

网络群体性事件中舆情导向 与政府回应的逻辑互动*

——基于“雪乡”事件大数据的情感分析

文 宏

内容提要:网络空间已成为公民参与公共事务的重要渠道,网民在舆情走向中是不可忽视的群体,政府希望通过回应和介入事件来缓解网民的情绪。但在实践操作中,政府回应未必能取得很好的效果,甚至会加剧冲突的爆发,将网民的舆论焦点引至政府自身,造成政府公信力的缺失。本文基于对“雪乡”事件舆情大数据的分析和挖掘,以过程视角对案例进行深入剖析,利用网民评论的主题和情感构建网络群体性事件中网民对政府回应的反馈图谱。本文将数据文本分为政府介入前后两个阶段分别展开分析和探讨,发现政府回应与网络舆情存在逻辑互动偏差。政府管控与网民诉求之间、政府回应滞后与首因效应之间、正当履职与信任缺失之间的偏差,都会导致舆情向消极方向发展。

关键词:政府回应 网络群体性事件 舆情导向 大数据 情感分析

一、问题的提出

随着网民规模的不断扩大以及社会事件在网络中的深度呈现,网民成为网络舆情治理中不可忽视的有生力量。据中国互联网络信息中心最新数据显示,截至2018年初,我国网民规模达7.72亿,普及率达到55.8%,其中,全国在线政务服务用户规模达到4.85亿,占总体网民的62.9%。^①可见,网络空间已经成为公民参与的重要场域,也是社会风险治理的主战场。现实中的突发事件常在网络空间引发广泛讨论,加之网民参与形式多样和传播工具低廉高效的特点,极易导致突发事件关注热度的提升而引爆网络舆情,事件性质也由网络围观转化为网络群体性事件,使社会矛盾凸显威胁社会稳定。因此,在网络空间治理中,政府必须密切关注网络舆论的动态变化,在网络群体性事件中发挥应有的引导作用,及时回应公众关切,平息社会矛盾与舆情争论,实现对网络群体性事件的有效治理。

然而,政府回应往往事与愿违,难以取得缓解冲突的效果,有时甚至进一步加剧冲突的爆发,

* 本文为国家社会科学基金重大专项课题“社会稳定风险及对策研究”(17VZL017)的研究成果。

① 参见中国互联网络信息中心(CNNIC)在京发布的第41次《中国互联网络发展状况统计报告》,2018年3月5日, http://www.cnnic.net.cn/hlwzfzyj/hlwxbzg/hlwjtjbg/201803/t20180305_70249.htm,2018年9月15日。

使事态升级。网络群体性事件的发生和演化具有突发性、动态性和不可预测性等特征,政府对网络群体性事件的认知和引导常常处于被动状态。通过梳理近年来的相关案例,笔者发现政府对于网络群体性事件越来越重视,在维稳需求下政府会主动予以回应,试图降低突发事件带来的危害,并将可能造成的社会风险限定在可控范围之内。但由于现实情境的复杂性和地方政府能力有限性之间的矛盾,政府回应的方式和效果呈现出显著差异,出现政府回应失效的现象,从而引发事件舆论的升温和恶化,衍生出针对政府的不满和论战,严重损害政府形象与公信力。在民众权利意识觉醒、公众政治参与意愿不断提升的时代背景下,如何引导民众有序参与、提升政府管理水平、增进民众对政府信任与认同,成为亟须关注的议题,处理网络群体性事件的方法则是透视政府能力的放大镜。因此,有必要针对网络群体性事件中的政府回应过程及其治理效果进行深度挖掘,理清其中的关键节点和发展逻辑,从而为网络群体性事件的治理提供理论参考和经验借鉴。

作为公共行政理论的核心概念之一,政府回应蕴含着公共管理的基本责任,要求“政府对民众关于政策变革的接纳和民众要求做出反应,并采取积极措施解决问题”^①。西方学者较早开展对政府回应的研究,并在应然层面构建了政府回应的理论框架,取得较为丰硕的成果。^②然而,由于西方的制度环境、政治文化与我国存在显著差异,国外关于政府回应的研究,并不能直接适用于我国具体情境。近年来,国内学者逐渐关注政府回应的议题和网络舆论现象,重视强调实践层面的案例分析。有学者通过多案例研究,探讨网络反腐事件的要素特征对政府回应的影响^③;有学者梳理政府回应的类型对网络群体性事件解决效果的影响^④等,这些研究为本文研究提供了一定的思路启发。但总体而言,以往大多数研究的关注层次,主要落脚于案例表层信息提取与编码分析。虽然不少学者在政府回应视角下关注网络群体性事件,但是这些研究大多存在一些局限性:第一,理论建构研究成果较多,缺乏基于大样本的实证研究,大数据分析并未得到足够的重视;第二,多数案例研究仅关注案例的表层信息,忽略了对舆情载体——文本内容的深入挖掘;第三,已有研究多从数据指标的构建和分析入手,未能重视案例全过程对于政府回应效果研究的价值。基于以上分析,本文拟从典型的网络群体性事件入手,运用大数据研究方法,从海量网民评论文本中深度挖掘舆情主题,分析不同时间节点上网民的情感倾向,还原案例事件起因、过程和结果,讨论政府回应的效果和存在的不当之处,以期对政府决策提供政策参考。

二、研究框架:核心概念的关联性构建

(一) 政府回应:事件演变的关键要素

关于网络群体性事件的概念内涵,有学者认为网络群体性事件可以包括线下群体性事件的线上舆情传播、互联网群体性事件^⑤,有学者认为网络群体性事件专指网络空间中发生的集群现

① 格罗弗·斯塔林:《公共部门管理》,第132页,上海译文出版社,2003年版。

② H. G. Frederickson, "Comparing the Reinventing Government Movement with the New Public Administration", *Public Administration Review*, Vol. 56, No. 3, 1996. M. A. Glaser, R. B. Denhardt, "Local Government Performance through the Eyes of Citizens", *Journal of Public Budgeting, Accounting & Financial Management*, Vol. 12, No. 1, 2000.

③ 文宏、黄之珏:《网络反腐事件中的政府回应及其影响因素——基于170个网络反腐案例的实证分析》,《公共管理学报》,2016年第13期。

④ 杨立华、程诚、刘宏福:《政府回应与网络群体性事件的解决——多案例的比较分析》,《北京师范大学学报》(社会科学版),2017年第2期。

⑤ 谢进川:《互联网与群体性事件研究述评》,《现代传播》(《中国传媒大学学报》),2010年第8期。

象^①。笔者认为,网络群体性事件是指在网络空间中,网民自发的、无组织参与的集群行为,其目的往往在于制造舆论、发泄不满。本文的研究对象包括网民群体行为和政府回应行为。在网络群体性事件处理过程中,政府的回应至关重要,一旦处置失当,极易产生新的舆情危机。

政府回应具有互动性,是依据政府管理职责对公众的需求和诉求做出积极反应和回复的过程^②,参照不同的标准可将政府回应划分为多种类型:按回应的客体对象可分为“对政治指令的回应、对行政程序的回应、对行政目标的回应、对单一公民的回应、对合作方的回应、对冲突需求的回应”^③;按回应方式可分为“话语性回应、行动性回应和制度性回应等”^④。网络群体性事件是在网络空间发生的新型群体性事件,具有信息放大、传播迅速、反体制、群集化^⑤等风险特征,如果政府对网络舆情缺乏识别和关注,极易导致事件升级甚至引发公共危机。在网络群体性事件中,政府回应主要遵循“表明立场——跟进调查——公布事实——事后问责”的行动路线,政府或运用电视新闻、网络新闻、线下听证会等传统信息发布方式,或运用官方微博、微信公众号等新媒体渠道与网民进行特定议题的回应与互动。此外,由于当前网络通讯具有高度便利、成本低廉的特征,网民与政府之间的互动过程通常具有反复性,网民往往会对政府回应进一步做出反馈,从而导致事件舆情产生新走向,延长网络群体性事件的生命周期。因此,本文对这一过程的研究将政府回应纳入到网络群体性事件的考察要件中,提出网民对政府回应的评价维度,强调政府多次回应的必要性。

(二)主题情感:舆情信息的可量化路径

虽然学界对网络群体性事件的概念存在争议,但无论是何种定义,网络群体性事件的发生都是建立在网络舆情集聚的基础上。因此,构建网络群体性事件的分析框架,需要研究者对网络舆情要素及其发展过程有缜密的考察和清晰的认知。^⑥ 主题及情感是表达网络舆情核心信息的重要载体^⑦,同时也是网络舆情的主要衡量因素^⑧。对主题、情感要素的把握,有助于强化我们对网络舆情状况的识别、监控以及对潜在网络舆情危机的预警和治理。^⑨ 可见,明晰网络舆情的主题及情感为认知网络群体性事件提供了可量化路径。

(三)主题挖掘:舆情演变的规律探寻

主题挖掘是指通过一系列的算法设定,从海量的文本内容中抽取重点信息,从而帮助人们判断样本主要内容、核心语意的过程。结合舆论场的时序变迁,主题挖掘能够高效地揭示舆情变化的主要规律,为进一步分析学科意涵提供落脚点。在大数据时代,研究不再局限于结构化信息,对非结构化、半结构化信息进行挖掘的自动文本分析技术得以迅速发展^⑩,机器学习的蓬勃发展推动了该方法的技术探讨。目前学界常用的主题挖掘方法可归纳为层次聚类、划分聚类、隐含主

① 姚伟达:《网络群体性事件:特征、成因及应对》,《理论探索》,2010年第4期。

② 何祖坤:《关注政府回应》,《中国行政管理》,2000年第7期。

③ Bryer Thomas A.,"Toward a Relevant Agenda for a Responsive Public Administration",*Journal of Public Administration Research and Theory*, Vol. 17, No. 3, 2006.

④ 李放、韩志明:《政府回应中的紧张性及其解析——以网络公共事件为视角的分析》,《东北师大学报》(哲学社会科学版),2014年第1期。

⑤ 刘瑾、刘伟:《近十年来国内学界的群体性事件研究:回顾与反思》,《甘肃行政学院学报》,2017年第5期。

⑥ 郭小安、王国华:《网络群体性事件的概念辨析及指标设定》,《情报杂志》,2012年第2期。

⑦ 黄微、李瑞、孟佳林:《大数据环境下多媒体网络舆情传播要素及运行机理研究》,《图书情报工作》,2015年第21期。

⑧ 曾润喜、杜焕霞、王君泽:《网络舆情指标体系、方法与模型比较研究》,《情报杂志》,2014年第4期。

⑨ 李光敏、张行文、张磊、杨朋英:《面向网络舆情的评论文本情感分析研究》,《情报杂志》,2014年第5期。

⑩ 孟天广、李锋:《网络空间的政治互动:公民诉求与政府回应性——基于全国性网络问政平台的大数据分析》,《清华大学学报》(哲学社会科学版),2015年第3期。

题模型以及统计指标提取关键词等。层次聚类^①包括分裂式、聚合式两种向度。分裂式是指首先将所有文本样本看作一个大类,再逐渐分裂成若干小类;聚合式则恰恰相反,首先认为所有文本个体自成一类,再根据相似度聚合成若干大类。层次聚类算法的优势在于无需预设主题个数,但仍需指定一个距离阈值作为终止条件以结束分裂、聚合过程。划分聚类法^②要求人为预先设定主题个数,将所有文本转化成词向量并计算距离,由距离远近判断文本所属的主题。划分聚类法的输出效果依赖于给定的主题个数,因此要求研究者做出合理科学的预判。隐含主题模型法^③同样要求研究者预先设定主题个数,该方法能计算每篇文档、每个词在各个主题中的分布概率并依此对海量文本进行分类。因此,科学的先验知识同样是保障隐含主题模型法输出效果的关键。统计指标提取关键词法^④,则是通过调用 TF-IDF、N-gram 等统计算法,计算海量文本中所有关键词的权重从而遴选能够有效代表文本语意的关键词^⑤。该方法较为简便且过程原理通俗易懂,但由于基于词汇层面的统计特征忽略了对语境因素的考察,因此抽取得到的关键词需通过人工识别予以二次确认方能适用于展现主题。

从当前的工业实践及学术研究来看,上述主题挖掘方法均得到了较为广泛的运用。然而,这些方法并非适用于任意场景。总的来说,目前尚未诞生完美的自动文本分析方案^⑥。鉴于本文所选取的文本材料为网民评论,其口语化、篇幅短的特征,增加了终止阈值的预设难度以及主题个数的不可确定性,采用单一的聚类方法容易导致主题分类缺乏精确度。因此,为了最大程度保证机器学习的精度与效率,本文选择 TF-IDF 关键词编码与隐含主题模型 LDA 相结合的方法进行主题挖掘^⑦。具体而言,首先采用 TF-IDF 方法遴选语料库关键词,回归关键词所属语境进行人工编码形成对话料的先验知识,构筑主题识别的维度数量;在此基础上,调用隐含主题模型 LDA 进行主题判别和输出;最后通过朴素贝叶斯算法构建自动文本分类器将所有评论归类生成独立语料库。

(四)情感分析:网民情绪的评判工具

情感分析指基于计算机算法对文本材料进行情感倾向评判的过程。目前实践中的情感分析主要包括两种技术路径,一是情感词典标记法,二是机器学习分类法。情感词典标记法^⑧的原理是通过标注具体的情感词汇(积极词汇、消极词汇),统计该词汇在文本中的出现频率,据此为文本累加(减)积分从而得到一段文本的情感得分。为保证情感分析的效果,该词典不但需要符合研究样本的表达习惯,也要求对接研究领域的专业知识,对研究者的先验知识提出了较为苛刻的要求。机器学习分类法^⑨的原理是经过语料库训练形成文本情感判断模型,从而对海量文本内容进行自动的情感得分赋值。该方法依赖语料库的人工标注,本质上属于“机器学习中的有监督学习”范畴的文本分类。为保障效果,选择用于训练的语料库应在语言结构、内容主题方面与

① 贾燕花、徐蔚鸿:《K-means 聚类和支持向量机结合的文本分类研究》,《计算机工程与应用》,2010年第22期。

② 贺玲、吴玲达、蔡益朝:《数据挖掘中的聚类算法综述》,《计算机应用研究》,2007年第1期。

③ 唐晓波、房小可:《基于文本聚类与 LDA 相融合的微博主题检索模型研究》,《情报理论与实践》,2013年第8期。

④ 黄承慧、印鉴、侯昉:《一种结合词项语义信息和 TF-IDF 方法的文本相似度度量方法》,《计算机学报》,2011年第5期。

⑤ 赵晓航:《基于情感分析与主题分析的“后微博”时代突发事件政府信息公开研究——以新浪微博“天津爆炸”话题为例》,《图书情报工作》,2016年第20期。

⑥ Grimmer Justin and Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts", *Political Analysis*, Vol. 21, No. 3, 2013.

⑦ 王婷婷、韩满、王宇:《基于“21世纪海上丝绸之路”文献的文本挖掘研究》,《统计与信息论坛》,2017年第11期。

⑧ 周咏梅、杨佳能、阳爱民:《面向文本情感分析的中文情感词典构建方法》,《山东大学学报(工学版)》,2013年第6期。

⑨ 荀静、刘培玉、杨玉珍、张艳辉:《基于潜在狄利克雷分布模型的多文档情感摘要》,《计算机应用》,2014年第6期。

实际测试样本保持相近。

考虑到网络语言在情绪表达方面与传统书面语言存在显著差异^①,词典分析的方法在捕捉情感讯号时,有可能存在偏差从而导致语意误判。因此,本文选择机器学习分类法,实现网民评论情感的精确表达;通过捕捉舆论全过程的情感分布,揭示情感变化趋势,并结合政府行为介入的前后变化,反映政府回应所引起的舆论场变迁,从而评判政府回应的效果。

(五) 研究框架:治理效果的认知基础

舆情具有生命力,网络群体性事件从发生到终结往往具有一定的生命周期,这既反映出对政府介入的现实需求,也为研究者提供了可识别、可考察的要素和载体。自然语言处理、机器学习技术的蓬勃发展,启发我们拓展学科研究的工具集合,大数据方法的广泛应用为层次丰富的大样本研究提供了技术支撑。同时,主题和情感作为文本内容的重要维度,亦能有效拓宽学科探索的边界,为理论探讨提供落脚点。基于以上分析,本文认为利用网民评论的主题、情感可以充分勾画网络群体性事件中网民对政府回应的反馈图谱,将数据文本分为政府介入前后两个阶段分别展开分析和探讨,运用大数据工具对政府回应效果进行有效考察,从而构筑对网络群体性事件中政府回应的清晰认知和优化路径。研究逻辑框架具体如图1所示:

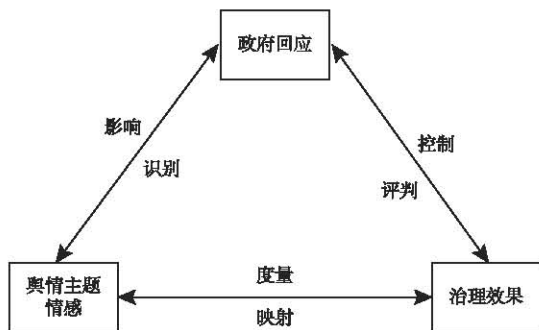


图1 研究逻辑框架

三、研究设计:雪乡事件的大样本呈现

(一) 案例选取

2017年12月29日,自媒体一木行在微信公众号平台发布了一篇名为《雪乡的雪再白也掩盖不掉纯黑的人心!别再去雪乡了!》的文章,讲述了作者本人在黑龙江雪乡旅游景点“赵家大院”旅馆的被宰经历,包括房间坐地起价、60元一桶泡面、扣留房费订金等细节。该文章经过社交媒体的裂变式传播引发了全国性关注,其阅读量在极短时间内突破十万峰值。

门户网站、新闻媒体等媒介的快速介入将舆情引向高峰,激发了网民的热烈讨论。随着舆情的不断扩散、升温,当地政府监管部门即黑龙江省森工总局大海林国有林管理局,于2018年1月3日接受新京报记者采访时称,发现舆情后大海林国有林管理局已开展调查,证实了赵家大院价格欺诈行为并采取了罚款、限期改正以及启动行业专项整治等措施。然而,“网帖后文中所说雪

^① 骆昌日、何婷婷:《网络语言的特点及其情感性意义》,《武汉理工大学学报》(社会科学版),2015年第2期。

乡售卖的桶面60元一桶及餐馆中酸菜炒肉丝几百元一盘等内容不属实”,该局相关负责人表示正考虑“走法律途径”。与此同时,有报道称,涉事旅店赵家大院对一木行展开了反击,称游客“就是想讹我钱”,网帖是勒索的工具,对此一木行公开实况录音、聊天记录等更详细的证据予以反击。值得注意的是,一木行的后续文章显示,事件发生后至1月2日,当地政府监管部门曾多次尝试联系当事人开展公关。至此,雪乡事件舆情达到最高峰,百度指数^①、新浪微博微指数^②显示,雪乡事件的搜索引擎指数、微博触达数等指标在2018年1月3日达到历史统计峰值,并保持近半月震荡。

政府回应的既定目标,是缓和舆情并实现对网络群体性事件的有效治理,然而在本案例中,政府回应导致了新的舆论战场出现并引发舆情的升温。尤其是大海林国有林管理局反驳价格指控并扬言走法律途径,以及私下联系当事人等争议性做法,使得舆论矛头直指政府主管机构,充分反映了网民对政府回应的不满和抵触。

综上所述,雪乡事件属于近期发生的热点网络群体性事件,该事件网络文本材料完整度较高,构成了充足的语料库,且文本特征适合进行主题挖掘和情感分析;案例中的政府回应对整体舆情的发展造成了可预见的影响,为我们在全过程视角下研究政府回应效果提供了可能。因此,本文选取“雪乡事件”作为典型案例,具备可行性与实效性。

(二) 分析流程与方法

1. 数据获取。考虑到门户网站对流量以及网民评论的集聚作用,本文选定在雪乡事件中资讯传播较活跃的新华网、凤凰网两大门户网站为数据源,通过python计算机语言编写的爬虫程序,对其中新闻报道下的网民评论进行抓取。

2. 数据预处理。首先进行数据清洗。当前各个商业门户网站都建有相应的反爬虫机制,网络通讯等不稳定因素也会对网络评论抓取结果产生消极影响,导致爬虫程序有可能对少量评论数据实施重复下载。因此,在正式分析之前,需要对数据进行去重操作。其次,考虑到发表评论的网民可能来自繁体字使用地区,因此本文对语料统一执行繁转简操作。此外,网民评论具有高度开放性,需要对评论中夹杂的html标签等无意义字符(@、[]、<br)进行清洗。我们一共得到18153条网民评论用于分析,时间跨度为2018年1月3日至2018年1月13日。

接着进行分词并去除停用词。汉语中词类与句法成分之间不是简单的对应关系,相较于天然分词的英语文章,中文文章在进入关键词指标提取之前需要进行分词工作。^③此外,为了保证关键词提取的效果,还需要剔除文本中无实义的词汇,如关联词、语气词等停用词。^④

3. 主题挖掘过程。本文采取TF-IDF统计算法、LDA隐含主题模型、基于朴素贝叶斯的文本分类法完成主题确定、可视化、文本归类的完整主题挖掘流程。

(1) 主题确定。TF-IDF法是通过计算词频(term frequency)和逆文件概率(inverse document frequency)这两项指标来实现关键词的判断提取和权重计算。其中,词频表示关键词在整个语料库中出现的频次,审查逆文件概率则是为了考察关键词在多条评论中的共现状况,基于两项指标算法匹配每个关键词的权重并排序。对于得到的关键词,笔者回归语境进行筛选,进一步剔除与

① 参见百度指数工具,2018年3月1日,http://index.baidu.com/? from = pinzhuan,2018年3月1日。

② 参见新浪微博微指数工具,2018年3月1日,http://data.weibo.com/index,2018年3月1日。

③ 采用中文分词系统“jieba”处理语料分词并生成词条矩阵。

④ 采用哈尔滨工业大学停用词表。

主题不匹配的词语,并在筛选后的关键词的基础上建立主题维度编码表(参见表1),挖掘出雪乡事件网民评论的主题维度。

表1 主题维度编码表(部分)

关键词	权重	所属语境	主题编码
东北	0.238	早年就有俗话在传“东北人都是黑社会”,全国民众要对此深信不疑啊!	地域攻击
雪乡	0.197	心心念念的雪乡,本来我很想去的,现在看到你的评论,不想去了,心凉到极点。	讨论旅游策略
旅游	0.094	多亏看到报导,胆子小,有威虎山的感觉,不去了,还是出国旅游吧。	讨论旅游策略
黑龙江	0.065	十几年前去过黑龙江,从此发誓不过山海关。	地域攻击
坑人	0.056	太坑人!要严惩!	评论事件要素
宰客	0.045	旅游景点不宰客的有几个呀。	反思旅游业乱象
看雪	0.037	看雪我只去北海道。	讨论旅游策略
哈尔滨	0.022	看雪,去哈尔滨就行了,其余的地方别去。	讨论旅游策略
日本	0.022	东北人的脾气应该是日本人宠出来的。	地域攻击
海南	0.022	海南三亚也是被这黑龙江人搞黑的,看来他们的本质就是黑的。	地域攻击

(2)主题可视化。我们承接上文关键词编码获得的5个主题维度,输入LDA隐含主题模型^①进行数据可视化。结果显示,5个主题相互间离散程度较高,呈现出良好的模型拟合效果。其中,topic1(地域攻击)、topic2(反思旅游业)、topic3(政府回应反馈)、topic4(事件要素)、topic5(旅游策略)分别覆盖了24.3%、22%、20%、19.9%、13.8%的文本信息,见图2。

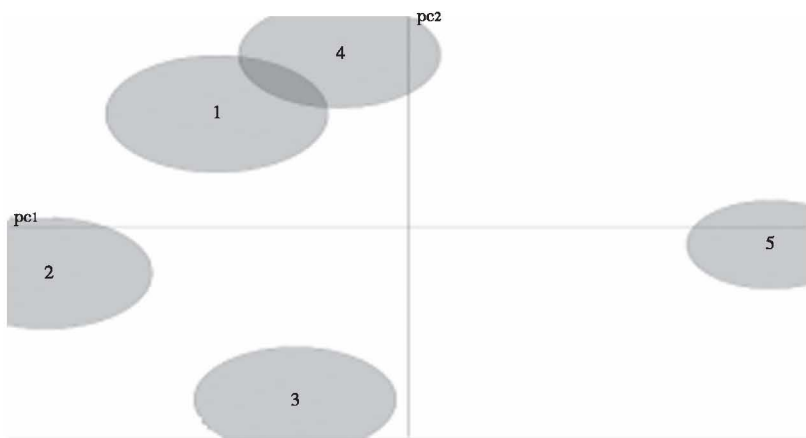


图2 主题维度分布象限

(3)主题归类。为了深入挖掘网民评论的结构内涵,笔者对1500条网民评论进行主题标注(地域攻击 T_1 、评论事件要素 T_2 、讨论旅游策略 T_3 、反思旅游业 T_4 、政府回应诉求/反馈 T_5),通过朴素贝叶斯算法训练文本获得分类器,最后调用分类器对剩余所有评论进行主题归类形成独立语料,见表2。

① D. M. Blei, A. Y. Ng, and M. I. Jordan. "Latent dirichlet allocation", *Journal of machine Learning research*, Vol. 1, No. 3, 2003.

表2 主题分类编码标注

语料	主题标注
黑龙江嘛,黑。	T1
三人炕,没毛病呀,涨价就不对了。	T2
多年前就听说雪乡是个坑,一直没去。有时间去北海道看雪!	T3
好像中华大地哪里都一样。	T4
人民的保护神你在哪里?!	T5

4. 情感分析过程。本文通过基于朴素贝叶斯(Naïve Bayes)的机器学习分类法,对雪乡事件的网民评论进行情感得分计算。首先,在分词后更新文档的基础上随机选取1500条网民评论作为训练集,逐条标注情感倾向并构建模型,接着调用模型对“雪乡事件”语料库中的所有评论进行情感倾向预测,输出0~1的数值作为情感得分(越接近0表示情感越消极,越接近1表示情感越积极)。对于结果,本文依照分值将情感分级,分别定义为: s_1 表示非常消极($0 \leq s_1 < 0.2$), s_2 表示比较消极($0.2 \leq s_2 < 0.4$), s_3 表示一般($0.4 \leq s_3 < 0.6$), s_4 表示比较积极($0.6 \leq s_4 < 0.8$), s_5 表示非常积极($0.8 \leq s_5 \leq 1$),共五个级别。表3为训练语料标注示范。

表3 情感分类编码标注

语料	情感评判(积极/消极)
讲的大实话,太坑了。	消极
你们把老天爷对你们的厚爱糟蹋了!	消极
敲一个好看的门,说:施主,路过此地,可否坐一下你的炕?哈哈哈哈哈	积极
反应还挺迅速,如果能未雨绸缪就更好了。	积极
北方哪儿都可以看雪,没必要去民风彪悍又不讲理的地方看,挨宰还不敢吭气。	消极
我一号才从哈尔滨玩完,感觉超棒,当地人热情,司机也能说哈哈哈哈哈,看雪不一定去雪乡啊,冰雪大世界就有,路上就有,市区也有。	积极
滚吧,出了问题事后监管管用。	消极

四、数据结果:回应效果的多维度分析

(一)样本的主题分布状况

政府回应介入前,舆情初期新闻报道的关注点在于还原事件细节,网民的评论主要围绕事件话题展开,按语料库所占比重排序依次为讨论旅游策略(30.49%)、评论事件要素(9.56%)、反思旅游业现状(5.24%)、政府回应诉求(1.75%)。此外,剩余52.96%的网民评论属于地域攻击主题,他们自发划分为“东北阵营”与“全国阵营”,展开了脱离于事件要素之外的交锋。此类语句包含较多极端情绪,具备较强的互动性,往往能够营造论战氛围从而加剧舆情扩散。对此,有学者研究表明^①,网络舆情中普遍存在地域攻击话语泛化现象,政府应重视并加以引导和治理。

按同样步骤对政府回应介入后的网民评论进行主题挖掘,政府行为成为舆情聚焦点之一。各主题在语料库中的比重依次为地域攻击、反思旅游业、政府回应反馈、评论事件要素和讨论旅

^① 徐顽强、谭伟:《网络新闻跟帖中地域歧视现象的现实解读与理性反思》,《电子政务》,2013年第9期。

游策略,所占比例依次为 41.30%、21.44%、15.94%、13.25% 和 8.07%。由此可推断,雪乡事件中的政府回应触发了网络讨论聚焦点的转移,除了地域攻击主题依然占最大比重外,其他主题的次序较前一阶段均有调整。反思旅游业、反馈政府回应等事后话题的讨论热度超过了事件要素、旅游对策等事中话题,显示出舆情的显著变迁。

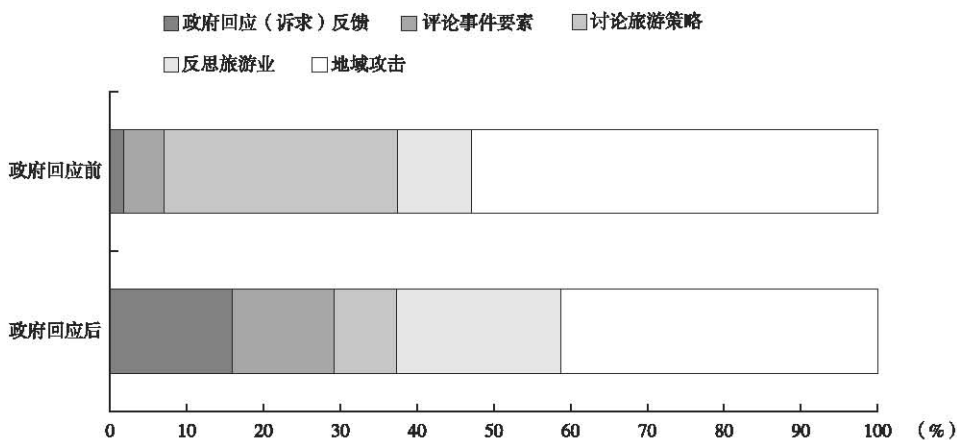


图3 主题分布比例

(二) 样本的情感倾向分布状况

总体而言,数据结果显示,无显著情感倾向评论(s_3)仅占语料总体的 14.33%,大部分网民评论具有情感倾向。其中,积极情感评论($s_4 + s_5$)为 36.77%,消极情感评论($s_1 + s_2$)则达到 48.89%,且非常消极(s_1)的比重达 34.13%,是所占比重最多的情感倾向,反映网民对雪乡事件的负向整体认知。

然而,将样本按政府回应介入前后划分,可观察到网民情感的分布存在显著差异。政府回应介入前,样本情感得分均值为 0.468,按照前文界定的情感级别属于无明确情感倾向的一般级别(s_3)。在此阶段,积极情感评论($s_4 + s_5$)占比最大,达到 51.1%,说明网民参与消极网络群体性事件的情绪未必是消极导向,舆情引导及治理存在基础可行性。

政府回应介入后,样本情感得分均值为 0.385(s_2),较政府回应介入前的情感得分有所下降,整体情感倾向落入消极范畴。具体而言,积极情感评论($s_4 + s_5$)比重大幅下降至 28.2%,消极情感评论($s_1 + s_2$)比重则上升至 55.8%,且其中非常消极组别(s_1)的变化幅度最为显著,比重达到 39.2%,较前一段阶段上升 14 个百分点。这说明政府回应对整体舆情的疏导和治理效果并不乐观,反而发生负向调节。

(三) 基于主题分类的情感挖掘

前文展示了雪乡事件网络舆情的主题及情感概况。由此可见,政府回应的介入导致网络舆情综合分布的变迁。为了深入探讨政府回应对网络舆情影响机理,需明确细分维度深挖文本,结合文本主题与情感的结构特征综合审视政府回应的作用过程。对此,本文采取有监督机器学习法将全体文本分为地域攻击(T_1)、事件要素(T_2)、旅游策略(T_3)、旅游业现状反思(T_4)、政府回应(T_5)五个主题维度,并输出政府回应介入前后的情感倾向差异。结果显示,议题聚焦点方面,除地域攻击外其他主题在政府介入后均存在议题点转移;情感变迁方面,除地域攻击、政府回应诉求(反馈)外,其他主题在政府介入后均存在消极化倾向(见图 5、图 6)。

1. 地域攻击与网民情感。围绕该主题展开的讨论未见明显的情感变迁,消极情感占比从政

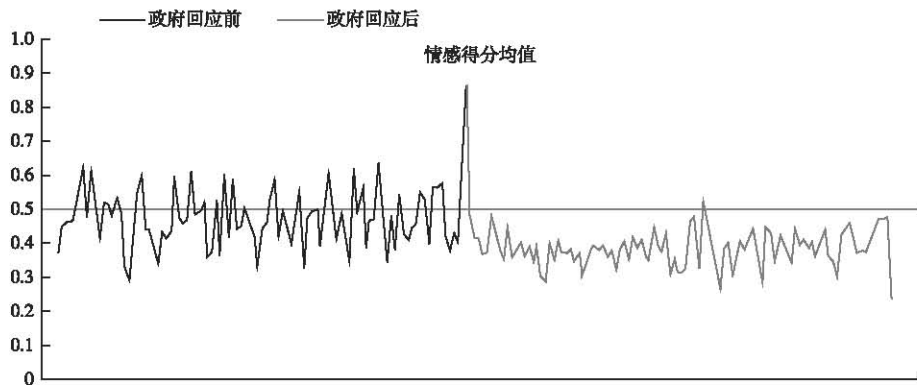


图4 情感趋势

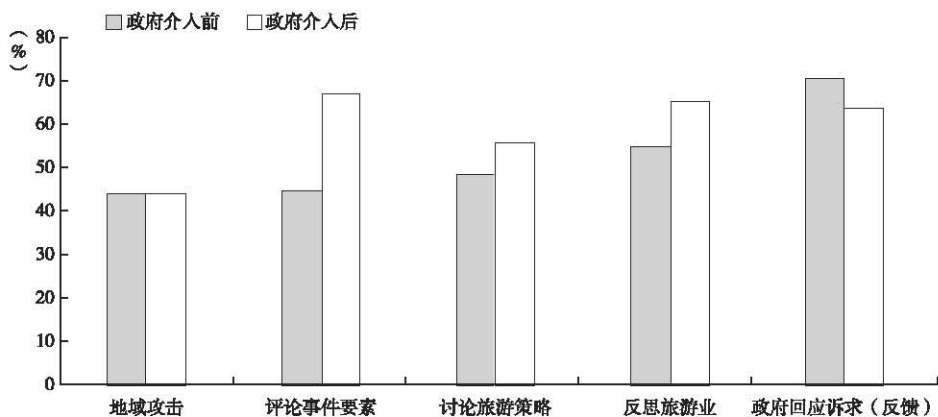


图5 情感主题结构(消极情感)

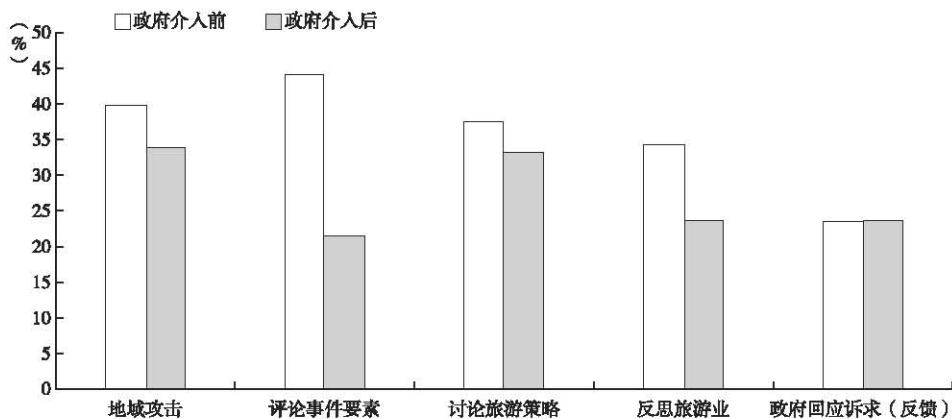


图6 情感主题结构(积极情感)

府回应前的 43.88% 略微增加至 43.92%，积极情感占比则由 39.81% 下降至 33.85%，说明政府回应对该类舆情的调节作用并不显著。具体内容方面，政府回应前后议题中心亦较为相似。外地网民将雪乡地理概念泛化为东北，进行针对东北人的地域攻击。此类语料包括直接采用网络

暴力语言对东北人进行人身攻击,以及陈述理由佐证人性评判,如称东北人常年在外“作恶”已经是全国人民的普遍认知。东北网民进行自我辩解并展开反击,称宰客现象并非东北特有,“不要片面攻击”。此外,地域攻击呈现差序格局,部分吉林、辽宁网民称“不要把东北人都带上,那是黑龙江”,表达了在面对地域攻击时东北群体内部的立场分化。

2. 事件要素与网民情感。该主题呈现出情感消极化转变,消极情感占比从44.62%大幅上升至66.97%;积极情感占比大幅下跌,从44.09%降至21.48%。审视具体内容发现,政府介入前,为原作者爆料行为“点赞”的评论构成了重要语篇,如有网民称“谢谢作者良心披露”,“感谢此文,了解真相”等,使整体舆情保持在中等水平。政府回应介入后,有关部门的调查结果披露了详实的宰客违法证据,并公布了涉事旅店的辩解和对原文内容的反击,因此引发了网络谩骂,如评价旅店老板“狡辩无耻至极”,“颠倒黑白,心真是黑透了”等,造成整体舆情急转直下。

3. 旅游策略与网民情感。该主题消极情感占比从48.57%上升至55.79%,积极情感占比从37.44%下降至33.23%。考察主题内部议题变迁可知,政府介入前后议题皆围绕反对出游雪乡、探讨替代旅游方案两个向度,其中受政府证实宰客事实及涉事旅店辩解、“不认错”的影响,后期网民评论夹杂了更多负面情绪化表达,从而导致情感消极化,如“这种地方就该让它自生自灭”“作茧自缚谁敢去”等。

4. 旅游业现状反思与网民情感。该主题消极情感占比从54.90%上升至65.11%;积极情感占比从34.31%下降至23.69%。其中,反思国内旅游业宰客现象及探讨行业经营之道是最为聚焦的两个话题。政府回应前,网民主要对国内旅游业整体现状表达不满,代表性语料包括“全国景点都一样”“天下乌鸦一般黑”;政府回应介入后,源于旅店言行的进一步披露,网民评论的焦点转移到行业经营模式的深入思考,称“杀鸡取卵,竭泽而渔”,“诚信经营才长久”。

5. 政府回应与网民情感。该主题情感变迁不显著,消极情感占比从70.59%下降至63.57%,积极情感占比则基本保持不变,仅从23.53%上升至23.72%,但在具体内容方面呈现出显著变迁。在政府回应介入前,网民评论探讨最多的议题点有两个方面,一是呼吁政府的介入,如“党和政府管不管”,倡导“政府监管部门有所作为”;二是直接批评政府监管不到位,如把旅游宰客问题看成是“有关部门的问题”,指出问题的关键“还是监管不力”。政府回应介入后议题中心发生偏转,少量评论对有关部门的惩治行动表达了认可,称“看到了主管部门的诚意”,“态度很中肯”。然而,大部分评论依然对政府回应的个别细节表达了质疑和不满,其中包括:一是针对事件调查结果的质疑,如称官方回应“60元是编号而非价格”是“不可能”的,商家为了应付检查“赶紧变便宜”,“等检查人员一走立刻就涨价”;二是对政府前期行政监管缺失的不满,如抱怨“处罚总在媒体的曝光后”,称“这一套见多了,不揭露不整顿”;三是对政府立场的怀疑,如称有关部门为了经济利益“庇护”宰客行为,企图“用行政力量强行洗白”挽救雪乡品牌。

五、研究发现:舆情事件的结构化认知

(一)事件要素聚焦下的舆情层次扩散分布

网民对网络群体性事件的关注议论点,并非仅局限于事件要素,而是围绕事件核心要素,呈现出由内向外扩散、层次逐渐提高的舆情话语主题格局。具体而言,网民评论呈现出从聚焦“微观剧情要素”(控诉宰客行径,声讨涉事旅店),到探讨“中观行业状况”(分享旅游策略,感慨旅游业乱局),再到阐发“宏观社会感知”(片面偏激的地域攻击)的议题空间,显示了网络群体性事件中网民的关注逻辑。

（二）政府回应影响下的事态阶段变迁

作为网民意见的载体,网络评论具有特定指向性,与网络资讯的内容要素紧密关联。而网络群体性事件作为动态的过程,其衍生出的网络资讯也存在内容和视角的变迁。研究表明,舆情初期网络资讯报道主要承担了事件曝光的职能;舆情中后期,则主要对政府回应进行跟进报道,讲述黑龙江省森工总局大海林国有林管理局的相关查处过程及结果。因此,网民评论主题要点也与之匹配,行成了从议论“事中”剧情到讨论“事后”政府所作所为的舆情变迁。

（三）地域攻击话语下舆情的恶化升级

网络地域攻击可以理解为由网络群体性事件中的地域属性触发了网民的地域刻板印象,从而引起不同地域网民超脱事件本身的非理性语言互动,通常伴随暴力语言的使用。虽然各大门户网站均倡议“文明发言”“理性发帖”,然而地域攻击现象仍然普遍存在于网络资讯的跟帖板块,甚至以地域攻击为主题的网帖常常成为热门网帖。本研究发现,政府回应介入前后,地域攻击一直占据着舆情主题分类的最大比重。其中,凭借数量优势,外地网民对当地地区网民展开围剿,造成舆情立场一边倒。另外,相关评论引来了较为热烈且频繁的网民互动,通常呈现出你来我往、争锋相对的攻击与反驳,使舆情在该主题范围内进一步集聚、增长。

（四）网民情感趋势下情绪的消极发展

纵观舆情发展的全过程,政府回应对整体舆情的疏导和治理效果并不乐观,反而导致舆情向消极化发展,消极情绪的占比较政府回应前显著增加,积极情绪的占比相应大幅下滑。与此同时,不同主题领域,对政府回应的情感反应也呈现不同。其中,地域攻击主题由于内容与政府回应匹配度较低,因此情感走向的变化幅度较不明显,这在一定程度上印证了政府回应对该类主题的治理失效。事件要素主题、旅游策略主题,牵涉较多政府介入的细节内容,因此,情感走向受政府回应调节表现为负相关。同样,旅游业反思主题涉及较多行业秩序、政府监管、公共服务等领域内容的探讨,其情感走向也随政府回应发生显著负向变动。而在针对政府回应的诉求、反馈主题领域,虽情感走向未发生显著变动,但仍可从具体内容方面观测到大部分评论针对政府回应的细节表达了质疑和不满。

六、事与愿违：政府回应与舆情导向的互动逻辑

通常而言,网络群体性事件大多被视为网络舆情的非理性集聚,可能会对社会秩序造成冲击,影响社会和谐稳定,因此政府倾向于采取管控和压制等应对手段,否定网络群体性事件的隐含意蕴。随着社会的不断发展,尤其是在互联网政治参与的普及,政府逐渐认可网络舆情聚集在民意表达、政民互动乃至优化公共服务等方面的重要价值。因此,政府对网络群体性事件的立场态度,逐渐由控制转变为疏导,强化政府介入和回应,力图实现网络群体性事件的善治。然而,在实践操作中,政府回应并非总能达到对网络舆情的正向调节,有时甚至会造成舆情转向消极化发展,以致“引火上身、事与愿违”。

（一）政府管控与网民诉求间的偏差

政府回应的合法性和必要性虽得到普遍认可,然而,在现实情境中,维稳工作仍是官员绩效评价体系中的重要组成部分,因此,在现行机制下,政府往往从成本的角度出发,采取消极策略回应网络舆情,从而管控、压制舆论的规模。然而,随着互联网浪潮的发展,网民参与网络群体性事件的期望,已从低层次的围观需求向更高层次的公共诉求转变,这个过程伴随着对良好政府回应性的希冀。本案例中,在舆情爆发的最初阶段,当地旅游监管部门积极捕捉识别舆情,体现了政

府部门对网络舆情的重视,政府及时采取了措施进行舆情管控。然而,在掌握事态进展后,政府并未第一时间表态或者公布调查结果,而是优先选择开展公关,派遣工作人员私下联系,意图说服当事人,引导并控制舆情导向。然而,在当事人拒绝私下和解,并在网络公开上述细节后,网民纷纷表示对政府“诚意”“立场”的不认可,用质疑和谩骂表达对政府回应方式的不满,正是这种维稳导向的管控思维,使网民议论的焦点转向政府。

(二) 回应滞后与“首因效应”间的偏差

政府回应的主动性是引导舆情走向的重要前提,同时也是达成网络群体性事件治理目标的关键要素。有研究表明,网络舆情极易受“首因效应”影响,舆情初始阶段塑造的“第一印象”,将很大程度决定舆情传播参与者的态度及立场,后续信息的调节作用将受到“首因效应”的影响,对事件的澄清、解释作用相应地加强或减弱。^①在网络环境下,信息不对称造成网民对网络群体性事件的认知壁垒,故事在黑箱中演绎,整体舆情容易受少数具有话语权的意见领袖牵引。政府回应时大多保持审慎的态度,以求不在公开信息中留下把柄。这种思路在以往的实践中本无可厚非,但随着信息社会的到来,政府需要转变思维,加快对突发舆情的处理。众多案例表明,政府回应时间的滞后性,已越来越难以适应瞬息万变的网络环境。对于网络中讨论的热点事件,政府回应的滞后首先会给网民带来不好的印象,且可能在被迫回应之后出现更大规模的混乱,如何找到最优点,是政府必须衡量的问题。

因此,为了确保网络舆情的可控性,实现网络群体性事件的治理,需要政府积极、主动、及时介入。在本案例中,在舆情爆发初期,当地政府监管部门虽然介入事件调查,却采取消极被动的方式进行回应。整个回应过程政府未出台官方声明,仅仅在相关媒体联系求证之后,于舆情爆发第六天委托第三方表明立场,发布相关事件调查结果。这种贻误时机的回应方式,最终造成政府在引导整体舆论导向的缺位,也暴露出政府回应不及时、不诚恳等顽疾。

(三) 正当履职与信任缺失间的偏差

在网络空间,由于信息传播具有低门槛、即时性、裂变性的特征,失去公众信任的政府,在面对网络公共事件时很容易陷入“塔西佗陷阱”^②。具体而言,在网络群体性事件中,部分网民倾向于对政府的介入、回应作消极理解,不关注政府的具体工作内容,只是一味质疑政府行为目的和用意。“塔西佗陷阱”的关键性诱因,是政府公信力的缺失,这与长久以来政府应对网络舆情时所表现出的不透明、不诚实、不客观等密切相关。本案例中,从政府介入前的网民评论中,即可观察到网民对政府处理宰客事件存在刻板印象,网民先入为主地认为政府会包庇宰客违法行为。此外,当地政府监管部门公布的调查处置结果也引起网民争议,被网民认为是政府监管不力、政府隐瞒事实糊弄公众,从而引起了消极舆情大幅集聚,展现公信力缺失条件下政府回应的无奈和乏力。

七、结论

本文基于对“雪乡事件”舆情大数据的分析和挖掘,以过程视角对案例进行深入剖析,阐释了网络群体性事件舆情的分布、变迁状况,探讨了政府回应与网络舆情存在的逻辑互动偏差,构筑了群体性事件认知的环境和范围,为当前网络群体性事件治理提供了详实的经验素材,为政府

① 叶皓:《突发事件的舆论引导》,第34页,江苏人民出版社,2009年版。

② 杨妍:《自媒体时代政府如何应对微博传播中的“塔西佗陷阱”》,《中国行政管理》,2012年第5期。

治理提供良好的目标导向。然而,在现实情景下,网络舆情作为一个涵盖多种要素影响的复杂系统,不同要素以及彼此之间的集群互动,会产生多重效应,导致网络舆情的互动变迁。因此,厘清上述复杂系统过程,明确政府介入要素在整个效应场中的作用及边界,通过舆情变迁揭示网络群体性事件治理效果至关重要。

鉴于网民评论是网络舆情研究的重要载体,而主题及情感作为网民评论的内容要素,能够承担网络舆情维度挖掘的任务,笔者认为对主题及情感的测度,能够较为全面地刻画网络舆情的分布、变迁。而对于网络舆情的要素范围、传播路径和关系网络等关键构件,则需从更多维度及视角展开分析,需要优化 TF-IDF、朴素贝叶斯模型等算法,消除机器学习程序在主题分类、情感赋分环节的误差,深入探讨经典算法内容及规则,有效解决自然语言处理及机器学习中的算法适用性、科学性等关键问题,推动跨学科交流,强化学科知识与案例情景的适配性改进,确保政府回应介入的准确性,推动大数据研究与实验研究、实证分析的结合,提升大数据计算的精确度和效率。

作者:文宏,华南理工大学公共管理学院(广东省广州市,510641)

(责任编辑:孟令梅)