

突发事件的网络舆情监测与恐慌度量系统

李季梅^{1,2} 陈宁³ 陈安^{1,4} 武艳南^{2,5}

- (1. 中国科学院科技政策与管理科学研究所,北京 100190; 2. 中国科学院研究生院,北京 100039;
3. 葡萄牙波尔图理工学院,波尔图 4200-072; 4. 河南大学应急管理研究所,河南开封 475004;
5. 山东大学管理学院,山东济南 250100)

摘要:突发公共事件不仅对人的生命财产造成损害,而且引起人们不同程度的心理变化,由此引发公众恐慌,并激发衍生突发事件,从而导致造成更大的灾难。互联网技术的发展使信息传播更加快捷,随之带来的负面影响也更加严重。本文在分析公众恐慌的起因及特点的基础上,通过网络舆情监测实现公众恐慌的度量,为应急管理提供有效的决策支持。

关键词:突发事件;应急管理;舆情监测;恐慌;度量;数据挖掘

中图分类号: TP311 文献标识码: A DOI: 10.3772/j.issn.1674-1544.2009.02.012

1 引言

近年来,突发灾难性事件频繁发生,从中国的 SARS、雪灾、汶川地震到美国的“9·11”事件、卡特里娜飓风、印度洋的海啸,对人类生命和财产造成了重大损失。由于通信和信息技术的发展,这些事件不仅影响到一个区域人们的生活,而且影响到跨区域、跨国界的人群,而影响的范畴也从物质形态发展到精神层面^[1]。随着社会系统的复杂化和信息技术的发展,重大公共安全隐患也越发变得隐蔽,一旦遇到条件合适就会爆发,造成灾难性的后果,同时也给公众心理造成伤害,使人群处于恐慌之中,进而影响着人群的行为决策,产生群体反应过度,甚至造成突发事件的发展扩大,使其影响范围和程度进一步扩大和加深,激发生突发事件的产生,造成更大的灾难。信息时代传播技术的发展也使信息的传播突破了“耳语相传”的局限,关于突发事件的各种消息通过手机、互联网等新兴途径快速传播,让人陷于“信或不信”的困惑。

群体性的恐慌常常会对政治的稳定、经济的

发展及社会的和谐造成很大的危害,甚至导致新的突发事件产生。例如,2005年吉林石化公司爆炸事故衍生出的哈尔滨市水危机事件。因此,建立政府危机管理的预警机制,增强对公共危机的可预见性,以及对公众心理的监控已经成为应急管理中的重要问题。本文首先分析了公众恐慌的起因及特点,然后介绍一个利用网络舆情分析实现恐慌度量与监测的系统,并通过实例说明该系统对应急管理的决策支持作用。

2 突发事件的公众心理问题研究

突发公共事件发生后,当人们感到原有处理方式难以应对时,就会产生心理危机,导致人的情感、认知和行为方面的功能失调,导致群体性的恐慌心理及行为,通常表现为焦虑、恐惧、求证心切、烦躁心强、盲目和从众行为等。灾难的不确定性、突然性和巨大的破坏性是恐慌产生的主要原因,对灾害的无知或者知之不多是恐慌发生的根本原因。而灾难的发生导致了生活的无序状态,原有的适应模式被打破,心理失去平衡,是恐

第一作者简介:李季梅(1986-),女,硕士研究生,主要研究方向是应急管理理论与方法。

基金项目:中国科学院科技政策与管理科学研究所所长基金“城市公众恐慌的特征机理及基于 Web 挖掘技术的度量模型与支持系统研究”(0700811J01),北京劳动保护科学研究所院地合作项目“城市脆弱性与基于保障率的机构弹性评估方法研究”(0700421901)。

收稿日期:2009年3月11日。

慌产生的直接原因。

对于突发事件造成的公众心理问题的研究可以追溯到英国心理学家 Clark 的“灾难心理认知偏差模型”^[2]。心理恐慌认知引发了对心理紧张刺激的正反馈，导致了人类对感知到的可能对自己造成伤害的信息产生误解。之后，很多学者对此模型做了改进和应用。例如，对 SARS 危机所造成的公众心理状态和心理行为的探讨^[3]，对不同人群、阶段心理危机的特点及危机干预和心理辅导方式的研究^[4]，对公众风险认知和大众媒体、政府关系的分析等^[5]。

现代社会是一个群体社会，信息技术快速发展，人口密度不断提高，信息网络成为群体沟通的重要手段。随着互联网的普及，网络的发展为民众内隐的情绪和态度提供了一个理想的表达途径，形成了多样化的网络舆情信息，为理解、监测和引导公众观点提供了一种新的途径。计算机网络的发展对群体极化效应有很强的促进作用，而发生在网络上的恐慌传播速度也会因为群体极化而加快。曾祥平等利用元胞自动机方法建立了网络舆论发展的激励模型^[6]，郑军研究了网络舆情监控的热点发现算法，并设计了 Web 舆情监控与上报系统^[7]，张珏等以中华网社区的 actual 数据为样本，利用神经网络预测 BBS 上帖子的回复数^[8]，余红构建了网络论坛舆论领袖筛选模型，以强国社区中日论坛为案例进行研究^[9]，以及利用网页挖掘技术鉴别极端分子活动的技术研究^[10]。

3 基于网络舆情的恐慌度量与监测系统

由于在突发事件的发生过程中，即使一些人有了恐慌的心理，但是由于多种原因，并不一定愿意在日常生活中直接表达出来，因此传统的问卷调查出现了困难，并且对于突发事件这样一个时效性很强的调查要求，传统的问卷调查设计、发放、回收、整理、分析过程将消耗很多时间，也就失去了在恐慌发展蔓延过程中及时跟踪调查的能力。因此，我们选择网络调查方法来实现对城市公众恐慌的实时追踪。当突发事件发生或关于恐慌的流言出现时，全国性或者地域性的论坛往往会

在较早的时间里出现这个话题的讨论。而话题的受关注程度以及讨论者的参与程度就反映了该突发事件的影响程度，也体现了城市公众恐慌的情况。联机分析挖掘方法可以帮助实现对城市公众恐慌现象的在线分析，利用前端的 Web 集成挖掘工具搜集网络信息平台上与恐慌现象相关的信息，如论坛中散播的大量还未经证实的消息等，经过汇总整理后存入数据库，通过挖掘算法进行分析处理，得到一系列反映公众心理的信息，将其用于恐慌传播状态和影响力的度量模型，从而实现对恐慌传播的监控^[11]。

本系统通过对互联网上论坛的发帖、回帖等相关信息进行自动抓取、主题检索、热点分析及跟踪，实现对公众心理恐慌的度量和监测。图 1 是公众恐慌度量与监测系统的处理流程。

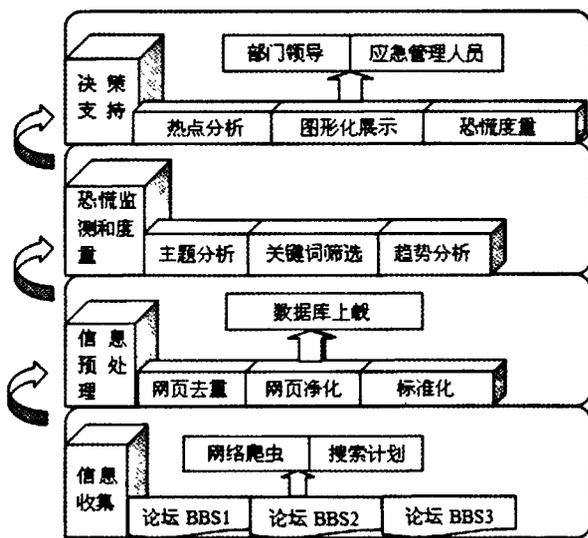


图 1 公众恐慌度量与监测的处理流程

该系统设计的目标是对论坛信息进行有效及时的分析与监控，按照不同的要求筛选信息，进行多样化的图形展示和模型度量，为各级领导和应急管理人员提供决策支持，以便在适当的时机采取适当的措施进行引导，避免出现大规模的恐慌事件，减少社会危害。该系统具备以下功能。

(1) 根据设定的关键词对帖子进行筛选。根据关注点的不同，当某一事件发生后，按照相关的信息设定查询关键词，筛选出特定时间和主题的帖子，观察其数量及比例。

(2) 友好的界面展示和多方式输出。分析主题

帖及回帖的标量和增量等信息的变化趋势,进而分析公众的恐慌程度,以多种方式输出。

(3) 发帖群体的信息分析。通过对发帖 ID 的分析,识别相关人员和帖子的有效性。

(4) 公众恐慌度量。结合模型和系统分析得出的数据,计算恐慌值。

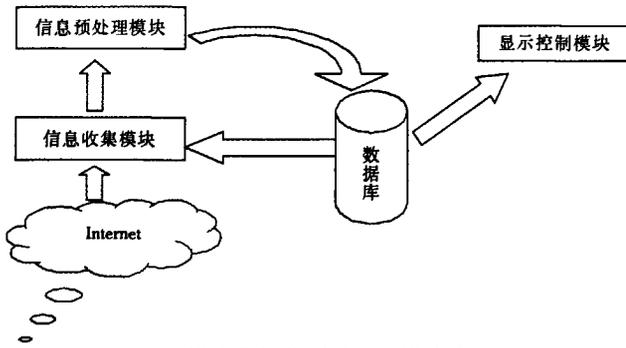


图 2 恐慌度量与监测系统的模块功能图

如图 2 所示,恐慌度量与监测系统主要包括信息收集、信息预处理、数据库和显示控制等模块。

信息收集模块:获取选定网页源码,利用 MetaProducts Corporation 公司研发的 Offline Explorer 将整个 Web 站点下载供离线浏览。

信息预处理模块:对网络采集到的数据进行整理分类,排除不在选定页面的网页,处理非规定格式的文档。

数据库:由于不同的论坛有不同的页面结构,应根据实际情况来设计数据库的结构。例如,针对“科苑星空”BBS 设置的数据库包含两个表。main 表记录论坛帖子的主要信息,topic_re 表记录主题帖和回帖的关系(表 1)。

控制显示模块:查看数据库的信息,通过选定关键词的运算,在相应的时间和字段内进行筛选,或者进行结果内筛选,筛选完成后点击数据分析图,选择相应的图标和类型进行查看,以多种方式进行输出(表 1)。

4 科苑星空 BBS “火车票”事件实证分析

为了验证该系统的有效性,我们以中国科学院的科苑星空 BBS(bbs.gucas.ac.cn)为对象,进行—64 万方数据

表 1 数据库字段及含义

表 main	
pid	帖子 ID,此字段为关键字
bid	版面 ID,标识论坛不同版面
type	是否为主题帖(主题贴为 1,回帖为 0)
content	帖子内容
userid	用户 ID,即发帖人的标识
ptime	发帖时间
title	主题,回帖的 title 字段即为对应主题贴的主题
表 topic_re	
topicID	贴子的 ID
reID	若为主题贴则此字段为其帖子 ID,若为回帖,此字段则为其相应主题的 ID

针对性的恐慌度量与监测,对结果进行了分析,并给出了相应的对策建议。科苑星空有其特有的会员群体即学生,也有特有的数据格式,故在实证过程中,我们采用随机访问的方式进行调查,再在论坛相应版面中搜索,最终选择了“寒假订票”作为关注的话题。

5.1 数据获取与恐慌度量指标

首先,确定数据来源,即在科苑星空 BBS 内限定某些版面。与火车票相关的帖子数量最多的版面为“科苑特快”、“交通信息”,因此,数据来源为这两个版面。其次,根据恐慌指标和 BBS 本身的信息特征,选取所要采集的数据类别。所选数据类别包括帖子的 ID、帖子主题、版面 ID、是否为主题帖、帖子内容、用户 ID、发帖时间等。然后,使用 Offline Explorer 下载论坛两个版面 3 层以内以“bbs.gucas.ac.cn”为起始地址的所有链接。爬虫自动按一定的文件格式进行保存,得到初始数据文件。最后,将这些文件中的数据按照固定格式导入数据库中。以 2009 年 1 月 10 日进行的数据获取为例,两个版面中所有帖子导入数据库后得到 3407 条记录。

为了研究火车票在中科院学生中引发的恐慌问题,我们利用现有数据,采用多指标方式进行评价。考虑到帖子总数对目标(火车票类)帖子的影响不同,将帖子总数作为基数加入模型,得到两个子指标:

$$S_t = \sum_0^t N_{pt} / \sum_0^t N_t \quad V_t = \sum_0^t N_{pt} - \sum_0^{t-1} N_{p(t-1)}$$

其中,t 为时间(相关帖子出现的第一天,t 记为

0)。 $\sum_0^t N_t$ 是截至 t 时间累计的帖子总数, $\sum_0^t N_{pt}$ 是截至 t 时间累计的目标帖子总数, $\sum_0^{t-1} N_{pt(t-1)}$ 是截至 t-1 时间点时的累计目标帖子总数。 S_t 是一个标量, 反映事件影响相关人群的比例; V_t 是增量, 测度事件影响的增加速度, 如果选用天为时间单位, 则为日发帖量。我们既可以得到以时间为自变量, S_t 及 V_t 各自的变化曲线, 也可以结合两个指标, 将恐慌值域划分成 4 个区域, 对应不同的恐慌程度。在图 3 中, S_t 、 V_t 是 4 个区域的划分界限。

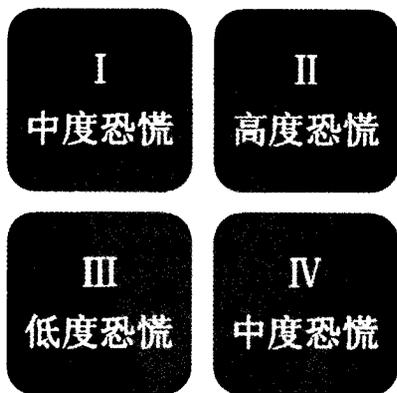


图 3 科苑恐慌指标矩阵

5.2 结果分析与对策

选择 2008 年 12 月 11 日至 2009 年 1 月 11 日为监测时间段, 在所获取的全部帖子数据中以关键词“订票”或“火车票”进行检索, 将检索到的结果进行数据分析。为了深入地了解中科院学生关于火车票的恐慌程度, 我们得出系列图表, 包括每日发帖标量图、每日发帖增量图、帖数累计总量

图、查询前后比例图、帖数按时间分布图、每日主题帖数与回帖数比较图、发帖 ID 分布图、同主题发帖统计图等, 并按照恐慌指标的两个维度对该数据进行恐慌度分析。比较帖子数的变化(图 4)和变化速度(图 5), 可以发现具有转折意义的时间点为 2008 年 12 月 12 日、19 日及 2009 年 1 月 9 日, S_t 的最大值出现在 2008 年 12 月 19 日, V_t 的最大值出现在 2008 年 12 月 12 日。根据科苑星空的特点, 将 S_t 设为 31%, V_t 设为 39, 得到图 6。

由图 6 可见, 恐慌只分布在低度恐慌和中度恐慌两个区域内, 并且只有 3 个时间点落在右下角的中度恐慌区, 说明“火车票”事件虽然在这段时间内在中科院成为热点问题, 但并没有造成过大的慌乱。在低度恐慌区有很多时间点的 S_t 与中度恐慌区 3 个时间点的值相差无几, 特别是 2008 年 12 月 19 日的时间点, 是 S_t 的最大值点, 可是还是落入低度恐慌区。从实际情况看, 订票期间只是提出买票申请, 在订票前会因无法预知而担心, 但订票之后就只能等待, 而在确定是否订票成功的时候, 有大批未订票成功的学生担心买不到票, 恐慌程度明显增高。

在低度恐慌区的时间点, 要提高警惕, 做好应对的准备。如果是订票期间, 比如 2008 年 12 月 19 日、20 日, 需要发布准确的订票程序、注意事项、公布订票结果的时间等。如果订票结果即将公布, 比如 2009 年 1 月 4 日、5 日, 则需及时辟谣, 发布准确消息, 稳定学生情绪。在中度恐慌区的时间点, 要积极应对, 采取适当补救措施。如果是订票期间, 比如 2008 年 12 月 12 日、13 日, 需要调查听取

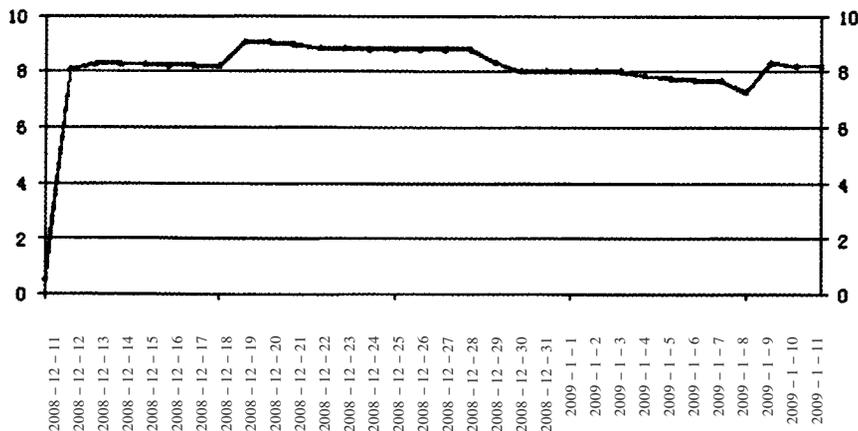


图 4 累计目标帖子数比例

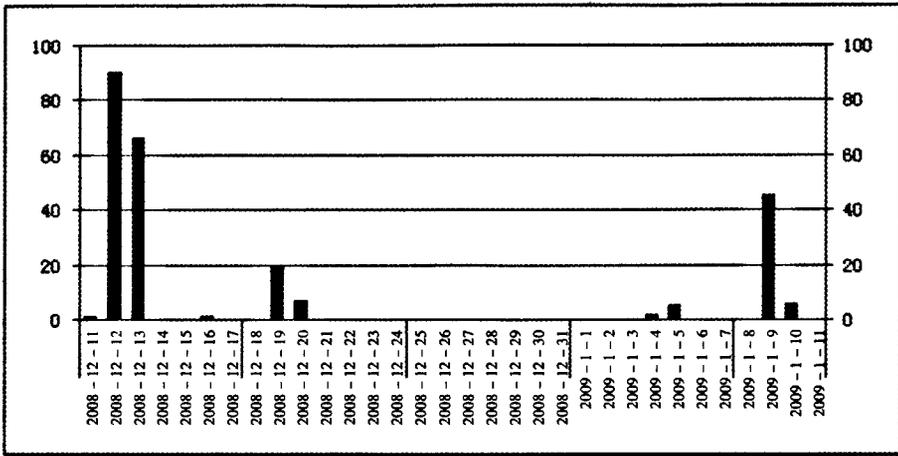


图 5 日发帖量

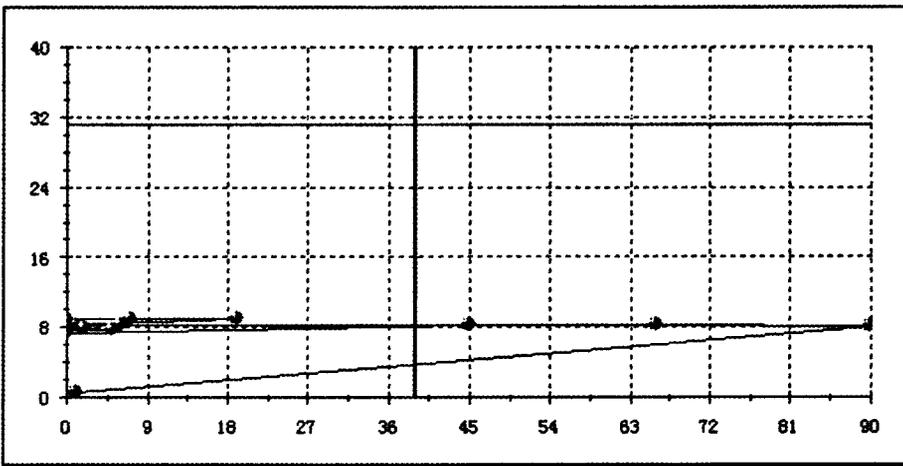


图 6 科苑恐慌指标矩阵

学生们建议,制订更可行的订票计划,尽可能增大订票成功的机率。如果订票结果已经公布,比如2009年1月9日,则可通过解释订票流程来安抚未订票成功学生的不安情绪,并给予有价值的信息帮助。

5 结论与展望

本文首先分析了突发事件引发的公众恐慌的特征及表现,从恐慌传播以及恐慌的程度考察了影响恐慌的因素,通过对网络舆情信息的监测得到反映公众心理的数据,利用恐慌度量模型,得到度量值。度量值和直观展现恐慌变化的曲线可以使政府以及相关部门更好地评测突发事件在某一个时刻对城市公众产生的影响,以期在突发事件发生时,在恰当的时间采取适当的措施防止恐慌的发展和蔓延,降低恐慌对社会带来的影响。本文

提出的基于网络的调查方法,通过对论坛、BBS等网站的数据挖掘可以得到一个大样本的数据,由于被调查者是主动地发表观点或者浏览,这就保证了数据的客观性。整个调查过程可以很快地实现数据的统计以及分析阐释,也可以实现实时的调查分析,因此可以为突发事件的应急管理提供决策依据。随着科技的进步和人们言论自由的提高,网络等信息交流方式必将带来更多的问題,同时也为我们提供更多的研究途径和监测渠道,在未来的研究中,应充分利用网络资源,为应急管理工作提供更多的信息支持。

下一步的研究工作可以从以下两方面考虑。一是模型本身可以进一步扩展,通过听取更多的心理学专家意见,进一步完善恐慌度量与监测模型,并在实际的案例应用中测试、改进模型。二是帖子的内容对恐慌的度量有重要的影响,需要利用文本识别技术和语义识别技术来区分贴子的内容。

参考文献

- [1] 陈安,等. 应急管理理论与方法 [M]. 北京:科学出版社, 2009.
- [2] D. M. Clark. A Cognitive Approach to Panic[J]. Behaviour Research and Therapy, 1986(24): 461 - 470.
- [3] 严亚琼 陈冬峨 等. 公众对 SARS 的理性认知与风险认知相关分析[J]. 中国公共卫生 2004 20(1): 5 - 6.
- [4] 王一牛, 罗跃嘉. 突发公共卫生事件下心境障碍的特点与应对[J]. 心理科学进展 2003 11(4): 387 - 392.
- [5] 谢晓非, 郑蕊. 风险沟通与公众理性 [J]. 心理科学进展, 2003 11(4): 375 - 381.
- [6] 曾祥平, 方勇, 袁媛, 等. 基于元胞自动机的网络舆论激励模型[J]. 计算机应用 2007(11): 2686 - 2689.
- [7] 郑军. 网络舆情监控的热点发现算法研究[D]. 哈尔滨工程大学 2007.
- [8] 张珏, 刘云. 基于 BP 神经网络的 BBS 上帖子回复数预测 [J]. 电脑与电信 2008(10): 28 - 29.
- [9] 余红. 网络论坛舆论领袖筛选模型初探[J]. 新闻与传播研究 2008(2): 66 - 75.
- [10] Yilu Zhou, Jialun Qin, et al. Collection of U. S. Extremist Online Forums: A Web Mining Approach[A] // Proceedings of the 40th Hawaii International Conference on System Sciences[C], 2007.
- [11] Minglu Li, An Chen. A Web Mining Based Measurement and Monitoring Model of Urban Mass Panic in Emergency Management[A] // Proceedings of 5th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) [C], 2008(4): 366 - 370.

Network Opinion Monitoring and Panic Measurement System for Public Incident

Li Jimei^{1, 2}, Chen Ning³, Chen An^{1, 4}, Wu Yannan^{2, 5}

- (1. Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190;
 2. Graduate University of Chinese Academy of Sciences, Beijing 100039;
 3. GECAD, Instituto Superior de Engenharia do Porto, Porto, 4200 - 072, Portugal;
 4. Institute of Emergency Management, Henan University, Kaifeng 475004;
 5. School of Management, Shandong University, Jinan 250100)

Abstract: Public incidents not only lead to great loss to human life and property, but also cause psychological changes of people, which probably trigger public panic, stimulate derivative events and even produce greater disaster. Internet technology makes the transmission of information more quickly, with the negative impact even more serious. This paper analyses the causes, characteristics and response of public panic, then introduces the public panic measurement system based on the monitoring of opinions on network to provide an effective decision support for emergency management.

Keywords: incident, emergency management, opinion monitoring, panic, measurement, data mining