

文章编号: 2096-1472(2017)-01-45-04

大数据环境下面向突发公共事件的微博用户情绪分析

鲁艳霞, 吴迪, 黄川林

(大连东软信息学院信息管理系统, 辽宁 大连 116023)

摘要:近年来突发公共事件频出,随着互联网的普及和大数据等信息技术的迅猛发展, Twitter、博客、微博等使得公众在突发公共事件发生后表达个体情绪更加便捷。本文以“天津港爆炸事件”为研究对象,首先利用爬虫工具收集微博内容,然后通过ROST CM内容挖掘软件进行中文词频分析,最后通过SPSS对微博情感进行分析统计。研究发现,公众情绪容易受到集群效应的影响,网民群体情绪的不稳定性会导致其行动的不确定性,政府或意见领袖的积极引导将会促进突发事件的良性发展。

关键词:大数据;突发公共事件;情绪;舆情引导

中图分类号: TP312.63 **文献标识码:** A

An Analysis of Micro-Blog Users' Emotions in Public Emergencies Based on Big Data

LU Yanxia, WU Di, HUANG Chuanlin

(Dalian Neusoft University of Information, Department of Information Management, Dalian 116023, China)

Abstract: With the development of internet and big data, public emergencies frequently occurred in recent years. The public are more likely to express individual emotional states via Twitter, blog, micro-blog, etc. This paper focuses on the micro-blog data in the Tianjin Explosion event. Firstly, micro-blog contents are collected through crawler tools. And then, the Chinese word frequencies in the micro-blog contents are analyzed through ROST CM. Finally, the analysis and statistics of the micro-blog users' emotions are conducted through SPSS. The research results show that public emotions are easily influenced by the clustering effect, the instability of cyber citizens' emotions can lead to the uncertainty in actions, and the active guidance from governments and opinion leaders will promote the healthy development of public emergencies.

Keywords: big data; public emergencies; emotion; public opinion guidance

1 引言(Introduction)

随着移动互联网技术的不断发展与成熟,再加上无线网络的覆盖明显提升,给人们的生活带来了便捷。与此同时,我国网络新闻用户规模为5.64亿,较去年增长8.8%,表明越来越多的公众对各类新闻事件给予更多的关注。据互联网发展状况统计报告显示,社交领域的微博使用率达到33.5%,越来越多的网民喜欢通过微博刷“存在感”。网络的普及性、移动性、草根性与匿名性等使得越来越多的民众喜欢通过网络表达对突发公共事件的观点和看法。微博用户对突发公共事件的关注相应会产生大量的、非结构化的微博内容,微博内容中涵盖了民众的情绪与情感,及时挖掘这些特征将有利于政府相关部门对舆情的监控以及谣言的控制,从而有效提高政府的公信力和应急管理能力。本文以“天津港爆炸事件”为研究对象,进行大数据环境下突发公共事件的微博用户情绪研究,以期使政府相关部门了解公众在事件发生过程中的情感传播状况,为避免公众情感的集聚和极化,提供有针对

性的信息,从而为政府引导策略的制定提供一定的理论和实践指引。

2 理论基础与研究现状(Theoretical basis and research review)

大数据环境下的微博用户之间存在着大量的信息交互,往往在对突发公共事件的反应中,自觉不自觉地就反映出群际情绪。群际情绪理论^[1]由美国心理学家Mackie提出,认为群际情绪是个体认同某一社会群体时的情绪体验,当个人认同某一群体时,对相关事物的评价将会带有情绪色彩。

同时,突发公共事件具有突发性、持续性、不确定性等特征,由于事件相关信息在短时间内并不一定能够清晰显现,因此部分微博内容存在着一定的片面性,不明真相或认知能力较差的网民极易受到情绪感染以至于产生不良的负面影响。情绪感染理论由McDougall提出,Hatfield^[2]给出了较为全面的定义,认为个人情绪感染到其他个体时,情绪会反过来影响他人,从而强化个人原本的情绪状态,最终导致某

种情绪在群体间同质化。

因此，及时、有效地对微博用户产生的微博内容进行分析，将会有效提升政府对突发公共事件的舆情监控。基于此，越来越多的学者近年来对突发公共事件的情绪与情感传播进行了相关研究。

刘志明等^[3,4]提出突发事件引发的群体情绪会引起次生群体事件的发生，研究了突发事件发生、发展过程中群体情绪尤其是负面情绪的演化规律并建立了相应的模型，对次生突发事件起到了较好的预警效果，但是文章并未针对大数据环境对微博情绪进行分析；李从东等^[5]从系统学和耗散结构理论角度对社会情绪进行了分析，结合系统动力学和元胞自动机模型模拟社会情绪，并且对社会情绪的稳定性进行了建模，从而对社会情绪的评估与预测提供了一定的理论和模型基础；赵卫东等^[6]从信息传播与群体行为角度研究了网民的情绪传播机制并构建了情绪传播模型，对于网民群体情绪演化规律进行了仿真分析，但是文献[5]和[6]并未从实际应用角度分析政府对微博情绪尤其是负面情绪的引导作用。

3 基于ROST CM的微博用户内容情绪分析(An analysis of micro-blog users' emotions based on ROST CM)

微博用户群体涵盖不同年龄、性别、教育背景等基本特征，具有一定的广泛性，因此微博用户内容也反应了对同一突发公共事件却可能会有不同的情绪与情感。羊群效应^[7]在网络舆情传播过程中是一个普遍存在的现象，使得个人观念或行为在特定的群体影响下与多数人相一致，如果不对网民群体内部的羊群效应加以引导，不仅会影响网络舆情的传播过程，还会影响舆情的最终演化方向，甚至可能对社会稳定造成威胁。

为了更好的分析突发公共事件发生后的微博用户情绪变化规律，本研究选取发生于新浪微博的“天津港爆炸事件”，分析单位是事件发生期间微博内容。该事件经国务院调查组认定是一起特别重大生产安全责任事故，属突发公共事件中的社会安全事件。该事件背景如下：

2015年8月12日23:30左右，位于天津滨海新区塘沽开发区的天津东疆保税港区瑞海国际物流有限公司危险品仓库发生爆炸。截至2015年9月11日，共发现遇难者总人数165人，8人失联，其中公安消防人员24人、民警11人、天津港消防人员75人、其他人员55人。鹰眼与情网显示该事件最先由微博曝光，8月12日晚微博用户陆续发布爆炸相关信息或图片引发网民关注，此后经由微博中明星媒体发表评论或者转发相关微博，随后相关事件的信息量呈爆炸式增长。

本文使用武汉大学沈阳教授研发编码的ROST CM文本挖掘软件对微博数据进行预处理以及后续分析，该软件可以实现对收集到的微博内容进行分词、词频统计、情感分析等操作。论文首先收集研究所需要的微博大数据，在对微博用户内容大数据进行预处理的基础上，将其导入ROST情感分析工具，从而对微博内容情感倾向性进行判断。该软件将微博内容的情感分为三类：积极、中立和消极，其中积极情绪和消极情绪还可以进一步的细分，其定义的情感值区间分别为一般积极(0, 10]、中度积极(10, 20]、高度积极(20, +∞)，一般消极[-10, 0)、中度消极[-20, -10]和高度消极(-∞, -20)。在情感倾向性判断的基础上，可以对不同类别的情感走势进行模拟和预测。

4 实证研究(Empirical research)

4.1 数据获取

本文选取新浪微博发表的博文为数据资源，以“天津爆炸”和“天津塘沽大爆炸”为关键词，利用GooSeeker数据采集工具抓取了2015年8月12日至2015年12月25日共135天的微博内容。GooSeeker是一款专业的网络爬虫软件，有垂直搜索、信息汇聚、个人信息检索等主要功能。根据研究需要，本文主要抓取了“用户名”“微博内容”“发布时间”“发布设备”“转发数”“评论数”“赞数”相关数据，部分数据如表1所示。

表1 天津爆炸微博数据一览表

Tab.1 The micro-blog data list of the Tianjin

Explosion event

用户名	微博内容	发布时间	转发数
新浪河南	今日凌晨一个女孩称自己父亲在爆炸中身亡，骗得3700多名网友现金打赏，天津爆炸事故罹难者家属并收取“捐款”。鉴于其真实性受到其他网友广泛质疑，故已将该账号关闭，并冻结了其收取的款项	2015-8-14 14:46:00	421
GE中国	通用电气(GE)公司宣布向天津港爆炸事故的救援工作捐赠价值300万元的医疗设备。设备将统一送往相关应急救助的医院和机构，为保障救援医院医疗设备正常运行，GE医疗工程师将协助医院安装设备，并随时待命，守护在生命前线	2015-8-14 16:50:00	55
火线英雄 真人秀	中国日报记者从天津消防武警总队得到消息，已有11名消防人员牺牲。致哀！为这些逆行而行的和平战士祈福、敬礼！希望失联的战士们平安归来	2015-8-13 10:49	55
婉婉弯弯 爱玩玩	一旦发生重大事件，广大人民群众能参与出力的就是献血、捐款和祈祷！！无论是实际行动支持，还是精神上的鼓励祈福，都是对伤者的帮助和祝福	2015-8-13 15:17	

4.2 数据预处理

由于同一个人发表的相同的博文信息属于重复数据，

会对情感分析结果的有效性产生影响，所以本文利用Excel工具，对微博内容进行去重处理，从而得到本实验的研究数据，涵盖微博内容共计38934条。

利用ROST CM挖掘软件对微博采集信息进行中文分词，接着利用ROST WordParser软件对分词后的数据进行中文词频分析，统计得出了网民反复讨论出现的热词，结果如表2所示。

表2 天津爆炸事件微博用户讨论热词及词频统计

Tab.2 The statistics of the hot words and word frequencies in the Tianjin Explosion event

热词	词频	热词	词频	热词	词频	热词	词频	热词	词频
天津	81074	你们	6045	生命	3990	现在	3463	政府	2837
爆炸	49223	希望	5835	祈福	3788	新闻	3339	灾难	2785
塘沽	31294	知道	4705	不要	3726	消防员	3153	英雄	2762
我们	12833	发生	4636	事件	3582	牺牲	3117	现场	2232
事故	6561	消防	4079	自己	3547	平安	2988	战士	2211

由表2可看出，在这种网民大规模参与的集群行为中，热词被多次反复提及和讨论。网民大多表达了对天津爆炸事件的震惊、对救灾人员的尊敬与祝福以及对相关单位的谴责。不同的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，讨论的过程中容易形成多个热点话题，而参与讨论某个话题的人越多，说明该话题的热度越大，微博舆论也就此形成。

4.3 数据分析

将38934条微博内容作为样本，可将“天津爆炸”事件粗略划分为“爆发期”“衰退期”“消亡期”三个阶段，如图1所示。

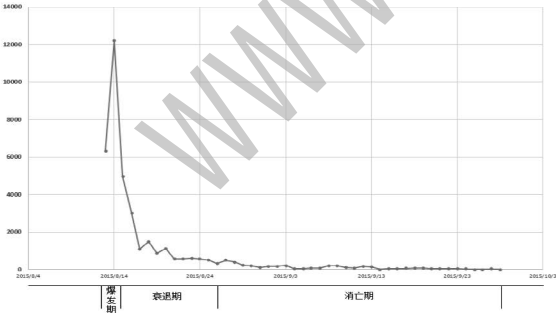


图1 “天津爆炸”事件演化规律

Fig.1 The evolution rule of the Tianjin explosion event

图1可看出，天津爆炸事件在8月12日发生后，微博相关评论迅速升温，在8月13日达到极值，之后关注量持续减少，最后形成长尾效应。

进一步统计可得到网民情感的整体分析结果，如表3所示。

表3 网民情感整体分析结果

Tab.3 The overall analysis of the cyber citizens' emotions

汇总情况	名称	数量	比例
总体情况	积极情绪	15590条	40.04%
	中性情绪	922条	2.37%
	消极情绪	22422条	57.56%
积极情绪分段统计结果	一般(0—10)	8924条	22.92%
	中度(10—20)	4102条	10.54%
	高度(20以上)	2564条	6.59%
消极情绪分段统计结果	一般(-10—0)	11485条	29.50%
	中度(-20—-10)	5667条	14.56%
	高度(-20以下)	2555条	6.56%

由表3可知，在抓取到的博文数据中，消极情绪占了最大比例(57.56%)，说明对于天津爆炸事件网民的整体情绪趋向于消极。

为了更好地看出人们的情感走势，本文对每天的情感值求平均值，得到了从2015年8月12日至2015年12月26日的每天不同类别情感值以及整体情感值的平均值，利用SPSS软件得出所有博文内容的情感值、积极与消极情绪情感值随时间序列走势如图2—图4所示。

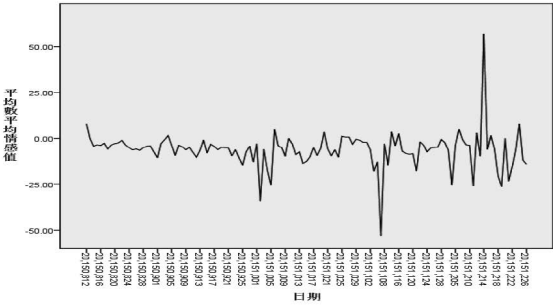


图2 网民情感整体时间序列走势

Fig.2 The time sequence trend of the cyber citizens' emotions

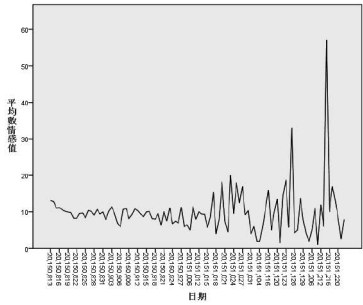


图3 积极情绪情感值时间序列走势

Fig.3 The time sequence trend of positive emotions

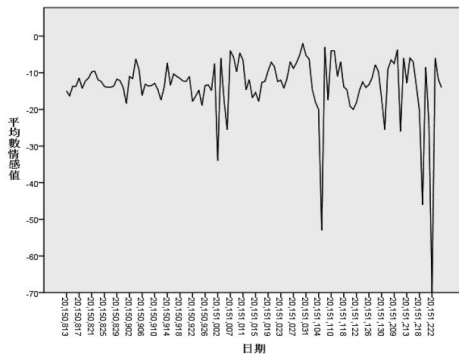


图4 消极情绪情感值时间序列走势

Fig.4 The time sequence trend of negative emotions

为了进一步验证该组数据的可靠性,导入SPSS 22.0中进行分析,结果表明:数据分布范围为 $[-127, 124]$,在理论值范围之内;均值为 -3.58 ,中位数为 -2.00 ,方差为 317.286 ,说明数据的离散程度较高;偏斜度 <0 ,峰度 >0 ,根据统计学中的判断标准,偏斜度 $=0$,峰度 $=0$ 时,分布呈正态,偏斜度 >0 时,分布呈正偏态,偏斜度 <0 时,分布呈负偏态,峰度 >0 时,曲线比较陡峭,峰度 <0 时,曲线比较平坦,说明本数据分布为负偏态(朝右偏),较陡峭。而当偏态在样本例数够大时可看作近似正态分布,因此本数据也可看作近似正态分布,符合自然界中正常的数据分布形态,表明本数据具有较高的可靠性。

5 结论(Conclusion)

通过对不同情感的趋势模拟,可回溯到相应的博文内容,从而发现引起情感波动的因素。

8月13日,网友@妖妖小精在13日上午创作了一幅漫画名为“世界上最帅的逆行”,感动无数网友;8月16日,李克强代表党中央与国务院,赶赴天津“8·12”爆炸事故现场,看望慰问救援官兵、消防队员和伤员及受灾群众,部署下一步救援救治、善后处置和安全生产工作;9月4日,天津港爆炸核心区积水废土已被外运;10月14日,天津滨海新区爆炸最小遇难者海葬;9月16日,检察机关公布了对事故中12名领导干部的刑事拘留措施和有关部门的责任认定初步结论,明确了相关部门的失职渎职行为。

通过以上过程分析可知,当政府出台新的有利于灾区救援工作的政策,媒体报道了关于网友为消防人员而作的画,或者政府领导亲自到灾区探望受灾群众时,网民的情感走势

出现波峰,说明政府工作良好有序地开展对舆情有积极的影响。当检察机关公布事故的责任认定,其他地区类似爆炸案的发生、当一些特殊的纪念日到来或者特殊事件发生时,网民情感走势会出现波谷,说明媒体的报道对舆情的影响至关重要。

同时,我们发现网民群体由于其情绪的高度不稳定性必然导致其行动方向的极度不确定性,独立思考能力和意志力薄弱的网民更容易受到集群效应的影响。虽然整个微博讨论过程中积极和消极情绪交替上涨,但安全事故类突发事件中消极情绪始终占据主流。

因此,为了避免集群行为在极端情绪的感染下走向“极化”,政府部门应尊重网络民意,及时把握网络舆论关注的社会问题;正面处理网络舆情事件,把握网络集群演变趋势;同时,应该加强对微博网络的监督,完善对造谣者、滋事者言论的监测和清理机制,充分利用微博平台建立政务信息平台。

参考文献(References)

- [1] 刘峰,佐斌.群际情绪理论及其研究[J].心理科学进展,2010,18(6):940-947.
- [2] Hatfield,Cacioppo,Rapson.Emotional Contagion[J].Current Directions in Psychological Science,1993(2):96-99.
- [3] 刘志明,刘鲁.面向突发事件的群体情绪监控预警[J].系统工程,2010,28(07):66-73.
- [4] 刘志明,刘鲁.面向突发事件的民众负面情绪生命周期模型[J].管理工程学报,2013,27(01):15-21.
- [5] 李从东,洪宇翔.面向突发事件的社会情绪稳定性建模方法研究[J].情报杂志,2014,33(01):146-151.
- [6] 赵卫东,赵旭东.突发事件的网络情绪传播机制及仿真研究[J].系统工程理论与实践,2015,35(10):2573-2581.
- [7] 陈福集,黄江玲.基于演化博弈的网络舆情传播的羊群效应研究[J].情报杂志,2013(10):1-5.

作者简介:

鲁艳霞(1977-),女,硕士,教授.研究领域:信息系统。

吴迪(1975-),男,硕士,副教授.研究领域:ERP,数据分析。

黄川林(1979-),女,硕士,副教授.研究领域:SAP开发技术。