

Social Computing
Methods and Applications



社会计算的 基本方法与应用



王飞跃 李晓晨 毛文吉 王 涛◎著



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

Social Computing
Methods and Applications



社会计算的 基本方法与应用



王飞跃 李晓晨 毛文吉 王 涛◎著



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

图书在版编目 (CIP) 数据

社会计算的基本方法与应用 / 王飞跃等著. —杭州:
浙江大学出版社, 2012.12(2013.5 重印)

ISBN 978-7-308-11200-0

I. ①社… II. ①王… III. ①计算机应用—社会科学
—计算—研究 IV. ①C32

中国版本图书馆 CIP 数据核字 (2013) 第 029433 号

社会计算的基本方法与应用

王飞跃 李晓晨 毛文吉 王 涛 著

丛书策划 许佳颖

责任编辑 许佳颖 王元新

封面设计 俞亚彤

出版发行 浙江大学出版社

(杭州市天目山路 148 号 邮政编码 310007)

(网址: <http://www.zjupress.com>)

排 版 杭州中大图文设计有限公司

印 刷 杭州日报报业集团盛元印务有限公司

开 本 710mm×1000mm 1/16

印 张 17.5

字 数 282 千

版 印 次 2013 年 5 月第 2 版 2013 年 5 月第 2 次印刷

书 号 ISBN 978-7-308-11200-0

定 价 69.00 元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行部联系方式: 0571-88925591; <http://zjdxcbbs.tmall.com>

前 言

本书是我们团队出版的第一部关于社会计算的专著，是团队成员十余年来不懈追求与探索的结果。借此机会，我谨从个人角度就“社会计算”这一新兴交叉学科的发展，做一简要回顾。

自己关于社会计算的想法，源自对计算实验的认识，而对计算实验的想法又源自 20 世纪 80 年代初在浙江大学力学系读硕士研究生的经历。当时我的专业是断裂力学，导师为王仁东教授。王教授十分强调研究的试验环节，然而，30 年前断裂力学的试验成本相对较高，而且非常费时，就是一次最普通的微裂纹扩展拉伸压缩的试验，往往也需要几个月的时间。可我最初选择的微裂纹下材料疲劳寿命的概率可靠性研究，按要求至少需要进行上百次这种试验方能得到有意义的结论，这几乎是不可能实施的。无奈之下，我提出了结合物理试验和数学模型的“计算实验”设想，希望以此在短时间里低成本地完成实验要求，并写了一个简单的建议呈给导师（此建议原稿一直保留，后收录于 2004 年我的力学论文集《弹性力学与板壳理论研究》之中）。后来，由于导师因病去世，我不得不从断裂力学转入计算力学，后又改为从事板壳和弹性理论的研究。完成硕士学业后，计算实验的想法也就不了了之。

回想起来，我能够从 30 年前的纯工程性理论研究转到今天的社会科学与计算方法结合的交叉领域，硕士学习结束时读的一本书起了很大的作用，这就是库恩的《科学革命的结构》。这是我硕士论文答辩后读过的第一本书。库恩所揭示的学人的作用、范式的转移，再加上格式塔心理学的影响，让我感觉到科学研究不再是理想中科学家探索真理的纯净过程，而似乎是一部大师们借助真理制造范式，利用个人、学术、政治及社会许多过去的影响，创造、领导科学潮流的历史。库恩对我的冲击，是感性非哲学的，就是自幼想象中罩在“科学”二字上面的“神圣”光环突然消失，再也找不回以前那种说不清、道不明的科学“神圣”感。我开始感到，许多过去认为是客观的物理定律在发现时也有很大的主观成分影响，科学的研究也是诸多研究中的一种，而献身于科学也不过是诸多献身中的一类。库恩的书，最终促使我从理论研究工作转向工程研发及其应用，也使得我重新认识文学、历史、社会学，甚至世俗的人际关系学。就是这些感性的想法，引导我开始大量地阅读经济与社会

科学方向的书籍,最终有了今天的境况。

我攻读博士的专业是计算机和系统工程,而我博士毕业后做的第一件事却是申请再读经济学的博士,被拒后无奈才去了亚利桑那大学教书。20世纪90年代初,在我的第一位来自国内的研究生的鼓动之下,我开始了Day Trader的股票交易尝试。没想到,七年多“疯狂”般的交易,没有把我引上金融之道,但短时间内必须阅读大量经济、社会、政治、新闻等材料并实时决策的经历,却将自己引入了语言动力学(Linguistic Dynamic Systems, LDS)和情报与安全信息学(Intelligence and Security Informatics, ISI)的新兴研究领域。这段磨炼,对于我今天的研究至关重要,是我能够进入社会计算等交叉领域的物理和精神基础。

这段时期,我得益于与亚利桑那大学商学院的陈忻君教授(Hsin-chun Chen)和社会科学院的史蒂夫·兰森(J. Stephen Lansing)教授的交流与合作。忻君是ISI的主要创始者,而史蒂夫是自然、人和社会交叉研究的世界领军学者。特别是后来史蒂夫和亚利桑那大学的其他几位研究人类社会学的同事,去圣塔菲研究所(Santa Fe Institute, SFI)分别担任研究员和管理者后,又给了我同SFI研究复杂性的许多著名学者交流与学习的机会,其中花旗银行Arthur博士关于人工股市的研究引起了我很大的共鸣,也使我对自然科学、工程技术与社会科学的交叉有了更加深刻的认识。1999年,我在亚利桑那大学领衔成立复杂系统高等研究中心(The Program for Advanced Research in Complex Systems, PARCS)就是这些合作交流的直接结果。十分巧合的是,刚一回国,中国科学院自动化研究所的戴汝为院士也正有成立复杂性研究机构的设想,后来就演变成为2000年在中国科学院自动化所成立的“系统复杂性研究中心”,戴老师亲自担任首任主任,我是副主任。当时中心确立的主要方向之一就是经济、社会,特别是人口复杂系统的研究。

然而,促使我明确提出“社会计算”这一名词的直接原因是发生在2003年的三件事。一是美国“9·11”事件后,西方各国对反恐情报研究极为重视,陈忻君教授择机倡议召开ISI研讨会。第一届会议就在亚利桑那大学召开,除美国国家基金会NSF之外,美国中央情报局CIA、联邦调查局FBI,还有国防部等情报机构皆派人参加,作报告并提供资助。2004年,又在图森召开了第二届会议,情况与上一届基本一样。然而,对于情报二字,当时心里总有顾虑,内心不希望自己的名字与“情报”一词联系起来,加上会议又涉及许多情报机构的专业人士,就更加担心了。而且,会议上交流的论文大多专注于计算方法的应用,使我感到了ISI还缺少真正的科学内涵和自己的方法体系。因此,我希望能用

一个新的角度来审视 ISI。这些考虑,是我提出“社会计算”的主要原因,目的就是要使情报研究进一步“中性”,使社会舆情的分析等也成为其主要的內容之一。2005年,我与忻君合力,终于把 ISI 发展成 IEEE 的国际会议,而社会计算也成了会议的主题之一。二是 2003 年和 2004 年参加了中国科学院组织的几次关于建立科学发展观理论体系的内部讨论会,使我更加明确了以人工社会对社会问题进行建模,将计算机作为社会实验室进行试验,再把人工社会与实际社会并举互动,实现平行执行管理与控制的思路。这一认识是我研究断裂力学时计算实验想法的扩展,也是自适应控制方法的发展,一定意义上缓解了因科学发展观问题一般无法做实验所导致的根本性矛盾。这就是后来的 ACP(人工系统+计算实验+平行执行)方法,当时得到了许多与会领导和学者的共鸣与支持。三是当时中国科学院复杂系统与智能科学重点实验室未来发展方向的迫切需要。2003 年初,我代表实验室首次参加了科技部组织的每五年一次的工程类国家重点实验室评估,评审专家纷纷认为实验室定位不清,不知“复杂系统”是什么,质疑我们研究了“复杂科学”的哪些问题,而且认为我们哲学讨论多,具体方法少。苦思数月并与室内多位研究人员,特别是与戴汝为院士讨论之后,我们最终认定 ACP 是一个方向,而复杂社会问题是一个突破口。当时,我与戴老师正参加由国家计划生育委员会组织,蒋正华、宋健和徐匡迪等领导和学者主持的国家人口发展战略的重大研究项目,因此提出了与计划生育委员会合作建立社会计算实验平台的建议,并与蒋正华和戴汝为两位老师撰写了一篇利用人工人口系统进行复杂人口问题研究的文章。2003 年,我完成了“从一无所有到万象所归:人工社会与复杂系统研究”一文,投《科学时报》,呼吁大家关注这一研究方向。由于初稿过长,不适于报纸上发表,修改后还是太长,最后只好变相作为“广告”性质的文章于 2004 年春发表。在这篇文章中,我首次提出“计算社会学”的名词和概念,显然是受到过去从事计算力学工作的影响。在此文的影响下,我们开始组织人员撰写基于人工交通系统研究城市交通堵塞问题的“973 计划”立项书,并于次年在上海召开首届人工交通系统研讨会,主要就是讨论“973 计划”的立项问题。

“社会计算”一词和 ACP 方法的正式提出是在 2004 年,相关文章可见本书附表 2 所列的目录,分别发表在我自己创办、后由周孟初教授主编的国际智能控制与系统杂志(*International Journal of Intelligent Control and Systems*, IJICS)和几份中文期刊上。不用“计算社会学”而用“社会计算”,是为了消除有些从事计算研究的科研人员的顾虑,因为他们可能不希望涉入社会科学的研究,而“社会计算”一词比较中性,更

容易被看成是关于一类计算方法的研究。2005年春,正值中科院信息技术学部和技术学部在河南省南阳市联合举办《科学·技术·人文》论坛,我受邀做了《社会计算:科学·技术·人文》的报告,这是自己关于社会计算的首次正式报告。之后,《科学时报》(现《中国科学报》)又组织了华中科技大学原校长杨叔子院士与自己的一次访谈,给了我又一次通过报纸宣传社会计算的机会。

在此之前,当我与学生一起上网查询时,不论是百度还是谷歌,都没有中文“社会计算”作为一个研究术语出现的情况,但英文“Social Computing”一词早在1994年就有人用过。这曾一度使我十分沮丧,但找来原文一查,发现文中虽用了“Social Computing”一词,可一开头就解释“Social Computing”是“Social Software”,即社会软件之意,如电子邮件、计算机支持的协作 CSCW 等,并不是我们所指的“社会计算”,即广义而言的“面向社会科学的计算理论和方法”,狭义而言的“面向社会活动、社会过程、社会组织及其作用和效应的计算理论和方法”。这一发现,使我释然,最后决定还是以“社会计算”而非“计算社会学”来命名这一新的研究领域。

为什么提倡社会计算研究呢?

以农业时代向工业时代过渡为例。工业时代中的各种“人造过程”,如现代工厂里的各种物理化学反应过程,其实在农业时代里都存在,但却是自然的过程,其强度、速度或规模一般都比较小,一旦大了,有的就是“自然灾害”了。“人造”自然过程的强度、速度和规模要求我们必须以工业化的生产方式进行管理,因此过程控制系统(PCS)、集散控制系统(DCS)和企业资源规划(ERP)系统等,都成了当今工业时代必不可少的“人造过程”。因为再按过去农业时代的方式进行生产管理,就会产生“人造”自然过程的爆炸、起火、污染和资源的浪费,而新的工业化管理的基础就是现代物理学和其他现代科学。

目前,我们的社会正从工业时代向智业或知业时代过渡。由于因特网、社会媒体,特别是微博等的出现,催生了许多“人工过程”,如人肉搜索、团购、“秒杀”等。其实,这些社会现象过去都有,但在无网络的年代,这些过程的速度、强度、规模和影响力往往很小,一旦大了就可能是“社会动乱”。可以预见,未来的“人工”社会过程之强度、速度和规模将要求我们必须以科学化的社会管理方式来治理社会,所以未来的社会过程控制(SPC)系统、集散社会管理(DSM)系统和社会资源规划(SRP)系统等,将变得必不可少。因为再按过去的方式管理社会,“人工”社会过程的“爆炸”、“起火”、“污染”和资源的“浪费”将不可避免。

而实现科学化的社会管理,落实“科学发展观”,保障“可持续和谐社会”的基础就是社会科学真正的“现代化”,从定性走向定量,从离线走向在线,从时滞走向实时,其核心就是社会计算方法或计算社会科学。一句话,社会系统本身已光速化、量子化了,社会科学也必须像百年前发生的物理科学革命那样,实现“量子化”、“光速化”、计算化。

所以,我们要研究并普及社会计算。

2006年,借成立国际计算机协会ACM北京分会和ISI国际合作团队评审的机会,我们在北京组织了第一次关于社会计算的研讨会,即“ACM Workshop on Societal Security and Computing”,同时申请香山会议,即后来的2007年第299届香山会议“社会计算的基础理论和应用”。之后在国内外,我们团队首先于2008年举办了IEEE Social Computing Workshop(SoCo)和中国科协学术沙龙“社会计算——社会能计算吗?”。2009年,由来实验室寻求合作的Justin Zhan博士推动并启动了IEEE International Conference on Social Computing和中国首届社会计算会议。此外,我们团队还于2006年和2007年与《中国计算机学会通讯》和《IEEE 智能系统》合作组织了中外最初的新意义下的社会计算专刊。

在美国,除社会软件的Social Computing研讨会,第一次真正意义下的相关会议是2007年底哈佛大学举行的“计算社会科学”研讨会。我曾受邀参加此次研讨会,可惜因为当年10月刚去波士顿参加了IEEE TAB会议并顺访了哈佛大学,无精力于12月再回去,故与此次会议失之交臂。2008年,美国空军等资助在亚利桑那州举行了“社会计算、行为建模和预测”的研讨会,同年还有海军资助的“社会计算与文化建模”的研讨会。2009年,美国《科学》杂志发表了哈佛研讨会的总结,大大地促进了计算社会学或社会计算的研究。同年,应美国海军全球研究办公室之邀,我去泰国,在第二届国际“社会计算与文化建模”研讨会上做了关于人肉搜索及网民社会运动群体(CeSMO)的主题报告。现在,社会计算已成为一个全球化的研究热点。

过去,总有人问我什么是社会计算。现在,很少有人再问我这个问题了。但我却要说,要十分明确地说清楚何谓社会计算非常不易,因为这是一个发展十分迅猛的新领域。

不管怎样,从文献而言,“社会计算”一词新在中文,旧在英文,但新意义下真正明确的社会计算研究源于中国。对此,我十分欢迎大家提供相关的反例。

值得高兴的是,在中国科学院研究生院(即今天的中国科学院大

学,简称国科大)的大力支持下,社会计算已于2012年成为跨“控制科学与工程”、“计算机科学与技术”和“管理科学与工程”三个一级学科的交叉学科,学科代码为“99J2”,拥有博士和硕士学位授予权,学位培养工作已依托国科大和中国科学院自动化研究所展开。本书的撰写,正是因为社会计算研究生培养的迫切需要。需要说明的是,在本书的基础上,我们还正在组织《社会信息处理》和《社会行为建模与管理》的撰写工作。

社会计算是一门年轻的交叉学科,其发展成熟还需要一段较长的时间,特别是近年兴起的解析学(Analytics)、大数据以及社会制造等的研究,无疑将给社会计算带来新的冲击。因此,实际上目前并不是撰写一本完整且稳定的社会计算专著的成熟时机。本书主要是我们团队近十年研究工作的一个阶段性的概括性总结。由于时间限制,书中的内容选择和安排不尽理想,文字处理更非精细,这些问题希望在今后的版本中作进一步改善。显然,本书只能起到抛砖引玉的作用,希望更多的学者和学生都参与到这一领域,出版更高水平的专著,促进社会计算的深入发展与广泛应用。

在此,我谨向在社会计算研究起步时给予我极大支持的陈忻君教授、Steve Lansing 教授、Jim Hendler 教授、戴汝为院士、吴宏鑫院士、张�钊院士、陈国良院士、陆汝钤院士、李国杰院士、郑南宁院士,以及 Wendy Hall 和 Tim Berners-Lee 等教授表示衷心的感谢。曾大军教授虽然因时间的原因没有参加本书的写作,但他为社会计算学科的成长付出了很多心血和努力,在此深表感谢。在过去的八年时间里,我们团队培养了十余名社会计算领域的博士和硕士研究生,相应的成果形成了目前研究和应用工作的核心与基础,特别是我从他(她)们身上也学到了许多新的知识。实际上,我们许多在读和毕业的研究生都为本书的完成作出了贡献,书中一些内容还引用了部分学生的论文及项目成果,在此深表感谢。需要特别致谢的是陈伟运、罗川、张清鹏、张长利、王友忠、苏鹏、谭章文、曾轲、王晓等。同时,衷心感谢科技部、国家自然科学基金委、中国科学院长期以来给予的大力支持!

最后,感谢浙江大学出版社陈晓嘉、许佳颖老师在本书出版过程中所给予的大力帮助。

王飞跃

2012年12月于北京中关村

目 录

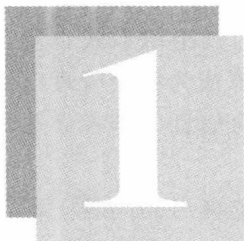
第 1 章 引 言	1
1.1 社会计算的研究背景	1
1.2 社会计算的历史和发展	2
1.3 社会计算的研究方法	7
1.3.1 基于人工系统的社会计算建模方法	8
1.3.2 计算实验与社会计算的分析与评估	9
1.3.3 平行系统与社会计算的实现	10
1.4 社会计算的主要研究与应用领域	10
1.4.1 社会数据感知与知识发现	11
1.4.2 个体与群体的社会建模	11
1.4.3 社会文化建模与分析	12
1.4.4 社会交互及其规律分析	12
1.4.5 决策支持与应用	13
1.5 本书的结构	13
参考文献	15
第 2 章 社会媒体数据获取与分析	18
2.1 社会传感器网络	19
2.1.1 社会传感器网络概念	20
2.1.2 社会传感器网络模型	21
2.1.3 社会传感器网络技术体系	23
2.1.4 小 结	24
2.2 行为知识抽取	24
2.2.1 行为抽取的意义和挑战	26
2.2.2 行为名称抽取与求精	28
2.2.3 行为知识抽取	29
2.2.4 讨 论	34

2.3	观点挖掘	36
2.3.1	词语级情感倾向性分析	36
2.3.2	句子级情感倾向性分析	38
2.3.3	篇章级情感倾向性分析	39
2.3.4	情感对象抽取	40
2.4	情感分析	40
2.4.1	基于字符串核的情感倾向性分析算法	41
2.4.2	从句子到篇章级的情感分析算法	43
	参考文献	50
第3章	社区发现	54
3.1	社区的基本概念	54
3.2	非重叠社区发现	56
3.2.1	层次聚类方法	57
3.2.2	图分割方法	58
3.2.3	G-N 算法及扩展	60
3.2.4	基于模块度优化的算法	62
3.2.5	谱分析算法	66
3.2.6	基于动力学的算法	68
3.2.7	标签传播算法	70
3.2.8	InfoMap 算法	71
3.2.9	局部社区发现算法	72
3.3	重叠社区发现	73
3.3.1	派系过滤算法	73
3.3.2	CONGA 算法	74
3.3.3	基于局部扩展的方法	75
3.3.4	基于边划分的方法	76
3.4	社区发现算法评价	76
3.4.1	模块度	77
3.4.2	强连通社区和弱连通社区	78
3.5	小结	79
	参考文献	79

第 4 章 社会网络建模与分析	83
4.1 社会网络分析的研究历史	84
4.2 社会网络上的传播过程	88
4.2.1 社会网络上的信息传播	88
4.2.2 社会网络上的传染病传播	89
4.2.3 社会网络上的传播动力过程	90
4.3 社会网络分析	92
4.3.1 社会网络的常用度量方法	93
4.3.2 社会网络拓扑特性	96
4.3.3 社会网络的基本模型	98
4.4 社会网络上的传播过程建模	100
4.4.1 社会网络上实际传播行为的实证研究	100
4.4.2 社会网络上的传播动力模型	102
4.4.3 社会网络上的传播动力过程的控制与决策	108
4.5 小 结	111
参考文献	111
第 5 章 面向社会群体的建模与分析	115
5.1 网群运动组织	115
5.1.1 研究背景	115
5.1.2 网群运动组织的基本概念与研究意义	116
5.1.3 研究现状及方向	119
5.2 网民群体的计算建模	121
5.2.1 主要智能体	122
5.2.2 智能体状态	123
5.2.3 智能体行为	124
5.2.4 智能体交互规则	125
5.3 人肉搜索案例与演化分析	127
5.3.1 人肉搜索案例	127
5.3.2 人肉搜索事件基本特征	128
5.3.3 人肉搜索群体的拓扑结构	131
5.3.4 HFS 网络的联通性与层次结构	134
5.3.5 核心参与者特征	141

5.3.6 人肉搜索群体演化分析	143
5.4 小 结	149
参考文献	149
第 6 章 社会行为分析与决策评估	153
6.1 行为分析与意图推理	154
6.1.1 行为分析研究现状	154
6.1.2 意图推理方法	163
6.2 复杂行为分析与预测	167
6.2.1 经典规划识别方法	168
6.2.2 多规划推理方法	169
6.3 行为建议与决策评估	175
6.3.1 问题定义	178
6.3.2 挖掘算法	183
6.3.3 讨 论	189
6.4 小 结	189
参考文献	190
第 7 章 社会计算应用	194
7.1 情报与安全信息学	194
7.1.1 网站信息监控、搜集与抽取	196
7.1.2 情报分析	199
7.2 互联网舆情计算	201
7.2.1 预警指标体系	202
7.2.2 舆情预警算法	205
7.2.3 应用案例:汶川地震	209
7.3 突发事件应急管理	214
7.3.1 关键核心技术	215
7.3.2 应用案例:重大传染病防控	217
7.3.3 改进模型	221
7.4 小 结	222
参考文献	223

第8章 从社会计算到社会制造	224
8.1 引 言	224
8.2 从减式制造到加式制造:历史与现状	226
8.3 3D 打印技术的应用、挑战与应对	229
8.4 从社会计算到社会制造:产业革命的基础与动力	232
8.5 社会制造的平行运营与管理	236
8.5.1 社会制造的人工社会	238
8.5.2 社会制造的计算实验	238
8.5.3 社会制造的平行执行	239
8.5.4 社会制造的支撑技术:云计算、物联网、大数据	240
8.5.5 社会制造的系统设计	240
8.6 小 结	243
参考文献	243
第9章 展 望	246
9.1 未来研究课题	247
9.2 社会计算发展线路图	248
附录1 王飞跃研究团队组织的社会计算会议及专刊列表	251
附录2 王飞跃研究团队发表的社会计算相关论文汇总	253
索 引	264



引 言

随着互联网技术的发展、计算资源和移动设备可获取性的增长、富媒体内容的普及以及随后的社会经济文化变迁,当代信息社会条件下的社会问题日益呈现出动态性、快速性、开放性、交互性和数据海量等特点,使得社会计算作为社会科学、管理科学与计算科学等的新兴交叉学科应运而生并迅速发展,成为处理网络化复杂社会系统的建模、分析、管理和控制等问题的有力方法和手段。

1.1 社会计算的研究背景

各种迹象表明,继物理计算和生物计算之后,社会计算可能成为科学计算研发的新焦点,并产生新的方向和领域(王飞跃,2006)。形成这一趋势的最根本性原因就是互联网的出现及其在社会各个层面和角落的不断深入与普及。每时每刻,世界各地数亿人游历于网上,工作和生活于真实的“虚”空间里,这一事实本身就是人类发展历史上的一个奇迹。而且,按照互联网研发先驱者的思路,这一奇迹也是一场规模空前的“社会计算”。虽然目前还很难预测互联网对人类生活的最终影响是什么,但可以肯定,人与人之间正在建立一种新型的“远程”社会关系,即从面对面的传统交流到数字操纵的网络交往,而这必将深刻地改变我们已有的社会模式。而且,随着基于网络的行为的不断演化和发展,其隐性的后果将是深远的,必将深刻影响未来人们的交往方式、相互关

系,以及社会的组织形式和活动机制。

在网络化社会的背景下,网上社会团体的组织及其影响的形成已变得非常容易,而且其动态变化更快、更难以预测,其组织形式也更广泛、更深不可测。正是基于这一背景,使得社会计算成为社会管理与发展必须面对且必不可少的科学工具,社会计算成为信息工作者必须面对的一个新的重要领域。

例如,网络化社会的社会安全就是一个越来越突出的问题。不仅普通人在生活和工作中利用网络,恐怖和犯罪组织也在利用网络。正如原子弹的出现改变了现代战争的概念一样,互联网的发展也深刻地改变着人们关于社会治安的传统概念。原子弹的威力在于其残酷程度极高,杀伤范围极广,完成破坏的时间极短,机动能力极强;同原子弹类似,通过网络发送危害社会的信息也具有成本代价极低、影响范围极广、完成时间极短、机动能力极强的特点。目前已发生的网络现象,从计算机病毒的传播,到普通人通过在网上的怪异行为而一夜成为“名人”,都说明网络社会的发展态势及其影响不容忽视。目前,这些网络现象还可以被看成是社会的“良性”肿瘤,但我们必须利用人文社会知识,深入进行关于数字网络化社会状态及其动态趋势的研究,为应付将来网络上可能出现的危害性更大、更广的“恶性”社会肿瘤做好准备。社会计算必将成为解决此类重大问题的关键技术之一。

1.2 社会计算的历史和发展

布什(Bush)、里克利德(Licklider)和恩格尔巴特(Englebart)是公认的互联网的主要先驱者,其中布什和里克利德还被认为是催生互联网的第一、二号人物。尽管这三位学者都没有亲自参与创立和发展现代互联网和万维网,但他们的思想和远见是后来许多互联网和万维网技术发明的源泉,而且目前的技术只是实现他们想法的开端,还有许多更有意义的设想目前没有实现。对本书而言,更重要的是介绍他们的人文思想和与社会计算相关的远见。

布什是现代科学史上的一位传奇性人物。第二次世界大战之前,布什作为美国麻省理工学院的一名研究人员,负责模拟计算机的研制。第二次世界大战时,他是卡内基梅隆大学的前身——卡内基学院的院

长。美国参战之后,他向罗斯福总统建议将科学家组织起来参加战时的军事研究,并成为国家国防研究委员会与后来的科学研究和发展局的负责人,协调指挥 6000 多名科学家从事军事研发,其中最著名的项目就是曼哈顿计划。第二次世界大战末期,布什提交了一份给美国总统的著名报告《科学:无边的新领域》(*Science: The Endless Frontier*),建议了后来成立的国家科学基金会(NSF)和国防部先进研究项目署(ARPA,也就是目前的 DARPA,直接组织了互联网的前身——阿帕网的实施)。

布什的工作为包括阿帕网在内的许多“Big Science”项目创造了条件。但布什的直接影响是他于 1945 年发表的科技散文《随便我们想》(*As We May Think*),特别是文中提出的通过关联加强人记忆能力的装置“Memex”,被认为是个人计算机的原始模型。正是读了布什的散文,鼠标的发明人恩格尔巴特才意识到利用计算机来管理信息在处理数字之外的无比潜力。后来,恩格尔巴特在里克利德的支持下,直接参与了阿帕网的实施。此外,超文本(Hypertext)的主要发明人纳尔逊(Nelson)也把他的工作归功于布什散文的启示。个人计算机的出现被看成是“Memex”的一种简化,而万维网被看成是其初级实现方式,现在流行的博客更能从布什的散文中看到影子。

尽管布什散文中的许多想法已在不同程度上得到了实现,但它所蕴含的宏大图景,即社会计算或计算的社会化,目前只能说是“初露端倪”,而且还没有被信息工作者充分认识和重视。虽然布什没有明确提出“社会计算”一词,但他散文的主要动机和思路都是围绕着以下三点展开的:①个人计算的集成化;②群体计算的可能性和重要性;③历史性计算的可能性和重要性。而如果把“集成深度计算”、“群体广度计算”和“历史经验计算”综合起来,恰恰揭示了社会计算的内涵。

提出社会计算概念的一个客观依据是:我们实际上可以把传统的限于语言层次和静态的知识,不管是书本上还是社会上、解析型还是经验型、历史的还是现实的,都数字化、网络化和动态化,并用于各种复杂社会问题的建模、分析和决策支持,而这也是布什的主要动机与目的。既然大家都公认布什的散文是催生互联网的原始动因,社会计算的思想是他写作散文的动力,那么我们也可以说,社会计算是互联网产生的原始动机。

里克利德和恩格尔巴特的工作,以及今天互联网所带来的现实,或许能使我们认识到社会计算也将是互联网所导致的必然结果。同布什