

网络舆情评估 EHA 三维指标体系的构建

朱 辉¹, 骆公志²

(1. 南京邮电大学 通达学院, 江苏 南京 210023)
(2. 南京邮电大学 经济与管理学院, 江苏 南京 210023)

摘要:随着网络在我国的兴起和发展,网络舆情的评估已经成为评价和预警社会形势的重要指标。通过对现有舆情研究的调研和分析,得出由舆情要素、舆情热度和舆情活性3个一级指标、7个二级指标以及14个三级指标共同构建的网络舆情评估的EHA三维指标体系,并对各级指标给出了详尽的分析和阐述。

关键词:网络舆情; EHA 三维指标体系; 舆情要素; 舆情热度; 舆情活性

中图分类号:G203 **文献标识码:**A **文章编号:**1673-5420(2013)02-0037-06

2012年7月,中国互联网络信息中心(CNNIC)发布了《第30次中国互联网络发展状况调查统计报告》。报告显示,截至2012年6月底,我国网民数量达到5.38亿,互联网普及率为39.9%。随着互联网的飞速发展,网络已经成为人们每天获取信息、表达观点、相互交流的媒介和平台。网络内容的生产者已经从过去单纯的网站制作人员转变成为网站和广大网络用户共同作用的混合群体。这种趋势使得互联网正在逐渐成为继报纸、广播、电视等大众传播和人际传播之外的社会舆论传播的新媒体。同时,这一新媒体的重要性也日益凸显,2012年8月发布的《中国社会舆情与危机管理报告(2012)》指出:由于网络社会的迅速发展,由新媒体作为首次曝光媒介的舆情事件正逐渐增加,2011年其占全年社会舆情热点事件的比重为65.2%^[1]。

与传统媒体舆论相比,网络舆论的主体和受众群体更为分散,舆情态势形成更为迅猛,可以

更为直接地反映社会现实,能够在相当程度上代表民意、民声。但是,网络舆情也存在着良莠不齐、虚假信息泛滥等严重不足。网民往往以隐匿的方式存在于虚拟环境中,现实世界里的责任与义务在网络世界中往往会有所缺失,难免会相信、传播甚至编造一些虚假不实、不负责任的信息。极端情绪和虚假的政治舆论的传播,有可能危害国家的安全和社会的稳定。

因此,近年来,网络舆情的变化趋势、网络舆情的管理和引导等问题已经受到了国家管理者的普遍关注,各国政府也已开始高度重视网络舆情的研究工作。时至今日,对网络舆情的研究也已经不仅仅局限于社会科学领域,而是更多地从自然科学的角度进行研究。

一、我国网络舆情指标体系研究现状

网络舆情作为一种公众情绪、态度和意见的反映,从形成到结束都处于动态的变化过程

收稿日期:2012-12-23

作者简介:朱 辉(1982-),女,讲师,主要研究方向为信息系统与网络的管理。

骆公志(1972-),男,副教授,博士,主要研究方向为粗糙集理论及其应用。

基金项目:教育部人文社会科学研究项目“不完备信息下网络舆情快速预警方法及应用——基于粗糙集理论”(11YJC630148);南京邮电大学校科研基金项目(NY210056)

中。网络舆情评估指标体系是进行网络舆情的整体态势分析和安全管理的基础,具有非常重要的理论指导和现实应用价值。在这一过程中,要发现对网络舆情出现、发展和消亡具有重要影响的因素,并连续不断地动态监测、度量及采集它们的信息,对当前的网络舆情做出评级分析并预测其发展趋势^[2]。在关于网络舆情评估体系构建的研究中,不少学者也针对性地提出了自己的观点。

谢海光从网络内容分析的角度提出了“十度”指标体系,从舆情的热度、重度、焦度、敏度、频度、拐度、难度、疑度、粘度和散度十个方面透视互联网内容和基本特点^[3]。吴绍中综合考虑了舆情内容、传播过程和传播受众三个关键要素,以此作为二级指标,分析出了11类三级指标,由此构成了网络舆情预警等级指标体系^[4]。戴媛从网络舆情的安全角度出发,结合政府舆情监控的战略需求,遴选出14个舆情主题,确定了传播扩散、民众关注、内容敏感性、态度倾向性这四个一级指标,在此基础上分别确定了第二级和第三级指标,由此构建了网络舆情安全评估的指标体系^[5]。

在以上指标体系的构建中,大都存在指标体系不完整,一级指标过于冗杂,末级指标的分析只是定性的认识,缺乏定量的研究等问题。部分指标也缺乏深度,有些指标难以进行明确计算^[2]。总结起来,大多存在不完整、太繁杂、难计算这三大问题。在本文的舆情评估指标体系的构建过程中,笔者正致力于解决这些问题。

二、我国网络舆情评估指标体系的构建

在网络舆情评估之前,需要先做好有效舆情信息的收集和汇总,这涉及舆情网络信息的来源和舆情汇总主体两个方面的问题,笔者将首先对这两个问题进行说明。

1. 网络舆情信息的来源

网络舆情汇集的渠道越来越多,信息也越来越复杂,这对网络舆情的监测和分析工作提出了越来越大的挑战,在网络舆情信息的收集过程中,应坚持抓主要网络舆情信息汇集渠

道,并且做到去粗取精、去伪存真。笔者认为,当前主要的网络舆情信息来源主要有以下五大渠道,如表1所示:

表1 主要的网络舆情信息来源及举例

网络信息来源分类	举 例
知名新闻网站、门户网站	人民网、新浪网、搜狐网
微博和博客	新浪微博、腾讯微博、搜狐微博、新浪博客
网络社区、BBS论坛	天涯社区、猫扑论坛、水木社区
社交网站 SNS	人人网、朋友网、开心网
视频网站	优酷网、土豆网

(1) 知名新闻网站、门户网站

知名新闻网站、门户网站等是网上信息发布和交流的重要平台。这些网站登载的网络新闻,往往反映了近期网络舆论的焦点,多数网络新闻可以通过评论的形式让网民更多地互动参与。

(2) 微博和博客

博客专注在特定的课题上提供评论或新闻,因此,通过对博客信息的抓取可以有效地辨析网民关注的热点问题。值得一提的是,微博近几年风靡全国,引领了大量用户原创内容的爆发式增长。强大的群功能、与手机互联互通的特性,使得微博成为最快捷的信息源,成为网上热点敏感话题的发源地、突发事件的扩散地。微博在所有新闻事件曝光渠道中首次曝光的比例已由从2010年的9.3%飙升至2011年的20.3%,跃居第一位^[1]。要抓取有效网络舆情,微博这一舆情汇集渠道是不可或缺的。

(3) 网络社区、BBS论坛

网络社区、BBS论坛是网络中最容易反映社情民意、最活跃的平台之一。网络社区由相对固定的网民群体组成,可以从各个侧面反映网民的思想动态与意见倾向。几乎所有的社会热点问题,都会在网络社区形成议论的热点。

(4) 社交网站 SNS

社交网站是通过关系建立的网络结构。由

于紧密的关系网的存在,爆炸性的新闻或信息在社交网站中传播速度惊人。此外,社交网站中的关系往往真实度比较高,也比较容易实现舆情信息的抓取和管理。

(5) 视频网站

如今,越来越多的新闻和爆料开始以视频等多媒体的方式被转载和传播,视频网站也必将成为网络舆情信息收集中不可或缺的一环。

2. 网络舆情主题的收集汇总

将采集到的信息进行整理,需要将信息按照主题罗列。常见的主题分类模型有:美国的“联邦政府组织架构”(EPA)中的业务参考模型 BRM(Business Reference Model)以及 TRS 政府信息公开目录服务平台中的主题分类等。但这些常用的模型都缺乏对于我国现阶段网络舆情特点的针对性。

纵观国内近期的研究,戴媛的研究中根据中国政府自身对舆情监控的战略需求,结合近几年发生的网络舆情重大事件的相关历史数据和互联网上事件披露多、政策评论多的舆情信息,以及政府公告和白皮书的相关内容遴选出的 14 大类舆情主题指标很有针对性,具有一定的参考价值。这 14 大类舆情主题是:社会经济类、政府管理类、自然灾害类、公共卫生类、重大事故灾害类、社会公共安全类、教育类、医疗类、三农类、交通类、和谐社会类、重大活动类、劳动就业类、社会保障类^[5]。

3. 网络舆情评估 EHA 三维指标体系的构建

在网络舆情信息收集汇总完毕之后,就可以对网络舆情进行评估了。在前文的研究中,笔者认为,已有舆情评估指标体系存在不完整、太繁杂、难计算这三大问题。对于舆情评估这一复杂多级的任务,舆情评估的指标体系必然是一个多级多维度的体系。而作为一个多维度立体化的指标体系,舆情评估的第一级指标应该尽可能的简明清晰。舆情评估的主要目标在于通过对于舆情主题内容、发展趋势、影响情况这三大方面的把握,以此评估和预警舆情态势并制定解决方略。因此,笔者认为,第一级指标应由舆情的基本情况(舆情要素 E)、舆情的综合影响(舆情热度 H)以及舆情的发展态势(舆情活性 A)这三个因素构成。

在确定这三个一级指标的基础之上,笔者通过对于网络舆情传播特点和传播流程分析的探讨,以网络舆情内容为主体,结合前人研究,逐一剖析出网络舆情评估的二级指标和三级指标,构建了我国网络舆情评估的 EHA 三维指标体系,如表 2:

表 2 我国网络舆情评估的 EHA 三维指标体系

第一级指标	第二级指标	第三级指标
舆情要素 (E)	舆情发布者 (EP)	主题署名度(EP1)
		舆情发布者影响度(EP2)
	舆情内容 (EC)	舆情主题类别(EC1)
		舆情内容敏感度(EC2)
舆情热度 (H)	舆情强度 (HA)	舆情内容倾向(EC3)
		累计 Web 页面数量(HA1)
		累计点击数量(HA2)
	舆情传播地理 分布(HD)	累计主题搜索数量(HA3)
		地理区域分布扩散程度 (HD1)
		舆情传播渠道 (HC)
舆情活性 (A)	舆情数量活性 (AQ)	Web 页面数量变化率(AQ1)
		点击数量变化率(AQ2)
		主题搜索数量变化率(AQ3)
	舆情渠道活性 (AC)	舆情渠道活性指数(AC1)

从表 2 中构建的指标体系可以看出,对某一特定的网络舆情信息进行评估,主要从三个方面进行:舆情要素(E)、舆情热度(H)和舆情活性(A)。舆情要素作为表征网络舆情信息最基本属性和信息指标,受到舆情发布者信息和舆情内容这两个二级指标影响。舆情热度是衡量舆情蔓延状况和表现程度的重要指标,主要从强度、舆情传播地理分布、舆情传播渠道三个角度进行汇总。舆情活性是描述舆情变化规律及未来发展的指标,通过舆情数量活性

和舆情渠道活性两个方面综合表述网络舆情的动态变化或趋势。这7个二级指标又分别由14个三级指标综合汇总得到,在下面的论述中笔者将对该指标体系的各级指标逐一说明并分析。

三、我国网络舆情评估 EHA 三维指标体系的解析

1. 网络舆情要素指标(Element)

舆情要素(E)是评价舆情基本属性和范畴的指标。舆情要素主要通过舆情发布者(EP)和舆情内容(EC)2个二级指标综合体现。

(1) 舆情发布者

舆情发布者指的是第一时间发布以及在第一时间转载舆情信息的网民。该指标关系到舆情信息的影响性及可信程度。该指标又受到两方面因素的影响,即:主题署名度(EP1)和舆情发布者影响力(EP2)。

主题署名度是对舆情信息发布者的署名情况的综合描述,可以用来衡量发布出的舆情信息的可信程度。

$$EP1 = \frac{\sum_{i=1}^n f_i}{n}$$

其中, f_i 为信息署名函数, $f_i = 1$ 定义为署名, $f_i = 0$ 定义为未署名。

主题署名度是同一主题下各个子信息署名情况加总后与总信息条目数的比值,取值在(0,1)范围内。主题署名度取值越高,说明信息的真实性越强,越容易被公众接受并关注。

舆情发布者影响度是对于舆情发布者权威性和影响程度的评价,与舆情发布者的活跃程度、受关注程度等相关。

$$EP2 = \frac{\sum_{i=1}^n (K_1 C_1 + K_2 C_2) f_i}{\sum_{i=1}^n f_i}$$

其中, C_1 是舆情发布者的发文总数, C_2 是文章的回复总数, K_1 和 K_2 分别是相应权值。

(2) 舆情内容(EC)

舆情内容是表征和衡量舆情内容和敏感程度的指标,通过舆情内容可以更好地了解舆情

的范畴属性,把握舆情问题解决的迫切、棘手程度。舆情内容由舆情主题类别(EC1)、舆情内容敏感度(EC2)和舆情内容倾向(EC3)这3个三级指标综合影响得出。

舆情主题类别是对于舆情内容的分类判断,在具体划分舆情主题的时候可以参考戴媛的分类方式^[5]进行。

网络舆情内容敏感度是指某一特定的网络舆情信息可能造成的危害程度,与评估者的意图和着眼点密切相关。判断和识别舆情信息的敏感程度是网络舆情信息安全的重要环节。网络舆情内容敏感程度是主观性判断指标,需要量化为定量指标后进行评估,评判结果可以通过专家问卷的方式确定。

舆情内容倾向是指某舆情主题所有条目信息内容的观点态度。可以利用分类算法将观点分为三类:反对、中立和支持,利用观点提取的算法计算指标值。在具体计算中可以参考叶慧敏提出的简单分类算法^[6],根据特征词汇库中的预设词,将一条信息中的每个词都赋予相应的值(-1,0,1),分别对应表示反对、中立和支持。在一个信息条目中,如果特征词汇的个数超过了给定的正面(或反面)阈值,则该条信息观点就被相应地定位为反对、中立和支持。

2. 网络舆情热度指标(Heatedness)

舆情热度(H)是衡量舆情影响程度的重要指标,偏向对于舆情即时影响状态的描述。通过舆情热度指标可以把握舆情已形成的规模、影响的范围和程度。舆情热度指标主要通过舆情强度(HA)、舆情传播地理分布(HD)、舆情传播媒介(HM)3个二级指标来表现。

(1) 舆情强度(HA)

舆情强度是表征舆情传播状态的重要指标,主要表征舆情传播的即时状态,通过累计Web页面数量(HA1)、累计点击数量(HA2)和累计主题搜索数量(HA3)这三方面三级指标综合得出。

累计Web页面数量指在一定的统计时期内某一网络舆情信息涉及的总的Web页面数,是对于舆情信息规模的一个绝对数据的把握,它是体现舆情信息强度的非常重要的指标之

一。该指标可以通过一定统计时期的 Web 页面数据的收集汇总获取。

累计点击数量指在一定的统计时期内网民就某一舆情信息相关的全部帖子的点击总量。如果网民对某一信息比较关注,他就会点击浏览相关信息,对于点击总量的累计产生贡献。因此,通过对于点击数量的累计,可以在一定程度上了解网民的关注程度。该指标可以通过一定统计时期内搜索该舆情信息累积点击数量来定量采集获取。

累计主题搜索数量指在一定的统计时期内就某一舆情主题的累计搜索总量,用来衡量网民对某一舆情主题的主动关注的程度。该指标可以通过一定统计时期内相关搜索引擎的主题词搜索量的汇总获得。

(2) 舆情传播地理分布(HD)

舆情传播地理分布表达的是相关网络舆情信息的空间分布和变化规律,通过地理区域分布扩散程度 HD1 这个三级指标来展示。主要通过对于舆情信息发布者及参与方的 IP 地址来查询和定位相关地址位置,用以判断某一网络舆情信息在地理位置上的流通过程,判断它的分布范围和扩散趋势。这一指标的评价结果可以通过特尔斐法来确定最终结论。

(3) 舆情传播渠道(HC)

舆情传播渠道描述的是舆情传播的方式即信息获取的来源,由舆情传播渠道指数 HC1 综合体现。舆情传播渠道指数受到舆情信息收集的来源比重及相关影响因子的影响。

$$HC1 = \sum k_1 P_1 + k_2 P_2 + k_3 P_3 + k_4 P_4$$

其中, P_1 是新闻网站转载百分比, P_2 是微博转载百分比, P_3 是 SNS 网站转载百分比, P_4 是其他网站转载百分比, $P_1 + P_2 + P_3 + P_4 = 1$; k_1, k_2, k_3, k_4 分别为相应的影响因子,可以由每年权威发布的各类舆情渠道的影响指数获得。

3. 网络舆情活性指标(Activity)

舆情活性(A)衡量的是舆情的变化趋势及未来可能发展的指标,由数量活性(AQ)和渠道活性(AC)2个二级指标构成。

(1) 数量活性(AQ)

数量活性表达的是舆情信息的数量方面变化趋势的综合指标,受到 Web 页面变化率

(AQ1)、点击数量变化率(AQ2)和主题搜索数量变化率(AQ3)三方面的综合影响。

Web 页面变化率是由累计的某一舆情信息的 Web 页面总数量在两个连续的统计时期的差值除以时间的变化量得出。

点击数量变化率是由累计点击某一舆情信息的所有相关帖子总数在两个连续的统计时期的差值除以时间的变化量得出。

主题搜索数量变化率是由累计搜索某一舆情主题总次数在两个连续的统计时期内的差值除以时间的变化量得出。

(2) 渠道活性(AC)

舆情传播渠道活性描述的是舆情传播渠道来源的综合变化规律,由舆情传播渠道活性指数 AC1 综合体现。舆情传播渠道活性指数受到各种舆情信息来源的活性及相关影响因子的影响。

$$AC1 = \frac{\sum k_1 Q_1 + k_2 Q_2 + k_3 Q_3 + k_4 Q_4}{\sum Q_1 + Q_2 + Q_3 + Q_4}$$

其中 Q_1 是新闻网站转载量变化率, Q_2 是微博转载量变化率, Q_3 是 SNS 网站转载量变化率, Q_4 是其他网站转载量变化率; k_1, k_2, k_3, k_4 分别为相应的影响因子,可以由每年权威发布的各类舆情渠道的影响指数获得。

四、结束语

本文通过网络舆情评估 EHA 三维指标体系的构建,将网络舆情的相关理论与技术研究有机地结合起来,为网络舆情分析及预警系统的实现提供了一个系统化的基础性平台。针对性地对苗头性、敏感性网络舆情信息的抓取、识别和分类,再从舆情内容、舆情热度、舆情活性三个维度对舆情进行评估以及及时做出反应,实现对网络舆情安全态势的最大程度掌握,防范负面网络舆情危害社会。

文中提出的网络舆情评估 EHA 三维指标体系由简单准确的 3 个一级指标,完整全面的 7 个二级指标,可以明确计算的 14 个三级指标构成,很好地解决了原有指标体系太繁杂、不完整、难计算这三大问题,可以广泛适用于舆

情评估工作中。

参考文献:

- [1] 中国社会舆情与危机管理报告(2012)[R]. 上海:上海交通大学舆情研究实验室,2012.
- [2] 王青,成颖,巢乃鹏. 网络舆情监测及预警指标体系构建研究[J]. 情报科学,2011(7):7-10.
- [3] 谢海光,陈中润. 互联网内容及舆情深度分析模式[J]. 中国青年政治学院学报,2006(3):95-100.
- [4] 吴绍中,李淑华. 互联网络舆情预警机制研究[J]. 中国人民公安大学学报(自然科学版),2008(3):38-42.
- [5] 戴媛,郝晓伟,郭岩,等. 我国网络舆情安全评估指标体系的构建研究[J]. 信息网络与安全,2010(4):12-15.
- [6] 叶慧敏,唐三平. 用于网上舆论观点抽取的几种算法[J]. 计算机应用研究,2006(5):26-27.

Establishment of EHA three-dimensional indicator system of network public opinion assessment

ZHU Hui¹, LUO Gong-zhi²

(1. Tongda College, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
2. School of Economics and Management, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: With the rising and developing of internet, the assessment of network public opinion has become an important indicator of evaluation and precaution of the social situation. Based on the survey and analysis of current study on public opinions, this thesis proposes a three-dimensional indicator system of Chinese network public opinion assessment established on 3 first-level indicators of opinion element, opinion intensity and opinion activity, 7 second-level indicators and 14 third-level indicators. Meanwhile, the thesis provides a detailed analysis and elaboration of indicators at all levels.

Key words: network public opinion; the EHA three-dimensional indicator system; opinion element; opinion intensity; opinion activity

(责任编辑:刘云)