

Social Network

社会网络

Big Data

大数据 下

企业舆情建模和管理

Enterprise Public Opinion

Management

梁循 杨小平 李志宇 编著

清华大学出版社



Social Network

社会网络

Big Data

Social Network
大数据下

企业舆情建模和管理

Enterprise Public Opinion

Management

梁循 杨小平 李志宇 编著

清华大学出版社

北京

内 容 简 介

本书综合了国内外的最新资料和作者的研究成果。通过研究社会网络上用户的行为理论,探索适用于社会网络大数据环境下的企业舆情挖掘方法,提出若干个典型的企业舆情发现与合理处置方法。首先提出形式化定义,为后续定量描述用户行为奠定基础,然后进一步应用到企业课题,并研究企业舆情管理优化问题。在基础模型方面,本书研究了基于文本与图像内容的企业舆情模型和基于网络结构的发现模型。在衍生模型方面本书探讨了社会网络用户行为和企业舆情管理优化方法的一些具体课题。

本书的读者可以是对社会计算感兴趣的专业人士,或是对社会化媒体挖掘感兴趣的商业界人士,也可作为计算机应用方向的教材或参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。
版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

社会网络大数据下企业舆情建模和管理/梁循,杨小平,李志宇编著.--北京:清华大学出版社,2016

ISBN 978-7-302-41054-6

I. ①社… II. ①梁… ②杨… ③李… III. ①企业管理—公共关系—舆论—研究 IV. ①F270

中国版本图书馆 CIP 数据核字(2015)第 173423 号

责任编辑:刘向威 李 晔

封面设计:文 静

责任校对:李建庄

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 刷 者:三河市君旺印务有限公司

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:170mm×230mm 印 张:11.75 字 数:194 千字

版 次:2016 年 2 月第 1 版 印 次:2016 年 2 月第 1 次印刷

印 数:1~2000

定 价:25.00 元

本书讨论了在社会网络大数据环境下,企业舆情的建模和管理问题。本书与作者先前出版的另外 11 本书籍《网络金融》、《数据挖掘算法与应用》、《互联网金融信息系统的设计与实现》、《电子商务理论与实践》、《网络金融信息挖掘导论》、《网络金融系统设计与实现案例集》、《互联网金融信息智能挖掘基础》、《支持向量机算法及其金融应用》、《金融数据挖掘——基于大数据视角的展望》、《面向社会化媒体大数据的社会计算》、《社会化商务理论与实践》和本书之间的关系见图 0-1。

本书的编写得到了中国人民大学科学研究基金项目(10XNI029)的支持。作者的一些学生也参加了本书的编写,这些同学包括张海燕、申华、施晓菁、马跃峰、马超等。

由于作者水平和时间的限制,书中一定存在不少缺点和错误,恳请读者批评指正。

编者

2015 年 12 月

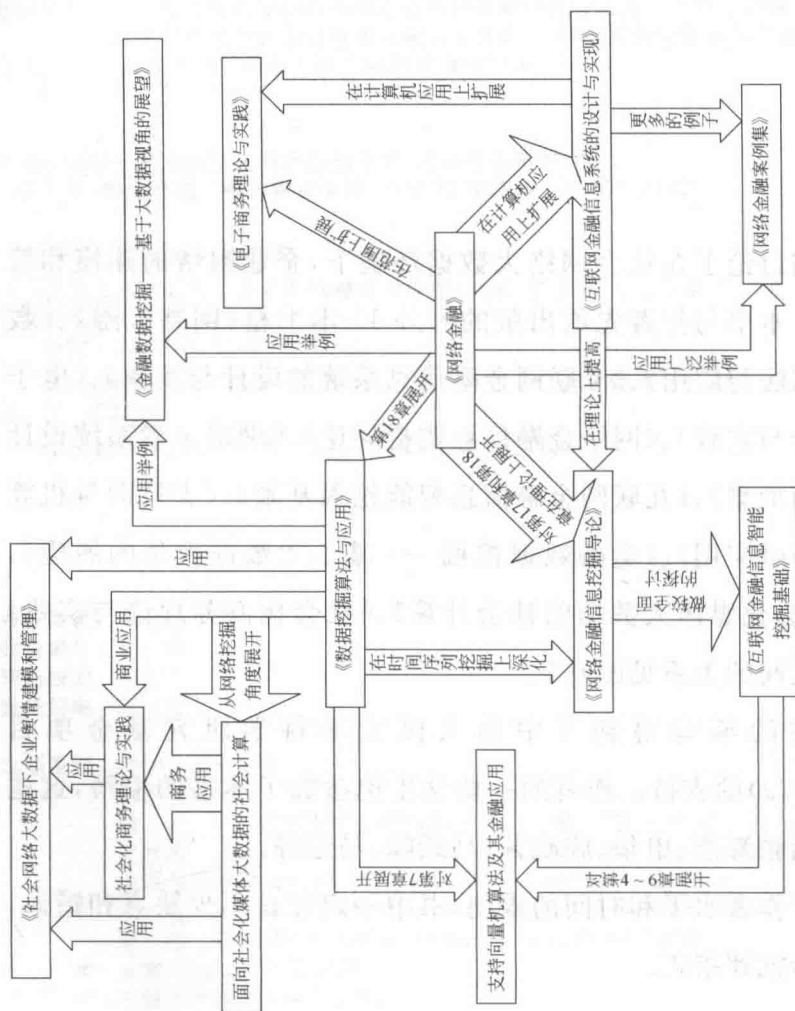


图 0-1 《网络金融》、《数据挖掘算法与应用》、《互联网金融信息挖掘导论》、《电子商务理论与实践》、《网络金融信息智能挖掘基础》、《支持向量机算法及其金融应用》、《金融数据挖掘——基于大数据视角的展望》、《面向社会化媒体大数据的社会计算》、《社会化商务理论与实践》和本书之间的关系

第 1 章 绪论	1
1.1 社会网络	1
1.2 社会网络大数据	4
1.3 社会网络计算	6
1.4 舆情、网络舆情与企业网络舆情	9
1.5 企业社会网络舆情的特点	12
1.6 企业社会网络舆情与国家层面舆情的联系与区别	14
1.7 企业社会网络舆情的研究意义	17
1.8 本章小结	20
思考题	20
第 2 章 互联网舆情分析的主要技术	21
2.1 引言	21
2.2 舆情信息抽取	23
2.3 关键词提取	25

2.4	摘要提取	26
2.5	文本倾向性分析	28
2.6	关联分析技术	31
2.7	主题检测和追踪	33
2.8	舆情热点发现和监测	38
2.9	本章小结	42
	思考题	43
第3章	社会网络中的用户行为	44
3.1	引言	44
3.2	基于用户行为的社区网络	46
3.3	社会网络中的“社交圈”与“兴趣圈”	49
3.4	社会网络用户的行为	51
3.5	本章小结	56
	思考题	57
第4章	企业网络舆情管理的模型	58
4.1	引言	58
4.2	企业在线舆情的分析预警管理模型	61
4.3	企业在线舆情的干预处置管理模型	65
4.4	本章小结	71
	思考题	71

第 5 章 数据平台和系统结构	72
5.1 数据获取	72
5.2 数据平台	75
5.3 系统结构	78
5.4 本章小结	80
思考题	81
第 6 章 企业网络舆情管理的计算机技术	82
6.1 基于文本内容的企业网络舆情管理的技术	82
6.2 基于图像内容的企业网络舆情管理的技术	100
6.3 本章小结	103
思考题	104
第 7 章 面向企业网络舆情的社会网络信誉及营销管理	105
7.1 面向企业网络舆情的社会网络信誉 平台构建方法	105
7.2 基于内容和交易网络结构的信任测度	108
7.3 基于企业网络舆情分析的企业网络 营销管理方法	109
7.4 本章小结	112
思考题	112

第 8 章 面向企业社会网络舆情管理的用户行为理论	113
8.1 引言	113
8.2 基础模型	116
8.3 衍生模型	127
8.4 本章小结	133
思考题	134
第 9 章 社会网络舆情大数据的分解算法	135
9.1 问题的环境和解决问题的思路及框架	135
9.2 大数据的分解模型	137
9.3 网络舆情大数据面临的挑战	140
9.4 网络舆情大数据发展方向的展望	142
9.5 本章小结	152
思考题	152
第 10 章 企业社会网络舆情管理方法	153
10.1 社会网络下 C2B 营销的实现及其对企业 业绩的影响	153
10.2 企业的社会网络个性化信息推荐	155
10.3 社会网络大数据环境下企业的开放式 信用管理	156

10.4 社会网络大数据环境下针对在线舆情服务 挽回管理措施对企业绩效影响评估	157
10.5 社会网络大数据环境下企业舆情管理方法 及其对在线舆情的影响	158
10.6 本章小结	159
思考题	160
第 11 章 展望	161
11.1 企业社会网络舆情给企业管理的挑战	161
11.2 企业社会网络大数据舆情管理的应对策略	166
11.3 本章小结	168
思考题	168
参考文献	169

第1章

绪 论

本章学习目标

- 学习社会网络、社会网络大数据、网络舆情与企业网络舆情的概念。
- 理解社会网络计算的研究内容。
- 了解企业社会网络舆情的特点。

1.1 社会网络

近年来,在新的功能、技术和标准的推动下,网络变得更加社会化和互联化,社交性网络平台得到了前所未有的迅猛发展。这些网络平台(包括移动网络平台),如,Facebook、Twitter、新浪微

博、微信、网易新闻等不仅聚集了大量用户,也为企业提供了获取巨大潜在客户资源的渠道。据美国调查公司 Unity Marketing 调查发现,社交网络用户数量已经占据了互联网用户总量的 40%。

社会网络(social network)是指社会个体成员之间因为互动而形成的相对稳定的关系体系。在社会网络中,人们形成了“社交圈”、“兴趣圈”等关系。从狭义上来说,社交圈是我们日常生活中与朋友、同学、同事之间的各种关系网络构成的一个人际圈子。从广义上来说,社交圈可以延伸为我们每个人的生活圈。

社交圈的存在让营销者更为欣慰。以人人网、Facebook 为例,大多数的注册会员都是来自于各个高校的学生,即便日后工作生活变迁,但是这些稳固的同学关系却不会消失。同样,在以白领为主要客户群体的开心网上,同事间的关系虽然相对于同学来说因为变化更大而淡薄一些,但是依然是一种相对稳固的人际关系(Poyry,2013)。互联网社交圈的建立与发展还取决于社交网站用户的交互方式(Backstrom,2006)。社交圈中不同身份的人的影响力也不尽相同(Stutzman,2006)。

社交圈对于企业进行网络营销来说无疑有着积极的促进作用,无论是品牌的创建或者是促销的推进,在稳定网络中的传播广度与深度,都会比大众网络更有优势。

社交圈的分享也可以分为两种:一种是消费者主动的分享;一种是营销者促使的被动分享。被动分享一般都是由网络营销者来促成的,他们往往会针对消费者、潜在客户的商品促销信息、订

单信息的分享、发布行为给予一定的奖励,以这种激励方式达到自己的促销、品牌传播目的。

兴趣圈从理论上讲,其范畴应该要比社交圈更大一些。相对于社交网站的社交圈为主的状况,微博、群组、知识分享平台、视频图片网站、团购、LBS 服务等,大多都是基于兴趣形成的社会化的关系网络。当然,广义上来说这种因为兴趣而产生的关系网络也属于社交范畴,但是单纯从社会网络的类别、数量上来区分的话,因为兴趣而形成的社会网络应用种类更多、品种更齐全。

常见的社会网络中,除了社交网站之外,还存在大量的其他网站。比如维基百科、百度百科为代表的知识分享平台,以 Twitter、新浪微博、腾讯微博为代表的微博平台,以 YouTube、优酷为代表的视频分享网站,以 Flickr 为代表的视频分享网站,以豆瓣、百度文库为代表的文档分享网站,以大众点评为代表的消费评论网站,甚至包括各种团购网站,以及以微信为代表的移动用户端上的平台,其实都属于兴趣圈网络的范畴。

以兴趣圈形成的网络具备一些特殊的特性,那就是社会网络的自优化特性。以最典型的新浪微博为例,对于一个商家或者公众意见领袖来说,普通的网民与之形成的关注同时是不相互关注的关系,从本质上来说并非是社交关系。同样地,这种大多数时候的单向互动也完全不符合社会网络的互动特性,而只是类似于更多的媒体中心进行的单向信息传递而已(Bouras,2004)。

对于商家来说,交互从来也都不是充分的,而这正是社交圈之

所以存在的原因和价值。社交圈可以分为陌生人购物分享网络,例如蘑菇街、美丽说,以及熟人圈社交网络例如微博、人人网等。对于商家,他们更关注的是这些人是否真正地对他们的商家和产品感兴趣,而对于每一个用户的深度交流,只能作为一个远景却未必能实现。而社会网络活动中的自优化功能会帮助商家找寻到最忠实的用户,也就是长期留存下来的稳定网络。

基于社会计算的领域视角,本章通过一种以 Web 2.0 思想为核心的社会计算模式对社会网络进行剖析。因为社会计算及社会网络都是相对较新颖的概念。本小节重点对社会计算和社会网络的相关概念及研究成果做出较为完善的阐述。

1.2 社会网络大数据

在实践中,社会网络的数据量非常大,形成了大数据。大数据指的是所涉及的资料量规模巨大,大数据这个术语最早期的引用可追溯到 apache org 的开源项目 Nutch。当时,大数据用来描述为更新网络搜索索引需要同时进行批量处理或分析的大量数据集。随着谷歌 MapReduce 和 Google File System(GFS)的发布,大数据不再仅仅用来描述大量的数据,还涵盖了处理数据的速度。

最早提出大数据时代到来的是全球著名管理咨询公司麦肯锡,麦肯锡称:“数据,已经渗透到当今每一个行业和业务职能领域,成为重要的生产因素。人们对于海量数据的挖掘和运用,预示

着新一波生产率增长和消费者盈余浪潮的到来。”

大数据一词由英文“Big Data”翻译而来,过去常说的“信息爆炸”、“海量数据”等已不足以描述这个新事物。全球著名管理咨询公司麦肯锡的报告《大数据:创新、竞争和生产力的下一个前沿》对大数据做了如下定义:大数据是指大小超出了传统数据库软件工具的抓取、存储、管理和分析能力的数据群。

大数据在物理学、生物学、环境生态学等领域以及军事、金融、通信等行业存在已有时日,近年来,随着互联网和信息行业的发展,大数据引起了更多人的关注。大数据在互联网行业指的是这样一种现象:互联网公司在日常运营中生成、累积的用户网络行为数据。这些数据的规模是如此庞大,以至于不能用 GB 或 TB 来衡量,大数据的起始计量单位至少是 PB(1000 个 TB)、EB(100 万个 TB)或 ZB(10 亿个 TB)。

在信息技术不发达的年代,存储设备的价格昂贵,数据的保存所付出的代价是非常大的。随着科技的进步,存储设备便宜了,数据可以在较低的成本下得到妥善的保存,用户自己产生的数据得到了重视。也正是由于数据的价值被重视了,因此越来越多的数据被持续保存,大数据由此产生。网络数据的十几年的保存产生了大数据。

大数据的首要特征是数据量大。截至 2009 年,美国几乎所有部门中每个雇员数量在 1000 人以上的企业所存储的数据平均值至少为 200TB,是美国零售商沃尔玛 1999 年数据仓库的两倍。很

多经济部门中,每个企业平均存储数据超过 1PB。欧洲组织 2010 年存储容量总计接近 11EB,大约为整个美国存储容量的 70%。全球企业 2010 年在硬盘上存储了超过 7EB 的新数据,消费者在 PC 和笔记本电脑等设备上存储了超过 6EB 新数据,而 1EB 数据就相当于美国国会图书馆中存储数据的 4000 多倍。数据容量增长的速度大大超过了硬件技术的发展速度,以至于引发了数据存储和处理的危机。大量的数据会被处理掉,比如医疗卫生提供商会处理掉他们所产生的 90% 的数据(包括手术过程中产生的几乎所有实时视频图像)。

此外,大数据不只是大。海量数据引发的危机并不单纯是数据量的爆炸性增长,还牵涉到数据类型的改变,也称为多样化。原来的数据都可以用二维表结构存储在数据库中,如常用到 Excel 软件所处理的数据,称为结构化数据。但是现在,更多互联网多媒体应用的出现,使诸如图片、声音和视频等非结构化数据占到了很大比重。有统计显示,全世界结构化数据增长率大概是 32%,而非结构化数据增长率则为 63%。预计未来用于产生智慧的大数据,往往是这些非结构化数据。

1.3 社会网络计算

社会网络计算,也称社会计算,是对社会网络的智能计算。它作为一个新兴跨学科的研究领域,目前还没有一个公认的定义。

不过,我们可以从社会计算出现的背景去剖析概念,将社会计算概括为“用社会化方法计算社会”,具体包含两层意思,即“为社会计算”和“用社会化方法计算”。

所谓“为社会计算”反映了社会计算研究与服务的对象是社会,包括虚拟网络社会和现实社会,以及从中抽象出来的人工社会。从这个角度来说,通过信息技术方法对社会数字轨迹进行分析,了解社会已经发生,监控正在发生和预测将要发生的事情,准确地把握社会的动态特征和运行规律,预测政策实施的可行性,为虚拟网络社会的科学管理和政府决策提供有效参考。

所谓“社会化方法”就是以“草根”客户为中心,并依靠“草根”客户的方法,是一种协同和群体智能的方法,一种从个体到整体,从微观到宏观的思维模式。许多事件都是由无数网民的“你一言我一语”和微不足道的微观行为最终发展成为一个重大的社会事件或浩大工程。如维基百科就是由无数网民微不足道的努力而完成的巨大的百科全书,这靠少数专家是无法完成的。从这个角度来讲,社会计算是一种群体智能的计算模式(Liang, 2012; 梁循等, 2014)。可以看出,社会计算本质就是对社会网络进行智能分析的过程。

社会计算的研究对象是社会,包括现实的物理社会和虚拟的网络社会(Moreno, 2004; Cebi, 2013)。前者主要指传统意义上的社会,如某国家或地区;后者主要指基于 Web 的虚拟网络社区。从广义来讲,整个 Internet 就是一个虚拟网络,但从狭义来讲,虚