

大数据、人工智能和 网络舆情治理

徐涵 著

非外借



WUHAN UNIVERSITY PRESS
武汉大学出版社

大数据、人工智能和 网络舆情治理

徐涵 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

大数据、人工智能和网络舆情治理/徐涵著. —武汉: 武汉大学出版社, 2018. 10

ISBN 978-7-307-20531-4

I. 大… II. 徐… III. ①数据处理 ②人工智能 ③互联网络—舆论—管理 IV. ①TP274 ②TP18 ③G219

中国版本图书馆 CIP 数据核字(2018)第 210159 号

责任编辑:杨晓露

责任校对:汪欣怡

版式设计:汪冰滢

出版发行: **武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷:北京虎彩文化传播有限公司

开本:720×1000 1/16 印张:17 字数:245千字 插页:1

版次:2018年10月第1版 2018年10月第1次印刷

ISBN 978-7-307-20531-4 定价:49.00元

版权所有,不得翻印;凡购我社的图书,如有质量问题,请与当地图书销售部门联系调换。

前 言

网络舆情的监测与治理是一项十分繁杂而系统的工程，同时涉及自然科学与社会科学，尤其涉及计算机科学、信息科学、语言学、传播学、社会学、管理学和心理学等多学科的交叉。针对网络舆情的研究具有很强的生命力，是当代学科新的增长点，对国民经济和社会发展具有重要的推动作用。

本书内容涉及大数据、人工智能与网络舆情的监测和治理。全书共分为5章，结构如下：第1章为绪论，阐述分析关于网络舆情的研究背景和意义，介绍大数据和人工智能在我国发展的现状，并进一步引出本书的主题，即研究基于大数据和人工智能的网络舆情分析。第2章主要介绍大数据的概念、大数据系统下批量处理数据、流式处理数据、交互处理数据和图数据四种不同形式数据的突出特征和各自的典型应用场景以及相应的代表性处理系统；介绍有关物联网的相关概念和技术；介绍云计算的概念，分析云计算相关技术，并且对云计算将面临的机遇和挑战进行分析。第3章从技术的角度，介绍人工智能发展简史、人工智能领域经常提到的几大大概念、机器学习相关算法、数据挖掘的概念及相关技术。第4章介绍网络舆情的分类、特点、传播阶段、传播特征、传播要素和传播模型，并综述分析借助于大数据、人工智能及数据挖掘实现网络舆情分析的自动化、智能化、精准化的手段和路径。第5章则通过分析两个具体案例，论证利用大数据和人工智能实现网络舆情治理的实效。

本书是一本主要给社会科学研究人员了解大数据及人工智能技术前沿和基础工作、参考如何将相关技术应用到网络舆情研究的书籍。虽然

本书从计算机学科的角度出发，介绍相关技术对网络舆情的影响，但考虑到读者层面的诸多因素，全书中，作者试图尽可能少地使用数学公式，此外书中涉及有关大数据与人工智能技术相关的计算机名词、概念、算法等，均没有作太深入的解释。对这一部分感兴趣的读者，可以参考相关数学或者计算机科学书籍。

书中不可避免地涉及大量外国研究人员名字、机构或概念。如果全部翻译成中文，则读者在日后需要进一步查阅相关文献资料时，可能会出现困惑或者陌生感，不利于进一步学习，因此，本书仅对一般已有相对固定翻译、大众耳熟能详的名称，如“图灵”“冯·诺依曼”等直接使用，其他人名或概念，则在翻译后保留英语拼写，或者直接使用英语名称。

书中各章末列出了主要的参考文献，在此对索引参考文献中的作者和出版机构表示感谢。

湖北大学知行学院的汪玉光老师、江汉大学外国语学院的刘畅老师对本书的初稿提出了很诚恳的意见和建议；作者指导的4位研究生也对本书提供了一定的帮助；作者在研究过程中，得到了华中科技大学新闻与信息传播学院领导和同事的大力支持；本书的出版，得到了武汉大学出版社尤其是王金龙老师和杨晓露编辑的大力帮助和支持，在此一并表示衷心的感谢。

本书为国家社会科学基金项目“基于大数据及人工智能技术的新时代网络舆情监测研究”（18CXW028）的阶段性成果。

大数据、人工智能、网络舆情相关学科发展极为迅速，方兴未艾，很少有人在上述学科领域内均有深刻理解与造诣。作者自认才学疏浅，水平有限，更由于时间及精力所限，书中难免有错误和疏漏之处，恳请读者批评指正，若能不吝赐教，更将不胜感激。

徐 涵

2018年7月于喻家山

目 录

第1章 概论	1
1.1 网络舆情的研究背景	1
1.2 大数据、人工智能和网络舆情	10
第2章 大数据、云计算和物联网	18
2.1 数据存储	18
2.1.1 二进制 (binary)	18
2.1.2 八进制 (octonary) 和十六进制 (hexadecimal)	20
2.1.3 进制转换	21
2.1.4 数据的存储单位	24
2.1.5 形象理解数据有多少	26
2.2 大数据	27
2.2.1 大数据起源及相关研究	27
2.2.2 大数据的概念及特征	30
2.2.3 大数据关键技术	33
2.2.4 大数据处理系统	39
2.2.5 大数据技术面临的挑战	55
2.2.6 大数据发展趋势	57
2.3 物联网	61
2.3.1 物联网的概念	64
2.3.2 物联网的系统架构	66
2.3.3 物联网的异构融合网络层	69

2.3.4	感知节点及终端	70
2.3.5	物联网关键技术	71
2.3.6	物联网面临的问题及发展理念	74
2.4	云计算	77
2.4.1	云计算的概念	78
2.4.2	云计算的交付模式和部署模式	80
2.4.3	云计算的特征	82
2.4.4	云计算的关键技术	83
2.4.5	云计算的机遇与挑战	87
2.5	本章小结	89
第3章 人工智能与数据挖掘		100
3.1	人工智能	101
3.1.1	广义人工智能	101
3.1.2	狭义人工智能	104
3.1.3	人工智能发展简史	107
3.1.4	人工智能学派简介	114
3.1.5	总体架构	116
3.1.6	研究内容与应用领域	116
3.2	机器学习	126
3.2.1	概念	126
3.2.2	演化	127
3.2.3	算法	128
3.3	人工神经网络	134
3.3.1	感知机	136
3.3.2	两层神经网络	138
3.3.3	效果	140
3.4	深度学习	142
3.4.1	深度学习的概念和特点	143

3.4.2	深度学习的基本思想	145
3.4.3	卷积神经网络 (Convolutional Neural Network, CNN)	148
3.4.4	循环神经网络 (Recurrent Neural Network, RNN)	149
3.4.5	长短时记忆神经网络 (Long Short-Term Memory, LSTM)	151
3.4.6	深度学习应用	152
3.5	数据挖掘	153
3.5.1	聚类分析 (cluster analysis)	154
3.5.2	预测建模 (prediction mode)	158
3.5.3	关联分析 (association analysis)	160
3.5.4	异常检测 (anomaly detection)	162
3.5.5	数据挖掘的发展趋势	163
3.6	本章小结	164
第4章	基于大数据和人工智能的网络舆情分析	170
4.1	舆情和网络舆情	171
4.1.1	基本概念	171
4.1.2	相近词义辨析	174
4.1.3	国外学术界对于网络舆情的研究	176
4.1.4	网络舆情要素	178
4.1.5	网络舆情传播阶段与传播模型	180
4.1.6	网络舆情分类	183
4.1.7	网络舆情特点	185
4.1.8	网络舆情传播特征	187
4.1.9	网络舆情传播要素	192
4.1.10	网络舆情传播模型	194
4.2	大数据与网络舆情	200
4.3	人工智能与网络舆情	210

4.3.1 舆情热点发现技术	210
4.3.2 舆情情感分析	215
4.3.3 舆情分析与处理	218
4.3.4 舆情预测	219
4.3.5 其他相关研究	224
本章小结.....	225
第5章 案例分析.....	241
5.1 空姐滴滴打车遇害案	241
5.1.1 事件简介	241
5.1.2 传播分析	242
5.1.3 情感倾向分析	248
5.1.4 社会情感	253
5.2 “罗一笑”事件	254
5.2.1 事件简介	254
5.2.2 传播分析	255
5.2.3 情感倾向分析	258
5.3 对策建议	264

第 1 章 概 论

1.1 网络舆情的研究背景

随着计算机技术、无线通信技术的飞速发展，互联网进入了一个全新的时代，成为了人们生活中不可缺少的组成部分。移动终端的方便快捷，使得信息的获取更为容易，且传播速度更快。与之伴随而来的，是数据的爆炸式增长。根据“*We Are Social*”和“*Hootsuite*”团队 2018 年 1 月发布的最新研究报告^[1]，截至报告发布时，全球互联网用户总数已经突破了 40 亿大关，证实了全球有一半的人口“触网”。由于网络高速地发展，全球各行各业领域数据量都飞速地增长，比如互联网、医疗业、物联网等，人们在日常生活和工作学习中制造了大量的数据，数据几乎呈现指数级增加。当今全世界一年产生的数据量甚至超过了历史上几万年来产生的数据量的总和。由此，有研究预测在不久的将来全世界的数据数量会很快地增长，到 2025 年，全世界的数据量预计会增加到 163ZB 左右^[2]。社会和民众的关注点已经转向“大数据”。未来世界上的人们还不知道会制造出何等巨大的数据量，但是可确认人类世界已经进入了“大数据”时代^[3]。绝大多数的商业数据库和历史数据库最早通常只是用来储存数据，不允许用户随意访问和提取数据。直至 20 世纪 90 年代，大型数据存储公司开始考虑为什么用户都想要从他们的数据库中提取数据这一问题，人们才开始研究应该如何开展对大数据的处理^[4]。所以，对海量数据进行合理的挖掘和运用可有效改变我们的

生活和工作方式，提高我们的决策效率，同时也是我们创造更多财富的保障。

与此同时，中国的互联网技术飞速进步，移动互联网的普及率也快速增长，尤其是近几年中国网民规模逐渐上升，使得互联网使用率也有大幅增加。同样也是在2018年1月，中国互联网络信息中心（CNNIC）发布了第41次《中国互联网络发展状况统计报告》^[5]。数据表明，截至2017年12月，我国网民规模达7.72亿，普及率达到55.8%（图1-1），超过全球平均水平（51.7%）4.1个百分点，超过亚洲平均水平（46.7%）9.1个百分点。全年共计新增网民4074万人（图1-1），增长率为5.6%，我国网民规模继续保持平稳增长，连续11年位居全球第一。互联网商业模式不断创新、线上线下服务融合加速以及公共服务线上化步伐加快，成为网民规模增长的极大推动力。信息化服务快速普及、网络扶贫大力开展、公共服务水平显著提升，让广大人民群众在共享互联网发展成果上拥有了更多获得感。报告还指出，我国手机网民规模达7.53亿，网民中使用手机上网人群的占比由2016年的95.1%提升至2017年的97.5%（图1-2）；与此同时，使用电视上网的网民比例也提高了3.2个百分点，达28.2%；台式电脑、笔记本电脑、平板电脑的使用率均出现下降，手机不断挤占其他个人上网设备的使用。以手机为中心的智能设备，成为“万物互联”的基础，车联网、智能家电促进“住行”体验升级，构筑个性化、智能化应用场景。移动互联网服务场景不断丰富、移动终端规模加速提升、移动数据量持续扩大，为移动互联网产业创造了更多价值挖掘空间。

由于智能手机能够随时随地连接网络，此外还有更丰富的移动互联网应用的出现，移动互联网用户持续高速增长、推动移动互联网流量持续高速增长。2017年1月至11月，移动互联网接入流量消费累计达212.1亿G比特，比上年同期累计增长158.2%，而中国的网民每个人平均每周花在网上的时间已经长达27小时，这相当于每位网民每天花费约4个小时上网^[5]。

如此多的互联网用户，加之如此长的接触互联网的时间，使得网络

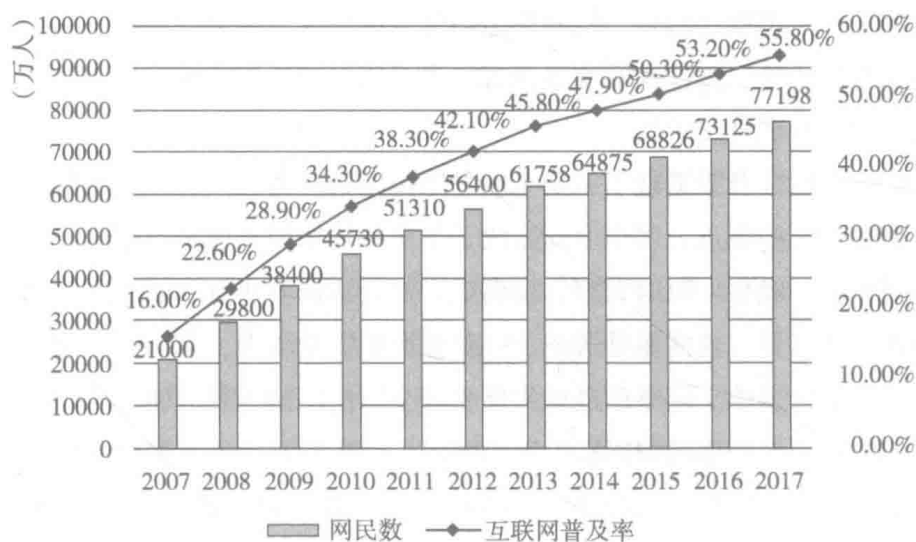


图 1-1 中国网民规模和互联网普及率

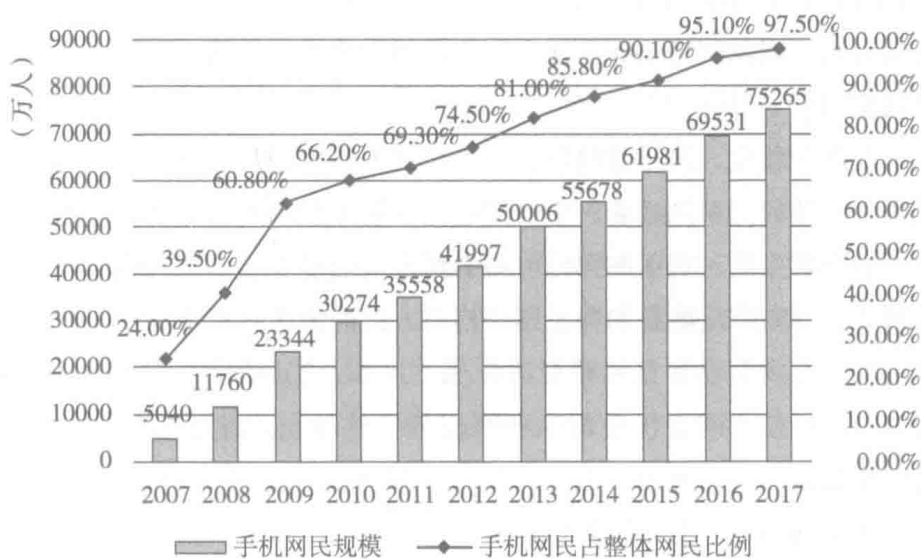


图 1-2 中国手机网民规模及其占网民比例

已经成为人们不可或缺的生活必需品，影响着人们生活的方方面面，是人们相互交流、表达意见的主要渠道。网络被公认为是继报纸、广播、

电视之后的“第四媒体”，民众知情权、表达权、参与权、监督权在互联网上已基本得到落实。网络媒体与传统媒体的不同之处在于，网络媒体具有发散性、开放性和交融性等特点的同时，近年来还研发了一些新的社交工具和社交手段。

这一时代背景造就了一个全新的网络舆情环境，人们随时随地都可以表达自己的观点。网络由此变成了孕育和传播舆论的发源地之一。网民对企业、民生、政府管理、反腐败、社会道德等热点问题在互联网上踊跃发表意见，这些意见形成了一种强大的舆论压力，其影响已经大大超过了传统媒体。网络已经成为反映社会舆情的最主要载体。

然而，在网络舆情迅猛发展的同时，信息的真伪并存使得网络舆情充满了不确定性和不可控性。网络具有的开放性和相对自由的宽松度，使得民众发言摆脱了社会权利体制的管制和限制，可以畅所欲言无所顾忌地表达个人的观点、立场、情绪，民意表达更为畅通。网络也由于其虚拟性带来了很大的安全隐患，发言者身份隐蔽，并且缺少规则限制和有效监督，因此网络很容易成为一些网民发泄不良情绪或者传递虚假信息的空间。

互联网时代，网络舆情传播的速度更加快捷，影响范围也更加广泛和深刻。任何关系到社会民生的问题都可能爆发较大的舆情，影响社会个人的正常生活，负面的舆论可能会使人们的思维或者情绪朝着坏的方向转变，而富有正能量的舆论则可指引人们的情绪或者态度往积极的方向发展，起到引导和宣传教育的作用。比如，2018年7月下旬，长春长生生物科技有限公司生产假疫苗事件曝光，经过社交网络的传播，在短短一周时间里，各类型的网络媒体报道量就已经达到573000次，社交网络上相关内容的转载、评论更是不计其数。该网络舆情最终引起了党和国家领导人的高度重视，习近平总书记和李克强总理专门作出批示，要求解决相关问题。

同时，我们也认识到网络舆情传播具有巨大的影响力，应该要采取合适的措施，降低负面、消极舆情的危害。消极、不真实的言论和评论有可能会危害到社会的安全，很多不法用户在网络中利用这种网络传播

方式的便捷性，进行不实新闻的传播、虚假广告的宣传甚至利用煽动性的言语从事非法活动等。这些网络中的信息和舆论如果不能及时地被制止和处理掉，就会像病毒一样迅速地扩散，影响人们的日常工作和生活。

在这样一个新时代，对网络舆情的分析监控治理工作显得愈发重要。

党的十八大以来，以习近平同志为核心的党中央高度重视党的新闻舆论工作。从2013年出席全国宣传思想工作会发表重要讲话，到2015年底视察解放军报社，再到2018年全国两会，习近平总书记多次对新闻舆论工作作出重要部署。以下为十八大以来，习近平总书记在多种场合关于“新闻舆论工作”话题的论述摘编。

谈新闻舆论工作的职责和使命：

做好党的新闻舆论工作，事关旗帜和道路，事关贯彻落实党的理论和路线方针政策，事关顺利推进党和国家各项事业，事关全党全国各族人民凝聚力和向心力，事关党和国家前途命运。

在新的时代条件下，党的新闻舆论工作的职责和使命是：高举旗帜、引领导向，围绕中心、服务大局，团结人民、鼓舞士气，成风化人、凝心聚力，澄清谬误、明辨是非，联接中外、沟通世界。要承担起这个职责和使命，必须把政治方向摆在第一位，牢牢坚持党性原则，牢牢坚持马克思主义新闻观，牢牢坚持正确舆论导向，牢牢坚持正面宣传为主。

——2016年2月19日，习近平在北京主持召开党的新闻舆论工作座谈会上强调。

谈新闻工作的舆论导向作用：

新闻舆论工作各个方面、各个环节都要坚持正确舆论导向。各级党报党刊、电台电视台要讲导向，都市类报刊、新媒体也要讲导向；新闻报道要讲导向，副刊、专题节目、广告宣传也要讲导向；时政新闻要讲导向，娱乐类、社会类新闻也要讲导向；国内新闻报

道要讲导向，国际新闻报道也要讲导向。团结稳定鼓劲、正面宣传为主，是党的新闻舆论工作必须遵循的基本方针。

——2016年2月19日，习近平在北京主持召开党的新闻舆论工作座谈会上强调。

做好网上舆论工作是一项长期任务，要创新改进网上宣传，运用网络传播规律，弘扬主旋律，激发正能量，大力培育和践行社会主义核心价值观，把握好网上舆论引导的时、度、效，使网络空间晴朗起来。

——2014年2月27日，习近平主持召开中央网络安全和信息化领导小组第一次会议时提出。

对中国人民和中华民族的优秀文化和光荣历史，要加大正面宣传力度，通过学校教育、理论研究、历史研究、影视作品、文学作品等多种方式，加强爱国主义、集体主义、社会主义教育，引导我国人民树立和坚持正确的历史观、民族观、国家观、文化观，增强做中国人的骨气和底气。

——2013年12月30日，习近平在中央政治局就提高国家文化软实力研究进行第十二次集体学习时指出。

我们正在进行具有许多新的历史特点的伟大斗争，面临的挑战和困难前所未有，必须坚持巩固壮大主流思想舆论，弘扬主旋律，传播正能量，激发全社会团结奋进的强大力量。关键是要提高质量和水平，把握好时、度、效，增强吸引力和感染力，让群众爱听爱看、产生共鸣，充分发挥正面宣传鼓舞人、激励人的作用。

——2013年8月19日，习近平在全国宣传思想工作会议上发表重要讲话。

谈新闻舆论工作坚持党性原则：

党的新闻舆论工作坚持党性原则，最根本的是坚持党对新闻舆论工作的领导。党和政府主办的媒体是党和政府的宣传阵地，必须姓党。党的新闻舆论媒体的所有工作，都要体现党的意志、反映党的主张，维护党中央权威、维护党的团结，做到爱党、护党、为党；都要增强看齐意识，在思想上政治上行动上同党中央保持高度一致；都要坚持党性和人民性相统一，把党的理论和路线方针政策变成人民群众的自觉行动，及时把人民群众创造的经验 and 面临的实际情况反映出来，丰富人民精神世界，增强人民精神力量。

加强和改善党对新闻舆论工作的领导，是新闻舆论工作顺利健康发展的根本保证。各级党委要自觉承担起政治责任和领导责任。领导干部要增强同媒体打交道的能力，善于运用媒体宣讲政策主张、了解社情民意、发现矛盾问题、引导社会情绪、动员人民群众、推动实际工作。

——2016年2月19日，习近平在北京主持召开党的新闻舆论工作座谈会上强调。

要坚持党管媒体原则，严格落实政治家办报要求，确保新闻宣传工作的领导权始终掌握在对党忠诚可靠的人手中。

——2015年12月25日，习近平视察解放军报社时指出。

党性和人民性从来都是一致的、统一的。坚持党性，核心就是坚持正确政治方向，站稳政治立场，坚定宣传党的理论和路线方针政策，坚定宣传中央重大工作部署，坚定宣传中央关于形势的重大分析判断，坚决同党中央保持高度一致，坚决维护中央权威。所有宣传思想部门和单位，所有宣传思想战线上的党员、干部都要旗帜鲜明坚持党性原则。

——2013年8月19日，习近平在全国宣传思想工作会议上发表重要讲话强调。

谈新闻舆论工作必须创新：

随着形势发展，党的新闻舆论工作必须创新理念、内容、体裁、形式、方法、手段、业态、体制、机制，增强针对性和实效性。要适应分众化、差异化传播趋势，加快构建舆论引导新格局。要推动融合发展，主动借助新媒体传播优势。要抓住时机、把握节奏、讲究策略，从时度效着力，体现时度效要求。

——2016年2月19日，习近平在北京主持召开党的新闻舆论工作座谈会上强调。

现在，媒体格局、舆论生态、受众对象、传播技术都在发生深刻变化，特别是互联网正在媒体领域催发一场前所未有的变革。读者在哪里，受众在哪里，宣传报道的触角就要伸向哪里，宣传思想工作的着力点和落脚点就要放在哪里。要顺应互联网发展大势，勇于创新、勇于变革，利用互联网特点和优势，推进理念、内容、手段、体制机制等全方位创新，努力实现军事媒体创新发展。要研究把握现代新闻传播规律和新兴媒体发展规律，强化互联网思维和一体化发展理念，推动各种媒介资源、生产要素有效整合，推动信息内容、技术应用、平台终端、人才队伍共享融通。

——2015年12月25日，习近平视察解放军报社时指出。

推动传统媒体和新兴媒体融合发展，要遵循新闻传播规律和新兴媒体发展规律，强化互联网思维，坚持传统媒体和新兴媒体优势互补、一体发展，坚持先进技术为支撑、内容建设为根本，推动传统媒体和新兴媒体在内容、渠道、平台、经营、管理等方面的深度融合，着力打造一批形态多样、手段先进、具有竞争力的新型主流媒体，建成几家拥有强大实力和传播力、公信力、影响力的新型媒体集团，形成立体多样、融合发展的现代传播体系。

——2014年8月18日，习近平主持召开中央全面深化改革领导小组第