Computer Science and Application 计算机科学与应用, 2018, 8(8), 1259-1269
Published Online August 2018 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2018.88136

Study on Spatio-Temporal Variation of Residents' Night Emotions of Ganzhou Based on Sina Micro-Blog Data

Xiang Du[1], **Xiaoji Lan**[2]

[1]West Campus, Jiangxi University of Science and Technology, Ganzhou Jiangxi
[2]School of Architectural and Surveying & Mapping Engineering, Jiangxi University of Science and Technology, Ganzhou Jiangxi
Email: 404452958@qq.com

Received: Aug. 5th, 2018; accepted: Aug. 18th, 2018; published: Aug. 28th, 2018

Abstract

In the process of China's development, more and more attention has been paid to residents' happiness. With the advent of the "Big Data" era, research data has changed from previous questionnaires to online data. Taking the residents as the research object and the residents' emotions as the focal point, this paper systematically explores the spatial and temporal variation of residents' night emotions with the support of the Sina micro-blog's data. Firstly, using the web crawler collects residents' micro-blog data, and then the microblog text sentiment analysis model is constructed to convert the microblog text into the emotion tendency. Finally, ArcGis software is used to visualize the residents' emotions in temporal, spatial and spatio-temporal three angles, and explore the changing rules and characteristics of residents' night emotions by combining the word frequency of micro-blog's text. Studies have shown that residents' emotional changes at night are mainly affected by factors such as hunger, fatigue, and sleep.

Keywords

Microblog, Big Data, Residents' Emotions, Spatio-Temporal Variation

基于新浪微博数据的赣州市居民夜间情感时空变化研究

杜　翔[1]，**兰小机**[2]

[1]江西理工大学西校区，江西　赣州
[2]江西理工大学建筑与测绘工程学院，江西　赣州

文章引用: 杜翔, 兰小机. 基于新浪微博数据的赣州市居民夜间情感时空变化研究[J]. 计算机科学与应用, 2018, 8(8): 1259-1269. DOI: 10.12677/csa.2018.88136

Email: 404452958@qq.com

收稿日期：2018年8月5日；录用日期：2018年8月18日；发布日期：2018年8月28日

摘　要

在我国发展的过程中，越来越注重居民的幸福感，随着“大数据”时代的到来，研究数据从以前的调查问卷变成了线上数据。本文以居民为研究对象，以居民情感为着力点，在新浪微博的数据支持下，系统全面地探索居民夜间情感的时空变化规律。首先通过使用网络爬虫采集居民微博数据，再构建微博文本情感分析模型将微博文本转化为情感倾向，最后使用ArcGis软件对居民夜间情感变化进行可视化，并结合微博文本词频，探索情感的变化规律和特征。研究表明居民在夜间的情感变化主要受饥饿程度、疲劳程度、睡眠等因素影响。

关键词

微博，大数据，居民情感，时空变化

1. 引言

智能手机的普及、web2.0 技术的日益成熟以及移动网络的迅速发展，使得社交媒体软件迅速发展和被使用。据统计，截至 2017 年 6 月，中国网民人数达到 7.51 亿，其中手机网民占 96.3% [1]，而新浪微博，截至 2017 年 9 月，活跃用户就达到了 3.76 亿[2]，占网民人数的一半以上。大量的用户在新浪微博平台上每天就会产生上万 GB 的数据，这些数据包括文字、图片、视频、地理位置和个人信息等数据。由于这些数据都是用户表达自己当时的情感、所想而产生的，他人干预的因素较小，因此数据能较为真实的反映出用户当时的状态，具有较高的挖掘价值。

Rotton [3]深入的研究发现，每天的气温、天气的干燥或者潮湿、雾霾程度都可以影响居民的情绪，让居民变得更加暴躁甚至具有攻击行为；Welsch [4] [5]则利用二氧化氮气体的排放量来评估环境污染的程度，对比了 54 个国家的数据，研究了居民的主观幸福感与环境污染的关系，研究结果显示，空气污染程度与居民幸福感程度成反比；刘华珍[6]从情感体验和人际关系出发，对萧山居民精神生活状态进行了深入研究，发现女性居民情感状态要比男性的好；陈映雪[7]通过微博数据分析居民情感与建成环境之间的联系；易善君[8]等人通过分析微博数据和空气数据，研究空气质量与居民情感之间的关系，得出空气中对居民情感影响较大的为颗粒物质与有刺激性气味的气体。上述都是研究居民情感与客观因素的关系，而本文则是注重研究居民情感在生活中自身的变化规律。

2. 数据采集及预处理

2.1. 数据采集

目前采集新浪微博数据的主要方法有两种，即新浪微博 API [9]和网络爬虫[10]-[15]。由于新浪微博

并没有将 API 的使用权限完全对开发者开放，而且限制了 API 的调用次数，导致使用新浪 API 无法大量采集微博数据，故使用网络爬虫获取数据。

本文设计并实现了基于 Pthon 的新浪微博爬虫，并获取了赣州市居民微博共 413,450 条，所获取的微博数据内容包括：用户 ID、昵称、用户所在地区、微博发布时间、微博发布位置的坐标、用户性别、用户年龄及微博文本等。为更好地将数据展示出来，现将部分数据从 MongoDB 数据库中导出到 Excel 表中，部分微博数据如图 1 所示。

2.2. 数据预处理

针对从微博上采集的数据，为保证数据的真实性、有效性、和可靠性，在研究数据开始前必须对所有数据进行整理和清洗。微博数据预处理按以下步骤逐步进行：

第一步：数据初步去噪。数据进行分割后，通过 MongoDB 查询语句，将有值缺失的数据进行处理。由于用户年龄、性别、等数据为随机且互相独立的，故无法通过使用插值法补齐，所以针对有值缺失的数据，一律删除，共删除 3184 条微博，剩余 410,266 条。

第二步：进一步去噪。经过初步去噪，保证了现有数据都为完整的数据，由于本文的研究对象为赣州居民，而非旅游者，所以，在微博数据采集时已经将用户所在地不是“江西 赣州”的微博进行剔除，现将所在地为“江西 赣州”但所发微博不在赣州范围内的微博数据剔除，以保证数据尽可能的准确，剔除后共保留数据 326,852 条。

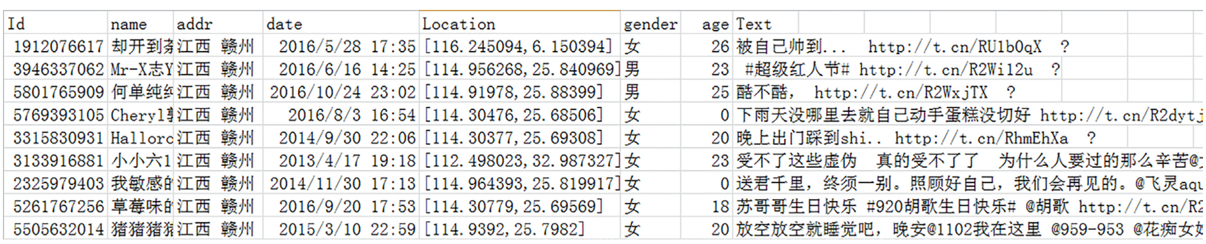
第三步：对微博文本进行精简。通过观察图 1 可以发现，几乎所有的微博文本都有几个共同的部分：一，几乎所有微博文本后都带有一个链接，且该链接与微博文本情感无丝毫关联；二，大部分微博中存在@符，在微博中@表示向某人说或是提到某人的意思，所以@后的部分文本是用户名，这些文字对微博情感也没有任何影响；三，部分微博中存在##符，在微博中##代表话题，话题也不能作为情感判断的语句。所以将微博中这三部分的内容全部清除后，可以得到一条更干净且表达情感更直接的微博文本。这样做不仅可以有效减少计算机处理的数据量而且还可以提高情感分析的准确率。

虽然经过了三个步骤的数据整理和清洗，但仍然有些无法被检测的错误信息存在。如微博发布位置的坐标，用户可以在发布微博时更改真实的位置信息从而使地理坐标失去真实性，但由于本文研究的数据量较大，因此认为这部分错误数据对研究结果的影响相对较小，因此不对此类数据做任何操作。

3. 微博文本情感分析模型构建与情感倾向计算

3.1. 情感分析模型评估指标

本文使用精确度(Precision)、召回率(Recall)、F-度量值(F-Measure)和宏平均(Macro-averaging) [16]作为情感分析模型的评估标准。计算公式如公式 1 至公式 4。



Id	name	addr	date	Location	gender	age	Text
1912076617	却开到	江西 赣州	2016/5/28 17:35	[116.245094, 6.150394]	女	26	被自己帅到... http://t.cn/RU1b0qX ?
3946337062	Mr-X志Y	江西 赣州	2016/6/16 14:25	[114.956268, 25.840969]	男	23	#超级红人节# http://t.cn/R2Wi12u ?
5801765909	何单纯	江西 赣州	2016/10/24 23:02	[114.91978, 25.88399]	男	25	酷不酷， http://t.cn/R2WxjTX ?
5769393105	Cheryl	江西 赣州	2016/8/3 16:54	[114.30476, 25.68506]	女	0	下雨天没哪里去就自己动手蛋糕没切好 http://t.cn/R2dyt
3315830931	Hallorc	江西 赣州	2014/9/30 22:06	[114.30377, 25.69308]	女	20	晚上出门踩到shi.. http://t.cn/RhmEhXa ?
3133916881	小小六1	江西 赣州	2013/4/17 19:18	[112.498023, 32.987327]	女	23	受不了这些虚伪 真的受不了了 为什么人要过的那么辛苦@
2325979403	我敏感	江西 赣州	2014/11/30 17:13	[114.964393, 25.819917]	女	0	送君千里，终须一别。照顾好自己，我们会再见的。@飞灵aq
5261767256	草莓味	江西 赣州	2016/9/20 17:53	[114.30779, 25.69569]	女	18	苏哥哥生日快乐 #920胡歌生日快乐# @胡歌 http://t.cn/R
5505632014	猪猪猪	江西 赣州	2015/3/10 22:59	[114.9392, 25.7982]	女	20	放空放空就睡觉吧，晚安@1102我在这里 @959-953 @花痴女

Figure 1. Partial microblog data
图 1. 部分微博数据

$$\text{Pricision}(c) = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{1}$$

$$\text{Recall}(c) = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{2}$$

$$\text{F-Measure}(c) = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \tag{3}$$

$$\text{Macro-averaging} = \frac{1}{n}\sum\nolimits_{c} \text{F-Measure}(c) \tag{4}$$

其中 TP (真阳性)是相关项目中正确识别为相关的，TN (真阴性)是不相关项目中正确识别为不相关的，FP (假阳性)是不相关项目中错误识别为相关的，FN (假阴性)是相关项目中错误识别为不相关的，c 为分类属性(为积极，中性，消极)，n 为分类数目(n 为 3)。模型的整体性能由宏平均值决定，宏平均值越高说明分类模型的整体性能越好。

3.2. 语料库的制作

本文通过从获取的微博文本中随机提取 4600 条作为待分类文本，通过去重，最终得到 4585 条，使用人工标注其情感倾向。由于人工对微博文本情感分析带入了个人感情色彩，有可能对同样的语句产生不同的理解从而导致不同的情感倾向分类，故而将这 4585 条微博文本同时交于三人分类，待三人都分类完成后再进行比较，如果针对某条文本三人分类一样，则取该值作为文本的分类，如果两人一样一人不一样，则取有两人一样的值作为文本的分类，如果三人分类都不一样，则交由第四人进行情感倾向分类(第四人并不知道前三人对文本情感倾向的分类情况)，之后再取有两人一样的值作为该文本的分类。通过这种方式可以保证每条微博文本至少有三人对它进行过阅读、分析，可以得到一个可靠的语料库。

经过人工文本情感分析，得到积极语料 2139 条，中性语料 1601 条，消极 845 条。由于是随机抽取，理论上应与实际的分布相吻合即积极约占 46.65%、中性约 34.91%、消极约占 18.44%。该语料库将用于微博文本情感分析模型分类效果的验证。部分语料库如图 2 所示。

3.3. 微博文本情感分析模型的构建

本文使用基于情感词典与语义规则的情感分析模型来计算微博文本的情感值，该模型的中心思想就是先将文本分成若干个语义完整的句子，之后提取每个句子中的情感词，根据情感词典和语义规则求出每个句子的情感值，最后再根据语义规则求出整个文本的情感值。

3.3.1. 基础情感词典的构建

本文汇总了清华大学李军中文褒贬义词典、台湾大学 NTUSD 简体中文情感词典和知网 Hownet 情感

Text	Emotion
这个人！！！骗子！！！被骗上贼船	-1
安定点吧。累死了真是的	-1
把持住没有把半月生活费买黄牛票，于是在场外听演唱会也不错，一个人照样听着high到爆啊爱死你了	1
今天被画了鬼脸 难以直视	-1
待我毛发及腰，去修剪一下！	0
〈须改变教育评价体系〉有评论指出，"中国式暑假"让孩子身心俱疲，让家长不胜其累，也让社会为教育的现状和未来感到困惑和担扰。	-1
作为人类降生于世，吃是我至高无上的使命！	1
回家几日，第一次见，窗前明月光。	0
3吨桶装冰淇淋出生，从产房送到活动中心…	0

Figure 2. Partial corpus data
图 2. 部分语料库数据

词典，通过去重合并后得到积极词汇 14,443 个(含英文)，消极词汇 18,299 个(含英文)，程度词 328 个(含英文)，其中积极词汇权值为 1，消极词汇权值为−1，又把程度词分为“最”、“很”、“较”、“稍”、“欠”五个程度，权值分别为 2.0、1.75、1.5、1.25、0.5。另收集转折词 15 个，否定词 19 个，基本的情感词典就初步构建完成。

3.3.2. 情感词典的修改更新

由于网络词语更新十分迅速，基础词典中存在着一些包含个人丰富情感的网络词语没有被收录的问题，从而导致该模型对部分带有新生网络词语的微博文本情感分析结果不准确，所以需要对基础情感词典进行一个更新与修改。将流行的网络情感词汇添加到情感词典中，如“嗨”、“嗨皮”、“囧”、“么么哒”、“巴适”、“给力”等等共 1231 个，将一些情感不明的词语从情感词典中删除如热、光顾、动态平衡、上等等共 239 个。

3.3.3. 语义规则

设文本 t 经过分句得到 n 个语义完整的句子分别为 S1、S2、……、Sn，语义规则如下：

一，如果句子 Si 中只包含情感词，例如“她漂亮”，其中“漂亮”为积极词汇，则整个句子也是积极倾向，所以这类句子的情感值计算公式如下：

$$w_i = \sum_{j=1}^{a} e_i \tag{5}$$

其中 w_i 为句子 Si 的情感值，a 是句子中情感词的数量，e_j 为第 j 个情感词的情感权值。

二，如果句子 Si 中不存在情感词，例如“我在吃饭”，该句子并没有任何情感倾向，所以这类句子的情感值计算公式如下：

$$w_i = 0 \tag{6}$$

三，如果句子 Si 中情感词前存在否定词，则情感倾向就发生了反转，如“她不漂亮”，其中“漂亮”是积极词汇，但在前加了一个否定词，则使整个句子变成了消极，所以这类句子的情感值计算公式如下：

$$w_i = \sum_{j=1}^{a} \left((-1)^{b_j} * e_i \right) \tag{7}$$

其中 b_j 表示第 j 个情感词前否定词的个数，若 b_j 为奇数则句子情感与情感词情感相反，若为偶数则句子情感与情感词情感相同。

四，如果句子 Si 中情感词前存在程度词，如“她非常漂亮”，其中“非常”加强了“漂亮”的程度，使得句子的积极程度更高，所以这类句子的情感值计算公式如下：

$$w_i = \sum_{j=1}^{a} \left(\prod_{f=1}^{k_j} d_{jf} * e_j \right) \tag{8}$$

其中 k_j 表示第 j 个情感词前程度词的个数，d_{jf} 表示第 j 个情感词前的第 f 个程度词的权值。

五，如果句子 Si 中情感词前即存在否定词也存在程度词，例如“她非常不漂亮”，其中“非常”加强了“不漂亮”的程度，所以所以这类句子的情感值计算公式如下所示

$$w_i = \sum_{j=1}^{a} \left((-1)^{b_j} * \prod_{f=1}^{k_j} d_{jf} * e_j \right) \tag{9}$$

六，如果 S1、S2、……、Sn 中出现了转折词且最后出现转折词的句子为 Sm，例如“今天好累啊，但是很开心。”，可知文本的主要情感是由转折词后的句子决定的即“很开心。”，所以文本 t 的情感值为最后出现转折词的句子及之后句子的情感值之和，计算公式如下：

$$S = \sum_{i=m}^{n} w_i \tag{10}$$

其中 S 为文本 t 的情感值。

七，如果 S1、S2、……、Sn 中都没有出现转折词，则文本 t 的情感值为所有句子的情感值之和，计算公式如下：

$$S=\sum_{i=1}^{n} w_i \tag{11}$$

通过以上规则，可以计算出文本的情感值，再通过判断情感值所在区间，得出文本情感倾向。由于本文是粗分类，所以将情感值大于 0 的判定为积极情感倾向，等于 0 的为中性，小于 0 的为消极情感倾向。

结合上诉情感词典和语义规则，基于情感词典与语义规则的情感分析流程如下：

第一步，文本分句。根据常用的标点符号，如“。、……，？！”等将文本分割成若干个句子。

第二步，分词。使用前文介绍的包含网络用语词典的 jieba 分词工具对每个句子进行分词。

第三步，情感词定位及句子情感值计算。根据情感词典，提取出每个句子中的情感词及情感词所在位置，并找出情感词前的程度词和否定词，计算出每个句子的情感值。

第四步，文本情感值计算。判断句子中是否存在转折词，如果存在则对最后出现转折词的句子及之后句子的情感值进行求和，结果作为文本情感值，如果不存在转折词对则所有句子的情感值进行求和，结果作为文本的情感值。

整体流程如图 3 所示。

3.3.4. 情感分析效果评估

由于本文研究的是粗分类即将文本情感倾向划分为积极、中性、消极，所以情感值大于 0 的即为积极，等于 0 的为中性，小于 0 的为消极。通过上诉文本情感分析模型和语料库得到评估数据如表 1 所示。

3.4. 情感倾向计算

使用上诉文本情感分析模型计算出所有微博的情感值，并转换为微博文本的情感倾向，情感倾向整体结果如表 2 所示，与语料库的比例较为接近。

4. 居民情感时空变化分析

4.1. 研究区域的提取

由于本文研究是基于微博数据进行的，如果区域中微博数量过少则无法反映居民的真实情感规律及特征，所以要将微博数据密度较高的区域提取出来作为研究区域。

首先利用 ArcGis 软件先将所有微博数据定位到空间，发现微博数据较为分散(如图 4(a)所示)，再使用核密度分析和自然间断点分级法将密度分为极低、低、中、高、极高五级，为了显示更加清晰，所以没有显示数据密度过低的区域(如图 4(b)所示)，发现只有一个区域微博密度极高，而其他区域微博密度都为低或极低，最后使用半径为 10 KM 的圆将密度高的区域完全选取，又因为极大部分数据都在章贡区内，所以对微博数据进行剪裁，经过剪裁得到研究区域(如图 4(d)所示)。

通过观察图 4 可以发现：一，微博分布与赣州水系分布基本相似；二，微博数据集中在章贡区内，其他区域微博密度极低，所以可以认为章贡区内居民的情感特点可以代表赣州市居民的情感特点。

4.2. 居民情感时空变化研究方法

为了了解居民的情感的时空变化情况，本文通过将居民情感时空变化与该时段下的高频词汇相结合，从而更加深入、准确的分析出导致居民情感时空变化因素，对居民情感在空间上的变化规律进行分析与总结。

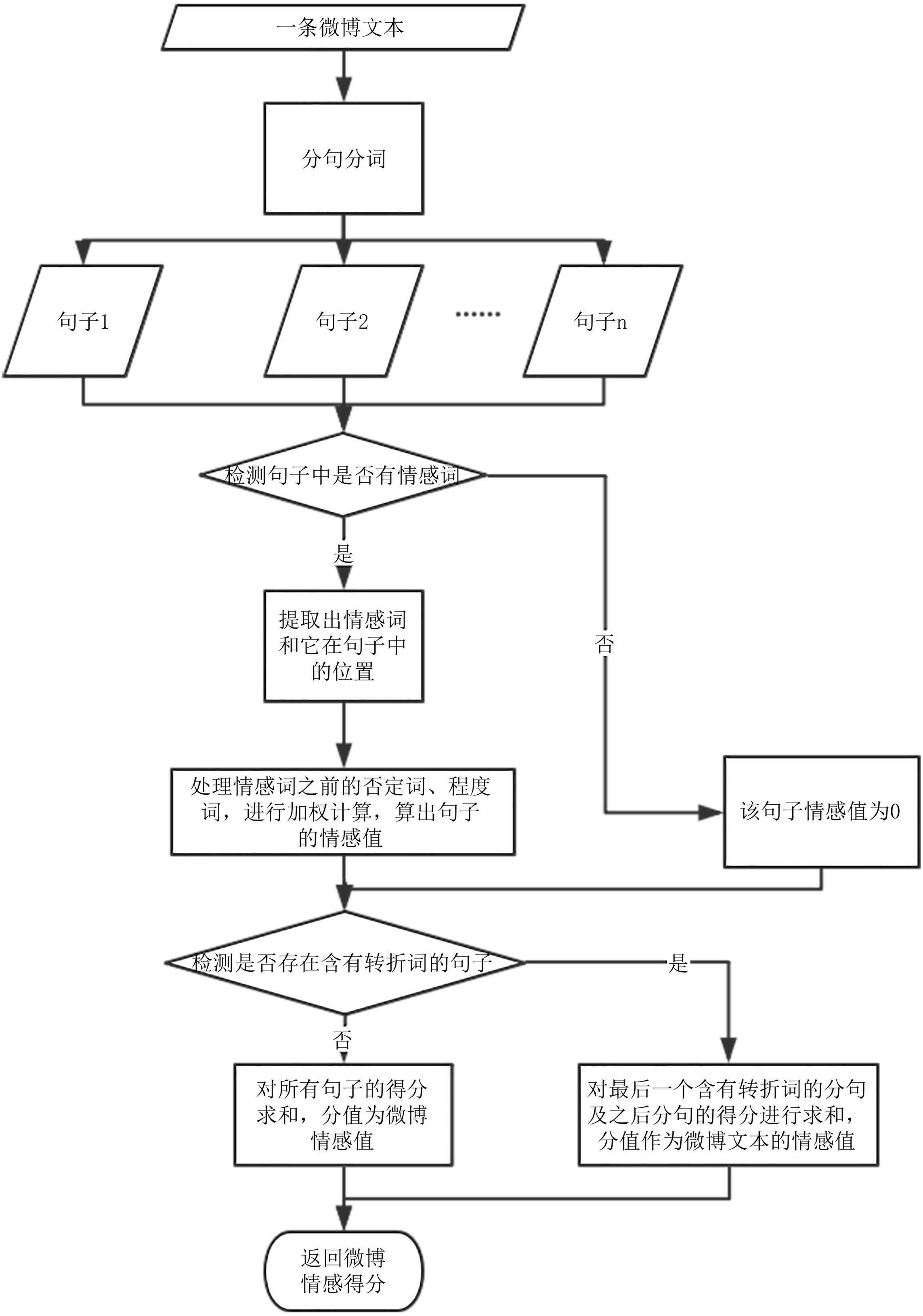


Figure 3. Flow chart of sentiment analysis model

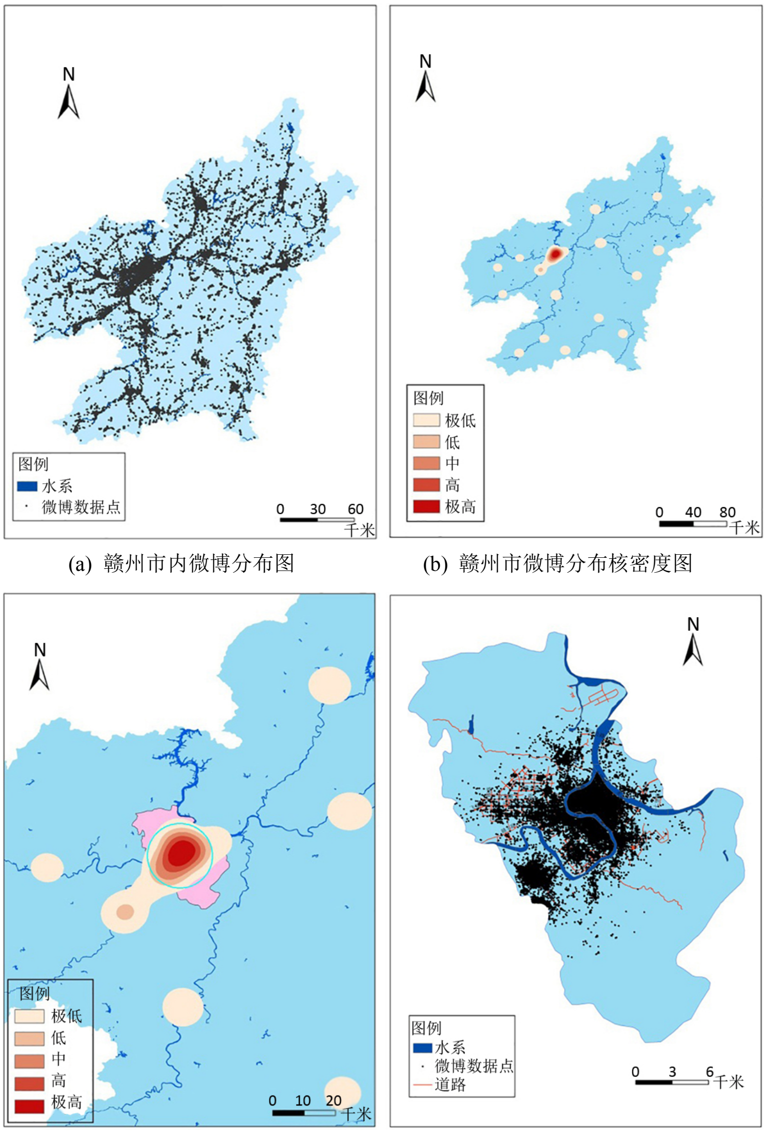
图 3. 情感分析模型整体流程图

Table 1. Assessment data of sentiment analysis model
表 1. 情感分析模型评估数据

准确率	F 度量值(积极)	F 度量值(中性)	F 度量值(消极)	宏平均
0.810905	0.789818	0.774278	0.581362	0.715152

Table 2. Proportion distribution of sentiment in microblog text
表 2. 微博文本整体情感倾向比例分布

情感倾向	情感值范围	数量	比例
积极情感	$(0,+\infty)$	160843	49.21%
中性情感	0	111097	33.99%
消极情感	$(-\infty,0)$	54912	16.80%



(a) 赣州市内微博分布图 (b) 赣州市微博分布核密度图
(c) 裁剪密度最高区域 (d) 裁剪后的微博分布图

Figure 4. Process chart of study area extraction
图 4. 研究区域提取过程图

关于居民情感的时空变化可视化方法，先使用 ArcGis 中的渔网工具生成 500*500 米的网格，之后计算网格内的情感均值，再利用后一个时间段的均值减去前一个时间段的，得到情感变化数据，最后根据数据值的大小分为三类：如果结果大小范围在之间[−0.1,0.1]，情感均值波动较小，则认为该网格情感均值较前一个时间段基本无变化，在(0.1,2]之间，则认为该网格积极情感较前一个时间段有增加趋势，在[−2,−0.1)之间则认为该网格积极情感较前一个时间段有减少趋势。最后通过给网格的三种分类分配不同的颜色进行可视化，得到居民情感时空变化图。

4.3. 居民夜间情感时空变化

本文将微博数据按照 24 小时划分，得到发布微博数与时的关系变化曲线如图 5 所示。

根据图可以发现从 17 时开始到 23 时布数呈上升趋势，且在 0 时微博发布数也较多。根据居民“朝九晚五”的作息规律和微博发布数，故将夜间时间范围定义在 17 时至 0 时。在夜间时段，居民能够自由活动，受到的干涉较少，其微博情感更有意义；在这个时间段内，居民发布微博数量较多，居民的情感表达更为丰富。

首先将 17 时至 0 时共 8 小时分为 4 个时间段，分别是：时段 1，17 时到 18 时(17 点 0 分 0 秒到 18

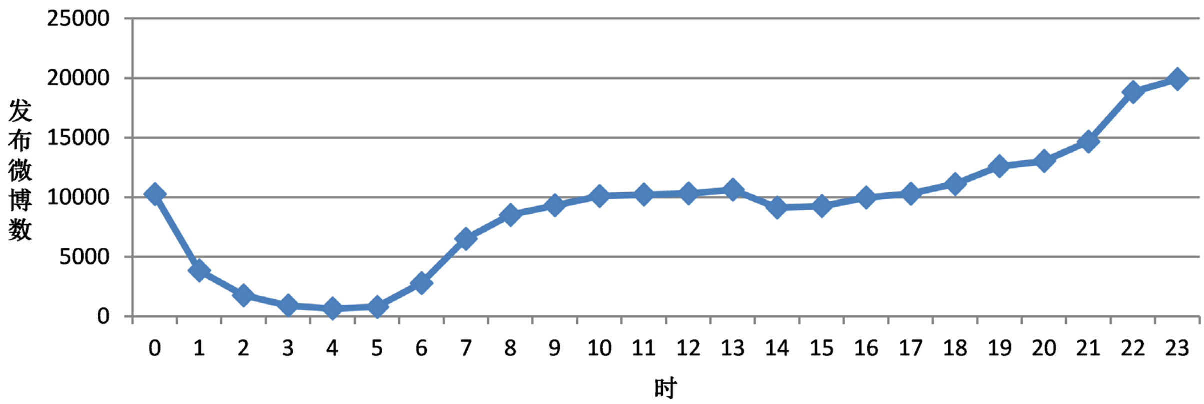


Figure 5. The relationship between the publish number of microblog and the time
图 5. 发布微博数与时的关系变化曲线

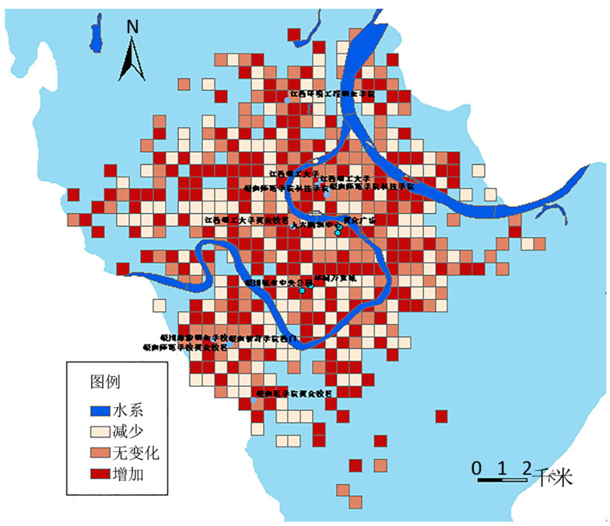


Figure 6. The variation of residents in period 1 to period 2
图 6. 时段 1 到时段 2 居民情感变化

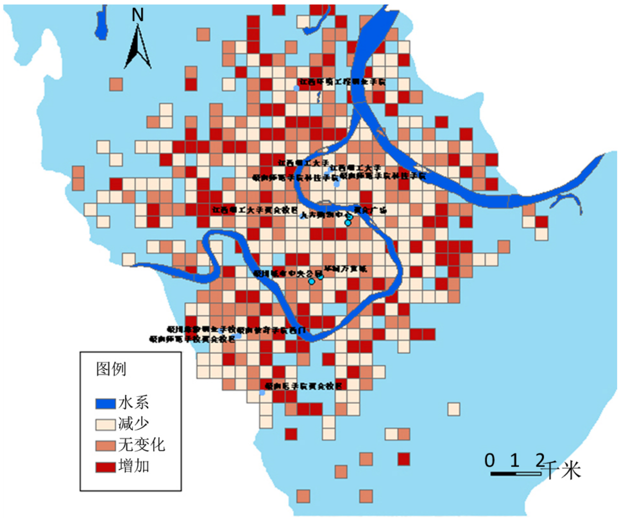


Figure 7. The variation of residents in period 2 to period 3
图 7. 时段 2 到时段 3 居民情感变化

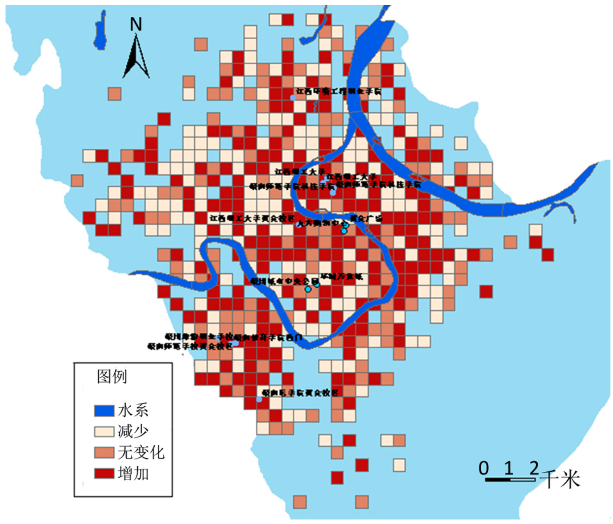


Figure 8. The variation of residents in period 3 to period 4
图 8. 时段 3 到时段 4 居民情感变化

点 59 分 59 秒)；时段 2：19 时到 20 时(19 点 0 分 0 秒到 20 点 59 分 59 秒)；时段 3：21 时到 22 时(21 点 0 分 0 秒到 22 点 59 分 59 秒)；时段 4：23 时到 0 时(23 点 0 分 0 秒到 0 点 59 分 59 秒)，再使用 ArcGis 得到居民夜间情感的时空变化图如图 6 至图 8 所示。

通过观察图 6 到图 8，可以发现以下变化特征：

1) 时段 1 到时段 2 (图 6 所示)，高校积极情感变化较为平稳且有增加趋势，在万象城、赣州市中央公园、九方购物中心、黄金广场等地居民积极情感增加趋势明显。在 19 时微博中“吃”、“喜欢”、“买”等词出现频率较高，可知 19 时段大部分居民处于吃饭或者购物阶段，而万象城和九方购物中心刚好能满足居民的这两项需求，所以居民情感不断增加，部分居民也会选择饭后去公园散步，所以黄金广场和赣州市中央公园积极情感也在增加，其他区域的变化特征则反映了居民情感的一种波动性。

2) 观察时段 2 到时段 3 (图 7 所示)，万象城、赣州市中央公园、九方购物中心、黄金广场等地居民积极情感较为平稳但有减少趋势，居民整体积极情感开始减少。在该时间段，微博中“累”、“心累”、

“困”等词的出现频率升高，导致居民积极情感出现减少，但在万象城、九方购物中心等地变化平缓，说明居民在这些地方更容易产生积极情感，受到“累”等影响较小。

3) 观察时段 3 到时段 4 (图 8 所示)，可以发现，居民整体情感又开始上升，在高校和住宅区尤其明显，但在万象城、九方购物中心等地居民积极情感有所下降。该时段微博中“晚安”等词的词频较高，说明大部分居民在此时段进入睡眠且通过微博表达了居民睡前的美好祝愿和对明天的憧憬，所以高校和住宅区积极情感上升，而万象城、九方购物中心等地微博数量在此时段已经较少，积极情感出现了下降。

5. 结论

通过结合微博高频词汇对居民夜间情感时空变化特点与规律进行了分析，总结起来有以下几点：

1) 居民夜间情感时空变化较为明显，整体的居民积极情感呈现先增、稍减、再增的趋势。

2) 居民在消费或是公园散步时积极情感比较容易产生且受到疲倦的影响较小。

3) 居民在准备休息时，会通过微博表达睡前的美好祝愿和对明天的憧憬，居民情感较为积极。

基金项目

2015 年赣州市社联社科课题(编号：15087)。

参考文献

[1] 中国互联网信息中心. 第 40 次中国互联网络发展状况统计报告[EB\OL]. http://www.cac.gov.cn/2017-08/04/c_1121427672.htm, 2017-08-04.

[2] 新浪微博数据中心. 2017 微博用户发展报告[EB/OL]. http://data.weibo.com/report/reportDetail?id=404, 2017-12-25.

[3] Rotton, J. and Frey, J. (1985) Air Pollution, Weather, and Violent Crimes: Concomitant Time-Series Analysis of Archival Data. *Journal of Personality and Social Psychology*, **49**, 1207-1220. https://doi.org/10.1037/0022-3514.49.5.1207

[4] Welsch, H. (2002) Preferences over Prosperity and Pollution: Environmental Valuation Based on Happiness Surveys. *Kyklos*, **55**, 473-494. https://doi.org/10.1111/1467-6435.00198

[5] Welsch, H. (2006) Environment and Happiness. *Ecological Economics*, **58**, 801-813. https://doi.org/10.1016/j.ecolecon.2005.09.006

[6] 刘华珍. 萧山居民精神生活状况研究[D]: [硕士学位论文]. 武汉: 华中科技大学, 2006.

[7] 陈映雪. 居民情绪与建成环境关系研究及规划应用初探[D]: [硕士学位论文]. 南京: 南京大学, 2015.

[8] 易善君, 李君轶, 李秀琴, 等. 基于微博大数据的空气质量与居民情感相关性对比研究——以西安市和上海市为例[J]. 干旱区资源与环境, 2017, 31(5): 39-44.

[9] 新浪. 微博 API [EB/OL]. http://open.weibo.com/wiki/, 2017-07-10.

[10] 廉捷, 周欣, 曹伟, 等. 新浪微博数据挖掘方案[J]. 清华大学学报: 自然科学版, 2011, 51(10): 1300-1305.

[11] 李俊丽. 基于 Linux 的 Python 多线程爬虫程序设计[J]. 计算机与数字工程, 2015,43(5): 861-863.

[12] 陈琳, 任芳. 基于 Python 的新浪微博数据爬虫程序设计[J]. 信息系统工程, 2016, 9(20): 97-99.

[13] 刘艳平, 俞海英, 戎沁. Python 模拟登录网站并抓取网页的方法[J]. 微型计算机应用, 2015, 31(1): 58-60.

[14] Gao, K., Zhou, E.L. and Grover, S. (2014) Applied Methods and Techniques for Modeling and Control on Micro-Blog Data Crawler. *Lecture Notes in Control and Information Sciences*, **452**, 171-188.

[15] Ling, X., Ling, J. and Bao, P. (2016) A Novel Crawler Based on Loginning Simulation for Weibo Social Network. *International Journal of Grid and Distributed Computing*, No. 9, 135-144.

[16] Figueiredo, F., Rocha, L., Couto, T., Salles, T., *et al.* (2011) Word Co-Occurrence Features for Text Classification. *Information Systems*, **36**, 843-858. https://doi.org/10.1016/j.is.2011.02.002

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8801，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：csa@hanspub.org