用每个评论对象的词频数占评论对象总词频的百分比作为推荐指数的权重 $w_i$ .同时,通过对景点评论数据进行情感分析,挖掘景点评论内容中,景点的某个评论对象的情感倾向,统计景点评论对象褒贬义情感信息的数量.最终构建的推荐指数数学模型为

$$T = \sum_k^i w_i \times ((m_i - n_i)/(m_i + n_i)). \tag{1}$$

式(1)中: $T$ 为景点的推荐指数; $i$ 为游客景点旅游的某个评论对象; $k$ 为评论对象的个数; $w_i$ 为某个评论对象的权重; $m_i, n_i$ 分别为某个评论对象包含正面和负面情感信息的数量.

3 实验结果与分析

采用的景点评论数据为:鼓浪屿评论数据 8 014 条;三坊七巷评论数据 8 009 条.实验景点案例知名度 and 关注度较高,旅游需求明显,关于鼓浪屿和三坊七巷的点评数据丰富.待推荐景点的 3 位试验者游客年龄层、职业均不相同.为简化起见,以 A,B,C 分别标识 3 位游客,其发表景点评论分别为 321,232,81 条.通过对 3 位游客的景点评论对象进行词频统计,结果如表 2 所示.表 2 中: $\eta$ 为占总频次百分比.

表 2 3 位游客评论对象的词频统计  
Tab.2 Frequency statistics of three visitors' comments

评论对象		A		B		C	
		频次	$\eta/\%$	频次	$\eta/\%$	频次	$\eta/\%$
饮食	口味	14	7.03	6	3.94	2	1.73
	餐饮服务	12	5.85	2	1.41	3	3.03
	饮食环境	9	4.49	4	2.34	2	2.02
住宿	位置	5	2.66	2	1.22	3	2.60
	住宿卫生	3	1.36	9	5.81	3	2.60
	住宿设施	3	1.48	11	6.94	3	2.45
	住宿服务	11	5.38	4	2.44	6	6.06
	住宿费	4	2.19	5	3.38	2	2.16
购物	商品	7	3.37	3	1.88	5	5.19
	购物环境	5	2.48	12	7.69	1	1.15
	购物服务	4	2.13	8	5.25	2	1.44
	物价	10	4.91	4	2.34	5	4.90

续表  
Continue table

评论对象		A		B		C	
		频次	$\eta/\%$	频次	$\eta/\%$	频次	$\eta/\%$
娱乐	娱乐服务	5	2.60	4	2.34	9	8.80
	娱乐设施	4	2.07	2	1.50	6	6.20
	趣味性	9	4.43	8	5.06	4	4.18
	性价比	9	4.32	7	4.50	1	1.30
交通	通行状况	8	4.14	2	1.13	2	2.31
	交通费	4	1.89	5	3.28	1	1.01
	交通安全	7	3.31	7	4.31	4	3.61
游览	景区服务	22	10.70	18	11.16	15	13.99
	景色	28	14.01	15	9.09	13	12.12
	观光价值	13	6.27	15	9.66	10	9.95
	票价	6	3.01	5	3.28	1	0.87

由表 2 可知:3 位游客都对景点的景区服务、景色和观光价值较为关注,涉及的评论数也较多,但他们对景点的其他评论对象却因个人旅游需求和喜好的不同存在差异,如 A 游客涉及对景点饮食的特征评价明显高于 B,C 两名游客。

通过对鼓浪屿和三坊七巷两个景点的评论数据进行细粒度情感分析,挖掘群体贡献的经验知识,最终得到的统计结果,如图 3 所示。由图 3 可知:鼓浪屿和三坊七巷的正面评价远多于负面评价。两个景点作为国家 5A 级景区,具有较高的知名度,可见分析结果与社会普遍的认可度存在一致性。

由图 3 反映的 2 个景点评论对象情感分析结果具体分布情况还可知:鼓浪屿作为滨海观光景点,经过多年的自然人文的沉淀,海洋、小资、特色建筑等已经成为其旅游形象的代名词;而三坊七巷作为历史文化名街,主打街巷文化,并有各种福州特色小吃,但是占地面积只有鼓浪屿的五分之一,并处于内陆,景点相对较少。因此,游客对 2 个景点的关注点并不均匀,评论主要集中在景点游览、景点饮食两个要素的评价。这说明游客比较注重景点游览和饮食的旅游体验。

将 3 位游客评论数据的统计结果和景点评论数据进行分析统计,并将结果代入推荐指数模型公式,可得 3 位游客的个性化景点推荐指数,如表 3 所示。由表 3 可知:基于在线评论的景点个性化推荐模型,可根据群体贡献的经验知识发现更值得游客游玩的景点,并推荐给游客。在推荐过程中,模型也根据游客个性化旅游的需求,得到不同的景点推荐结果。案例中,对游客而言,鼓浪屿观光景点较多。B,C 两名游客对景点的游览价值需求明显,故鼓浪屿的推荐指数更高。而 A 游客虽然对景点的游览需求最大。但是对景点的饮食需求仅次于游览需求,为 17.37%。三坊七巷不仅有各种观光景点,而且有各种福州特色小吃,如肉燕、永和鱼丸等,因此,对于 A 游客三坊七巷的推荐指数更高,相比鼓浪屿更值得推荐。

4 结束语

综合利用自然语言处理技术和领域本体构建技术,有效地将群体智慧和个人旅游需求结合,为游客

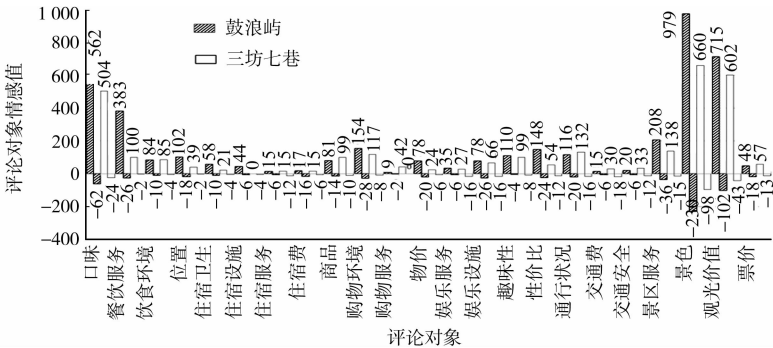


图 3 景点评论数据情感分析结果分布图  
Fig. 3 Distribution chart of attraction review data sentiment analysis results

表 3 景点推荐指数计算结果  
Tab. 3 Attractions recommended index calculation results

游客	景点	推荐指数
A	鼓浪屿	0.667 64
	三坊七巷	0.695 47
B	鼓浪屿	0.659 79
	三坊七巷	0.603 80
C	鼓浪屿	1.201 21
	三坊七巷	0.627 98

提供旅游景点的推荐. 提出的推荐策略能为游客景点选择提供参考, 对其他景点的个性化推荐具有借鉴和研究意义. 下一步的工作是考虑更多的旅游推荐影响因素, 尝试解决旅游推荐的冷启动问题.

## 参考文献:

- [1] ZHU Feng, ZHANG Xiaoquan. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics[J]. *Journal of Marketing*, 2010, 74(2): 133-148. DOI: 10.1509/jmkg. 74. 2. 133.
- [2] SCHUCKERT M, LIU Xiaowei, LAW R. Hospitality and tourism online reviews: Recent trends and future directions[J]. 2015, 32(5): 608-621. DOI: 10.1080/10548408. 2014. 933154.
- [3] FANG Bin, YE Qiang, KUCUKUSTA D, *et al.* Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics[J]. *Tourism Management*, 2016, 52: 498-506. DOI: 10.1016/j. tourman. 2015. 07. 018.
- [4] 乔向杰, 张凌云. 近十年国外旅游推荐系统的应用研究[J]. *旅游学刊*, 2014, 29(8): 117-127. DOI: 10.3969/j. issn-1002-5006. 2014. 08. 012.
- [5] 周玲. 信息超载综述[J]. *图书情报工作*, 2001(11): 33-35. DOI: 10.3969/j. issn. 0252-3116. 2001. 11. 009.
- [6] GRETZEL U, YOO K H. Use and impact of online travel reviews[C]// *Information and Communication Technologies in Tourism*. Innsbruck: DBLP, 2008: 35-46. DOI: 10.1007/978-3-211-77280-5\_4.
- [7] TRUSOV M, BUCKLIN R E, PAUWELS K. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site[J]. *Journal of Marketing*, 2009, 73(5): 90-102. DOI: 10.2139/ssrn. 1129351.
- [8] ZHAO Xinyuan, LIANG Wang, XIAO Guo, *et al.* The influence of online reviews to online hotel booking intentions [J]. *International Journal of Contemporary Hospitality Management*, 2015, 27(6): 1343-1364. DOI: 10.1108/IJCHM-12-2013-0542.
- [9] 郭潇. 在线评论对旅游预订意向影响的实证分析[D]. 广州: 华南理工大学, 2010.
- [10] YU Hong, HATZIVASSILOGLU V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences[J]. *Pediatrics*, 2003, 116(3): 58-59. DOI: 10.3115/1119355. 1119372.
- [11] TURNEY P D. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews[C]// *Meeting on Association for Computational Linguistics*. Stroudsburg: [s. n. ], 2002: 417-424. DOI: 10.3115/1073083. 1073153.
- [12] CAO Nieqing, CAO Jingjing, LIU Panpan, *et al.* Sentimental preference extraction from online reviews for recommendation[C]// *International Conference on Internet and Distributed Computing Systems*. Berlin: Springer International Publishing, 2015: 294-303.
- [13] XU Hailing, WU Xiao, LI Xiaodong, *et al.* Comparison study of internet recommendation system[J]. *Journal of Software*, 2009, 20(2): 350-362.
- [14] 孟祥武, 刘树栋, 张玉洁, 等. 社会化推荐系统研究[J]. *软件学报*, 2015, 26(6): 1356-1372. DOI: 10.13328/j. cnki. jos. 004831.
- [15] YANG Xiwang, STECK H, LIU Yong. Circle-based recommendation in online social networks[C]// *ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2012: 1267-1275. DOI: 10.1145/2339530. 2339728.
- [16] 李鑫, 刘贵全, 李琳, 等. LBSN上基于兴趣圈中社会关系挖掘的推荐算法[J]. *计算机研究与发展*, 2017, 54(2): 394-404. DOI: 10.1145/2339530. 2339728.
- [17] 黄金顺, 黎慧, 唐素勤, 等. 基于领域本体的 Web 旅游信息抽取[J]. *计算机与信息技术*, 2010(12): 32-37.
- [18] MARRESE-TAYLOR E, BRAVO-MARQUEZ F. A novel deterministic approach for aspect-based opinion mining in tourism products reviews[J]. *Expert Systems with Applications*, 2014, 41(17): 7764-7775. DOI: 10.1016/j. eswa. 2014. 05. 045.
- [19] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. *软件学报*, 2010, 21(8): 1834-1848. DOI: 10.3724/SP. J. 1001. 2010. 03832.
- [20] 杨立公, 朱俭, 汤世平. 文本情感分析综述[J]. *计算机应用*, 2013, 33(6): 1574-1578.

(责任编辑: 钱筠 英文审校: 吴逢铁)