引文格式：殷鹏, 何若楠, 倪敬峰, 等. 基于微博用户情感分析的城市幸福感影响因素与空间分异研究[J]. 地理空间信息, 2023, 21(5): 18-21. doi:10. 3969/j.issn. 1672-4623.2023.05.005

基于微博用户情感分析的城市幸福感影响因素与空间分异研究

殷　鹏[1]，何若楠[1]，倪敬峰[1]，徐晨曦[1]，魏倩枝[2]，万红莲[1,3,4]
（1. 宝鸡文理学院 地理与环境学院，陕西 宝鸡 721013；2. 山东桓台县荆家镇人民政府，山东 桓台 256406；
3. 宝鸡文理学院 灾害监测与机理模拟陕西省重点实验室，陕西 宝鸡 721013；
4. 陕西省重点中国特色社会主义理论体系研究中心（宝鸡基地），陕西 宝鸡 721013）

摘　要：以微博“2020 中国最具幸福感城市榜单”话题为切入点，利用网络爬虫技术抓取微博用户评论数据；利用文本主题生成模型（LDA）和 ArcGIS 分析了微博用户视角下幸福感城市的情感响应空间分异和影响因素，为榜单城市和国内其他城市的可持续高质量发展提供参考。结果表明：①全国各地对榜单的响应呈东高西低的趋势，响应热度较高的区域均为榜单城市所在省份及其周边，并以榜单城市为中心呈距离衰减现象，旅游与美食发展较好的城市容易引起更多地区的响应；②地域独特性，微博用户对长沙大多呈正面响应，对郑州大多呈负面响应，青岛、西安、杭州的情感响应省外地区高于省内，微博用户对榜单城市的印象是主要影响因素；③旅游和美食对城市幸福感具有显著正面影响，而房价和工资则具有显著负面影响。

关键词：城市幸福感；情感响应；空间分异；LDA；ArcGIS
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2023）05-0018-04

Research on Influencing Factors and Spatial Differentiation of Urban Happiest Based on Microblog Users Emotional Analysis

YIN Peng[1], HE Ruonan[1], NI Jingfeng[1], XU Chenxi[1], WEI Qianzhi[2], WAN Honglian[1,3,4]
(1. School of Geography and Environment, Baoji University of Arts and Sciences, Baoji 721013, China;
2. People's Government of Jingjia Town, Huantai 256406, China;
3. Shaanxi Key Laboratory of Disaster Monitoring and Mechanism Simulation, Baoji University of Arts and Sciences, Baoji 721013, China;
4. Shaanxi Provincial Key Research Center for Socialism with Chinese Characteristics(Baoji Base), Baoji 721013, China)

Abstract: Taking the topic of "2020 List of China's Happiest Cities" on Sina Microblog as the starting point, we used Web crawler technology to capture microblog user comment data, and used LDA model and ArcGIS to analyze the spatial differentiation and influencing factors of emotional response of urban happiness from the perspective of users in microblog, which could provide reference for the sustainable and high-quality development of listed cities and other cities in China. The results show that ①the responses of all parts of list are generally higher in the east and lower in the west. The regions with higher response heat are all the provinces and surrounding areas where the list cities are located, and the distance attenuation phenomenon is shown centering on the list cities. Cities with better tourism and food development are likely to cause responses from more regions. ②The place is unique. Most users in microblog respond positively to Changsha and negatively to Zhengzhou. The emotional responses of Qingdao, Xi'an and Hangzhou are higher in most areas outside the province than in the province. The microblog users' impression of listed cities is the main influencing factor. ③Tourism and food have a significant positive impact on urban happiness, while house price and salary have a significant negative impact on urban happiness.

Key words: urban happiness, emotion response, spatial differentiation, LDA, ArcGIS

党的十九大报告指出，当前需要形成有效的社会治理、良好的社会秩序，使人民获得感、幸福感、安全感更加充实、更有保障和更可持续。城市对居民的生活质量和幸福感具有显著影响[1]，提升城市幸福感尤为重要。提升城市幸福感需厘清其影响因素，可通过分析公众对城市幸福感的情感响应实现。在对居民城市幸福感的研究中，Cattaneo M D[2]等发现住房条件对居民幸福感具有显著影响；Krueger A B[3]等研究认为居民整体幸福感与日常活动中体验的累加有关；Schwanen T[4]等在此基础上得出日常活动中的陪伴会影

收稿日期：2021-12-14；修回日期：2022-01-27。
项目来源：教育部人文社会科学研究规划基金资助项目（18YJA810004）；陕西省科技厅软科学研究计划资助项目（2022KRM028）。
第一作者简介：殷鹏（1997—），硕士研究生，主要研究领域为可持续发展与人工智能 GIS 研究，E-mail: yopo_kg@sina.com。

响幸福感的结论；Lehnert K[5]等则系统地考察了锻炼对幸福感的影响；在国内的研究中，刘义[6]等以广州为例，研究了邻里建成环境对幸福感的影响；康雷[7]等结合调研数据发现，邻里社会环境对低收入居民主观幸福感具有重要影响；张体委[8]通过调查问卷研究发现，居民收入水平和地区富裕程度对城市幸福感具有显著正面影响。上述研究对城市幸福感影响因素进行了深入分析，但缺少对多种影响因素的比较研究；且数据均为调查问卷，数据体量较小，难以体现社会主流民意，问卷发放范围较小，结果普适性较弱，而基于大数据的城市情感分析能解决上述问题。近年来国内涌现出一批基于大数据的社会、经济、灾害问题研究[9-13]，但鲜有利用大数据研究城市幸福感的文章。因此，本文以微博“2020中国最具幸福感城市榜单”话题为切入点，结合GIS空间可视化技术与人工智能自然语言处理（NLP）领域的情感分析、主题挖掘技术，从微博用户视角分析最具幸福感榜单城市的情感响应空间分异特征及其影响因素，弥补城市幸福感情感响应研究的空白，挖掘榜单城市形象，探析榜单城市幸福感的主要影响因素，为榜单城市和全国各城市的高质量发展提供参考。

1　数据来源与研究方法

1.1　数据来源与预处理

2020年11月由新华社《瞭望东方周刊》评选的“2020中国最具幸福感城市榜单”话题通过新浪微博发布，迅速成为用户讨论的热点话题。该榜单经过大数据采集、问卷调查、材料申报、实地调研、专家评审等环节严格遴选，最终成都、杭州、宁波、广州、长沙、南京、郑州、西宁、青岛、西安10座城市上榜，话题引起了大批微博用户的评论。本文通过爬虫程序抓取话题评论数据，然后在Excel中对数据进行删除重复项、转换、重新排列等操作，共筛选出20 875条有效信息，包含评论者ID、昵称、性别、地区、评论时间和评论内容等。评论者遍布34个省、自治区、直辖市、特别行政区以及海外地区，其中男性6 374人、女性14 501人，具有位置信息的数据17 170条。为研究各城市情况，本文利用Excel高级检索功能将10个榜单城市的评论数据划分为10类。

1.2　研究方法

1.2.1　情感分类与模型预训练

NLP情感分类通过对网络文本语义情感打分，达到依据情感信息分类文本的目的。SnowNLP是一个由Python语言编写的开源中文文本处理工具，通过中文的网络购物语料进行了预训练，具有情感分类模块，数学模型为贝叶斯模型。该模块能对中文文本语义情感打分，分值在0~1之间，大于0.5，表明文本语义情感偏向正面；小于0.5则偏向负面；越靠近0或1分类结果越可靠。为提高该模型对特定文本集的精度，需对其进行预训练。因此，对评论进行正负面标记，重新训练SnowNLP模型，人工验证准确率为90.4%。利用训练后的模型对各榜单城市的评论文本集进行情感打分并分类，通过情感分数集“中位数”来反映各榜单城市情感响应的一般水平。

1.2.2　主题挖掘

文档主题生成模型（LDA）是人工智能领域中自然语言处理的热门技术，能帮助提取文本数据的主题词汇，已应用于各领域[14-15]，其数学模型是朴素贝叶斯模型，为非监督学习模型。使用LDA之前需对评论文本进行分词，自然语言处理工具包（FoolNTLK）是开源的以深度学习算法双向长短期记忆（LSTM）网络构建的中文分词处理Python包。利用FoolNTLK对评论文本进行分词，再利用停用词处理分词结果，并构建检索词频率矩阵。

$$p(W,Z,\theta,\varnothing;\alpha,\beta)=\prod_{i=1}^{K}P(\varnothing_i;\beta)\prod_{j=1}^{M}P(\theta_j;\alpha)\prod_{t=1}^{N}P(Z_{j,t}|\theta_j)P(W_{j,t}|\varnothing Z_{j,t}) \tag{1}$$

式中，K为主题数量；W为文档数量；N为全部词数量；α为不同主题在文档中分布的先验权重；β为单词在主题中分布的先验权重；θ_j为文档j中主题分布概率；$\varnothing_i$为主题i中单词分布概率；$Z_{j,t}$为t单词在j文档中对应的主题数量；$W_{j,t}$为文档j中单词t出现的次数。

利用LDA进行可视化过程中，Sievert C[16]等提出了式（2）。

$$r(w,k|\lambda)=\lambda\log(\varnothing_{kw})+(1-\lambda)\log\left(\frac{\varnothing_{kw}}{p_w}\right) \tag{2}$$

式中，k为某个主题；w为单词；r为单词与主题的关联程度；$\varnothing_{kw}$为在主题k中单词w的概率；λ为主题k下单词w与其全局出现概率的相对权重，当$\lambda=1$时，主题下关键词关联程度为该主题出现概率，当$\lambda=0$时，主题下关键词关联程度为其全局出现概率。

通过调整参数λ提取情感分类后文本主题的关键词，再利用关键词词汇估计主题含义来完成主题挖掘[17]。

2 研究结果与分析

2.1 情感响应空间分异

2.1.1 话题响应分析

从话题热度来看，全国各地对“最具幸福感城市榜单”的响应呈东高西低的趋势，局部集中在四川、广东、河南、山东、江苏、浙江、湖南等省份，响应热度较高的区域为榜单城市所在省份及其周边，并以榜单城市为中心呈距离衰减趋势。分析其原因，从第三次全国人口普查数据来看，响应热度较高区域均为我国人口较多省份，说明人口空间分布对响应热度空间分布具有影响；城市幸福感与亲情、人际关系等隐性因素有关[18]，受地方依恋影响，话题较易引起榜单城市及其周边地区用户的响应。

由全国各地对话题响应的情感得分可知，大多数地区情感响应呈中性，说明新媒体用户对榜单城市的评价较客观；青海、湖南和江西的情感响应呈正面，研究发现青海和湖南榜单城市的房价均为其正面评论主题词汇，稳定的房价显著提升了该省新媒体用户的情感响应，地方依恋较强，而江西与湖南早年就有人口交互性，两省关系较好，空间依赖性较强；河南、宁夏、山东的情感响应大多呈负面，研究发现银川作为《中国媒体大调查》公布的2018—2019年最具幸福感榜单城市，此次落选使响应呈负面，山东的新媒体用户围绕着济南与青岛话题讨论激烈，济南与青岛的国家中心城市争论已达3年之久[19]，具有话题延续性，整体响应偏负面，河南各地级市对郑州的素质文明颇有意见。

2.1.2 榜单城市分析

对各榜单城市来说，响应热度聚集在省内，并在空间上呈距离衰减现象。其原因在于，宁波人口少于其他城市，且第一次上榜，微博用户关注度较弱；成都、长沙、西安位于我国腹地，其旅游和美食领域发展较好，空间上较易引起多个省份的响应。

全国各地对各榜单城市情感响应差异显著，响应热度为0的地区没有情感得分，其中各地区对长沙大多呈正面响应，对郑州大多呈负面响应。研究发现，长沙作为新一线城市，城市宜居、经济实力和稳定的房价受到了大多数地区的认可。中国“地域歧视”地图显示，河南的文明环境受到大多数地区的质疑，且已在微博媒体中形成了一种负面刻板印象[20]。榜单城市所在省份的响应也不尽为正面，河南的新媒体用户对郑州的情感响应多为负面，房价、工资、环境和素质文明是负面评论热点；对于山东、陕西、浙江的榜单城市，多数省外地区情感响应得分高于省内。除房价和物价因素外，山东围绕着济南与青岛的话题讨论激烈，陕西对西安的素质文明和雾霾环境展开讨论，浙江多反映压力大且工作加班严重，互联网企业996制度对国民身体素质和幸福感造成不良影响[21]，杭州的互联网产业虽给杭州带来了经济发展，但也影响了微博用户对杭州的情感响应。

2.2 城市幸福感影响因素比较分析

利用LDA对各榜单城市正负面主题进行挖掘发现（表1），各榜单城市的幸福感影响因素具有差异性。根据统计结果得出城市幸福感的影响因素可归纳为旅游、美食、宜居、发展、交通、教育、养老、房价、包容、设施、节奏、气候、物价、交通、地铁、素质、雾霾、压力和工资（图1）。由排序结果可知，旅游和美食为正面主题共现高频词汇，表明二者对城市幸福感具有显著正面影响。心理物理学理论表明，环境的刺激将影响人们感觉、知觉与判断，城市旅游与美食直接刺激感官中的视觉与味觉，容易达到刺激阈限[22]，旅游与美食也是体验性评价的主要影响因素。房价和工资为负面主题共现高频词汇，表明二者对城市幸福感具有显著负面影响，房价和工资通过主观社会地位来影响幸福感[23]。

表1 榜单城市正负面主题词汇

分类	西安	西宁	长沙	郑州	杭州	广州	宁波	成都	南京	青岛
正面主题	旅游生活	旅游交通	低房价	购物方便	一线旅游	优质教育	文明城市	美食外卖	物价便宜	旅游发展
	美食教育	工作教育	就业医疗	交通便利	交通便利	旅游包容	养老宜居	养老宜居	旅游环境	交通便利
	发展设施	美食房价	安全包容	高速发展	养老适宜	美食宜居	书藏丰富	旅游基建	发展机会	美食宜居
	房价物价	四线苦寒	交通堵塞	高房价	高房价	高房价	高房价	高房价	高房价	高房价
负面主题	地铁拥挤	基建环境	工资低	雾霾严重	加班严重	快节奏	房贷压力	雾霾严重	工资低	消费过高
	素质文明	工资低	气候多变	工资低	压力大	地铁拥挤		交通堵塞		工资低
	雾霾严重		教育低质	素质文明	物价高	压力大				

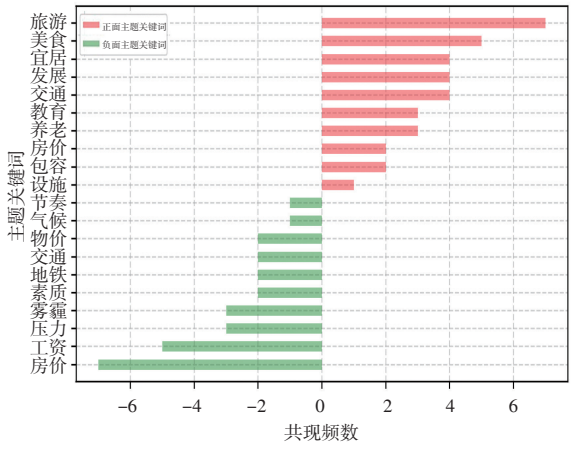


图1　正负面主题词汇共现

3　结　语

降低或解决各城市负面主题词背后可能涉及的问题，保持并加强各城市正面主题词涉及的领域，维持正面情感较多区域并改善负面情感较多区域印象才能使各榜单城市更具幸福感。

1）全国各地对榜单的响应整体呈东高西低趋势，响应热度较高区域为榜单城市所在省份及其周边，并以榜单城市为中心呈距离衰减趋势。人口因素和地方依恋影响了话题响应热度的空间分布，旅游与美食发展较好的城市容易引起更多地区的响应。

2）情感响应特征反映出地域独特性。微博用户对长沙大多呈正面响应，长沙的宜居性、经济和房价是主要影响因素；对郑州呈负面响应，微博用户的刻板印象是主要影响因素；青岛、西安、杭州的情感响应省外地区高于省内，青岛与济南之间的矛盾、西安的素质文明水平以及杭州996加班严重是主要影响因素。

3）微博用户对于主题词汇的讨论程度反映出当地的经济水平以及居民对城市的情感期望和意愿。由主题词汇共现频率来看，旅游和美食对城市幸福感具有显著正面影响，房价和工资对城市幸福感具有显著负面影响。

由于微博用户多为青年群体，受限于数据源，可能无法很好地统计儿童、年长者等群体，结论多适用于青年群体，未来将采用多源数据对城市情感响应问题进行补充研究。

参考文献

[1] 顾楚丹,王丰龙,罗峰. 中国城乡居民幸福感的差异及影响因素研究[J]. 世界地理研究,2021,30(1):179-191

[2] Cattaneo M D, Galiani S, Gertler P J, et al. Housing, Health, and Happiness[J]. American Economic Journal: Economic Policy,2009,1(1):75-105

[3] Krueger A B, Kahneman D, Fischler C, et al. Time Use and Subjective Well-being in France and the U. S. [J]. Social Indicators Research,2009,93(1):7-18

[4] Schwanen T, WANG D G. Well-being, Context, and Everyday Activities in Space and Time[J]. Annals of the Association of American Geographers,2014,104(4):83-85

[5] Lehnert K, Sudeck G, Conzelmann A. Subjective Well-being and Exercise in the Second Half of Life: A Critical Review of Theoretical Approaches[J]. European Review of Aging and Physical Activity,2012,9(2):1 059-1 079

[6] 刘义,刘于琪,刘晔,等. 邻里环境对流动人口主观幸福感的影响:基于广州的实证[J]. 地理科学进展,2018,37(7):986-998

[7] 康雷,张文忠. 邻里社会环境对居民幸福感的影响:以北京典型低收入社区为例[J]. 地域研究与开发,2021,40(3):50-55

[8] 张体委. 收入水平、收入差距与主观幸福感研究基于六省份CGSS 2017调查数据的分析[J]. 地域研究与开发,2021,40(3):31-36

[9] 陈子豪,张毅,刘瑜,等. 基于社交媒体感知城市旅游活动类型:以苏州市为例[J]. 地理与地理信息科学,2020,36(2):54-61

[10] 耿硕璘,周旗,李明洁,等. 新媒体用户灾害认知和响应的时空分异及影响因素研究:以山东寿光洪灾为例[J]. 灾害学,2020,35(2):210-216

[11] 千士齐,万红莲,王宝琪,等. “大数据”时代下宝鸡智慧旅游评价与发展策略研究[J]. 河南科学,2018,36(9):1 494-1 500

[12] 靳鑫,王成武,汪宙峰,等. 基于微博签到的成都市夜经济分布特征研究[J]. 地理空间信息,2021,19(8):13-18

[13] 王艳东,刘森保,姜伟,等. 利用社交媒体数据探测商业区顾客空间分布[J]. 地理空间信息,2018,16(2):6-8

[14] 刘启华. 基于LDA和领域本体的竞争情报采集研究[J]. 情报科学,2013,31(4):51-55

[15] 王树义,廖桦涛,吴查科. 基于情感分类的竞争企业新闻文本主题挖掘[J]. 数据分析与知识发现,2018,2(3):70-78

[16] Sievert C, Shirley K E. LDAvis: A Method for Visualizing and Interpreting Topics[C]//Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces,2014:63-70

[17] WANG B, LIU S B, DING K, et al. Identifying Technological Topics and Institution-topic Distribution Probability for Patent Competitive Intelligence Analysis: A Case Study in LTE Technology[J]. Scientometrics,2014,101(1):685-704

[18] 倪鹏飞,李清彬,李超. 中国城市幸福感的空间差异及影响因素[J]. 财贸经济,2012,33(5):9-17

[19] 杨百会. 济南、青岛国家中心城市之争尘埃落定？[J]. 中国经济周刊,2020(20):76-78

[20] 吴海舒. 媒体微博对“河南人”刻板印象的构建与传播:以“人民日报”和“头条新闻”为例[D]. 锦州:渤海大学,2018

[21] 魏咏梅,张硕. 互联网行业“996”现象成因及对策研究[J]. 华北电力大学学报(社会科学版),2019(6):70-77

[22] 韩君伟,董靓. 基于心理物理方法的街道景观视觉评价研究[J]. 中国园林,2015,31(5):116-119

[23] 易成栋,任建宇,高璇. 房价、住房不平等与居民幸福感:基于中国综合社会调查2005、2015年数据的实证研究[J]. 中央财经大学学报,2020(6):105-117