引文格式：吕蓓茹，彭　玲，陈嘉辉，等. 基于社交媒体的森林火灾舆情信息脉动分析[J]. 地理信息世界, 2021, 28(3):61-66.

基于社交媒体的森林火灾舆情信息脉动分析

吕蓓茹[1,2]，彭　玲[1]，陈嘉辉[3]，陈若男[1,2]，葛星彤[4]

（1. 中国科学院　空天信息创新研究院，北京　100094；2. 中国科学院大学　资源与环境学院，北京　100049；
3. 北京林业大学　信息学院　北京　100083；4. 北京林业大学 林学院　北京　100083）

作者简介：
吕蓓茹(1996—)，女，内蒙古包头人，地图学与地理信息系统专业硕士研究生，主要研究方向为多源时空大数据智能分析
E-mail:
lvbeiru@qq.com
通信作者：
彭玲(1965—)，女，湖北武汉人，研究员，博士，博士生导师，主要从事遥感影像智能分析与决策支持研究工作
E-mail:
plqiqi@126.com
收稿日期：2021-02-05

【摘要】2019年和2020年四川省发生了两起大型森林火灾，受到政府高度重视，也在网络上引发了广泛讨论。为了呈现森林火灾后微博文本中蕴含的舆情信息，有效地了解舆情，掌握规律，对四川凉山前后两起重特大森林火灾发生后的舆情进行了数据挖掘和对比分析。使用核密度、地理探测器方法对两起森林火灾舆情时空扩散和空间分异进行了研究，使用LDA主题提取模型、朴素贝叶斯、词云方法对两起火灾舆情进行主题提取、情感分析和可视化表达。研究结果表明：时空扩散和主题分布上，四川省两起重大森林火灾舆情具有较强相似性；空间分异上，两起火灾舆情空间分异与区域经济发展水平显著相关；情感演变上，重复发生同类灾害事故对于网民的情绪冲击明显。
【关键词】关键词：森林火灾；社交媒体；舆情分析；地理探测器；LDA文档主题生成模型
【中图分类号】P208　　【文献标识码】A　　【文章编号】1672-1586（2021）03-0061-06

Analysis of Public Opinion Information Pulsation of Forest Fire on Social Media

LV Beiru[1,2], PENG Ling[1], CHEN Jiahui[3], CHEN Ruonan[1,2], GE Xingtong[4]

(1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. School of information science &technology Beijing Forestry University, Beijing 100083, China; 4. School of forestry Beijing Forestry University, Beijing 100083, China)

Abstract: In order to excavate the public opinions contained in Weibo text, this study conducted a comparative analysis of the public opinion after two serious forest fires in Liangshan. Leveraging the nuclear density and geographic detector methods are used to investigate the temporal-spatial diffusion and geographic differentiation of the two fires, and the LDA theme extraction model, naive Bayes and word cloud method are used for the analysis of theme distribution, sentiment classification and visual expression of public opinion on Weibo in this study. The results show that the two major forest fires in Sichuan Province have strong similarity in temporal and spatial diffusion and subject distribution. The geographical distribution of the two fires is significantly related to the regional economic development level. In terms of sentiment evolution, compared with the 2019 Muli forest fire, netizen's negative feelings about the 2020 Xichang forest fire significantly increased in a short time after the fire broke out.
Key words: forest fire; social media; public opinion analysis; geodetector; linear discriminant analysis(LDA)

0　引　言

随着大数据、云计算、互联网、智能终端的高速发展，社交媒体已经成为互联网中最活跃的应用类型之一。社交媒体凭借其巨大的用户基数、快速的信息传播速度、高效的互动能力，已经成为突发热点事件、社会舆情信息传播的重要载体。2020年2月10日，习近平总书记在北京调研指导疫情防控工作时做出“要加强舆论引导工作”的重要指示[1]。政府相关部门应及时了解舆情发展状况和公民态度，根据处置进展动态发布信息，促进信息通畅和民心凝聚。因此，从社交媒体数据中挖掘舆情信息，快速分析舆情热点主题、发展态势、空间分布特征、时间演变规律，已是政府部门的关注点和刚性需求。

目前基于社交媒体数据的舆情分析研究已有一定的基础。国外学者Nair等使用机器学习方法，对2015年洪水灾害的推特文本进行了分类，并提取了网络意见领袖[2]；Alfarrarar jeh等提出地理情感分析方法，并基于多源社交媒体数据在两例突发自然灾害事件上进行了实验[3]。国内学者张庆民等通过熵权法建立了城市公共安全舆情评价的多属性决策模型[4]；刘继等[5]构建了网络舆情基本特征挖掘体系，将机器的定量计算和决策者的定性分析相结合，

构建舆情智能监测机制。彭玲团队也通过社交媒体数据进行舆情脉动分析研究[6]，实现在突发事件发生前后，挖掘舆情时空传播规律及网民话题关注度与情感波动程度。

在前人的工作基础上，本文将针对森林火灾这类突发事件进行社交媒体数据的舆情信息挖掘与分析研究。在中国，新浪微博是一个提供微型博客服务类的社交网站，通过用户发布、转发微博等方式实现网络舆情信息扩散，拥有庞大的用户群体，月活跃用户数量已经达到了5.5亿[7]，是社会舆论传播和发酵的主要平台之一。凉山州两起森林火灾的微博话题讨论热度分别达到了133万和37万，因此本文将选用新浪微博数据，通过挖掘微博数据中潜在的时间信息、地理位置信息、情感信息，并基于地学分析与机器学习方法，在发生森林火灾后进行网民重点关注主题挖掘、空间尺度上的舆情分异规律分析、时间尺度上的舆情演变特征挖掘等研究。期望能够为突发事件应急管理科学决策提供辅助技术支撑，帮助相关部门及时地了解舆情的发展趋势，及时发布有针对性的灾情信息，有侧重点地展开民众情绪疏导工作。

1 数据获取及预处理

首先在研究对象上，本文选取了凉山州两起森林火灾进行研究：

2019年3月30日18时，四川省凉山州木里县雅砻江镇立尔村发生森林火灾，3月31日下午，30名救火人员突遇山火爆燃后失联，火灾总计造成31位救火人员牺牲。

2020年3月30日15时，四川省凉山州西昌市突发森林火灾，3月31日凌晨，21名救火人员在去往泸山背侧火场指定地点集结途中失联，火灾总计造成19名救火人员牺牲。

这两起森林火灾均发生在四川省凉山州，时间和事件性质高度相似，扑救过程中均发生多名救火人员牺牲，引起政府的高度重视，也同时在网络上引发了广泛讨论。

本文采用爬虫技术，根据微博关键词功能，将木里森林火灾和西昌森林火灾分别作为关键词爬取了从火灾发生日2019年3月30日至7月1日、2020年3月30日至7月1日两起火灾微博数据以及相应微博评论数据，并根据微博作者用户名获取作者所在地，通过百度地理编码API获取用户对应地理位置。最终经过数据清洗，两起火灾对应的微博正文数据量分别达到了10401条和15645条，微博评论数据分别达到了13341条和19673条。其中凉山木里森林火灾数据少于凉山西昌森林火灾的原因在于事件已经过去一年，部分微博已经被作者删除。此外人口数据、GDP总量数据均来自国家统计局（人均GDP根据分省GDP总量和人口数量计算得到）公布的2019年、2020年全国统计数据。

2 微博舆情时空扩散规律及空间分异

本文以微博数量为研究对象，基于数学统计及核密度方法分析了舆论扩散在时间和空间上的分布特征。基于地理探测器探究了人口、经济、距事发地距离3个因素与空间分布特征间的相关性，尝试探索造成微博空间分异性的原因。

2.1 基于核密度的舆情时空扩散分析及可视化

舆情扩散具有周期特性，伴随着突发事件信息的发布而发生，通过转发、评论而扩散，最终也会因事件热度的下降而消亡[8]。为了发掘森林火灾事件在微博中扩散的时间规律，本文对获取的微博舆情数据时间进行了时序统计分析，两起火灾均在事件发生后15天完成了第一轮舆情的爆发式增长和基本平息，随后由于火场复燃或新的森林火灾又引发了短暂的舆论高峰，如2019年6月15日木里再次发生森林火灾、2020年4月21日凉山州再次发生火灾等（图1）。

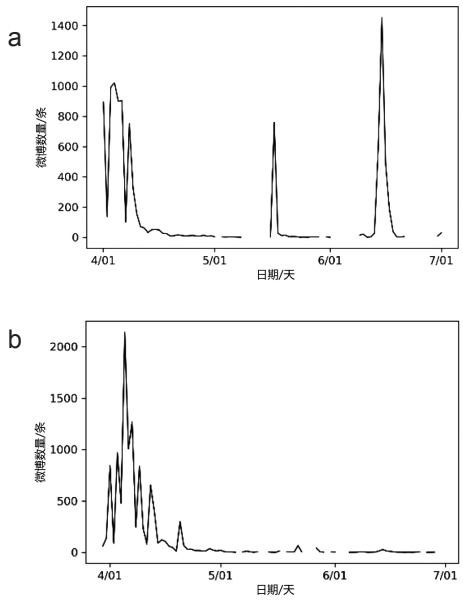

图 1 2019木里森林火灾(a) 2020西昌森林火灾(b) 微博数量演变
Fig. 1 Quantitative evolution of Weibo about 2019 Muli forest fire (a) and 2020 Xichang forest fire (b)

核密度分析用于计算要素测量值在指定邻域范围内的单位密度，能够直观反映离散测量值在连续区域内的分布

情况，对于点对象，其核密度曲面与下方平面所围成的空间体积近似于此点的测量值。在时间扩散研究的基础上，为进一步探究林火舆情的空间扩散规律，本文使用核密度方法对3个月内参与讨论的微博用户根据其所在地进行舆情空间扩散分析。分析结果显示，整体上舆情分布热点主要集中在四川省成都市、长三角、珠三角、京津冀、港澳特别行政区附近，在四川省周边地区及东部其他省份相对均匀分布，在新疆、西藏、青海、甘肃、内蒙古几个西北省份分布很少。从响应速度上看，4月1日第一批快速响应热点与最终舆情集中分布热点相同，到4月15日第一轮舆情平息，整体分布接近最终舆情分布结果。

通过对两起森林火灾舆情时空扩散分析，可以直观发现，虽然两起森林火灾的发生地均为四川省凉山州，但凉山州并未成为微博舆情的最大热点地区，四川省内的舆论热点出现在省会成都市。从全国范围来看，其余几个热点都分布在人口密度比较大、经济发展水平较高的北上广、东南沿海一带，整体舆情分布以东南沿海一带以及距离火灾发生地点较近的四川省周边几个省份为主。

2.2 基于地理探测器的空间分异规律分析

为进一步探究微博舆情空间分异的背后驱动力，本文采用地理探测器对微博数量与人口、经济、距离事发地距离进行相关性探究。其核心思想是利用统计学方法并基于假设，如果某个自变量对某个因变量有重要影响，那么自变量和因变量的空间分布应该具有相似性[9-10]，进行空间分异相关性度量。地理探测器包括分异及因子探测器、风险区探测器、生态探测器、交互作用探测器4种。分异及因子探测器用于探测因变量Y的空间分异性，从而探测某一个因子X在多大程度上解释了属性Y的空间分异，通过q值[9]进行度量，公式如下：

$$q=1-\frac{\sum_{h=1}^{L}N_h\sigma_h^2}{N\sigma^2} \quad (1)$$

式中，N与σ^2分别为研究区域中的分区单位数和区域微博数量的方差，微博数量Y由L层组成（$h=1,2,\cdots,L$）；q的值严格在[0，1]之内，q=0表示Y与X之间没有耦合；q=1表示Y完全由X决定。

p值(Probability)即概率，是进行检验决策的一个依据，反映某一事件发生的可能性大小。统计学根据显著性检验方法得到的p值，一般以$p<0.05$为有统计学差异。

本文主要研究不同人口、经济、距事发地距离几个因素对微博数量分异影响程度，从而判断其相关性。因此选用地理探测器中的因子探测器[11]进行分析（表1）（表2）。

表 1 2019木里森林火灾微博数量分布因子探测结果

Tab.1 Factor detector results of Weibo numbers about 2019 Muli forest fire

	人口	GDP 总量	人均 GDP	与凉山州距离
q 值	0.167	0.196	0.482	0.268
p 值	0.414	0.856	0.002	0.652

表 2 2020西昌森林火灾微博数量分布因子探测结果

Tab.2 Factor detector results of Weibo numbers about 2020 Xichang forest fire

	人口	GDP 总量	人均 GDP	与凉山州距离
q 值	0.232	0.136	0.778	0.188
p 值	0.795	0.647	0.010	0.918

地理探测器的因子探测结果显示，微博数量分布与人均GDP有较强联系，q值分别为0.482和0.778，而p值则分别达到了0.002和0.010。微博数量分布与人口、GDP总量、距离凉山州的距离3个因子则没有显著相关。

结合地理探测器与核密度分析结果，表明微博数量受到人均GDP显著影响，参与微博讨论的用户在京津冀、长三角、珠三角、港澳特别行政区等经济高度发达地区有集中分布，整体由东南沿海经济发达区向西北欠发达地区递减。除事发地四川省外的其他地区，微博发布数量与距事发地的距离没有显著相关。研究表明参与微博讨论的网民大部分集中在经济发达地区，因此在进行舆情分析时，除了关注事发地附近的舆情动向外，也需要格外关注人口集中的经济发达地区的舆情动向。

本文通过时序分析、核密度分析及地理探测器对舆情的时空扩散规律进行了分析，但以上分析都是基于相关微博数量来反映舆情热度，没有研究微博所蕴含的语义信息。为了进一步了解舆情讨论主题，下面在前续研究基础上进行了舆情可视化表达及主题分析。

3 微博舆情可视化表达及主题挖掘

本文以微博语义为研究对象，以词为研究单位，基于机器学习方法对两起森林火灾进行舆情主题可视化表达与主题挖掘，以揭示民众重点关注对象。

3.1 基于WordCloud的微博舆情可视化

为了获取两起火灾发生后网民所关心的核心主题，采用WorldCloud方法进行词频的可视化表达。

首先根据两起火灾的微博数据分别构建了各自高频词云。词云可视化结果显示，高频词主要分为3个类型：地点相关、救火人员相关、救火工作相关。高频地点词主要是火灾发生地地名，如四川、凉山、西昌、木里等；救火人员相关词主要有消防员、英雄、牺牲、烈士等；救火工作相关词主要有地方、扑救、灭火、烟点等（图2）。

a b

图 2 2019木里森林火灾(a)与2020西昌森林火灾(b)高频词云

Fig. 2 Word cloud of Weibo about 2019 Muli forest fire (a) and 2020 Xichang forest fire (b)

3.2 基于机器学习的微博舆情主题挖掘

在词频可视化结果基础上，采用LDA算法进行关键主题挖掘。在机器学习领域中，文档主题生成模型LDA (Latent Dirichlet Allocation)占有重要地位，常用来挖掘大数据环境下语料库中隐藏的主题信息。为进一步研究两起森林火灾事件主题分布，对两起火灾发生日至当年7月1日全部微博正文数据进行了主题提取。主题提取方法使用gensim中的类实例化LDA主题模型，对数据清理后的文本进行分类训练。

新浪微博限制每条微博可以发表140字，当超过140字时会提示变为长微博。本文所用的微博平均字数为71字。根据实验数据分析，一条微博的主题个数一般在1～20个，故拟定1～20区间内的整数作为候选主题数。通过对困惑度(perplexity)评价指标来确定文档最优主题数，困惑度常用来度量一个概率分布或概率模型预测样本的优劣程度，可用于调节主题个数[12]，计算公式如下：

$$Perplexity(W \mid M) = \exp - \frac{\sum_{m=1}^{M} \log p(W_m \mid M)}{\sum_{m=1}^{M} N_m} \quad (2)$$

式中，M为微博正文数量；W为微博正文中的每个词，W_m表示来自微博正文m中的词W；$p(W_m|M)$表示正文M中出现词W的概率；Nm表示每条微博正文m中的单词数；N同式(1)。困惑度一般随着潜在主题数量的增加呈现递减规律，困惑度数值越小，意味着主题模型的生成能力越强[13]。

在两次火灾微博数据中，通过困惑度计算得出主题数量在1～20区间内的困惑度数值。横轴为LDA主题模型中的潜在主题数，纵轴显示LDA主题模型的困惑度。两种数据的折线图变化态势大致相同，随着主题数的增加，总体上困惑度呈现先降后升再降的态势（图3）。一般认为，当主题结构平均相似度最小时，对应的LDA主题模型最优。因此，选择困惑度数值相对较小且主题数相对较少的主题数作为LDA主题模型训练最优模型参数[14]。困惑度的局部极小值点出现在主题数为6的模型选择上，根据上述原则对于西昌森林火灾与木里森林火灾两个话题均选取6作为LDA主题模型的主题参数值。

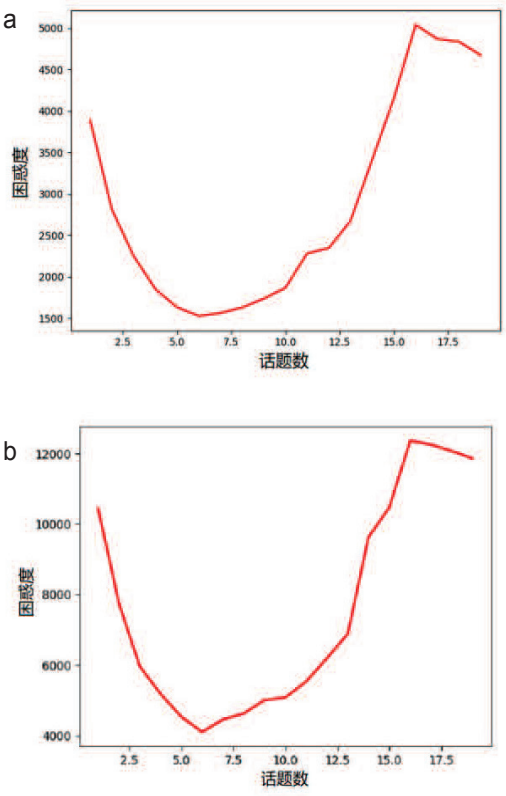

图 3 2019木里森林火灾(a)与2020西昌森林火灾(b)微博主题模型困惑度

Fig. 3 Perplexity of topic model of Weibo about 2019 Muli forest fire (a) and 2020 Xichang forest fire (b)

在确定最优主题数后，将分词后的数据用于LDA主题模型训练，得到“主题-词”以及“文档-主题”两个概率分布。通过“主题-词”分布，可确定各个主题包含的高频词，并结合这些高频词归纳主题内容。利用 LDA主题模型训练得到6个主题结果（表1），且各个主题均选取词频最高的前5个中文词（表3）（表4）。

表 3　2019木里森林火灾“主题-词”分布
Tab.3 Distribution of "topic-word" about 2019 Muli forest fire

主题 0	消防	凉山	视频	火灾	警方
概率	0.065	0.038	0.029	0.028	0.025
主题 1	木里	森林	牺牲	扑火	消防队员
概率	0.111	0.043	0.037	0.031	0.031
主题 2	四川	森林	勇士	牺牲	火灾
概率	0.097	0.092	0.065	0.050	0.048
主题 3	四川省	森林	办公室	人民政府	提供
概率	0.060	0.051	0.045	0.045	0.044
主题 4	原因	视频	火灾	蔓延	山火
概率	0.032	0.029	0.029	0.026	0.025
主题 5	森林	消防	火灾	木里	四川
概率	0.063	0.041	0.037	0.023	0.019

表 4　2020西昌森林火灾“主题-词”分布
Tab.4 Distribution of "topic-word" about 2020 Xichang forest fire

主题 0	英雄	西昌	前行	火灾	离开
概率	0.043	0.043	0.038	0.033	0.026
主题 1	森林	火灾	四川	西昌市	拍摄
概率	0.072	0.056	0.052	0.044	0.036
主题 2	存在	办公室	灭火	火灾	工作
概率	0.075	0.061	0.057	0.043	0.041
主题 3	四川省	凉山州	森林	事故	火灾
概率	0.1	0.046	0.043	0.035	0.03
主题 4	调查	沉默	面积	武警战士	处理
概率	0.181	0.113	0.035	0.025	0.016
主题 5	山火	火情	现场	蔓延	逼近
概率	0.049	0.031	0.028	0.028	0.027

通过“主题-词”分布可以看出，各个主题关键词都占有较大概率值，符合微博主题特点，即微博用户在某一特定话题空间下的评论用词习惯趋于相同。经过主题词提取后可以发现两起火灾的“主题-词”分布表中地名出现的概率较高，高度关注的三大主题为火灾情况、救火人员和政府救援决策。

4　微博舆情的情感分析与可视化

本文以微博语义为研究对象，以词义为研究单位，基于机器学习方法对两起森林火灾进行情感词的可视化，并对其中蕴含的网民情绪的演变规律进行探究。

4.1　基于WordCloud的微博舆情情感词可视化

使用哈工大情感词典对微博评论数据的情感词进行了筛选，根据词频制作了情感词云。整体上看，两起火灾微博情感词积极性居多，如平安、英烈、致敬等，表达了对救火人员的敬意。此外存在一些高频情绪词语，如教训、痛心、不幸等，表达了对救火人员牺牲的惋惜和痛心。

图 4　2019木里森林火灾(左) 与2020西昌森林火灾(右)情感词云
Fig. 4 Emotion word cloud of Weibo about 2019 Muli forest fire (left) and 2020 Xichang forest fire (right)

4.2　基于朴素贝叶斯的情感指数时间演化分析

由于微博正文数据中存在大量的不具备情感倾向的新闻数据，会对情感分析产生干扰，而微博评论数据则几乎全部为网民个人观点看法表达，具备较为清晰的情感倾向。因此选取了微博评论作为情感分析的研究对象，使用SnowNLP提供的朴素贝叶斯情感分析方法对两起森林火灾的微博评论数据进行了情感分析，情感指数是正面情绪的概率，分布在(0,1）区间上。根据日期分组计算了每天情感指数平均值（图5）（图6）。结合§2.1节中的微博数量时序可以发现，两起火灾均在15天后达到第一轮舆情的基本平息，两起火灾的微博舆情在情感演变上具有一定差异。木里火灾第一轮舆论爆发期内，整体舆情情感指数稳定在0.5以上，整体舆论情感偏向积极，偏重于对救火人员的感激及祝愿。而后续5月份和6月份再次发生火灾后，网民情绪出现大幅波动，负面情绪明显增多。对于西昌火灾，由于事件时间、地点与一年前的木里火灾高度相似，又一次造成多名救火人员牺牲，在第一轮舆论爆发期内，网民整体情感指数低于前一年的木里火灾，具有更多负面情绪。而在5月份后，由于西昌火灾没有发生复燃等事件，网络舆论的情感趋于中性和正面，整体稳定在0.5以上。整体而言，在林火事件初次发生时，舆情情感比较平稳正面，而当性质类似的事件重复发生，则会导致舆情中的负面情绪明显增多。

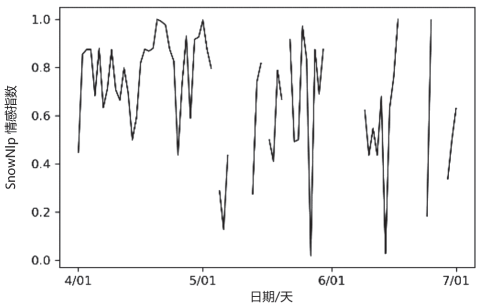

图 5　2019凉山木里森林火灾情感指数
Fig. 5 2019 Muli forest fire sentiment index

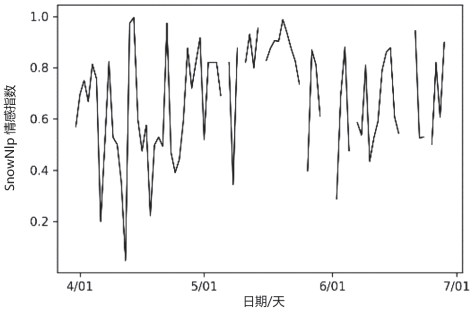

图 6　2020凉山西昌森林火灾情感指数

Fig. 6 2020 Xichang forest fire sentiment index

5　结束语

本文对四川省两起森林火灾事件的舆情进行了时空扩散及分异、主题提取、情感态度方面的综合演变规律脉动分析。与此前针对单一事件的单一语义或空间规律的研究不同，本文选取了两起性质相似的事件，同时对其时空和语义信息进行了挖掘和对比分析。研究表明，由于两起森林火灾均发生在四川省凉山彝族自治州，救援过程中都造成人员牺牲，因此两起森林火灾在舆情的时空扩散、空间分异、主题及情感表达方面都具有较强相似性。两起森林火灾舆情均在爆发后15天内基本平息，舆情热点和快速响应点主要分布在事发地及几个经济高水平发展的区域，舆情分布与区域经济水平具有较大相关性，说明参与微博讨论的主力群体集中在经济高水平地区。主题上关注事发地、消防救火人员安全和救火处置合理性。情感上主要表达对救援人员的致敬和牺牲英雄的惋惜。此外，两起火灾也表现出一定差异，火灾发生初期，西昌火灾舆情中网民消极情绪整体多于此前的木里火灾，说明由于凉山州又一次发生大规模森林火灾并造成人员伤亡，导致网民的消极情绪增多；木里火灾事件结束后同年5、6月份再次发生火灾也导致了舆情大规模反弹，消极情绪迅速增多，而次年西昌森林火灾在5月之后基本没有发生大型复燃，网民情感基本稳定在中性及积极区间，态度情绪向好。整体而言，森林火灾首次发生时，网民情绪偏向中性积极，但当事件重复发生，则会引起网民消极情绪的增加。

本文针对森林火灾舆情进行多角度分析，揭示出一定客观规律。但存在不足，在时空分异方面，由于统计数据获取不足，地理探测器结果只细化到省级行政区，没有精确到市县行政区；在情感分析方面，采用SnowNLP提供的朴素贝叶斯中文情感分析模型，情感指数计算时精度有限，后续可以进一步完善提高分析粒度和评价精度。

参考文献

[1] 许宝健. 做好打赢疫情防控阻击战的舆论引导工作[N]. 学习时报, 2020-02-14(1).

[2] NAIR M R, RAMYA G R, SIVAKUMAR P B. Usage and Analysis of Twitter During 2015 Chennai Flood Towards Disaster Management[J]. Procedia Computer Science, 2017, 115:350-358.

[3] ALFARRARJEH A, AGRAWAL S, KIM S H, et al. Geo-Spatial Multimedia Sentiment Analysis in Disasters[C]// 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2017.

[4] 张庆民, 王海燕, 吴春梅, 吴士亮. 基于熵权-离差聚类法的城市公共安全舆情评估[J]. 中国安全科学学报, 2012, 22(9): 147-152.

[5] 刘继, 李磊. 大数据背景下网络舆情智能预警机制分析[J]. 情报杂志, 2019, 38(12):92-97, 183.

[6] 彭玲, 池天河, 姚晓婧. 智慧城市脉动分析理论与实践[M]. 北京:科学出版社, 2017.

[7] 新浪财经. 微博第一季度财报[EB/OL]. Available: http://finance.sina.com.cn/roll/2020-05-19/doc-iirczymk2482234.shtml

[8] 王林, 王可, 吴江. 社交媒体中突发公共卫生事件舆情传播与演变——以2018年疫苗事件为例[J]. 数据分析与知识发现, 2019, 3(4):42-52.

[9] WANG J F, LI X, CHRISTAKOS G, et al. Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China[J]. International Journal of Geographical Information Science, 2010, 24(1):107-127.

[10] WANG J F, HU Y. Environmental Health Risk Detection with GeogDetector[J]. Environmental Modelling and Software, 2012, 33 : 114-115.

[11] 王劲峰, 徐成东. 地理探测器:原理与展望[J]. 地理学报, 2017, 72(1):116-134.

[12] HUANG L, MA J Y, CHEN C L. Topic Detection from Microblogs Using T-LDA and Perplexity[C]//2017 24th Asia-Pacific Software Engineering Conference Workshops. 2017

[13] 关鹏, 王曰芬. 科技情报分析中LDA主题模型最优主题数确定方法研究[J]. 现代图书情报技术, 2016(9):42-50.

[14] 曾子明, 王婧. 基于LDA和随机森林的微博谣言识别研究——以2016年雾霾谣言为例[J]. 情报学报, 2019, 38(1):89-96.