

重大突发公共卫生事件下的公众情感演进分析:基于新冠肺炎疫情的考察^{*}

边晓慧¹ 徐童²

¹(安徽大学管理学院 合肥 230039)

²(中国科学技术大学计算机科学与技术学院 合肥 230027)

摘要:【目的】对重大突发公共卫生事件背景下公众在社交媒体中的情感表达进行分析,揭示疫情期间公众情感的时空差异、不同主题下的情感演化以及情感的跨地域扩散。【方法】利用主题模型提炼潜在话题与关键词群,从全局视角和主题视角探究公众情感演进趋势,并使用社交传播模型描述公众情感的跨地域扩散。【结果】疫情期间公众以积极情感为主,消极情感呈现“恶”的情感主导、“惧”的情感先发、“哀”的情感反复等特征;疫区距离与经济水平导致公众情感存在空间差异。同时,情感表达及演化趋势因受时空变化、主题/事件区别等影响而具有一定规律的差异。此外,公众情感的地域扩散强度受空间关系和疫情严重性的双重影响。【局限】面向纯文本信息,无法对多模态信息如视频、图片等进行综合性分析。【结论】重大突发公共卫生事件下,公众在社交媒体上的情感表达及演化趋势受时空差异、主题差异等影响,并存在一定的地域扩散规律。这提示疫情防控要结合特定时期、特定地域采取差异化策略,关注不同主题类型对情感的关联性影响,同时注重疫情防控与舆情监控的区域统筹与合作,以实现舆情的积极引导和公众情感的有效疏解。

关键词: 突发公共卫生事件 微博舆情 情绪演化 主题分析 时空分析

分类号: G203

DOI: 10.11925/infotech.2096-3467.2021.0711

引用本文: 边晓慧, 徐童. 重大突发公共卫生事件下的公众情感演进分析:基于新冠肺炎疫情的考察[J]. 数据分析与知识发现, 2022, 6(7): 128-140. (Bian Xiaohui, Xu Tong. Evolution of Public Sentiments During COVID-19 Pandemic[J]. Data Analysis and Knowledge Discovery, 2022, 6(7): 128-140.)

1 引言

突发公共卫生事件具有突发性、急剧性、不可预测性等特征,往往会对人民群众的生命健康造成严重威胁,并引发公众恐慌、不满、愤怒等复杂情感^[1]。当下,社交媒体已成为突发事件舆情传播的重要平台,并为突发事件相关信息的迅速传播及复杂、多元情感的产生与演化提供了渠道。以2020年初的新型冠状病毒肺炎(简称新冠)疫情为例,其社交媒体参与度远高于2003年的非典疫情。例如,微博平台

的上亿用户通过微博关注疫情最新情况、获取防疫服务、参与公益捐助,微博不仅成为疫情信息传播的主战场,也成为公众情感的集散地。因此,在突发公共卫生事件背景下,通过挖掘社交媒体海量数据实现对公共舆情情感的追踪识别和演化分析,不仅有助于甄别公众情绪和实施心理疏导,也可为把握事件发展轨迹、改善应急管理效果提供计算支持^[2]。

鉴于此,本研究对微博平台中的新冠疫情相关博文数据进行挖掘,以了解疫情期间的公众情感演

通讯作者(Corresponding author): 边晓慧(Bian Xiaohui), ORCID: 0000-0002-1583-971X, E-mail: bianxh@ahu.edu.cn。

^{*}本文系国家自然科学基金项目(项目编号: 16CZZ025)的研究成果之一。

The work is supported by the National Social Science Fund of China (Grant No. 16CZZ025).

进规律。需要注意的是,此次新冠疫情持续时间长,波及范围广,疫情涉及热点事件及话题众多,舆情倾向性也随事态发展而改变。同时,不同地区用户情感受疫情影响程度也有较大差异,并存在明显的跨地域传播。这都使得单一围绕特定时间-地点的信息切片,或仅面向单一主题的舆情分析无法完整反映公众舆情全貌。针对这一背景与挑战,本文利用概率主题模型提炼潜在话题,从全局视角和主题视角分析比较公众情感演进趋势,并采用社交传播模型描述公众情感的跨地域扩散,最终获得面向重大公共卫生事件背景的全景化、细粒度、差异性舆情演进分析。

2 文献综述

2.1 突发公共事件下社交媒体情感的特征与类型

在网络时代,越来越多的信息通过社交媒体进行传播。不同于传统传播将理智和情感分置于公共空间和私人空间,社交媒体打破了公私界限,使得情感代替理智成为舆情的中心^[3]。在紧急突发状况下,公众越来越多地通过社交媒体获取和分享各类信息^[4],用户的情绪与情感也伴随事件信息的披露与传播而不断集聚、演化。

突发事件背景下,社交媒体用户的情感表达具有若干典型特点:一是情感多变性。突发事件下,社交媒体用户情绪深受不同性质事件信息的影响,伴随多元信息的持续披露,公众情感也处于波动和演化之中^[5]。二是消极情感更易传播。由于突发公共事件具有较强的负面属性,更易引发网民的恐惧、悲伤、焦虑、愤怒等负面情感^[6]。三是情感极化。突发事件信息通常不能在短时间内清晰显现,而各种谣言又极易传播,这些片面、虚假的信息极易驱动用户产生极端态度和情绪^[7]。

从类型来看,突发公共事件下的情感研究着重关注愤怒、恐慌等情感。愤怒是当愿望不能实现或行动受到挫折时引发的一种紧张而不愉快的情绪^[8]。在突发事件情境下,愤怒情感极易在社交媒体上发酵升温,这是因为突发公共事件在性质上即属于违背个体预期和目标的一种非期望行为,同时就其传播而言,愤怒也被发现比其他情绪更容易感染用户^[9]。恐慌则是人们遇到突发危机时的一种应

激性反应^[10]。在突发事件刺激下,恐慌主要是由于人们对所面临事件的危险性、可控性存在不确定或模糊的理解,进而引起在应对过程中的心理不适^[11]。群体恐慌情绪给突发事件的救援和事后处理工作带来极大挑战^[12]。

2.2 突发公共事件下社交媒体情感的时间演化与空间差异

重大突发事件有持续的过程,这导致用户对待事件的情感存在发展演化。已有研究集中于从时间维度探讨面向突发事件的用户情感阶段划分,或基于特定的突发公共事件分析公众情感演化的规律和特征,如安璐等^[13]对寨卡病毒感染疫情的微博传播进行分析,区分了各个不同阶段的主导情绪类型及其演变趋势;叶勇豪等^[14]以温州动车事故为案例构建微博情绪变化走势图,发现不同“拐点”对应不同性质的事件,进而引发不同的情感;鲁艳霞等^[7]以天津爆炸事件为例,发现突发公共事件中,不同时间阶段有不同的网民参与某类话题,从而形成从不同侧面刻画该事件的集群行为。

相比于时间演化的分析,有关社交媒体情感空间差异的研究较少。解释水平理论(Construal Level Theory, CLT)提供了分析情感空间分化的有效框架。这一理论从社会距离概念出发,认为人们的认知解释取决于与认知客体的心理距离感知,人们会对感知到的远距离事件形成高水平解释(做出支持性、积极性的正面判断),而对感知到的近距离事件形成低水平解释(做出拒绝性、消极性的负面判断)^[15]。对于传染性疾病而言,疫情中心位置与人群的远近意味着不同的感染可能性。距离越近,人们感受到的危险性和威胁性就越高,消极的情感反应就会更强烈。在有关研究中,钱铭怡等^[16]对非典疫情期间北京居民认知和情绪状态的研究发现,非典肺炎与人们的空间距离减少时,人们则越多地体会到紧张、恐惧、悲观、无助等负面情绪;van Lent 等^[17]对埃博拉疫情的研究也发现,空间和社会距离均影响公民对埃博拉流行病的关注和个体情绪;但是,张放等^[18]基于新冠疫情期间微博文本面板数据的检验,未发现疫情心理空间距离对公众情绪的固定影响,空间距离的影响未得到验证。

2.3 突发公共事件下社交媒体的舆情主题与情感关联

网络情感不是凭空产生的,而是具有深刻的社会背景^[19]。在重大突发公共事件中,常常存在多元的话题,这些话题是民众对突发事件关注点的集中反映,成为网络舆情形成的基础。突发事件发展中,话题内容与强度也在持续变化^[20],而不同事件/话题引发的情感在性质与程度上存在很大差异。在社交媒体中,网民短期内的情感波动通常和突发事件中某一主题的走向紧密相关^[13],因此解释情感的变化必须重视从海量数据中进行主题挖掘,关注不同的主题及其与特定情感的关联,进而寻求舆情主题与情感演变之规律^[21],反之,如果不考虑演化的主题维度,则无法对波动的情感类型进行有效的解释和监测。

目前有关突发事件主题与情感演化的研究大多着眼于不同主题对情感在不同阶段演化的分析性解释,如任中杰等^[22]以天津“8·12”事故为例,分析舆情情感演变4个阶段的评论词云,发现了消防员英雄牺牲、事故追责、质疑领导管理部门、缅怀评论等4类相关主题;在对寨卡病毒事件的研究中,安璐等^[13]构建融合主体与情感特征的突发事件微博舆情演化分析方法,对寨卡疫情每一阶段下的聚类主题及对应的情感进行分析,揭示各个阶段的主题关联及情感扩散规律。总体上,目前对主题与情感的关联特征的研究还较为不足,这在一定程度上是因为多数突发事件的周期较短,主题也较为集中,导致有限主题与有限情感的关联缺乏区分度。而此次新冠肺炎疫情具有时间持续性、主体广泛性、问题复杂性等特征,使得深入挖掘多元疫情主题与不同情感的关联成为可能。

3 全局视角下的公众情感演化

3.1 数据来源与处理

新浪微博是目前中国使用最广泛的网络社交媒体之一,具有较好的用户代表性、信息时效性和话题覆盖度。本文采用新冠肺炎疫情期间的微博数据作为探讨重大突发公共卫生事件下公众情感演进的研究对象。

实验数据通过微博官方提供的开放平台API^①进行抓取:以疫情期间官方媒体辟谣的若干热门微博作为种子,获取参与这些微博话题讨论的所有用户,进而获取这些用户在指定时间段内的所有微博。通过这种方式,可以保证每位用户至少参与了一次疫情相关的讨论,确保必要的密度。最终获取的数据包含2020年1月1日至3月23日期间20 131名用户所创作的1 160 422条微博,每条微博记录包括用户所发表的内容和所转发微博的原内容、发表或转发的时间,以及部分用户所填写的地理位置信息。

为实现微博的疫情相关性判定,采用基于关键词匹配的方法。具体而言,选取20个与疫情高度相关的关键词,具体列表及所属分类如表1所示,微博文本中出现任一关键词即判定该微博与疫情相关。最终在全部1 160 422条微博中,与疫情相关的微博共有296 288条,占比25.5%,可见新冠肺炎疫情在从爆发到缓和的较长一段时间内都是微博上的核心话题。

表1 判断疫情相关微博的关键词列表

Table 1 Keywords for Judging the Relevance of the Epidemic-Related Microblogs

关键词类别	关键词列表
直接相关	武汉、疫情、肺炎、新冠、冠状病毒、口罩
事件相关	野味、海鲜市场、红十字会、复工、医疗队、卫健委
人名相关	钟南山、李文亮、李兰娟、高福
诊断相关	确诊、疑似、发热、体温

为实现微博情感分类和分析工作,采用大连理工大学信息检索研究室开发的“中文情感词汇本体库”^[23]对微博进行情感判定。在296 288条疫情相关微博中,具有情感属性的微博有284 449条,占有疫情相关微博的96%,有关7种情感的总体分布如表2所示。

整体来看,两种积极情感“乐”和“好”的比重分别为5.86%、66.87%,表明新冠疫情蔓延背景下公众情感仍以积极情感为主。5种消极情感中,“恶”的情感占比为14.66%,反映了微博用户看待新冠疫情的主流消极情感;“惧”的情感占比为8.15%,意味着面

①<https://open.weibo.com/wiki/%E5%B8%A8%E5%8D%A1>API.

表 2 疫情相关微博的情感分布

Table 2 Sentiment Distribution of the Epidemic-Related Microblogs

情感大类	情感子类	微博数量	微博比重
乐	快乐、安心	16 665	5.86%
好	尊敬、赞扬、相信、喜爱、祝愿	190 212	66.87%
怒	愤怒	1 637	0.58%
哀	悲伤、失望、内疚、思念	9 741	3.42%
惧	慌、恐惧、羞	23 177	8.15%
恶	烦闷、憎恶、贬责、妒忌、怀疑	41 708	14.66%
惊	惊奇	1 309	0.46%

对极具感染性的新冠肺炎,公众有较为明显的“恐惧”心理;“哀”的情感占比为 3.42%,证实了疫情期间表达悲哀、哀悼也是重要的情感类型;最后两种消极情感是“怒”和“惊”,分别占比 0.58%、0.46%,均低于 1%,表明这两种情感在此次新冠疫情事件中并不是主流。

3.2 疫情相关微博用户情感的时间演化

由于“乐”和“好”的情感均表达一种正面的评价和态度,将两者累计并绘制疫情期间积极情感比重的时序变化如图 1 所示。结合原始数据可以发现,在积极情感主导的疫情期,用户情感也存在三个明显的低点:一是 1 月 3 日出现第一个积极情感低谷(占比 58.3%),这一天武汉市卫生健康委员会通报发现 44 例不明肺炎病例,并向世界卫生组织和周边国家通报疫情;二是 1 月 9 日积极情感比重降到最低谷(占比 54.3%),当天武汉出现首例新冠肺炎死亡病例;三是 1 月 20 日积极情感再次降到 60% 以下(占比 58.3%),此日钟南山院士确认了新冠肺炎“存在人传人现象”,并传出武汉加强管制、预备“封城”等消息,导致积极情感再次下跌。整体上,积极情感的三个低点均出现在疫情爆发早期,表明突发事件初期公众情感更可能出现消极倾向且波动较大,到后期逐渐趋于稳定,表明应重视早期对公众情感的引导疏解。

相比于积极情感,消极情感的演进更加复杂。“哀”、“惧”、“恶”三种主要消极情感的微博数量时序变化如图 2 所示,可以发现:

(1)“惧”的情感增长早于其他情感,从 1 月 19 日开始迅猛增长,并在 24 日达到顶点,随后呈现下降趋势,这表明面对未知的、传染性极强的突发肺炎,

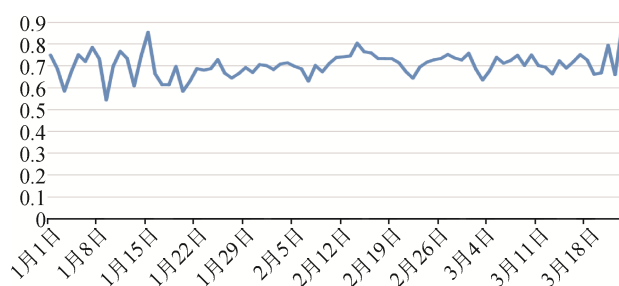


图 1 积极情感微博比重的时序变化

Fig.1 Temporal Change in the Proportion of Positive Emotion Microblogs

特别是出现感染死亡病例尤其医护人员死亡病例后,公众恐惧感极速增加,但随着政府团结抗疫的积极回应和有效防控政策的实施,“惧”的情感也很快回落。

(2)“恶”的情感持续处于较高水平,其中有三个值得注意的高点,一是 1 月 25 日前后,随着新冠肺炎感染与死亡病例的持续增长,以及政府初期应对疫情的不规范行为,民众对抗疫不力的不满情绪增长到第一个峰点;二是 2 月 1 日前后,双黄连抑制新冠引发抢购、红十字会物资分配混乱等负面话题广泛传播,引起公众较为普遍的质疑与不满,“恶”的情绪集中爆发至顶点;三是 2 月 7 日前后,李文亮医生病情恶化去世的消息引发广泛关注,对“吹哨”医务人员训诫问题再次激起公众的不满和批评。

(3)“哀”的情感也在起伏中增长,面对新冠肺炎感染死亡人数的增加以及各地抗疫人员牺牲等事件,公众哀伤的情感也不断加剧,并在 2 月 7 日因李文亮医生去世引发广泛哀悼而达到顶峰。

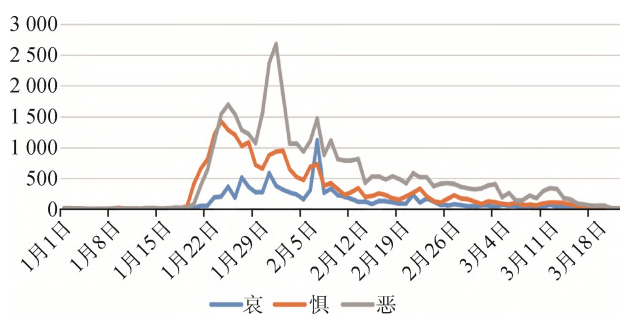


图 2 消极情感微博数量的时序变化

Fig.2 Temporal Change in the Number of Negative Emotion Microblogs

3.3 疫情相关微博用户情感的空间差异

(1) 疫情中心-边缘地带用户的情感差异

从地理空间来看,疫情类突发公共卫生事件存在疫情中心地带,并向四周蔓延传播,因此可以预期不同疫区距离的用户对疫情的关注程度、关注重点以及情感分布等存在差异。本研究以除去港澳台地区的31个省级行政区为研究对象,以湖北作为疫情中心地带,湖北相邻省份为邻近地带,通过湖北邻省与湖北间接相邻的省份归为间隔地带,其余省份为疫情边缘地带,具体分层如表3所示。

表3 基于地理距离的省份分层

Table 3 Province Stratification Based on Geographic Distance Level

分层类别	分层依据	包含省份
核心层	湖北本省(1个)	湖北
邻近层	湖北邻省(6个)	安徽、重庆、陕西、江西、湖南、河南
间隔层	湖北邻省的邻近省份(13个)	山东、浙江、江苏、甘肃、贵州、四川、山西、宁夏、内蒙古、福建、广东、广西、河北
边缘层	其他省份(11个)	黑龙江、吉林、辽宁、海南、上海、北京、天津、新疆、青海、云南、西藏

从疫情相关微博比重来看,核心层湖北的疫情相关微博占比达32.32%,显著高于邻近层(23.22%)、间隔层(23.74%)和边缘层(25.87%)。6

种情感在不同地理距离层类分布的雷达图如图3所示,“乐”和“好”的积极情感大体呈现从核心层、邻近层到间隔层、边缘层递减的趋势,表明处于疫情中心地带的微博用户比边缘地带用户有更多的积极情感。此外,“怒”、“哀”的消极情感也在核心层、邻近层较高,而在间隔层、边缘层较低,与之形成对比的是,“惧”、“恶”的情感则大体呈相反的变动趋势,这表明相比于核心地带用户,疫情边缘地带的用户表现出对疫情更多的担忧和不满。

事实上,在作为疫情中心的湖北,居民获取疫情信息更加充分、及时,特别是国家积极投入疫情防治,大力号召各地区团结支援湖北抗疫,都使得湖北地区用户对疫情普遍持有积极的情感;但另一方面,湖北省和武汉市的防疫、抗疫不力也激发了公众“恶”的情感,同时疫情初期的确诊患者死亡、医务人员感染等也使得“哀”的情感多于其他地区。而距离湖北最远的边缘层省份,有关疫情的信息传播相对滞后,导致公众在面对地方严格防控举措与有限疫情信息通报时,更容易因不确定性放大“惧”的情感,同时因疫情未能及时防治或防治不力等主流舆情产生“恶”的情感。

(2) 经济发达-落后地区用户的情感差异

经济发展程度一方面可能影响各地微博用户对于疫情的关注程度,另一方面也大致决定了各地区

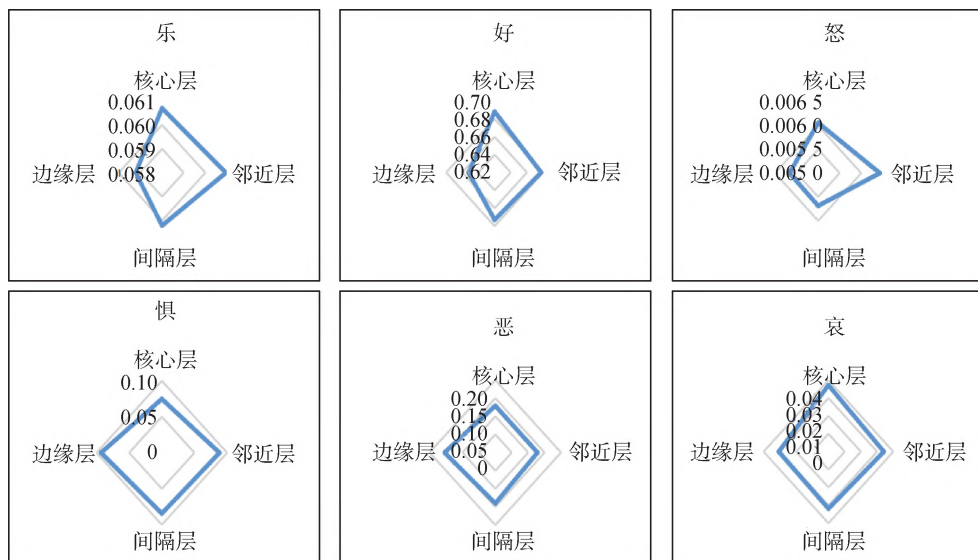


图3 6种情感在不同地理距离层类的比重变化

Fig.3 Proportion Change of the 6 Emotions at Different Geographic Distance Level

医疗基础设施、政府治理水平等抗疫能力。本研究按 2019 年各省(市)人均 GDP 将研究对象分为发达、中等、落后三类,如表 4 所示。6 种情感在不同经济发展水平的比重变化如图 4 所示。

表 4 基于经济发展水平的省份分层
Table 4 Province Stratification Based on Economic Development Level

经济水平分层	人均 GDP 区间(元)	包含省份
发达省份(10 个)	164 197~70 733	北京、上海、江苏、浙江、福建、广东、天津、湖北、重庆、山东
中等省份(11 个)	67 926~53 269	内蒙古、陕西、安徽、湖南、海南、辽宁、河南、四川、新疆、宁夏、江西
落后省份(10 个)	49 381~33 058	西藏、青海、云南、贵州、河北、山西、吉林、广西、黑龙江、甘肃

①就疫情相关微博数量来看,发达省份的微博数量最多,占比达 25.55%;中等省份的微博数量次

之,占比为 23.95%;落后省份的微博数量最少,占比为 21.72%,证实了随着经济水平提高,对疫情的关注也有所增加。

②从地域差异来看,经济发展水平越高的地区欢乐的情感越少;但对于“好”的情感而言,呈现 U 型变化趋势,这意味着经济发达地区和经济落后地区对抗疫的信心都强于经济水平中等地区,但前者可能由于对政府抗疫防疫能力的理性自信,后者则是由于较少的疫情关注或对抗疫的自我乐观。

③在消极情感上,随着经济水平逐步提高,“怒”和“哀”的情感轻微增加,但“惧”的情感有较为明显的下降,“恶”的情感则是经济水平中等、发达地区高于落后地区,这从总体上印证了经济发达程度可以在一定程度上缓解心理恐惧,用户有更少的担忧,但同时也增加了对疫情防控的批评态度,产生更多因关注疫情防治不力导致的“恶”、“怒”等负面情感。这也表明,在情感对象上,经济落后地区用户更着重于对疫情本身的情绪感受,而经济发达地区用户逐渐转向外在相关的非疫情情绪客体。

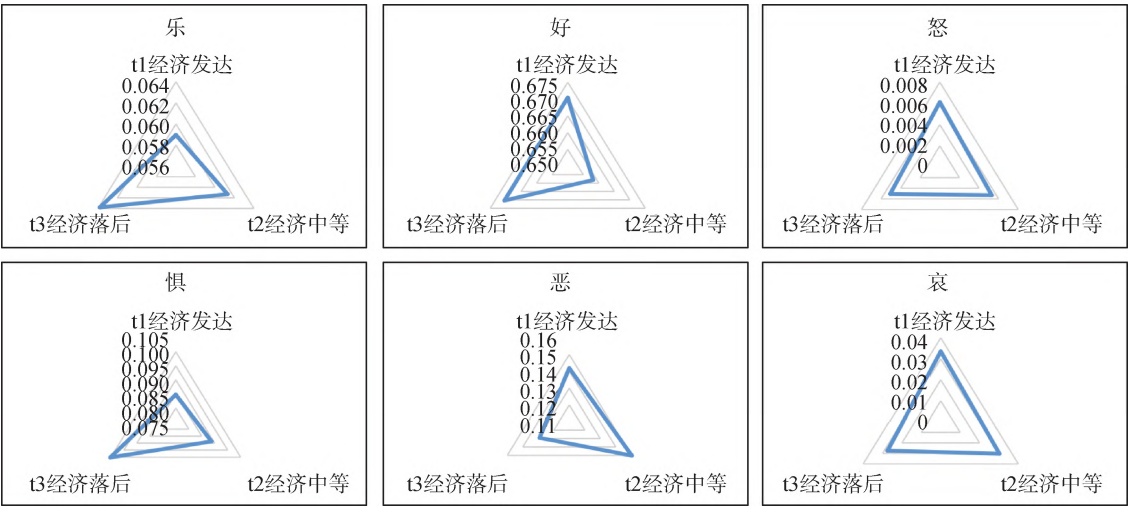


图 4 6 种情感在不同经济发展水平层类的比重变化

Fig.4 Proportion Change of the 6 Emotions at Different Economic Development Level

4 主题视角下的舆情演进与情感分析

此次疫情涉及话题众多且复杂多变,各个话题在时空分布与相互关联上存在明显不同,因此单一的全局视角可能因其笼统性而中和了各种情

感差异。为此,借助概率主题模型提取疫情相关微博中的主题,探讨主题视角下的公众情感演化规律。

4.1 基于概率模型的主题抽取方法及结果

为实现有效的主题抽取,采用基于隐马尔科夫

模型^[24]的开源分词工具结巴分词^①,并采用哈尔滨工业大学设计的停用词表^②删去停用词,以避免其导致主题失去区分度。在完成文本预处理后,采用隐狄利克雷分布模型^[25]生成主题及相应的关键词表。通过对不同取值进行多次尝试并权衡结果中的无关主题数与重复主题数后,最后选定主题数为50,其中2个主题(18号、45号主题)较为杂乱无章,其他48个主题均较为明确且与疫情直接相关。

根据微博文本与主题关键词的匹配,量化计算其与每个主题的相关程度并归一化处理,将归属度超过0.1(即50个主题对应平均归属度0.02的5倍,以保证具有较显著的归属性)的主题判定为该条微博相关的主题。所得主题类型及对应的统计数据如表5所示。

从表5可以看出,类别的热度大体上与主题的泛化程度相关,即越泛化的主题,如官方公告或日常交流等,往往吸引更多的关注;相反,特定事件如火神山医院建设、红十字会捐赠风波等,虽然也能吸引

表5 事件主题类别、编号与对应的微博比重

Table 5 Microblog Proportion of Different Event Topics

主题类别	主题编号	微博数量	主题相关性
政府公告与疫情通报	6、10、12、38、40、48	87 433	0.148 3
疫情日常交流	1、4、9、29、31、33、34、47、50	36 383	0.171 3
热点事件讨论	2、11、15、35、39、43、49	4 547	0.164 3
疫情求助	7、17	3 743	0.120 8
物资捐赠与公益活动	5、8、26、36、41	2 053	0.111 9
医护人员、专家组报道	3、14、32、44、46	1 834	0.131 8
治疗方案及科研进展	13、20、23	1 109	0.134 6
政府行政管理报道	19、21、22、24、27、30、37	731	0.145 8
海外疫情	16、28、42	313	0.128 1
视频附属文字	25	46 074	0.128 6

关注,但明显存在量级上的差异。为便于讨论,本文将所有主题大致区分为事件性主题(对应热点事件讨论)与一般性主题(其他全部类别主题),两类主题的部分示例及其词云图如图5和图6所示。



图5 官方通告和微博求助相关的主题关键词

Fig.5 Keywords Related to Official Announcement and Seeking Help



图6 热点事件相关的主题关键词

Fig.6 Keywords Related to Hot Events

①<https://github.com/fxsjy/jieba>.

②<https://github.com/goto456/stopwords>.

4.2 主题时空分布与多维度关联

(1) 主题时空分布

从时间分布来看,大多数主题基本集中在1月下旬到2月中旬。这是因为2月中下旬后,随着湖北之外各地病例数普遍增长放缓,广大网民对疫情的关注度也逐渐降低。但是个别主题,如涉及海外疫情的主题16、讨论新冠疫情科研进展的主题23等,由于其自身内容的特性而持续较长时间。

受讨论范围与主体等因素的影响,不同主题的时间跨度、热度波动也有很大区别。总体而言,一般性主题相对时间跨度较长、波动平缓,而事件性主题则明显跨度较短且波动幅度较大。一个极端的例子是围绕大理扣留口罩事件的主题43,相关微博基本集中在事件曝光的2月6日当天,在短暂高峰后即迅速下降。显然,此类主题的时间跨度基本上与事件关注度密切相关,如果再无后续报道,则很难延续热度。同时,事件性主题的热度也受关联事件的影响。以围绕韩红基金会和红十字会的主题2与主题8为例,2月1日韩红基金会宣布停止接受捐款,同时武汉红十字会爆出公务车前往仓库提取口罩事件。此时,两个主题形成关联并各自引发了讨论。2月20日前后韩红基金会和被举报和调查的事件,导致红十字会相关的主题8产生了小规模的热度回流。可以认为,这一回流主要由于主题2的联动反应所导致。显然,对热点事件之间的相互作用也应给予关注。

就空间分布而言,大多数一般性主题并不具有显著的地区特性。当然,在一些与疫情就医、日常生活等相关的主题上,湖北本地的关注度呈现更高的态势。例如,医院物资求援相关的主题17、复工复学讨论的主题19等,湖北的关注度是其他省份的两倍左右。与此对应的是,北京及一些对外交流较为频繁的省份如黑龙江、浙江等,在海外疫情相关主题上的关注度要远高于其他省份。少数具有地域特色的主题也体现出明显的地区差异。例如,有关大理征用口罩事件的主题43引起了被征用口罩的重庆市的特别关注,其比重远高于其他省份;而另一当事方云南对此却关注寥寥,形成鲜明的对比。

总体而言,主题的空间分布差异化程度较低。除去部分主题对应的微博数过少影响判断之外,微博用户所填写的地理位置信息与其真实的地理位置

间可能存在的差异也会影响这一结果。

(2) 主题关联分析

热点事件主题之间的相互作用也应受到关注^[26]。由于新冠疫情涉及热点话题众多,同时各个话题在时间轴上又存在相当程度的相互覆盖。因此本文进一步从语义相关、共现关系两个维度揭示不同主题之间的潜在关联。就语义相关而言,各个主题之间的相关类型和相互作用机制存在一定差异,大致可以分为三类:

①“因果关联”型。以主题17与主题26为例,前者关于各大医院有关防护物资的求助,而后者关于医院接收爱心物资,两者的相关性高达0.415,同时P值小于0.01。这表明如果两个主题在语义上具有一定的因果关联,即一个主题的讨论会引发另一个主题的响应,那么两者在趋势上将具有高度的相关性。

②“并列关系”型。这一类情况较为复杂。以主题7和主题17为例,两者都可视为与求助相关的主题,但因求助的主体(医院或个人)需求差异较大,因此难以形成互动,导致两者几乎无法体现任何的相关性。形成对比的是主题2和主题8,网民在讨论韩红基金会时往往会以红十字会作为对比,反之亦然,导致两者虽然在微博数量上存在巨大差距,但在趋势上具有高度一致性,其相关系数高达0.707,同时P值小于0.01。这表明,对于“并列”的主题,应根据其内在逻辑上的关联进行判断,而不能简单根据其字面意义的相关性加以评判。

③“此消彼长”型。这一类主题难以通过相关性分析加以判别,但结合时间演化趋势可以发现其规律。以主题16和主题41为例,前者主要讨论国际疫情发展,后者则主要关注海外华人支持国内抗疫。从相关性分析上看,两者并不具有显著的相关性。但从时间来看,主题16直到2月下旬日韩疫情的爆发才开始逐渐受到关注,而主题41仅在1月底、2月初的时段内出现。显然,两者具有明显的时间差,但这又并非源于两者之间的因果关联,而是随着事态发展促使焦点逐渐从一方转向另一方所致。

本文采用共现关系^[27](即同一条微博中同时出现多个主题)构建主题之间的关联网络,进一步揭示主题之间的关联。从结果来看,一般性主题往往关

联广泛,但其关联对象较为混杂而且缺乏明确语义逻辑。相比之下,事件性主题的关联往往局限在语义相关的话题上。以野味相关的主题15为例,其与科研进展(尤其是揭示野生动物和新冠病毒关系的科研进展)相关的主题23关联最为密切。又如,有关李文亮医生的主题39,不仅与谣言和辟谣相关的主题37密切相关,同时还与表达对医务工作者敬意的主题46密切相关。由此可见,事件性主题的共现性更多表现在与事件本身存在逻辑关系的主题上,这将有助于加深对语义关联的探索。

4.3 主题视角下的公众情感倾向与演化

从整体上看,一般性主题的情感趋势波动相对较为平缓。以政府公告为主题10在疫情高峰期,即1月30日至2月10日期间所对应的“好”“怒”“哀”“惧”“恶”5种情感变化如图7所示,另外两种因所占比重较低不再赘述。总体而言,主题10对应的情感基本接近基准值,仅在部分节点如红十字会争议的2月2日和李文亮医生去世的2月7日,对应的“怒”和“哀”的情感各自有一个短暂的提升,但又很快恢复标准数值。由此可见,不具有明显倾向的一般性主题可采用基准情感加以估计。

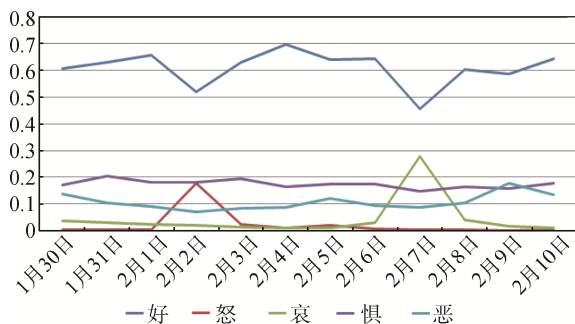


图7 主题10(政府公告)的情感倾向

Fig.7 Emotion Tendency for Topic 10 (Government Announcement)

而具有明显情感倾向的一般性主题,虽然波动也不明显,但其范围存在一定偏差。以病例通报相关的主题38为例,“惧”的情感明显偏高并随着病例增多持续上升,如图8(a)所示。而与为抗疫加油相关的主题50则相反,“好”的情感持续偏高,如图8(b)所示。显然,具有情感倾向的一般性主题,应在基准值上加入特定情感偏差,才能更为准确地

描述。

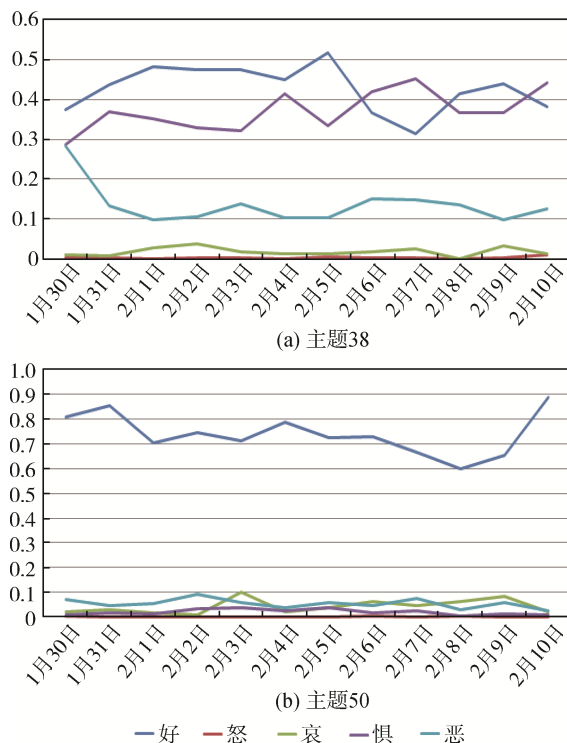


图8 主题38和主题50的情感倾向

Fig.8 Emotion Tendency for Topic 38 and Topic 50

相比之下,由于各种突发事件都会对情感产生较大影响,事件性主题的情感演进规律也更难描述。以红十字会相关的主题35为例,如图9(a)所示,相较于一般性主题,该主题的情绪波动明显更大,而且与突发事件有着明确的对应。例如,2月1日“好”的情感下降,对应公务车去红十字会仓库提取口罩的争议事件;而2月5日“好”的上升与“恶”的下降,则对应湖北红十字会副会长免职的处理。围绕李文亮医生的主题39,呈现更加剧烈的情绪波动,如图9(b)所示,“好”的情感一直处于较低的状态,而主体情感由“恶”转向“惧”,再转向对于李医生去世的“哀”,最终又回到“恶”。在这一时段中,各种新闻甚至谣言都对整体情感产生了影响。因此,热点事件驱动的主题其情感演进趋势会更难预测,不仅需要实时突发事件有准确的剖析,还需要对各种情感之间的转化作用,如由“哀”转“恶”或“怒”的趋势加以把握,才能实现准确描述。

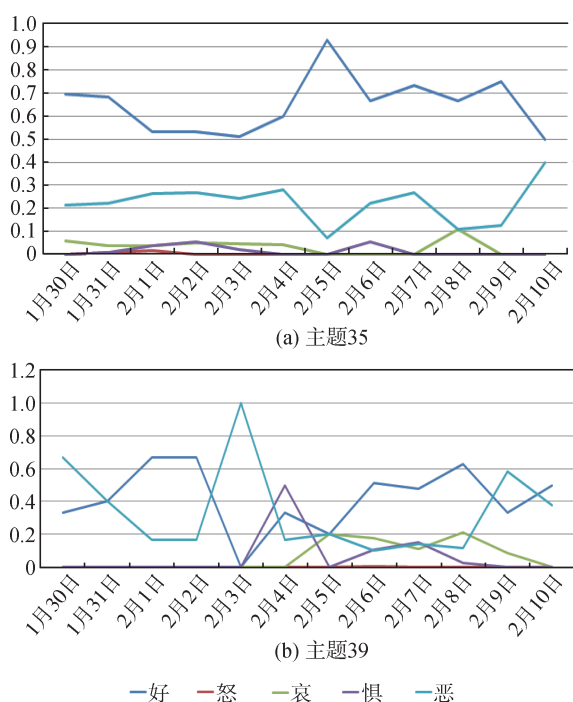


图9 主题35和主题39的情感倾向

Fig.9 Emotion Tendency for Topic 35 and Topic 39

5 传播视角下的地域情感扩散

相较于其他重大公共事件,疫情的特殊之处在于其跨地域扩散,疫情发生地距离用户所在地越近或越容易传播到所在地,该地疫情对用户情感的影响可能越强烈。基于这一假设,本节对公众情感地域传播进行建模分析,从而更为完整地描述疫情下的公众情感演进趋势,并借此实现对公众情感演进较为粗糙的预测。同时,通过分析不同地域间情感的相互影响强度,也可揭示各地疫情的内在联系提供新的线索。

本文采用并修正了社交传播建模中常用的独立级联模型^[28],以7种情感中可能最受疫情蔓延影响的“惧”为例,以省份为对象进行传播分析。具体而言,本文的传播模型基于以下两点合理假设:

(1)某地在某日的恐惧情感与本地疫情的增长幅度成正比。

(2)某地的恐惧情感受到外省疫情的影响,而这种影响既与外省前日的疫情严重程度(直接影响外省恐慌情绪和传播至本地的可能)相关,也与该省对本省的影响强度相关。

基于上述假设,本文将传播模型定义如公式(1)所示。

$$\hat{e}_i^t = (1 - \alpha) \cdot e_i^{t-1} \cdot \log \left(1 + \frac{c_i^t + 1}{c_i^{t-1} + 1} \right) + \alpha \cdot \left[1 - \prod_{j \neq i} (1 - w_{ji} \cdot e_j^{t-1}) \right] \quad (1)$$

其中, \hat{e}_i^t 为对于省份 i 在第 t 日的恐惧情感的估计值。该公式由两部分组成:前半部分对应假设(1),即恐惧情感既是前日恐惧情感 e_i^{t-1} 的延续,又与本地疫情的增长幅度成正比(c_i^t 为省份 i 在第 t 日的病例数。考虑到部分省份前期病例数为0,采用此种设计以避免分母为0的情况,并采用 \log 函数以平滑增长幅度较快造成的影响);后半部分对应假设(2),即恐惧情感同时还受到外省疫情的影响。考虑到当下交通的便利性,默认所有省份之间均可相互影响,并采用 w_{ji} 描述省份 j 对省份 i 的影响强度。最后,采用参数 α 对两部分进行加权求和。

在通过公式(1)对某省恐惧情感演化趋势进行估计时,有两项参数需要确定,即加权参数 α 和各省之间的影响强度 w_{ji} 。为此,引入损失函数,通过最小化估计误差来训练影响强度参数,并根据损失函数的最小值确定较为合适的加权参数,如式(2)所示。

$$\min_w \sum_{i,t} (e_i^t - \hat{e}_i^t)^2 \quad (2)$$

本文采用全国除港澳台地区外31个省级行政区2020年1月23日至2020年3月2日的病例数据,与相应时间内各省微博中“惧”情感所占比重进行联合分析。通过最小化估计误差,估计加权参数 α 约为0.6,可见本地疫情发展趋势对于用户恐惧情感的影响处于较为主要的地位。从结果来看,传播模型中两部分单独使用时的估计结果均方根误差高达约4.265,而基于 $\alpha=0.6$ 的加权结果可将均方根误差降至约2.002,证明了传播模型各个模块联合使用的有效性。同时,如果删去 \log 部分(即忽略本地疫情增长幅度影响),均方根误差升至约2.429, α 下降至约0.35,说明本地疫情增长幅度确实影响恐慌情绪产生,删去该部分削弱了模型的合理性。

本文还对不同省份之间的影响强度进行了分析。总体而言,相对影响强度较高的省份仍以邻近省份为主。以疫情中心湖北为例,其对外影响强度最高的5个省份分别为安徽、重庆、北京、上海和湖

南,除去作为全国政治、经济中心且人口流动频繁的北京、上海两地外,其他三个省份分别为湖北的东、西、南三个方向的相邻省份。另一方面,排名前10的江苏、山东、浙江等省份不仅与湖北间接相邻,同时属于疫情出现较早且受影响较重的省份(如浙江温州一度成为疫情重灾区),直接相邻的河南却由于其“硬核抗疫”而排名靠后。此外,较为偏远且受疫情影响较小的青海、宁夏、西藏等省份处于最后的位置。其他省份也具有类似规律,如安徽对外影响最强烈的为山东,重庆对外影响第二强烈的为四川(第一为北京)等。更进一步,考虑到安徽疫情前期以皖北为主,皖南相对较少,因此对于北方邻省山东的影响要超过南方的湖南、江西等邻省(排在第三、四位)也可得到合理的解释。相比之下,青海、西藏等边远省份对外输出的影响强度就较为杂乱无章,没有明显的空间规律。综上,公众情感的地域扩散以区域上的邻接为基础,同时兼顾疫情强度与交通便利性等因素。

6 结 语

本文以新浪微博舆情为研究对象,从全局视角、主题视角和地域扩散三个层面,分析重大突发公共卫生事件背景下公众在社交媒体中的情感表达,揭示了若干具有规律性的情感演进模式。

整体来看,疫情期间公众情感以积极情感为主,但受事件影响会出现低谷期,且多发生在疫情不明朗的早期爆发阶段,因此必须加强对早期阶段公众情感的调节引导。在消极情感中,总体存在“恶”的情感主导、“惧”的情感先发、“哀”的情感反复等特征,这提示面向传染性疫情一方面要重视“惧”的情感疏导,另一方面应格外关注对作为消极情感主体的“恶”情绪的疏导与防范,如重视对突发事件的追因、追责,加强突发事件应急管理的规范性和专业性等。另外,研究也发现公众情感随不同疫区距离、经济发展水平呈现一些明显差异,这意味着疫情期的情感疏导还必须考虑到地方实际,结合不同时期、不同地域采取有差别的舆情引导与情绪疏导策略。

从主题视角来看,研究证实不同主题的情感表达及演化趋势存在差异。其中,一般性主题和事件性主题在时间跨度长短、热度波动和地域分布等方

面存在规律性的差异,尤其是事件性主题因其内容的地方性导致的空间热度差异。同时,基于语义与共现的主题关联分析也发现了主题演化的协同性与相互作用,尤其是事件性主题基于内在逻辑关系导致的关联。

最后,以“惧”情感为例进行的传播建模分析,验证了各省份之间存在相互的情感影响,且影响强度决定于区位因素和疫情严重性。这再次证明重大突发疫情的防控与治理必须加强区域统筹与合作,各自为政甚至“邻避”思维已无法应对现代突发公共卫生事件的新挑战。

本文的局限和未来研究方向主要在以下方面:

(1)本文主要面向纯文本信息,未来将尝试对多模态信息如视频、图片等进行综合性分析。

(2)本文在相关性判断、情感判断等环节采用了简单的关键词匹配,未来将尝试引入自然语言处理工具进行更准确的判别。

(3)受数据制约,未能对4月以后的疫情发展,尤其是国外疫情发展所带来的舆情演化进行追踪分析,未来将围绕这一问题展开后续工作。

参考文献:

- [1] An L, Yu C M, Lin X, et al. Topical Evolution Patterns and Temporal Trends of Microblogs on Public Health Emergencies [J]. Online Information Review, 2018, 42(6): 821-846.
- [2] Palen L, Anderson K M. Crisis Informatics—New Data for Extraordinary Times[J]. Science, 2016, 353(6296): 224-225.
- [3] 卢嘉,刘新传,李伯亮. 社交媒体公共讨论中理智与情感的传播机制——基于新浪微博的实证研究[J]. 现代传播(中国传媒大学学报), 2017, 39(2): 73-79. (Lu Jia, Liu Xinchuan, Li Boliang. The Spreading Mechanism of Reason and Emotion in Public Discussions on Social Media—An Empirical Study Based on Sina Weibo[J]. Modern Communication(Journal of Communication University of China), 2017, 39(2): 73-79.)
- [4] Reuter C, Spielhofer T. Towards Social Resilience: A Quantitative and Qualitative Survey on Citizens' Perception of Social Media in Emergencies in Europe[J]. Technological Forecasting and Social Change, 2017, 121: 168-180.
- [5] 陈业华,张晓倩. 网络突发群体事件网民群体情绪传播模型及仿真研究[J]. 情报科学, 2018, 36(3): 151-156. (Chen Yehua, Zhang Xiaoqian. Research on Netizen Group Emotion Contagion Model and the Simulation Under Network Group Emergencies [J]. Information Science, 2018, 36(3): 151-156.)

- [6] 杨阳, 王杰. 情绪因素影响下的突发事件网络舆情演化研究[J]. 情报科学, 2020, 38(3): 35-41.(Yang Yang, Wang Jie. The Evolution of Emergency Network Public Opinion Influenced by Emotional Factors[J]. Information Science, 2020, 38(3): 35-41.)
- [7] 鲁艳霞, 吴迪, 黄川林. 大数据环境下面向突发公共事件的微博用户情绪分析[J]. 软件工程, 2017, 20(1): 45-48.(Lu Yanxia, Wu Di, Huang Chuanlin. An Analysis of Micro-Blog Users' Emotions in Public Emergencies Based on Big Data[J]. Software Engineering, 2017, 20(1): 45-48.)
- [8] 张雪, 陈安. 基于网络新闻评论的群体愤怒情绪监测与度量[J]. 科技促进发展, 2010(9): 44-49.(Zhang Xue, Chen An. The Monitoring and Measurement of Mood of Anger Based on Network Review of News[J]. Science & Technology for Development, 2010(9): 44-49.)
- [9] Fan R, Zhao J C, Chen Y, et al. Anger is More Influential than Joy: Sentiment Correlation in Weibo[J]. PLoS One, 2014, 9(10): e110184.
- [10] 李聪. 问题疫苗事件微博传播中的情绪与表达[D]. 武汉: 武汉大学, 2019.(Li Cong. Emotions and Expressions in Weibo Communication of Vaccine Events[D]. Wuhan: Wuhan University, 2019.)
- [11] 陈曦, 费奇, 李炜. 基于计算实验的公众恐慌研究初探[J]. 华中科技大学学报(社会科学版), 2009, 23(2): 34-37.(Chen Xi, Fei Qi, Li Wei. A Preliminary Research on Urban Mass Panic Based on Computational Methods for Experiment[J]. Journal of Huazhong University of Science and Technology(Social Science Edition), 2009, 23(2): 34-37.)
- [12] 王春雪, 吕淑然, 杨凯. 突发事件中恐惧情绪感染概率研究[J]. 中国安全科学学报, 2015, 25(9): 14-19.(Wang Chunxue, Lyu Shuran, Yang Kai. Research on Probability of Fear Contagion in Emergency[J]. China Safety Science Journal, 2015, 25(9): 14-19.)
- [13] 安璐, 吴林. 融合主题与情感特征的突发事件微博舆情演化分析[J]. 图书情报工作, 2017, 61(15): 120-129.(An Lu, Wu Lin. An Integrated Analysis of Topical and Emotional Evolution of Microblog Public Opinions on Public Emergencies[J]. Library and Information Service, 2017, 61(15): 120-129.)
- [14] 叶勇豪, 许燕, 朱一杰, 等. 网民对“人祸”事件的道德情绪特点——基于微博大数据研究[J]. 心理学报, 2016, 48(3): 290-304.(Ye Yonghao, Xu Yan, Zhu Yijie, et al. The Characteristics of Moral Emotions of Chinese Netizens Towards an Anthropogenic Hazard: A Sentiment Analysis on Weibo[J]. Acta Psychologica Sinica, 2016, 48(3): 290-304.)
- [15] Trope Y, Liberman N. Construal-Level Theory of Psychological Distance[J]. Psychological Review, 2010, 117(2): 440-463.
- [16] 钱铭怡, 叶冬梅, 董葳, 等. 不同时期北京人对 SARS 的应对行为、认知评价和情绪状态的变化[J]. 中国心理卫生杂志, 2003, 17(8): 515-520.(Qian Mingyi, Ye Dongmei, Dong Wei, et al. Behaviour, Cognition and Emotion of the Public in Beijing Towards SARS[J]. Chinese Mental Health Journal, 2003, 17(8): 515-520.)
- [17] van Lent L G, Sungur H, Kunneman F A, et al. Too Far to Care? Measuring Public Attention and Fear for Ebola Using Twitter[J]. Journal of Medical Internet Research, 2017, 19(6): e193.
- [18] 张放, 甘浩辰. 疫情心理时空距离对公众情绪的影响研究——基于新冠肺炎疫期微博文本面板数据的计算分析[J]. 新闻界, 2020(6): 39-49.(Zhang Fang, Gan Haochen. On the Influence of Temporal and Spatial Distance from Epidemic on Public Sentiment: A Computational Analysis Based on Panel Data of Weibo Text About COVID-19[J]. Journalism and Mass Communication Monthly, 2020(6): 39-49.)
- [19] 唐超. 网络情绪演进的实证研究[J]. 情报杂志, 2012, 31(10): 48-52.(Tang Chao. Empirical Research on the Evolution of Online Emotion[J]. Journal of Intelligence, 2012, 31(10): 48-52.)
- [20] 单斌, 李芳. 基于 LDA 话题演化研究方法综述[J]. 中文信息学报, 2010, 24(6): 43-49.(Shan Bin, Li Fang. A Survey of Topic Evolution Based on LDA[J]. Journal of Chinese Information Processing, 2010, 24(6): 43-49.)
- [21] 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[J]. 图书情报工作, 2016, 60(20): 104-111.(Zhao Xiaohang. The Study on Government News Release in the Era of Post-Microblog Based on Sentiment Analysis and Subject Analysis: A Case Study of the “Tianjin Explosion” on Sina Microblog[J]. Library and Information Service, 2016, 60(20): 104-111.)
- [22] 任中杰, 张鹏, 李思成, 等. 基于微博数据挖掘的突发事件情感态势演化分析——以天津 8·12 事故为例[J]. 情报杂志, 2019, 38(2): 140-148.(Ren Zhongjie, Zhang Peng, Li Sicheng, et al. Analysis of Emotion Evolution of Emergencies Based on Weibo Data Mining: Taking “8·12 Accident in Tianjin” as an Example[J]. Journal of Intelligence, 2019, 38(2): 140-148.)
- [23] 陈建美. 中文情感词汇本体的构建及其应用[D]. 大连: 大连理工大学, 2009.(Chen Jianmei. The Construction and Application of Chinese Emotion Word Ontology[D]. Dalian: Dalian University of Technology, 2009.)
- [24] 刘云中, 林亚平, 陈治平. 基于隐马尔可夫模型的文本信息抽取[J]. 系统仿真学报, 2004, 16(3): 507-510.(Liu Yunzhong, Lin Yaping, Chen Zhiping. Text Information Extraction Based on Hidden Markov Model[J]. Acta Simulata Systematica Sinica, 2004, 16(3): 507-510.)
- [25] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [26] 李纲, 陈思菁, 毛进, 等. 自然灾害事件微博热点话题的时空对比分析[J]. 数据分析与知识发现, 2019, 3(11): 1-15.(Li Gang, Chen Sijing, Mao Jin, et al. Spatio-Temporal Comparison of Microblog Trending Topics on Natural Disasters[J]. Data Analysis and Knowledge Discovery, 2019, 3(11): 1-15.)

- [27] Xu T, Zhu H S, Zhao X Y, et al. Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1285-1294.
- [28] Kempe D, Kleinberg J, Tardos É. Maximizing the Spread of Influence Through a Social Network[C]//Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2003: 137-146.

作者贡献声明:

边晓慧:提出研究问题,总体研究设计,论文撰写、修改及最终版本修订;

徐童:参与研究设计,采集和处理数据,撰写、修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: tongxu@ustc.edu.cn。

[1] 徐童. WeiboData.rar. 包含 20 131 个用户微博博文数据文件。

[2] 徐童. WeiboProfile.rar. 包含 4 个用户微博档案数据文件。

收稿日期:2021-07-16

收修改稿日期:2022-02-26

Evolution of Public Sentiments During COVID-19 Pandemic

Bian Xiaohui¹ Xu Tong²

¹(School of Management, Anhui University, Hefei 230039, China)

²(School of Computer Science, University of Science and Technology of China, Hefei 230027, China)

Abstract: [Objective] This study analyzes the social media posts during the COVID-19 pandemic, aiming to reveal the temporal and spatial differences of public opinion, the sentiment evolution under different circumstances, as well as the trans-regional spreading of the public sentiments. [Methods] Firstly, we utilized the Latent Dirichlet Allocation (LDA) model to generate the latent topics and related keyword groups, which also analyzed public sentiment evolutions from the perspectives of global and individual topics. Then, we described the trans-regional spread of public sentiments based on the social spread model adapted from the classic Independent Cascade Model. [Results] The new model summarized the general rules of the temporal evolution and spatial difference, as well as the impacts of distance to the epidemic centers and the financial levels. We also found two different types of topics indicating reasons for popularity and sentiment differences, as well as multi-view connections among these topics. The strength of trans-regional sentiment spread could be affected by both regional distance and epidemic situation. [Limitations] The new framework could not process the multimodal data. [Conclusions] The proposed model helps the local government make better strategies according to specific conditions, and pay more attention to the impacts of related events. They should also strengthen regional cooperation and coordination for controlling pandemics and monitoring public sentiments.

Keywords: Public Health Emergency Weibo Public Sentiment Sentiment Evolution Topic Analysis Spatio-Temporal Analysis