

大数据时代的社会舆情治理： 何以可能与何以可为

王仕勇

(重庆工商大学文学与新闻学院, 重庆 400067)

摘要:社会舆情处于大数据时代,话语表达及其治理呈现出复杂的局面。网络技术让舆情生成与表达数字化;让舆情快速广泛传播,实现信息互动与共享、关系建立与拓展、情绪传播与感染;使舆情表现形式多元化、参与主体多元化、舆论发展方向模糊化。大数据技术使计算机处理海量信息成为可能,原来沉寂、孤立、分散的原子式存在的数据,能够发声,这是社会舆情治理面临的重要机遇。大数据技术可以拓展社会舆情治理领域,提高社会舆情治理能力,丰富社会舆情治理手段。如何满足人们对更全面、更有价值的信息数据的追求,如何让数据“发声”,是大数据背景下社会舆情治理面临的挑战,这种挑战包括思维及数据的有效整合、挖掘利用与信息安全等。

关键词:大数据;社会舆情;社会治理;网络社会;镜像世界

基金项目:国家社会科学基金项目“网络圈群舆情共振现象的发生机制及治理研究”(20BSH152)。

[中图分类号] D669

[文章编号] 1673-0186(2021)012-0084-012

[文献标识码] A

[DOI 编码] 10.19631/j.cnki.css.2021.012.007

正如舍恩伯格所说,“大数据开启了一次重大的时代转型”,“大数据正在改变我们的生活以及理解世界的方式”^[1],随着网络新技术对传统舆论格局的根本性变革影响,网络虚拟空间与现实生活空间的边界日益模糊,社会舆情的治理已经很难用网络虚拟治理或现实社会治理简单进行划界。面对信息过载,人们开始运用新的思维方式审视这些庞大的数据信息。

社会舆情,一般指一定时期和范围人们对社会现实的主观反应,表现为思想、心理、情绪等。在大数据时代,无论是思想还是心理,抑或意见、情绪等,都因无处不在、无时不在的网络化影响变成了数据信息,社会舆情大多都变成了网络舆情。今天,我们说社会舆情治理,很大程度上指向了网络舆情治理。社会舆情处于大数据时代,话语表达及其治理呈现出复杂的局面。

一、社会舆情:数字技术的强大魔力

网络(尤其是媒体平台)自是今天公众表达诉求的重要载体和渠道。因而网络舆情成为社会舆情的重要组成部分。传统社会舆情,我们通过街巷、广场去观察捕捉,通过民意调查、问卷抽样、访谈

作者简介:王仕勇,重庆工商大学文学与新闻学院院长、教授、硕士生导师。

等方式去感知获取。在大数据技术背景下,社会舆情的呈现展示出与传统媒体占主导地位时代不一样的场景。

(一)生成与表达:网络技术让舆情数字化传播

网络已经成为当今社会得以组织、协同和互动的重要媒介。今天,远程学习、网络购物、微信朋友圈、二维码支付等,都依靠数字技术。作为人们态度、意见、情绪或看法表露的舆情,也越来越依靠数字化技术进行传播。现实社会中,人们对于具体社会现象或社会问题的议论及观点很容易上传到网络,网络上的讨论或看法也会很快向现实社会传播,这种基于数字流动技术的传播,已成为新媒体时代的常态。中国人民大学陈力丹教授认为舆论包括显舆论、潜舆论和行为舆论等。在大数据时代,这些舆论的形态都被打上了数字的烙印。

显舆论:言语表达数字化。在网络空间,借助键盘、虚拟键盘、投影键盘、手写板录入、语音识别录入、手势识别录入等方式,所有的言语表达都变成了可以被电脑识别和记忆的数字文本。在今天这个数字化社会里,无论是文字处理器、短信,还是微信之类的社交网络,数字化工具的输入功能让我们快捷地处理文本,使我们的言语表达不再瞬间即逝,具备了显示、记忆、存储、传播等多种功能。当文本不再只以书本形式存在,还可以以数字版本形式存在时,这意味着经由现代数字处理技术后的言语表达留下了痕迹。

潜舆论:情绪表达数字化。潜舆论,本是一种不易捕捉、较为零碎含糊的情绪。由于人的内心深不可测、瞬息万变,在互联网出现以前,要对潜舆论进行定量或定性的描述是比较困难的。网络传播技术的出现,让情绪的描述变成现实。“因为互联网产品基于计算机代码,从理论上来说,无论什么形态的信息都可以还原量化。”^[2]在互联网上,网民的言论不仅可以被记载,而且可以把只言片语间蕴藏的情绪变成确切的可讨论的对象。这些情绪,可能是宣泄,可能是零星的讨论中的态度,均变得可查可考,我们可以从“赞成”“反对”“批评”“困惑”“愤怒”等情绪中,看到一种被延迟被扩散的意见形态。互联网还为我们开发了各种各样的表情符、表情包,还有“囧”“槩”等生僻文字,以及各式各样的网络新词,这些都成为网民情绪表达的替代符号。

行为舆论:行为表达数字化。行为舆论,指的是人们在一定合意的基础上采取的行动,是“舆论在精神领域实现改造社会的现实延伸”^[2]。在现实社会中,行为舆论表现为抗议、游行、示威等。在网络社会,网络聚集形成的群体性事件,网络围观、网络话语抗争都可视为一种行动。网民以“发帖、灌水、加精、置顶”等方式对公共部门和公职人员“涉腐”“涉富”“涉权”等问题进行集中炮轰,草根网民在互联网上刮起群体性的问责狂潮,均属于行动舆论。日常社会的行为舆论往往随着事件的发生、发展、高潮、结束而演绎,要回溯这些行为舆论,只能依靠当事人、旁观者、目击者的共同回忆还原。在互联网虚拟场域中,行动舆论已经被数字化,我们可以利用 Cookie 技术、爬虫技术等,搜集、汇总、还原互联网上行动舆论从萌发到散场的轨迹,可以找到轨迹发展的时间路径,找到高潮的时间节点和形成高潮的缘由。不仅如此,我们还可以从网络上的行动舆论的影响范围,勾勒出一个事件的行为效果。

(二)喧嚣的氛围:网络技术让舆情快速广泛传播

社会学家认为,人际关系网中的节点是一小群一小群地聚集在一起的。六度分隔理论认为,和任何一个陌生人的联系最多只需要通过五个中间人。今天的舆情信息传播,把单个行动的主体连接起来,实现了舆情的社会性传播,即关系传播和情感传播。这种传播,导致互联网成了一个信息混杂、舆论喧嚣的阵地。

社会舆情的信息互动与共享。社会舆情本质上是信息,网络舆情本质上是数字化的信息。社会舆情叠加网络技术,使信息的互动与共享成为传播的基本特点。在网络上,信息的生产与传播变得非常容易,传播者与受传者的界限被完全打破,无论哪个国家或地域,只要有可以与互联网相连接的计算机,信息便可以跨越地理区隔,实现互联互通。有了网络连接,受众成为信息的传播者,可以借助邮件、BBS 等交互工具,实现即时反馈;文字、图片、音频、视频等信息,依靠数字技术实现综合信息传播,二进制元编码解决了不同形态的信息交换、分配与存储问题,实现了用户之间的无障碍交流。

社会舆情的关系建立与拓展。齐美尔认为,网络中的行动者,不仅仅是一个点,而且是复杂的网络关系^[3]。社会网络分析就是研究行动者及其之间关系以及他们之间的相互影响。也就是说,一个行动者的社会网络,就是其人际关系网和与其有交往互动关系的所有人形成的集合。在网络中,每个网民都有自己的社会网络,如网络邮箱中每一个联系对象,论坛中的交流对象、SNS 中的好友、微博中的关注对象及“粉丝”、微信中的朋友圈等。社会网络好比一张地图,标示出所有与节点相关的连结。这些星罗棋布的节点,可以是个人或组织,社会网络把各种关系中的人组织串联起来,这种联系可能是非常紧密的,如家庭朋友关系,也可能是素不相识的但并不是毫无关联的产生着琐碎复杂的联系。社会性网络使我们通过熟人的熟人进行社交网络拓展,将现实中的社交圈子转移到网络上,还可以根据自己的爱好拓宽个人社交网络。

社会舆情的情绪传播与感染。网络技术改变了人们情感表达以及经由情感形成的关系。在网络空间,民众对一些具有刺激性的社会敏感事件,往往会有明显的感性表达,比如尖锐的批评、激烈的抱怨或热情的赞美,如果这种情感表达是适度的,是明智理性的,也会让其他表达趋于明智或理性;如果这种表达是偏激的,过度情绪化的,则很容易引发盲目从众行为。《美国科学院院报》发表的研究成果指出,PNAS 的研究人员研究了 689 003 名英语为母语的 Facebook 用户,发现社交网络上普遍存在情绪传染现象,情绪状态能通过情绪传染的过程在人们之间转移,人们会不知不觉地带有他人的情绪。网络中“围观”式的社会参与行为,很容易被跟帖者的情绪感染。各大网站开设“我要评论”“我来说两句”“网友跟帖”留言区,目的是与受众形成互动,满足受众关注与参与心理,这种“围观”行为是一种看客式的参与行为,易被简单化思维主导,易被情绪感染。从某种角度讲,这种情绪感染具有仪式动员的特征,会导致相似的情感流动进而形成特定的社群。从个体情绪、群体情绪再到社会情绪传播,贯穿其中的是情绪表达的流动曲线。勒庞在《乌合之众:大众心理研究》中指出:“在群体中,每种感情和行动都有传染性,其程度足以使个人随时准备为集体利益牺牲他的个人利益。这

是一种与他的天性极为对立的倾向,如果不是成为群体的一员,他很少具备这样的能力。”^[4]

(三)繁杂的数字数据:网络技术让舆情呈现新变化

舍恩伯格在《大数据时代:生活、工作与思维的大变革》中说,“在 2007 年,所有数据中只有 7% 是存储在报纸、书籍、图片等媒介上的模拟数据,其余全部是数字数据”^{[1]12}。互联网产生的数字数据,来源于互联网交易、移动终端、各种网络设备和传感器、社交媒体等。繁杂的数字数据,颠覆了我们以往对于舆情的传统认知。在大数据时代,社会舆情具有了很多新的特征。

一是舆情表现形式多元化。网络时代的舆情信息,来源于新闻评论、BBS 论坛、博客、播客、微博、聚合新闻(RSS)、新闻跟帖及转帖、网站、平面媒体、微信等,可谓铺天盖地。这些信息,由于自身观察、观点的局限性,往往差异很大,甚至完全对立。快捷、多元的信息传播,复制粘贴后无限次传播的信息,以及被多种形态转化的信息,都杂陈在网络上,舆情表现形式的多元也就在情理之中了。不仅如此,音频、视频、图片、地理位置信息等,多类型的数据信息,仿佛能相互支撑、相互印证,但真正以人工面对的时候,却很容易让我们束手无策。

二是参与主体多元化。在互联网上,有一句经典的表述:“没有人知道你是一条狗。”舆论主体的匿名性,也即参与主体的多元化,表现得极为明显。渊博的学者、具有很大影响力的网络大 V、各种职业背景的草根网民,甚至是一个刚刚学会敲键盘的小孩,都可能在网上生产信息。对于一个网络事件,我们几乎能看到社会各个阶层的人参与发声,表达观点看法,参与主体的多元化导致信息蕴含价值的多元化。互联网上,参与主体的多元化导致对社会问题的不同看法。表现在,主要发表负面内容的黑色地带,发表能代表主流价值观的红色地带,发表思想观念较为消极、庸俗但不反动内容的灰色地带,不同的地带实质上反映出参与主体多元形成的不同思想文化。

三是舆论发展方向模糊化。网络社会舆情的一个突出特点是舆论发展方向模糊化。舆情从产生到扩散都依赖于互联网,贴吧、论坛、微博、微信、秒拍等都可能成为舆情的高发地带,舆情的发源地较为模糊不易把握。由于网络中布满了多个节点,这些节点可能是关键节点,也可能是弱关系节点,但无论哪一个节点,都可能成为影响舆情走向的节点。即使一个舆情已经发生了,但哪里会产生次生舆情,在什么时候产生次生舆情,次生舆情朝哪个话题方向分岔,也很难预测和把握。同时,网络上各种信息混杂,议题生成具有自发性,也导致舆论发展方向模糊化。

二、数据可以“发声”:大数据背景下社会舆情治理的机遇

社会舆情是民情的动态反映,是民意的集中体现,也是现代社会重要的政治资源。在大数据时代,如何利用大数据理念、技术、方法等做好社会舆情治理,是政府必须面对的重大问题。“大数据发展的核心动力来源于人类测量、记录和分析世界的渴望”^{[1]3},由于计算机将模拟数据转换成用 0 和 1 表示的二进制码,实现了信息的数字化,这为实现数据的数据化奠定了基础。大数据就是让万物数据化,让计算机处理庞大海量信息成为可能。原来沉寂、孤立、分散的原子式存在的数据,变得能够

发声,这是社会舆情治理面临的重要机遇。

(一)“一切皆可量化”:拓展社会舆情治理领域

大数据时代,数据无处不在。道格拉斯提出“万物皆可量化”,得到了验证。当我们对社会舆情涉及的所有内容都可以用数据描绘出来,我们就可以对社会舆情拥有跟以往不一样的深切了解和把握。

一是社会舆情的所有形态数据化。文字、图片、音频、视频等是社会舆情的基本表现形态。对不同形态的海量数据进行采样、存储、共享和分析,得到包括文、图、音、像、表等多源、多元数据,通过对这些数据进行专业化处理,挖掘其潜在价值,实现社会舆情数据化。大数据不仅把所有舆情形态数据化,还实现了微博、微信公众号、客户端、视频、电子报刊等海量数据的整合。这些数据,其实是舆情在萌发、发展、消落过程中的有价值的信息。这就为舆情的研判、处置等提供了量化的依据。

二是社会舆情的地理方位数据化。人类生活 80%的信息与地理位置有关。人和事物的地理方位是信息的组成部分。从地理方位我们可以预测或判断一个舆情事件可能在哪个地理区域发生,可能在哪个地方引发高潮;从地理方位我们可以判断不同地域的人群因何种原因积极或消极关注某一舆情,以及舆情发展的空间分布及走向。社会舆情变动中的每个参与者都有着自己的行动轨迹,这种行动轨迹在大数据时代借由标准的标记体系和收集记录的工具,实现地理信息的标准化和量化。在社会舆情运用方面,我们可以通过智能手机的定位功能收集位置信息,可以运用一些应用程序获取位置信息,可以利用 GPS 导航装置、无线路由器的信号强度、汽车上的无线传感器等收集用户地理位置数据。我们甚至可以通过数据服务商发现交通拥挤情况,发现一个地方人口的聚集程度。今天,人们往往在不知情的情况下,被收集着地理位置信息,如使用 iPhone、安卓、微软手机的人,操作系统其实都在收集地理位置信息。麻省理工学院人类动力学实验室主任亚历山大·彭特兰通过分析每个人去了哪里、见了谁,成功地区分出了感染了流感的人群,而且在感染者知道自己已经患病之前就作出了区分^{[1][18]}。地理大数据相当于桥梁,精准的空间位置服务可以实时感知道路拥堵、给城市“治病”、控制人类疫情、改变城市生活质量,甚至助力美国总统选举^[5]。

三是社会舆情关涉的交流沟通行为数据化。社会舆情的数据采集不只是不同舆情形态数据、地理方位数据,更重要的是舆情生成发酵及发展中的人与人之间的信息沟通与交流行为产生的数据。在互联网上,我们点击浏览、发帖跟帖、查看点赞等网络行为,其实就是沟通交流行为,这些行为都会留下痕迹。这些痕迹从数据数量的维度来看,构成了社会舆情大数据的重要来源。人具有社会性,人的沟通交流行为产生的数据,是一种社会行为数据。这种行为数据,虽然碎片化,但却是作为社会舆情主体的人的个性化特征和社会属性的体现,它折射了网民网络行为的多样性和复杂性,是网民作为社会成员的网络表达,通过这些具有社会意义的数据挖掘,我们可以了解和解释社会行为与社会心态。

现实社会治理和网络社会治理是社会舆情治理的两个领域。在大数据技术产生前,我们要寻找能打通两个领域治理的桥梁和纽带是很困难的,因为现实和虚拟社会产生的数据很难实现共享,它们之间要建立联系、实现意义交换是一件比较困难的事情。当文字、图片、视频、音频、地理方位、沟

通交流变成数据,网络就成为现实的“镜像”。现实世界和镜像世界彼此映照和投射,为社会舆情的治理领域从以往单纯的现实或网络治理中解放出来,实现了两个领域的协调共治。戴维·杰勒恩特在《镜像世界:或者当软件将整个宇宙装到鞋盒里的那天,会发生什么?意味着什么?》一书中,提出随着计算能力的增强和无所不在的链接,人类将用数据流和算法创造出一个真实世界的微缩模型,人类通过对信息的管理,理解和控制世界^[6]。因为大数据,我们可以以前所未有的丰富细节与深度,观察和跟踪真实世界。因为真实世界在互联网那里能找到倒影,这个倒影包含了“真实生活中的社会、机构和家庭结构”^[7]。

(二)信息超载到精准获取:提高社会舆情治理能力

美国学者比尔·科瓦奇、汤姆·罗森斯蒂尔在《真相:信息超载时代如何知道该相信什么》一书中提出,我们处在一个用户主导的新媒体时代,“我们会比以往获得更多信息,同时也更容易困惑;我们会更容易看见真相,同时真相也更难获得”^[8],这是我们面临的信息超载的现实。大数据为解决信息超载的困惑找到了一把科学的钥匙,实现了信息的精准获取和真相的科学发现。大数据为我们提高社会舆情治理能力提供了新的机遇。

大数据技术可帮助实现突发事件事前预警。舆情事件爆发前,往往会经历一个苗头期,只有随着传播速度加快、范围变广、意见倾向逐渐一致才会形成。因此舆情苗头期的预警对于舆情事件的处置非常重要。舆情预警包括舆情态势判断、预警等级判定和舆情走向研判三个方面的工作。大数据技术可以根据某一主题发布的信息量、回复量和浏览量计算网络舆情参与度,可以根据信息扩散情况计算网络社会舆情的波及度,可以自动采集处理监测的舆情信息,实现社会舆情监测、预警系统与各类舆情数据终端的无缝链接。大数据技术还可以通过对内容敏感性和传播态势等的分析,预判舆情等级;可以通过传播中焦点转移、观点变化、舆论场各方力量对比变化,研判舆情走向。

信息采集技术可提高舆情数据搜集效率和范围。传统舆情采集,大多只能采取抽样的方式,这种小数据、小样本采集分析,往往导致人们对其科学性的怀疑。大数据的一个突出特征是目标永远是全体数据,拒绝抽样分析,从庞大的数据中寻找相关关系,寻找关联,突出效率。大数据采集的海量数据是相互关联、相互印证的,从而使搜集的信息在时间维度和空间维度上建立了逻辑联系,使舆情事件的呈现更加客观。由于决策依据建立在数量庞大的信息基础上,即使有一些虚假或模糊的信息,也不会影响其真实性和科学性。研究认为,93%的行为是可以预测的,如果将事件数字化、公式化、模型化,其实多么复杂的事件都是有其可以预知的规律可循,事态的发展走向是极易被预测的^[9]。

数据存储和挖掘技术可帮助提高舆情分析、追踪和研判能力。无论是传感器采集的数据,还是移动设备、社交媒体上的数据,这些非结构数据的完整存储,为社会舆情的科学全面分析奠定了基础。大数据技术可以对某舆情话题在互联网上是否引起了关注及关注的量级与趋势进行舆情诊断;可以从舆情传播路径,识别传播的节点尤其是关键节点;可以模画网民情感倾向,提炼网民对某话

题的主要态度、关注面与侧重点,通过数据实现情感提炼;可以对某一话题的受众进行画像,分析人群特点。数据挖掘,指的是对过去的数据进行查询和遍历,并找出过去数据之间的潜在联系,帮助人们从大量的数据中智能、自动抽取出隐含的、事先未知的、具有潜在价值的知识大数据^[10]。大数据挖掘,往往按照舆情信息的主题进行,采取对信息进行分组聚类的方式,把相似的记录放在一个聚类里,根据发布信息的IP地址、语气、指向等划分不同类别。比如针对特定的时间维度和空间维度,依据数据积累的历史经验进行舆情诊断;利用预测模型、机器学习等技术,比对历史数据对未来或不确定的舆情事件进行预测评估。利用大数据可进行舆情态势和影响的研判。我们可通过技术手段,分析舆情观点的影响程度和人群,判断舆情趋势;我们可将网站新闻数据、论坛数据、博客数据、微博数据进行比对,分析不同舆情热点在不同职业、不同地域、不同年龄段、不同人群中的传播情况,为进行舆论引导提供科学依据。美国中央情报局通过抓取海量数据来追踪恐怖分子和监控社会情绪,首席技术官格斯·汉特称,在“阿拉伯之春”中,大数据分析可以了解多少人和哪些人正在从温和立场变得更为激进,并“算出”谁可能会采取对某些人有害的行动^[8]。

(三)从静态收集到动态跟踪:丰富社会舆情治理手段

变动的时产生变动的信息,要收集变动的信息,没有大数据技术是很难的,因此要找到合适的社会舆情治理手段也是比较困难的。一旦世界被数据化,只要具备数据分析工具和必需的设备,我们就可以更快更大规模地处理数据,进行社会舆情的科学治理了。

第一,大数据技术促进信息公开,防止谣言滋生和传播。网络为谣言的滋生与传播提供了新的土壤,同时大数据技术又为谣言的识别与遏止提供了手段。重要信息的缺失容易导致谣言产生。大数据技术,可以通过敏感词汇的辨别判断和建立算法模型提取谣言的属性及传播特性,把辟谣信息定向推送给受众,及时遏止谣言传播。2018年8月,中国互联网联合辟谣平台整合全国40余家辟谣平台3万余条数据,构建了对网络谣言“联动发现、联动处置、联动辟谣”的工作模式,实现对网络谣言“清存量、控增量、断传播”。

第二,大数据技术重构政府舆论主动权。利用大数据技术科学监测、分析并正确引导舆情,是创新社会管理、维护社会稳定的一个重要议题。大数据技术,可以帮助政府利用其强大的数据挖掘能力和关联分析能力,整合不同舆情形态和不同渠道的舆情,挖掘发现舆情同民意间的内在联系,有序推进舆情管理。同时,政府还可以凭借数据资源和整合能力优势,联合有关舆情信息生产部门或平台,建立不同类别的舆情大数据管理共享系统,实现及时把握和跟踪舆情动态,推动舆情治理创新。

第三,大数据技术可增强舆情引导效度。信息数据的海量性、动态性、无序性与人们关注时间及关注能力的有限性构成了舆情管理的基本矛盾。大数据技术可以梳理出复杂零乱信息之间的价值联系,可以勾勒出舆情信息之间的传播链条,把看似偶然的舆情信息甚至是突发事件舆情信息的必然性找寻出来,避免社会舆情管理的“盲人摸象”效应,增强舆情引导效度。大数据技术可以对海量信息重构,使舆情管理从片面化走向立体化、纵深化,使舆情引导从主体的单一化走向协同

化全局化,避免传统舆情引导中由于数据库资源的缺乏与各自为政导致的引导片面与偏颇。在引导的路径和方法上,大数据更加注重从信息的社会联络和人的社会交往中发现重点,关注焦点,使舆情引导更加科学。例如,利用独有的分布式网络技术的“微舆情”,可以实现对互联网上舆情相关数据源的完整采集,通过中文智能分词、自然语言处理、正负面研判等大数据处理技术,发现涉及用户的舆情信息,及时通过手机客户端、电子邮件、私信等方式进行报警。这就是大数据时代舆情引导效度的体现。

第四,大数据促进社会舆情治理协同联动。大数据技术产生之前,政府作为社会舆情治理主体,在工作中扮演了全能、全责的形象。大数据技术要求政府各部门、相关企业或组织之间信息系统、数据系统互联互通,要求各自应用软件兼容和用户界面一致,解决各自为政、互设壁垒的问题,解决数据孤岛问题,实现共存共享。大数据技术要求社会舆情治理从单中心治理走向多中心治理,实现舆情数据快速汇集、交流互通和高效利用,实现舆情管理内容从预警到决策环节环环相扣。

三、如何让数据“发声”:大数据背景下社会舆情治理的挑战

大数据技术对舆情治理产生了革命性影响。探寻数据之间的内在关联性,让数据为社会舆情有效治理提供可能与创新路径,即让数据“发声”非常重要。在数据规模最大的中国,如何满足人们对更全面、更有价值的数据信息的追求,如何让数据“发声”,是大数据背景下社会舆情治理面临的挑战。

(一)社会舆情治理思维转变面临挑战

大数据时代,我们面临的不再是随机样本,而是全体数据;我们不可能做到精确,因为我们不得不接受混乱;我们很难准确挖掘出信息之间的因果关系,因为大数据更多告诉我们哪些信息是密切相关的。这就是我们面临的社会舆情治理思维挑战。

一是从随机样本到海量数据的挑战。以随机抽样的方式获取和处理小量数据,以部分代替或者推测整体,是传统数据搜集与处理的方式。这种方式,由于数据采集过程中的人为偏差、数据样本的精确性、数据量小等问题,会造成部分偏差。网站、App、感应设备等新型数据源能提供更全面的数据,系统日志采集、网络爬虫或网站公开 API 等方式,是典型的大数据时代获取海量数据的方式,这将大大降低全体数据获取的成本。聚类分析、可视化分析、关联性规则、预测性分析、深度学习、数据挖掘等,提升了数据处理和分析效率。大数据时代的数据样本,已经无限接近于总体。从随机抽样向海量数据转变,从以随机样本代表整体向全体数据研究转变。在舆情治理方面,必须面对这一新的转变和挑战,因为海量数据意味着舆情信息更难以掌握,舆情会更加复杂多变,要求我们舆情监测的手段和方法必须适应大数据技术。

二是从绝对精确到接受混杂的挑战。大数据时代,随着数据采集方式和处理方式的发展,人们对容错规则放宽,越来越多的数据要求我们接受混杂现实。数据量的增加,可能会造成部分错误数

据进入数据库或数据集,从而使得原有数据发生混乱。大量结构化与非结构化的数据交织在一起,呈现一种混杂的态势。这意味着人们不得不接受数据不精确,不得不寻求更多数据背后更多的价值意义。混杂的数据意味着可能存在虚假的舆情信息。大数据并不是绝对客观的,它依赖于运用大数据的人的价值取向,一旦被别有用心政治力量或经济力量操纵,这种表面运用大数据处理舆情信息的结果,极有可能炮制出“伪舆情”,放大“杂音”或“噪音”。

三是从关注因果关系到关注相关关系的挑战。小数据时代,舆情数据信息量少,事物间的因果关系相对好找。大数据时代,数据规模不断扩大,更新速度不断加快,人们无法考量数据之间的因果关系。大数据技术开始注重相关关系。对相关关系的研究并不意味着将放弃或抛弃因果关系,相反相关关系可以作为因果关系的基础。相关关系对现象表面、浅层的规律探索,通过对可能相关的事物研究,在此基础上,可进行因果关系的分析。这种从关注因果关系到关注相关关系的转变,要求社会舆情治理从传统的“灭火式”管理走向“防火式”治理,要求从以前形成舆情危机开始发布信息、引导舆情,转向在舆情危机形成前进行舆情的关联分析、级别划分、聚类分析以及倾向性分析,将舆情危机发生的可能性降到最小。

(二)舆情数据与其他相关数据有效整合的挑战

大数据技术下的社会舆情预警、监测、处置的科学性建立在数据开放、互联互通互通的基础上。如果舆情数据和大数据池里的其他数据不能建立有机联系或强联系,那么这些舆情信息就是孤立的。

第一,“舆情信息孤岛”问题如何破解?我国是数据规模最大的国家,但这些数据极其分散。在政府层面,网信办、公安局、人社局、卫生局等政府职能部门都有自己的数据库、应用软件,但这些都是各自独立的体系,就像大海里的岛屿一样,彼此之间没有通道或桥梁可以联系,数据库之间不能互通互联;在企业层面,尤其是大型企业,其拥有的数据也存在于不同的数据仓库中,且这些数据技术互不相同,互设壁垒,不相通约。今天,一个事件的真实性版本很多,相关信息各有价值取向,这使得全面深刻地关注和分析舆情事件会变得越来越困难。因此,必须克服“信息孤岛”,运用大数据技术,建立网络舆情自动分析系统,避免因数据源不全面、数据源缺失造成的重要舆情信息监测缺失,要求政府、企业、社会单位之间的信息系统、数据源统一技术标准,共享舆情数据,实现社会舆情的多元共治。

第二,缺席的“声音”如何发现?按照陆学艺先生的“十阶层理论”,产业工人阶层、农业劳动者阶层(含农村外出务工人员)、城乡失业者半失业者阶层数量庞大,但囿于网络硬件与网络素养限制,他们很难通过网络诉求自身阶层利益。因此,单凭技术体系构筑的大数据平台仍然无法获取真正意义上的“全部数据”。要掌握真正的社会“大舆情”,必须把线上线下的数据整合起来,线下数据能弥补一部分社会群体在网络上缺失的数据,能打捞起“沉没的声音”,这样也才会增强舆情的真实性,更好地体现民意,防止错误决策。做到线上线下数据的整合,其实就是把网络舆情与现实社会动态的深层次关系挖掘出来,实现网络舆情治理与现实社会治理的有机结合、紧密联动。

第三,舆情数据可用性低质量的问题如何解决。由于大数据来源渠道多样,形式多样,利用爬虫技术收集到的数据并不能直接运用,还需要进行清洗、甄别及相应的转换。大数据技术对舆情数据的可用性和质量提出了更高要求。要提高清洗和甄别阶段数据的可用性,就必须高度重视大数据预处理阶段的工作,要运用 SQL 脚本和解释器语言编写脚本把数据转化成方便处理的数据类型,对数据进行清洗和去噪,以提高数据的有效性。大数据的爆炸性增长,劣质数据的存在将极大降低数据的可用性。据悉,美国企业信息系统中 1%~30% 的数据存在各种错误和误差,美国医疗信息系统中 13.6%~81% 的数据不完整或陈旧^[12]。在我国,大数据技术刚刚起步,数据的可用性和质量不够更是我们必须面临的挑战,这需要政府层面做好相关工作,这就是舆情数据的预处理工作。

(三) 社会舆情数据挖掘利用与信息安全的挑战

智能终端无处不在,网络传输随处可行,社交网络频繁互动,这些都在随时随地产生数据,这也为数据挖掘奠定了好的技术基础。但大数据技术的运用,从国家安全、个人安全的层面上讲,也带来更多安全风险。

首先是社会舆情数据挖掘、利用的权限和范围。社会舆情数据不仅是海量数据,更是复杂和敏感的数据。在新的技术面前,谁可以挖掘利用,如何挖掘,挖掘利用的范围边界在哪里,这些都是我们需要认真谋划和回答的问题。大数据技术本身无所谓好坏与善恶,但如何使用、谁去使用就涉及数据安全问题了。当前,社会舆情大数据技术的应用,必须解决有法可依的问题,必须建立和完善相关机制。当前,我国的法律条文还不能解决大数据时代背景下的信息安全问题,即使有一些信息安全方面的法律,也基本散落在某些特定行业的管理规定中。同时,大数据时代各类数据的获取易得性增强,成本很低,越来越多的机构、个人由于运用的目的不一样,所持的价值理念千差万别,通过数据挖掘和分析得出的结论也就会不一样,这会导致社会舆情管理的难度越来越大。“数据经常脱离数据拥有者的控制范围活跃着,这就对数据需求合规性和用户授权合规性提出新的要求,包括数据形态和转移方式的合规性。”^[13]唯有建立完善的法律法规体系,才能确保舆情大数据的安全。

其次是数据开放与隐私的边界。今天,任何网上的痕迹都被记录,通过这些痕迹找到个人隐私非常容易。大数据时代,每个用户在网上产生的数据具有累积性和关联性,即使单点信息没有暴露隐私,但大数据的关联和集成技术可以对多点信息进行汇聚处理分析,从而精准锁定信息,将个人信息全面分析出来。如何在推动数据全面开放、应用和共享的同时,有效地保护公民、企业隐私,实现开放与隐私保护的平衡,是大数据时代的一个重大挑战。当前,由于大数据安全标准体系不完善不健全,加之隐私保护的技术和法律法规缺失,导致数据开放与隐私保护很难做好协调。同时,无论是大数据标准开源软件 Hadoop,还是大数据依托的数据库基础 NoSQL,其本身均存在数据安全隐患。这些都增加了个人或单位信息泄露的风险。一些知名网站密码泄露、系统漏洞导致用户资料被盗取,个人敏感信息泄露,这些事件警醒我们,加强大数据网络安全建设势在必行。

再次是舆情数据开放与政治安全、意识形态安全。“一方面,大数据技术如提灯女神般点亮了人类前行的方向;另一方面,它也似开启了潘多拉魔盒,放出了反噬人类的万千魔怪。”^[14]大数据时代,信息和数据呈现出指数增长的态势,一旦这些快速增长的舆情数据被别有用心的人或利益集团控制,就很容易导致谣言快速扩散,风险日益加剧,严重影响国家意识形态安全。在网络空间,以前需要较长时间生成、发展的风险与威胁,在网络社会只需要极短的时间就可以实现从量变到质变。当前,世界大型数据库主要分布在欧美等发达国家,这些数据库自然有着西方价值观导向,我们在接受这些数据服务的同时,也容易受到西方价值观念的影响。更大的不安全还在于,在网络核心技术方面,我们远远落后于欧美发达国家,很多依靠进口的软硬件极易留下嵌入式病毒、隐性通道、可恢复密钥的密码等,存在较大的安全漏洞,这就为西方国家利用大数据技术分析敏感数据提供了技术上的通道与便利,极大影响我国意识形态安全^[15]。

最后是社会舆情治理与大数据技术理性。“理性是使技术彰显其现实力量的助推器。技术在人类社会生活大放异彩的背后,其实是理性精神的胜利。”^[16]大数据确实为我们进行社会舆情预警、监测、处置提供了技术支撑。但我们不能陷入“技术万能”的误区,迷信和依赖数据,要杜绝技术与人文生活之间的分裂。我们要在社会舆情治理中既充分利用大数据技术,又关注数据筛选、分析过程中的对人的尊严、价值和命运的关切。人民日报媒体技术股份有限公司总经理叶蓁蓁认为,人类对于优质内容的追求以及媒体对人类的人文关怀是任何技术手段包括大数据和人工智能等都无法替代的^[14]。社会舆情治理,必须重构以人为尺度的大数据价值观。数据始终是冷冰冰的,只有体现了人的尺度才会具有温度。数据筛选、清洗的过程,实质上应该是去粗取精、价值提取的过程。

参考文献

- [1] 维克托·迈尔-舍恩伯格,肯尼思·库克耶.大数据时代:生活、工作与思维的大变革[M].盛杨燕,周涛,译.杭州:浙江人民出版社,2013:1.
- [2] 陈力丹,林羽丰.再论舆论的三种存在形态[J].社会科学战线,2015(11):174-179.
- [3] 刘易斯·A.科瑟.社会学思想名家[M].施人,译.北京:中国社会科学出版社,1990:199.
- [4] 古斯塔夫·勒庞.乌合之众:大众心理研究[M].冯克利,译.北京:中央编译出版社,2005:17.
- [5] 徐红.地理空间大数据助力智慧城市建设[EB/OL].http://news.xinhuanet.com/city/2017-04/02/c_129523459.htm.
- [6] Gelernter D. Mirror Worlds: On the day software puts the Universe in a Shoebox: How it Will Happen and What It Will mean[M]. Oxford University Press,1993.
- [7] 贾利军,许鑫.谈“大数据”的本质及其营销意蕴[J].南京社会科学,2013(7):15-21.
- [8] 比尔·科瓦奇,汤姆·罗森斯泰尔.真相:信息超载时代如何知道该相信什么[M].陆佳怡,孙志刚,译.北京:中国人民大学出版社,2014:2.
- [9] 艾伯特·拉斯洛·巴拉巴西.爆发:大数据时代预见未来的新思维[M].马慧,译.北京:北京联合出版公司,2017:2.
- [10] 翟云.数据挖掘视域下的网络舆情监测与引导[N].学习时报,2014-03-17(010).

- [11] 尹亚辉.大数据时代网络舆情传播形态与引导战略[J].新闻知识,2013(12):61-62.
- [12] 李建中,刘显敏.大数据的一个重要方面:数据可用性[J].计算机研究与发展,2013(6):1147-1162.
- [13] 奇异网.浅析大数据时代信息安全面临的挑战与机遇[EB/OL].[2016-12-03].<http://news.idcquan.com/news/101009.shtml>.
- [14] 李昊远.运用大数据技术增强国家意识形态安全[N].新华日报,2017-05-04(011).
- [15] 吴家庆,曾贤杰.大数据与意识形态安全[N].光明日报,2015-10-14(013).
- [16] 邹广文.技术时代的人文关怀[N].光明日报,2016-04-07(001).
- [17] 唐红丽,段丹洁.大数据时代媒体需要数据化表达[N].中国社会科学报,2016-07-29(001).

Social public opinion governance in the era of big data: What is possible and what can be done

Wang Shiyong

(College of Literature and Journalism, Chongqing Technology and Business University, Chongqing 400067)

Abstract: In the era of big data, the expression and governance of social public opinion presents a Complicated scenes. Network technology makes public opinion generation and expression digital; it can spread public opinion quickly and widely, to promote information interaction and sharing, relationship building and expansion, emotional transmission and infection; to make public opinion expression form diversified, participants diversified, public opinion development direction blurred. Big data technology makes it possible for computers to process massive amounts of information. The original silent, isolated, scattered and atomic data can be heard, which is an important opportunity for social public opinion governance. Big data technology can expand the field of social public opinion governance, improve the ability of social public opinion governance, and enrich the means of social public opinion governance. How to meet people's pursuit of more comprehensive and valuable data and information, and how to make data "voice" are the challenges faced by social public opinion governance in the context of big data, including thinking, effective data integration, mining and utilization, information security, etc.

Key Words: Big data; social public opinion; social governance; network society; mirror world

(责任编辑:易晓艳)