



开放科学
(资源服务)
标识码
(OSID)

台风眼效应中的突发事件舆情数据分析方法

高影繁¹ 李岩¹ 浦墨¹ 郑明¹ 翟云²

1. 中国科学技术信息研究所 北京 100038;
2. 中央党校(国家行政学院)电子政务研究中心 北京 100089

摘要: 心理学家用“台风眼效应”表征人们离危机事件中心的距离越远、反应越强的心理现象。突发事件发生以后,网络舆情数据通常会呈现铺天盖地、汹涌澎湃之势。在这种情况下,如何去除舆情数据中的心理、情绪等主观信息,为民众中的不同层次主体呈现客观的事件描述,缓解由于信息不对称所带来的台风眼效应问题成为了一个有价值的研究课题。本文结合科技文本的短语识别方法,通过突发事件典型特征词与科技类短语的联合检索实现对舆情信息的过滤,利用可视化主题聚类 and 热点分析工具实现了对突发事件舆情数据的客观描述,为台风眼效应中信息不对称问题的解决提供了新思路。

关键词: 台风眼效应; 信息不对称; 短语识别; 舆情信息过滤

中图分类号: G35

Analysis Method of Public Opinion Data in Emergency from Typhoon Eye Effect

GAO Yingfan¹ LI Yan¹ PU Mo¹ ZHENG Ming¹ ZHAI Yun²

1. Institute of Scientific and Technical Information of China, Beijing 100038, China;
2. E-Government Research Center, Party School of the Central Committee of the C.P.C.(Chinese Academy of Governance), Beijing 100089, China

Abstract: Psychologists use “typhoon eye effect” to express the psychological phenomenon that people who are farther away from the center of emergencies would have stronger reaction. After the occurrence of emergencies, the network public

基金项目: 国家自然科学基金项目“非均衡概念漂移网络舆情大数据流挖掘模型、算法与评价机制研究”(61672178), 中国科学技术信息研究所重点work项目“上市公司年报数据库建设及服务系统研发”(ZD2020-12)。

作者简介: 高影繁(1974-), 博士, 副研究员, 研究方向: 智能信息处理、大数据挖掘, E-mail: gaoyingf@istic.ac.cn; 李岩(1994-), 硕士, 实习研究员, 研究方向: 数据挖掘算法研究; 浦墨(1987-), 硕士, 助理研究员, 研究方向: 情报分析方法; 郑明(1985-), 博士, 助理研究员, 研究方向: 上市公司创新能力评价; 翟云(1979-), 博士, 副教授, 研究方向: 数字政府和大数据治理。

opinion data usually presents overwhelming and surging trend. In this cases, how to remove the subjective information such as psychology and emotion from public opinion data, present objective event description for different levels of subjects in the public, and alleviate the typhoon eye effect caused by information asymmetry has become a valuable research topic. In this paper, we combine the phrase recognition method of scientific and technological text, filter public opinion information through the joint retrieval of typical characteristic words and scientific and technological phrases of emergencies, and identify the objective description of public opinion data of emergencies by using visual topic clustering and hot topic analysis tools. This paper provides a new way to solve the information asymmetry problem in typhoon eye effect.

Keywords: Typhoon eye effect; information asymmetry; Phrase recognition; Public opinion information filtering

引言

“台风眼效应 (typhoon eye effect)”是心理学中的一种现象。台风眼为台风中心气压最低之处,这里微风少云,而产生大浪的区域通常距离台风中心几十公里。心理学家用这种现象表征人们离危机事件中心的距离越远、反应越强的心理现象。突发事件发生以后,会有悲伤、感动、愤怒、压抑等各种情绪聚集交织,作为社会“晴雨表”和“风向标”的网络舆情数据在突发事件中通常会呈现铺天盖地、汹涌澎湃之势。在这种情况下,信息的不对称很容易成为维持网络舆情“台风眼”持续存在的外在能量。信息不对称的重要成因是舆情防控涉及大量的专业知识,与民众常识性认知之间存在差异。这样就会在不同的网络主体之间形成隔离带,许多信息局限在各主体隔离带内部,无法在不同的主体之间交换与互动,这就造成了一些问题上形成内部明晰稳定而外部模糊嘈杂的网络舆情“台风眼”^[1]。在突发事件面前,为了解决由于信息不对称造成的台风眼效应问题,如何通过对舆情信息的过滤分析,去除心理、情绪等主观信息,为民众中的不同层次主体呈现客观的事件描述,是本文要解决的主要问题。

1 突发事件的舆情分析方法述评

1.1 定性分析法

定性分析法主要可以划分为理论分析法和案例分析法两类。利用理论分析法的研究主要是从利益相关者的视角出发,构建突发事件舆情分析模型^[2],分析突发事件舆情关注点演化^[3]等;案例分析法则主要是以具体突发事件为分析载体,分析突发事件舆情的形成过程、结构体系及演化模式等^[4-6]。

1.2 定量分析法

基于数量统计的舆情分析法主要是通过统计分析涉及突发事件舆情事件的文章数、评论数、关键词等信息,进而从舆情的发展态势、波动情况、网民议论焦点、热点话题等方面对舆情的整体态势、焦点、热点等进行分析。基于内容分析的舆情分析法则需要借助机器学习、数据挖掘等方法从关键词、文章内容、转发内容以及评论内容等方面深入分析突发公共事件舆情的内容,通过分析关键词、情绪词的变化,实现舆情的热点主题挖掘、网民类型划分、意见领袖主题词情感分析以及主题演化等内容的分析。基于社会网络的舆情分析主要是利用社

会网络分析的方法对舆情的传播路径、网络结构进行定量化测度,以挖掘与识别舆情关键节点并解释其内在的结构特征与演变规律^[7-8]。基于层次分析法的网络舆情分析主要用于突发事件舆情风险的评估与预警研究,一般在分析突发事件舆情影响因素的基础上构建舆情的风险预警指标体系,并利用层次分析法、ABC分类法或灰色统计法相结合的方法计算各风险指标的权重系数,进而获得舆情的高风险因素,筛选出重点监测目标。

1.3 仿真分析法

学者们在分析突发事件舆情时所采用的仿真分析方法主要有系统动力学、BP神经网络及传染病传播模型等方法。在系统动力学的舆情分析方面,目前有学者通过利用系统动力学建模的方法从网民自身因素、媒体环境因素以及政府情绪引导控制因素三个影响网民情绪演化的主要方面对网民的情绪演化进行了分析^[9];也有学者通过构建系统动力学模型对突发事件后不实信息传播引起的舆情进行了分析,并采用 Vensim PLE 软件对模型进行仿真,分析了事件公共度、事件敏感度、网民质疑度、政府公信力四个方面对舆情热度的影响,对舆情态势的引导和控制给出了对策建议^[10]。BP神经网络的突发事件舆情分析主要用在风险监测、评估及危机预警等方面,如有学者在构建震灾舆情风险评估体系和网络谣言预警指标体系的基础上,将BP神经网络模型用于震灾舆情风险监测评估和网络谣言危机预警中,实现了震灾舆情风险评估方法的优化和突发事件网络谣言风险的定量评判^[11];传染病传播模型目前主要用于

舆情演化分析,如有学者通过构建突发事件舆情演化的传染病传播模型,研究了网民在舆情传播中的模仿传染机理和行为,分析了突发事件舆情演化传播因素对传播规律的影响^[12-13]。

上述方法能够根据舆情数据的特点从不同角度进行分析,但对本文的分析目标——引起台风眼效应的网络舆情信息不对称问题却不是很适合。本文认为,造成不同网络主体信息不对称的原因是网络舆情数据噪声过多,心理、情绪等主观信息的表达使得分析结论发生了偏移,因此噪声数据过滤是本文分析方法的基础和前提。在过滤后的数据集上进行事件主题分析,就可以实现为民众中的不同层次主体呈现客观事件描述的研究目标。

2 结合事件关键短语的舆情数据分析方法

数据体量大是网络舆情数据的典型特征,对于一个热点突发事件来说,几亿或是几十亿条评论都是很常见的。在这种情况下,需要根据分析目标过滤噪声数据。为了将心理、情绪等公众的主观信息过滤掉,本文采用了将科技类文本数据纳入检索策略,将筛选后的科技类词汇与突发事件相关的核心词汇一起进行联合检索,这样就可以将数据分析的目标对象集合缩小,也为突发事件的客观呈现奠定了基础。

2.1 科技文本的短语识别方法

科技类文本数据引入检索接口的途径是科技类短语的识别。本文采用关键词列表来对科技类文本进行短语识别。步骤如下:

(1) 对每篇科技类文本, 利用关键词列表对其进行分词, 关键词列表来源于期刊论文库, 选取了期刊论文的关键词字段, 经过整合去重, 得到关键词集合。分词方法可以采用替换分词工具词库的方式进行。本文中采用较为常用的 jieba 分词工具。具体方法如下:

使用上面得到的关键词集合作为分词词典, 替换 jieba 分词工具自带词典进行分词处理。考虑到 jieba 自带的分词词典规模约为 35 万, 而本文所用的关键词集合规模在 1200 万左右, 在规模上是原词典的 34 倍, 直接使用关键词集合作为分词词典并初始化 jieba, 可能导致程序崩溃。由于 jieba 中采用了基于汉字成词能力的隐马尔科夫模型(HMM), 用以识别未登录词, 本文将去除该模型以保证分词效率。经过此步骤的处理, 能够较为高效、准确地识别出科技类文本中的短语。

(2) 对步骤(1)得到的短语在论文库的期刊标题和摘要中进行信息统计, 根据 TF-IDF 方法对获取的短语进行排序, 其中 tf 指识别到的短语在论文库标题和摘要中出现的词频, idf 则指识别到的短语在论文库标题和摘要中的逆文档频率值, 通常的计算方法是用论文库的文档总数除以论文库中包含该词语的文档数量, 再将得到的商取对数。

(3) 按 TF-IDF 方法获得短语序列后, 用人工的方法对其中不重要的词进行过滤, 得到候选科技短语用于下一步的舆情数据过滤。

2.2 结合科技短语的舆情数据过滤方法

对于突发事件来说, 都有一个典型表达词, 比如“汶川地震”、“印尼海啸”、“马

航 MH370”等, 这个表达词代表着官方和民众对一个事件的公共认知, 如果单纯以这个事件表达词在舆情数据平台上进行检索的话, 将会产生海量的检索结果。为了减少噪声舆情信息对数据分析带来的影响, 本文采用结合科技短语的舆情数据分析方法。即将 2.1 节获得的来源于科技文本的短语或者短语组合, 与突发事件典型表达词一起作为舆情检索平台的入口词进行检索, 形成面向本文分析目标的舆情数据语料集。

2.3 突发事件主题分析方法

对 2.2 节得到的舆情数据语料集进行处理:

①要进行中文分词及停用词过滤等基础数据处理; ②采用 TF-IDF 方法对抽取到的舆情语料关键词按照权重进行排序; ③取前 2000 个舆情语料关键词, 采用人工的方法进行过滤, 最终保留 1000 个关键词; ④用步骤③得到的舆情关键词在舆情语料中进行统计分析, 构建共现矩阵; ⑤采用分析工具对舆情关键词共现矩阵进行聚类及热点分析, 得到聚类分析图和舆情热点图。

3 实验及结果分析

3.1 数据及工具来源

(1) 科技数据来源

本文实验采用的科技类数据为科技部重点研发计划重点专项项目申报指南。重点专项是国家重点研发计划组织实施的载体, 聚焦国家重大战略任务、围绕当前国家发展面临的瓶颈和突出问题, 是以目标为导向的重大项目群。

在已发布的重点研发计划重点专项中，与本文的分析目标——“诺如病毒”事件最相关的专项是“蛋白质机器与生命过程调控”（以下简称蛋白质专项），实验中我们选择了2016-2019年度的蛋白质专项项目申报指南数据（每年1份），为PDF格式的非结构化文本型数据。如图1为蛋白质专项2019年的申报指南样例。

“蛋白质机器与生命过程调控”重点专项 2019年度项目申报指南

为提升我国蛋白质研究水平并推动应用转化，按照《国家中长期科技发展规划纲要（2006—2020年）》（国发〔2005〕44号）的部署，根据《国务院关于深化中央财政科技计划（专项、基金等）管理改革的方案》（国发〔2014〕64号），科技部、教育部、中国科学院等部门组织专家编制了“蛋白质机器与生命过程调控”重点专项实施方案。专项围绕我国经济与社会发展的重大战略需求和重大科技问题，结合国际蛋白质研究的前沿发展趋势，开展战略性、基础性、前瞻性研究，增强我国蛋白质机器研究的核心竞争力，产出一批国际领先、具有长远影响的标志性工作，实现重点领域对国际前沿的引领，在原创性基础和理论研究中取得突破，为人口健康、生物医药、农业与环境、生物安全等领域提供理论支持和技术方法支撑。

图1 蛋白质专项指南示例

（2）突发事件舆情数据来源

舆情数据来源于清博开放平台，该平台提供微信、微博和APP两微端的数据接口，为不同领域客户提供数据，满足个性化需求，合作模式便捷，能够实现信息分享、数据分析、数据报告等。本文实验所用微博微信数据的时间区间是2019年1月1日至2019年12月31日，分析的目标突发事件为2019年爆发的公共卫生事件——“诺如病毒”事件。

舆情数据检索方式：以“诺如病毒”+“蛋白质专项抽取到的短语”作为清博微博微信平

台的检索入口词，通过清博平台提供的数据接口共获得数据182232条。

（3）分词词典及工具

本文实验采用的分词工具为在中文分词领域性能较好的jieba工具，用于科技类短语识别的关键词词典来自于论文库，对其中的关键词字段进行去重、噪声清洗等处理后，得到约1200万条关键词构成的分词词典。

（4）突发事件主题分析工具

本文采用的突发事件主题分析工具为VOSviewer软件，是荷兰莱顿大学科技研究中心于2009年开发的一款基于JAVA的用于构建和可视化文献计量网络的免费软件，侧重科学知识的可视化、图形展示能力强。

3.2 实验结果

（1）蛋白质专项指南科技短语抽取结果

依据本文2.1节所述的短语识别方法，对蛋白质专项指南按指南方向进行了短语识别，限于篇幅，本文截取了部分指南方向的短语识别结果，如图2所示。

（2）舆情数据主题词选取结果

依据本文2.3节所述的突发事件主题分析方法，对检索到的舆情数据集进行分词和关键词权重计算，对选取的前2000个关键词进行人工筛选，最终保留1000个词汇形成词共现矩阵。部分共现矩阵示意图3。

（3）“诺如病毒”事件主题词聚类图

采用VOSviewer工具，以图3所示的共现矩阵为VOSviewer的输入数据，得到“诺如病毒”事件主题词聚类图，如图4所示。

重点专项名称	专项指南方向	短语识别结果
蛋白质机器与生命过程调控	细胞生命活动相关的蛋白质机器*	新蛋白质; 检测技术; 自由基; 成像技术; 高通量; 内质网; 细胞迁移; 动态变化; 蛋白质; 调控机制; 分子基础; 高分辨率; 催化过程; 磁共振; Cell; 复合物; 单分子; 线粒体; 活细胞; 功能研究; 生物膜; 外泌体; 药物靶点; 相关蛋白质; Chem; 生理功能; 细胞器; Biolog; 重大疾病; 作用机制
	肿瘤微环境对蛋白质机器的影响和调控*	Cell; 免疫细胞; DNA; 天然免疫; PNAS; 信号通路; 调控机制; 过程和机制; 肿瘤细胞; 肿瘤微环境; 肿瘤发生; 微环境; 转录调控; T细胞; Communication; 作用机制; 蛋白质; 相关蛋白质
	蛋白质膜转运的分子机制*	动态调控; 基于结构; 内质网; 离子通道; 冷冻电镜; 蛋白质; 相关疾病; 膜蛋白; 调控机制; 结构基础; 分子基础; 结构解析; 虚拟筛选; Cell; 分子结构; 线粒体; 分子机制; 蛋白复合物; 靶向药物; 结构生物学; 相互作用蛋白; 复合体; PRL; 结构和功能
	植物特有蛋白质机器的分子机制*	Cell; 蛋白复合物; 根瘤菌; 分子机制; 关键蛋白质; 幼苗出土; 三维结构; 内质系统; 机械压力; 生命过程; 结构和功能; 信号转导; 蛋白质
	蛋白质翻译器的调控*	核糖体组装; 细胞内; 三维结构; 结核分枝杆菌; 蛋白质; 环状RNA; 调控机制; RNA; 核糖体; 结构基础; Cell; 复合物; 功能研究; 相关蛋白; 分子机制; 病原菌; 生命过程; 药物靶点; 非编码RNA; 相互作用蛋白; 翻译后修饰; 反式翻译; 生物学功能
	神经干细胞发育与细胞命运决定中的蛋白质机器	Cell; 复合物; 调控机制; 细胞命运决定; 结构与功能; 染色质修饰; 染色质; 关键蛋白质; 分子机制; 表观遗传修饰; 表观遗传; 神经系统发育; 表观遗传调控; 神经系统; 染色质重塑复合物; 结构和功能; 神经干细胞; 蛋白质
	RNA-蛋白质复合物与生命过程的调控*	染色质; lncRNA; 蛋白质; 基因沉默; RNA; 关键蛋白质; RNA介导的基因沉默; 染色质修饰; 染色质结构; 分子机制; 稳定性; DNA甲基化; 结构生物学; 非编码RNA; miRNA; 染色质高级结构; 生物学; 结构与功能; 转录调控; 生物学; 作用机制; 细胞生物学
	控制重要组织器官的系统发育与重塑的蛋白质机器	磷酸化; 蛋白质组学; 调控网络; 蛋白质; 神经环路; 相关疾病; 调控机制; RNA结合蛋白; RNA; 精子发生过程; Cell; 精子发生; 神经发育; 突触囊泡; 结构生物学; 神经元; 作用机制; 男性不育; 无精子症; 神经精神疾病; CRISPR
	高分辨率冷冻电镜在结构生物学中的应用	冷冻电镜技术; 生物大分子; 结构解析; 高分辨率; 数据采集; 样品制备; 结构生物学; 冷冻电镜; 蛋白质
	磁共振技术在结构生物学中的应用	相互作用; 膜蛋白; 调控机制; 动态结构; 蛋白质动态结构; 生命过程; 结构生物学; 核磁共振; 蛋白质翻译后修饰; 规模化; 复合物; PNAS; 变化规律; 亚细胞器; 相关蛋白质; 蛋白质组; 完整精肽; 肿瘤微环境; 细胞信号转导; 蛋白质复合物; 蛋白质组学; 靶向药物; 信号转导; 数据库; 蛋白质
	新一代蛋白质组学分析技术研究	相互作用; 生物学; 膜蛋白; 高通量; 化学生物学; 活细胞; 分子机制; 病原菌; 原位研究; 细胞信号转导; 作用机制; 信号转导; 蛋白质
	化学生物学在蛋白质机器研究中的应用	检测技术; 单分子荧光; 透射电镜; 成像技术; 高通量; 活细胞; 纳米器件; 蛋白质
依托大科学装置的新技术和新方法研究	动态变化规律; 相关疾病; 结构基础; 高精度; 蛋白质相互作用; 调控网络; 结构和功能; 蛋白质	
计算生物学的新技术、新方法及应用	分子机理; 调控作用; 先导化合物; 微环境; 靶向药物; Neddylation; 蛋白质	
基于蛋白质机器的肿瘤和免疫类疾病防治	Cell; 细胞死亡; 信号通路; 神经元; 分子机制; 分子基础; 先导化合物; RIPK1; 神经退行性疾病; 药物靶点;	

图2 蛋白质专项指南短语识别结果示例

症状	诺如病毒	患者	儿童	传播	食物	病毒	孩子	传染病	学校	疾病	流感	通风	消毒	
症状	1.20445	0.690714	0.744713	0.634497	0.517581	0.489485	0.465208	0.4471	0.402831	0.349678	0.3235	0.31499	0.305409	0.283955
诺如病毒	0.690714	1.17517	0.615605	0.499817	0.572683	0.553043	0.539879	0.445442	0.439051	0.405677	0.335245	0.35708	0.313921	0.336402
患者	0.744713	0.615605	1.09274	0.636766	0.513214	0.480663	0.432042	0.414682	0.412632	0.356517	0.334416	0.333492	0.323316	0.263009
儿童	0.634497	0.499817	0.636766	0.95168	0.562899	0.383984	0.390226	0.392926	0.426068	0.34954	0.311859	0.39239	0.333011	0.282595
传播	0.517581	0.572683	0.513214	0.562899	0.79757	0.554472	0.572088	0.486465	0.568778	0.487205	0.38181	0.448161	0.387051	0.387264
食物	0.489485	0.553043	0.480663	0.383984	0.554472	0.78564	0.547159	0.519882	0.417672	0.357925	0.375834	0.32422	0.346253	0.346953
病毒	0.465208	0.539879	0.432042	0.390226	0.572088	0.547159	0.74753	0.498682	0.440129	0.381978	0.362541	0.288564	0.288898	0.357444
孩子	0.4471	0.445442	0.414682	0.392926	0.486465	0.519882	0.498682	0.68084	0.494242	0.388006	0.397524	0.377005	0.428177	0.315551
传染病	0.402831	0.439051	0.412632	0.426068	0.568778	0.417672	0.440129	0.494242	0.66360	0.481269	0.298312	0.515627	0.42503	0.32072
学校	0.349678	0.405677	0.356517	0.34954	0.487205	0.357925	0.381978	0.388006	0.481269	0.55601	0.374975	0.487257	0.348375	0.45965
疾病	0.3235	0.335245	0.334416	0.311859	0.38181	0.375834	0.362541	0.397524	0.298312	0.374975	0.52793	0.369045	0.399068	0.324153
流感	0.31499	0.35708	0.333492	0.39239	0.448161	0.32422	0.288564	0.377005	0.515627	0.487257	0.369045	0.51962	0.50893	0.319734
通风	0.305409	0.313921	0.323316	0.333011	0.387051	0.346253	0.288898	0.428177	0.42503	0.348375	0.399068	0.50893	0.48836	0.35265
消毒	0.283955	0.336402	0.263009	0.282595	0.387264	0.346953	0.357444	0.315551	0.32072	0.45965	0.324153	0.319734	0.35265	0.47384

图3 舆情数据关键词共现矩阵示例

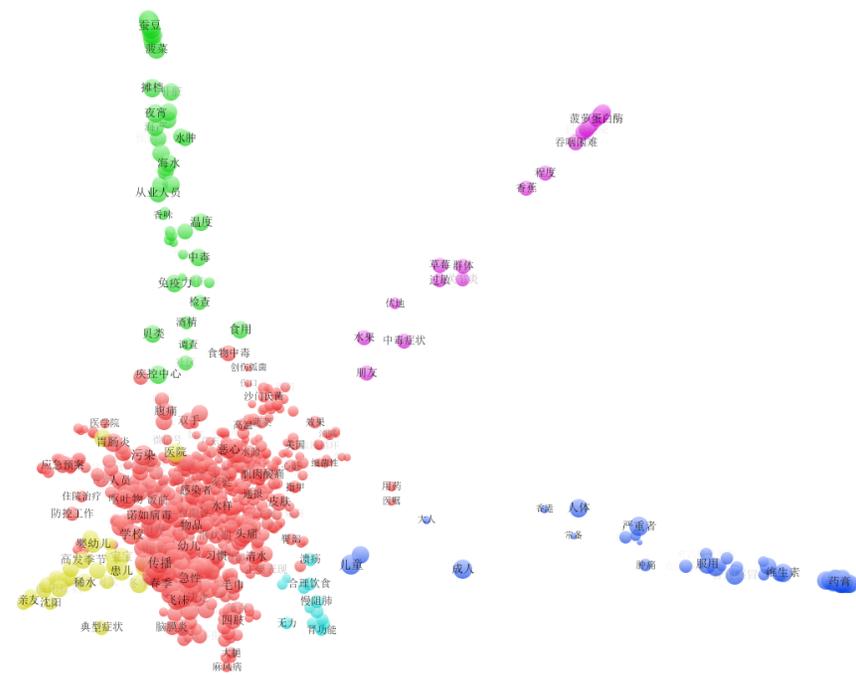


图4 “诺如病毒”事件主题词聚类图

(4) “诺如病毒”事件舆情热点图
采用 VOSviewer 工具, 以图 3 所示的共现

矩阵为 VOSviewer 的输入数据, 得到“诺如病毒”
事件舆情热点图, 如图 5 所示。

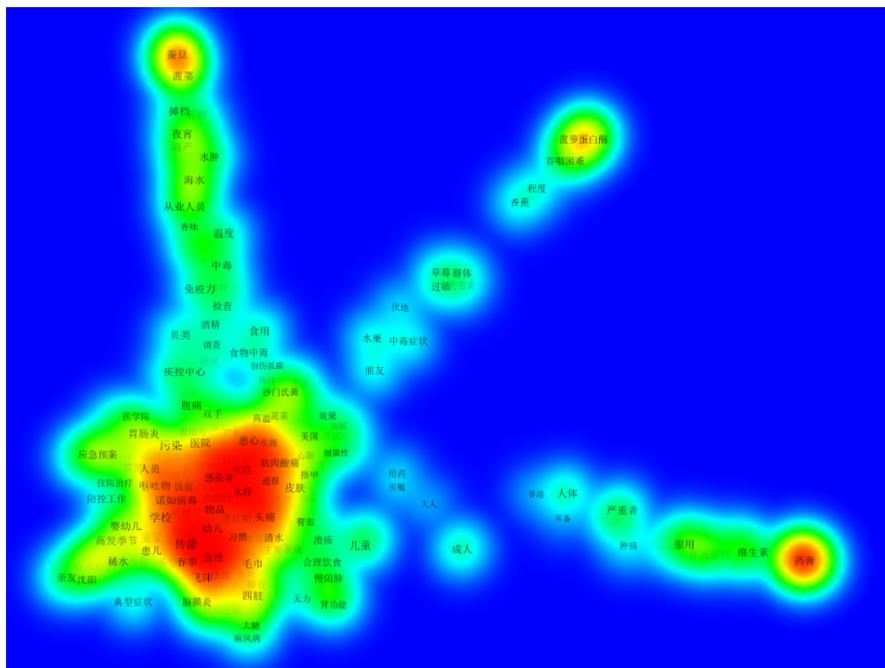


图 5 “诺如病毒”事件舆情热点图

3.3 实验结果分析

(1) 从图 2 的蛋白质专项指南短语识别结果可以看出, 对应于蛋白质专项的不同指南方向, 能够较为准确地识别出科技类短语。如指南方向“细胞生命活动相关的蛋白质机器”中, 识别出了“自由基”、“高通量”、“内质网”、“细胞迁移”等与指南方向关系密切的科技类短语; 指南方向“蛋白质翻译机器的调控”中, 识别出了“核糖体组装”、“三位结构”、“环状 RNA”、“核糖体”等重要短语。但是, 受指南文本篇幅短、短语出现次数少等因素限制, “检测技术”、“作用机制”、“结构和功能”、“相互作用”等没有指南方向区分能力的通用词汇无法去除, 这样会使得舆情数据的检索中引入更多噪声。

(2) 分析图 4 的“诺如病毒”事件主题词

聚类结果, 可以较为清晰地读出颜色所代表的类别信息:

- 红色 -- 诺如病毒感染途径及主要表现;
- 淡蓝色—诺如病毒可引起的其它危害;
- 绿色 -- 诺如病毒主要传染源;
- 黄色 -- 婴幼儿感染的舆情关注点;
- 深蓝色 -- 诺如病毒感染的治疗。

遗憾的是由于噪声的关系, 有些主题词的聚类结果可读性不强, 如紫色词的类别属性不是很清晰, 与其他颜色类别主题有交叉。

(3) 分析图 5 的“诺如病毒”事件舆情热点图, 图中的红色区域展示出了“诺如病毒”事件的热点表达, 如: “春季”、“飞沫”、“幼儿”、“呕吐物”、“肌肉酸痛”、“恶心”、“自限性”、“潜伏期”“学校”等; 绿色区域是“诺如病毒”事件的一般表达, 如“胃肠炎”、“合理饮食”、

“住院治疗”等。从“诺如病毒”事件本身来说，图5中的热点词和一般性词汇也与事件的实际情况较为一致。

4 结论

心理学家用“台风眼效应”表征人们离危机事件中心的距离越远、反应越强的心理现象。突发事件发生以后，网络舆情数据通常会呈现铺天盖地、汹涌澎湃之势。在这种情况下，去除舆情数据中的心理、情绪等主观信息，为民众中的不同层次主体呈现客观的事件描述，缓解由于信息不对称所带来的台风眼效应问题成为了一个有价值的研究课题。本文结合科技文本的短语识别方法，通过突发事件典型特征词与科技类短语的联合检索实现对舆情信息的过滤，利用可视化主题聚类 and 热点分析工具实现了对突发事件舆情数据的客观描述，为台风眼效应中信息不对称问题的解决提供了新思路。从实验结果来看，科技类文本的短语识别方法是有效的，对应于蛋白质专项的不同指南方向的短语识别结果较为准确，但也存在由于指南文本篇幅短、短语出现次数少等因素导致的通用词汇无法去除、舆情数据的检索中引入更多噪声等问题。从“诺如病毒”事件主题词聚类和舆情热点图可以看出，本文方法对突发事件主题词聚类和热点的分析效果是可接受的，验证了本文方法的有效性。但由于噪声数据的存在，也发生了主题词类别属性不明确、类别交叉等问题，需要在后续工作不断对方法加以改进。

参考文献

- [1] 杨建义.[战“疫”说理]强化显政,着力化解涉疫舆情“台风眼”[EB/OL].(2020-02-27)[2020-03-05].https://theory.gmw.cn/2020-02/27/content_33598059.html
- [2] 罗闯,安璐,徐健,李纲.突发事件网络舆情关注点演化研究——基于利益相关者视角[J].图书馆学研究,2018(16):36-42.
- [3] 郑昌兴,苏新宁,刘喜文.突发事件网络舆情分析模型构建——基于利益相关者视阈[J].情报杂志,2015,34(4):71-75.
- [4] 李翠敏.公共安全危机事件网络舆情的协同疏解研究——以“长生疫苗事件”为例[J].情报杂志,2018,37(11):110-115+102.
- [5] 杜洪涛,王君泽,李婕.基于多案例的突发事件网络舆情演化模式研究[J].情报学报,2017,36(10):1038-1049.
- [6] 张云,刘玲.SNA视域下体育突发事件网络舆情的形成及结构分析[J].西南民族大学学报(人文社科版),2018,39(8):141-146.
- [7] 赵蓉英,王旭.突发事件网络舆情关键节点识别及导控对策研究——以“大贤村遭洪灾事件”为例[J].现代情报,2018,38(1):19-24+30.
- [8] 王晰巍,文晴,赵丹,王楠,阿雪.新媒体环境下自然灾害舆情传播路径及网络结构研究——以新浪微博“雅安地震”话题为例[J].情报杂志,2018,37(2):110-116.
- [9] 叶琼元,兰月新,王强,夏一雪,杨谨铖.面向突发事件的网民情绪演化系统动力学模型研究[J].情报杂志,2017,36(9):153-159+105.
- [10] 袁国平,许晓兵.基于系统动力学的关于突发事件后网络舆情热度研究[J].情报科学,2015,33(10):52-56.
- [11] 张宇,傅敏,罗加蓉.震灾网络舆情风险监测指标及其评估方法[J].重庆大学学报(社会科学版),2018,24(6):33-44.
- [12] 陈福集,黄江玲.突发事件网络舆情演化传播模型研究[J].情报科学,2015,33(12):8-12+19.
- [13] 马颖,丁周敏,张园园.食品安全突发事件网络舆情演变的模仿传染行为研究[J].科研管理,2015,36(6):168-176.