ISSN 1000-1239/CN 11-1777/TP
50(12): 2483-2491, 2013

社会计算:大数据时代的机遇与挑战

孟小峰[1] 李 勇[1,3] 祝建华[2]
[1](中国人民大学信息学院 北京 100872)
[2](香港城市大学媒体和传播系 香港 999077)
[3](西北师范大学计算机科学与工程学院 兰州 730070)
(xfmeng@ruc.edu.cn)

Social Computing in the Era of Big Data: Opportunities and Challenges

Meng Xiaofeng[1], Li Yong[1,3] and Jonathan J. H. Zhu[2]
[1](*School of Information, Renmin University of China, Beijing* 100872)
[2](*Department of Media and Communication, City University of Hong Kong, Hong Kong* 999077)
[3](*College of Computer Science and Engineering, Northwest Normal University, Lanzhou* 730070)

Abstract With the rapid development of information technology, especially sweeping progress in the Internet of things, cloud computing, social networks and social media, the era of big data is coming. As a data-intensive science, social computing is an emerging thing that leverages the capacity to collect and analyze data with an unprecedented breadth, depth and scale. It represents a new computing paradigm and an interdisciplinary research and application field. A broad comprehension of major topics involved in social computing is important for both scholars and practitioners. In this paper, we give a brief survey of the various research fields in social computing. We present key concepts and analyze state-of-the-art of the field. The article not only sheds insights on social computing, but also affords conduit for future research in the field. Social computing has two distinct trends: One is on the social science issues, such as computational social science, computational sociology, social network analysis, etc; The other is on the use of computational techniques, such as social use, hedonic use and generative use. Finally some new challenges ahead are summarized, including interdisciplinary cooperation and training, big data sharing for scientific data mashups, and privacy protect.

Key words social science; social computing; social networks analysis; computational social science; big data

摘 要 信息技术的飞速发展,特别是物联网、云计算、社交网络、社会媒体以及信息获取技术的进步,数据正以巨大的速度迅速增长和积累,大数据时代已经到来.社会计算作为一种数据密集型科学,在收集和分析数据的广度、深度以及规模上都产生了巨大的影响,社会计算作为一种新的计算范式,产生了一个新的跨学科研究与应用领域,其广阔的研究内容与应用已引起了学术界和工业界的广泛关注.分析了社会计算产生的历史背景及概念、研究现状及大数据带来的机遇,综述了社会计算不同的研究领域,主要有2个发展趋势:一个面向社会科学,包括计算社会科学、计算社会学、社会网络分析等;一个面向技术应用,包括社交应用、娱乐应用、生产应用等,这2种发展趋势同时又相互影响.最后讨论了社会计算研究领域存在的挑战,包括跨学科合作与训练的问题、科学研究中大数据共享问题以及隐私保护.

收稿日期:2012-06-27;**修回日期**:2013-09-25
基金项目:国家自然科学基金项目(61379050,91024032,91224008,91124001,91324015);国家"八六三"高技术研究发展计划基金项目(2012AA011001,2013AA013204);中国人民大学科学研究基金项目(11XNL010);2012年甘肃省高校基本科研业务费专项资金项目
通信作者:李 勇(facingworld@126.com)

关键词 社会科学;社会计算;社会网络分析;计算社会科学;大数据

中图法分类号 TP3-05

物联网、云计算、社交网络、社会媒体以及信息获取技术的飞速发展,数据正以前所未有的速度迅速增长和积累,数据是人类社会最重要的财富,大数据时代的到来,为研究人类社会动态和模拟社会问题带来了前所未有的机遇.计算机科学家、社会学家等各领域学者开始关注大数据对社会、经济、科学研究等方面带来的巨大价值.美国圣塔菲研究所(Santa Fe Institute)、谷歌研究院(Research at Google)、惠普社会计算实验室(HP Social Computing Lab)等跨学科研究机构和哈佛、斯坦福、康奈尔等大学,开始用复杂性科学来描述社会系统中的复杂现象,提出了复杂适应系统等一系列新理论,用计算机作为研究复杂性科学的基本工具,开创了计算社会科学等新的研究方法,社会计算开始进入人类社会("social computing"一词也有文献译成"社会化计算").

本文从社会计算的本质出发,对现有的研究中若干关键问题进行分析总结,讨论这一研究领域存在的机遇和挑战.简要介绍了社会计算产生的历史背景及其定义,对社会计算领域进行分类研究,介绍了大数据为社会计算带来的机遇,分析社会计算研究现状的基础上讨论了其面临的挑战.

1 社会计算产生的历史背景

社会科学作为独立的学科是在19世纪末才出现[1],这是对工业化所带来的挑战作出的回应.由于大规模的社会结构、社会组织的出现,人们的社会联系越来越频繁和多样,导致社会冲突加剧,社会管理和控制的难度空前增大,传统的那种靠少数社会精英拍脑袋决策管理社会的方式已经过时,社会科学的产生和发展成为历史的必然.社会科学的主要学科有经济学、社会学、政治学等,社会学研究的是市民社会以及以市民社会为基础形成的社会组织,是关于社会良性运行和协调发展的条件以及机制的综合性具体社会科学[2].

自然科学、社会科学和人文科学只是学术建制意义上的区分,它们之间总是密切联系.自然科学的根本目的在于发现自然现象背后的规律,20世纪50年代之前,社会科学与自然科学相对独立,跨学科研究较少.如图1所示,社会科学与自然科学由一条学科河分隔,河的左岸是以科学计算为核心研究范式,以系统科学、控制论、人工智能等作为研究方法的自然科学;右岸是心理学、经济学、传播学、社会学、政治学等社会科学.

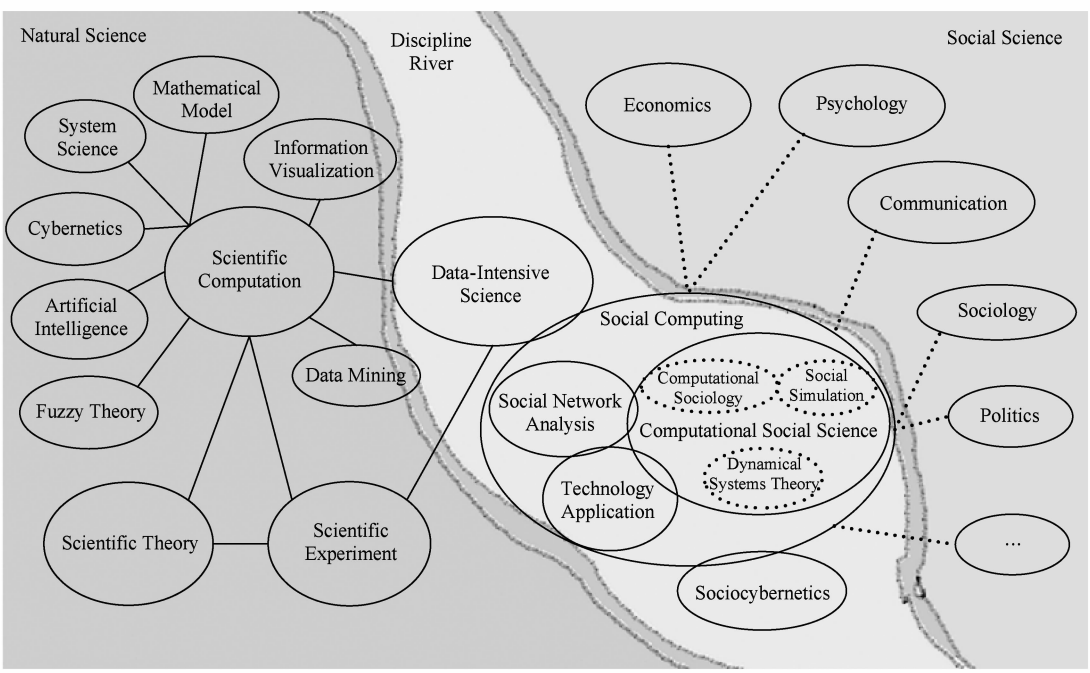

Fig. 1 Classification of disciplines.

图1 学科分类[3]

20 世纪 70～90 年代,人类进入后工业化(post-industrialization)时代,信息革命改变了一切,经济、政治、文化的全球化融合达到新的阶段,科学发展变得越来越快,越来越复杂.大工业和高科技为人类创造了非常丰富的财富,但同时也把人与自然的关系、人与人的关系置于一个危险的境地,社会不平等、阶级冲突、社会异化、种族冲突、政治革命、宗教冲突、国家冲突、环境恶化等社会问题突出,人类越来越重视对自身前途与命运的价值关怀.为了准确地研究社会问题,经济学、社会学等学科都试图形成一套完整的定量分析研究手段,用严谨的数学方法对问题进行描述和求解,但由于现实世界是一个多变量复杂系统,很难像物理学那样用数学公式精确地描述和求解问题.在这种共同的背景下,自然科学、社会科学开始走到一起,任何一门科学都开始意识到自己的相对性,意识到与其他学科密切关联[4].如图 1 所示,建立在自然科学与社会科学之间,位于学科河之上的跨学科社会计算研究逐渐兴起,这些学科都有数据密集型的特点,包括社会网络分析、计算社会科学、社会控制论等.

2 社会计算的定义

Parsons 在 1949 年主持美国社会学学会工作时,在他的努力下建立了哈佛大学社会关系系,尽管其结构功能主义理论(structural functionalism)广受批评,但他开创了跨学科研究的先河.Parsons 的理论另一个重大影响是对欧洲社会学在控制论和系统科学方法的影响[5].

互联网的缔造者、心理学家 Licklider 在担任美国国防部高级研究计划署信息处理技术办公室第一任主任期间,于 1960 年发表了《人与计算机的共生》[6]一文,创造了因特网的原始设想,第 1 次提出了交互计算概念,他认为人类将会有一种网络将世界上所有计算机联成一体,人们可以使用地理上很远的计算机,获取任何计算机中的数据,使用很多计算机来做一件事,可以互相共享资源、平衡负载,为全世界的用户服务.在 Licklider 等人的推动下,计算机网络作为一种通信设施于 20 世纪 60 年代末开始发展[7].

1978 年,社会学家 Hiltz 等人出版了《网络国家:人类通过计算机交流》[8]一书,这是最早描写网络社区社会学的著作.网络社区通过计算机网络将人们互联,人们在网络平台上相识、工作、讨论、争论、协作.该书最早预测到未来虚拟社区以及对社会、政治、法律等方面的影响,也最早预测到网络的爆炸式发展以及隐私问题、匿名问题、远程办公、在线政治活动等.书中还提到了审查制度、监管制度、成瘾问题、IP 和盗版、Email、美国邮政服务的消亡以及工作、政治、法律的改变.该书的出版引起了计算机科学家、管理科学家、政治家、社会学家、医生、计算机狂热人士等一大批人的兴趣,被称之为计算机会议上的“圣经”[9].

20 世纪 90 年代中期是个人计算向社会计算转型的时期.个人计算关注个体使用信息技术,社会计算通过小型的社会群体或大型的社区用户一起协作使用信息技术.社会计算对社会和经济具有深远的影响,尽管不用质疑其影响,但也很难定量研究其影响的程度,只能在跨学科的范围内保持其一定的领域评判.

1994 年,社会计算的概念第 1 次出现,Schuler 认为[10]:“社会计算可以是任何一种类型的计算应用,以软件作为媒介进行社交关系的应用”;Dryer 等人[11]将社会计算描述为一种理论概念,包括科学和技术 2 方面:“人类使用计算技术进行的社交行为和交互行为所产生的相互作用”;Wang 等人[12]认为社会计算是:“信息技术和通信技术等促进了人类社会的研究和社会动态发展”;文献[13]认为,社会计算是指使用信息系统作为社会交互的场所,并使用信息系统作为数据收集和处理的空间,社会计算是在虚拟场所中的感知、交流和协作,社会计算需要把计算设备作为人与人之间交流的媒介,需要将人机交互设定成为一个社会实践的环境,将理解社会过程作为交互系统的一部分工作.

近年来,科学家不仅发现人类大脑和生命是计算系统,而且发现整个世界也是一台计算系统,因此哲学界产生了计算主义思潮.计算主义认为:“人所处的整个世界是由算法控制,并且按算法确定的程序进行演化.宇宙是一部巨型计算装置,任何自然事件都是在自然规律作用下的计算过程.事物的多样性是因为算法的复杂度不同而产生的不同外部表现”[14].虽然计算主义导致许多哲学家的质疑[15],但计算的确已经渗透到经济学、物理学乃至科学研究的各个领域,计算已成为人们认识自然、生命、思维、社会的一种普遍的观念和方法.如果没有计算机,就没有当代科学的突飞猛进发展,就不会有社会的快速进步,复杂性科学就不会出现.

基于以上分析,本文认为社会计算是使用系统科学、人工智能、数据挖掘等科学计算理论作为研究方法,将社会科学理论与计算理论相结合,为人类更深入地认识社会、改造社会,解决政治、经济、文化等领域复杂性社会问题的一种理论和方法论体系.

3 社会计算主要研究领域

社会计算不仅是一种技术而且是一种社会现象,社会实践所处的环境决定了技术系统的社会特点.社会计算形成了2种发展趋势:一种是面向社会科学的社会计算;另一种是面向技术应用的社会计算,这2种发展趋势同时又相互影响.

3.1 面向社会科学的社会计算

3.1.1 社会网络分析

社会科学的网络概念非常广泛,主张采用多理论多层次方法研究传播以及其他形式的组织和社会网络[16].文献[17]研究了包括人在内4个不同动物种群的42个网络以及多种关系内涵、多种社群规模,尝试研究这些网络"表面存在差异,结构是否相似"这一问题.社会网络分析(social network analysis)[18-19]的目标是研究大型动态复杂网络,特别是人类社会系统的结构和交互模式等.社会网络分析研究的主题包括社会流动、健康行为、疾病扩散的关键节点、连锁董事、在线社区分析等.

社会网络分析所采用的研究与分析方法主要是基于代理的模型、理论物理、图论等.在Milgram[20],Watts[21]等人开创的小世界(small world)现象研究基础上,Barabási等人在无标度网络(scale-free networks)[22]特性上有多个发现,他们发现复杂网络中的联接符合幂律分布(power-law distribution),有些节点联接稠密,而大多节点联接很稀疏,这些发现说明复杂网络尽管结构复杂但并不是随机的,符合一定的宇宙主宰原则,在混沌中存在着秩序.社会网络分析领域经典的研究成果还有强弱关系(strong and weak ties)[23]、结构洞(structural holes)[24]、信息级联(information cascades)[25]等.

3.1.2 计算社会科学

社会学中早就有一个计算社会学(computational sociology)[26]分支.计算社会学不仅与传统的系统科学、控制论和复杂性科学交叉,并且跨越社会科学多个领域,如经济学、生态学、社会网络组织、人口学中的小群体动力学、环境以及城市规划等.计算社会学的主要研究领域是社会模拟(social simulation),使用计算机建立人工实验环境,研究复杂社会系统,模拟方法包括基于方程的模型(equation based modeling)和计算模型(computational modeling).主要计算方式是在个体代理或多代理系统观点下进行社会模拟[27].

计算社会学从20世纪50年代出现,早期使用离散事件模型等方法,此后40年时间一直试图使用计算机科学方法进行社会研究,但如同文献[27]指出的:"从社会学家使用计算机起,计算机模拟在社会学中扮演了一个重要的,但是第2位的角色.模拟方法得出的结果显得苍白,远离主流社会学的方法,因此这种方法总是不被社会学家们所欣赏".

最近10年之内,随着互联网的兴起和在线实时数据的公开易得,来自政治学、经济学、语言学、传播学、人类学等社会和人文学科的研究者联手计算机、物理、数学、控制等科学技术界专家,兴起了规模更大、参与更广的计算社会科学(computational social science)[28].计算社会科学的核心技术是数据挖掘[29],使用机器智能从大量复杂真实数据集中发现有趣模式和知识,在数据的驱动之下,使用统计学、机器学习、模式识别、数学模型等方法,进行探索式的知识发现和数据管理,数据源包括数据库、Web等以及动态地流入系统的数据.数据挖掘对于社会学家分析复杂社会系统产生的大量数据有很多好处,可以分析数据质量,可以聚焦于社会过程和关系,可以处理非线性、有噪音、概念模糊的数据等.与依赖计算机模拟的计算社会学相比,现在流行的计算社会科学依赖海量和实时的网络数据,由社会和人文科学各学科广泛参与,并与科学计算界紧密合作,所以研究成果及影响力均不可同日而语,一个例证就是《Science》、《Natrue》、《美国科学院年报》(PNAS)等世界顶级期刊经常刊登计算社会科学的最新研究成果.

3.2 面向技术应用的社会计算

面向技术应用的社会计算将社会科学中的一些概念或理论融入技术应用,例如社区、社会网络、社会心理学等,从而推动应用的快速健康发展.技术应用也有助于社会交互,获取的数据可以更好地分析社会交互的计算模型,包括从小规模的动态交互到大规模的社会演变.面向技术应用的社会计算经历了群件(groupware)、社交软件(social software)、社会媒体(social media)等几个阶段[30].

群件在20世纪70年代提出,主要在学术机构等有限的范围内使用.群件是指一套协作技术,其目

标是支持协同交互,例如 eies 系统(electronic information exchange system)[9],在这方面的研究产生了 2 个主要的应用成果:一个是计算机支持的协同工作(computer supported cooperative work);另一个是计算机支持的协同学习(computer supported collaborative learning).

20 世纪 90 年代末,随着互联网泡沫的破灭,为了应对信息技术市场面临的困境,以 Web 2.0 作为基础的社交软件成为主要的解决方案.互联网泡沫为社交软件的产生提供了前提,但互联网在家庭中的普及也同样重要.2000 年之后,面向商业的公共应用,支持群体交互的在线社交软件迅速发展,并于 2002 年达到顶峰.

2005 年,随着 Web 2.0[31] 的蓬勃发展,社会媒体开始涌现.社交软件向社会媒体的演变,主要是由面向技术的交互模式转变为面向人的交互模式,强调交互的动态性和社会实践的影响,这些应用通过用户自发的交互,对内容和服务进行生产、发布、消费,例如博客、播客、维基百科、社会网络站点等.近几年来,随着智能手机等设备的普及,移动社会媒体发展迅速.

面向技术应用的社会计算可以分为 3 个维度:社交应用(social use)、娱乐应用(hedonic use)和生产应用(generative use)[32],这 3 个维度并没有严格的界限,也是交叉影响.

3.2.1 社交应用

社交应用主要通过博客、微博、论坛等社会媒体维持社交关系,这方面的研究主要包括在线社会网络分析、社会媒体挖掘和社区发现[33]等.社会网络分析已有大量研究工作,分析社会网络的影响力,发现参与的机会,共享用户对特定的话题、品牌、产品的看法.社会媒体挖掘和社区发现探测网络环境中的社区结构,发现内聚的子群,对于定量分析社会群体演化、预测用户行为有重要意义.社会媒体对个人或企业的声誉、信任等方面的管理也有重大意义.

3.2.2 娱乐应用

娱乐应用主要通过共享媒体、社会新闻、社会书签、维基百科、在线游戏等平台,方便人们共享知识、享受便利.这方面的工作有协同标记(collaborative tagging),通过用户对 Web 内容打标签,促进社交导航和共享表达;协同过滤(collaborative filtering),通过大量的社会媒体数据,在推荐系统中利用用户偏好来预测其喜好的话题或产品.情感计算(affective computing)[34-35]将心理学、认知科学等融入科学计算,通过社会媒体累积的海量数据,分析人在特定时间段情绪随时间的变化趋势;环境感知计算(context aware computing)在移动计算应用中用不同的设备和服务来发现环境信息并从中获益,例如用户位置、用户行为、附近的人等.

3.2.3 生产应用

生产应用主要指创建和共享协作内容,为生产生活提供便利,例如计算机支持的协同工作、智能交通、危机应急响应、商业应用、健康管理等.已有大量研究成果和应用,例如,众包(crowdsource)通过大量的参与人协作解决难题,形成了人肉搜索[36-37]等社会现象;IBM 的智能交通项目[38]从用户的智能手机等不同的传感设备上收集路面、收费站、交叉路口、桥梁等基于位置的数据,判断道路运行模式,基于这些数据分析,为用户提供交通信息,预警拥堵状况,避免交通事故;将社会心理学中的社会惰化(social loafing)理论应用到计算机支持的协同工作的设计之中,利用行为理论解释人们在网络环境中为什么做事不按经验性规则或程式化模式,预测不同设计方案的结果[39];通过装配有 GPS 和传感器的智能设备,有效及时地分享环境感知的信息,为人类在面对灾难时作出正确的判断和提供决策支持[40]等.

4 大数据时代的社会计算

科学实验、科学理论和科学计算是人类探索自然、研究社会的 3 种基本范式.大数据[41]时代的到来在数据收集和分析的广度、深度以及规模上都产生了前所未有的影响.社会计算是一种数据密集型科学研究范式(data-intensive science)[42],近年来,随着大数据的影响不断深远,越来越引起学术界和工业界的高度关注.在大数据条件下,传统的以计算为中心的理念要逐渐转变到以数据为中心,形成数据思维.

所谓数据思维,Schonberger 指出就是在处理数据时要作到三大转变[43].第 1 个转变是在大数据时代可以分析更多的数据,甚至是与之相关的所有数据,而不再依赖于采样.社会科学研究社会现象的总体特征,采样一直是主要数据获取手段,信息技术的普及让人们意识到这其实是一种人为限制.使用所有数据可以带来更全面的认识,可以更清楚地发现样本无法揭示的细节信息.正如 Watts 所期望的[44],借助于大数据和计算机分析技术,21 世纪的社会科学可能实现定量化的研究,从而成为一门真正的科学.

第 2 个转变是不再追求精确度. 与银行、电信等行业的精确计算需求不同,社会计算是对社会动态的反映,当拥有海量即时数据时,绝对的精准不再是追求的主要目标,适当忽略微观层面上的精确度,会让社会科学在宏观层面拥有更好的洞察力.

第 3 个转变是不再热衷于寻找事物间的因果关系,而应该寻找相互之间的相关关系. 社会科学中的因果关系是概率性的,只能研究原因的结果(effects of causes),而不是结果的原因(causes of effects),相关关系也许不能准确地说明一个社会现象发生的原因,但是它会揭示其发展过程.

5 社会计算研究现状

社会计算作为一种新的研究范式,已引起了国内外学术界、工业界的普遍关注. 2012 年 11 月 15 日,第 4 届全国社会计算学术会议(China National Conference on Social Computing, NCSC2012)在中国人民大学举行,来自国内外计算机、管理学、经济学、新闻学、社会学、物理学等学科以及工业界 44 名专家学者发表演讲,全国 37 所大学和研究机构超过 300 人次参会研讨. 在这次会议上,形成了鲜明的两种观点,以技术为主导的学科和工业界普遍对大数据时代的社会计算持乐观的态度,而社会科学领域学者持过度谨慎的看法.

5.1 技术乐观派

计算机、物理学、经济学等学科学者以及工业界普遍认为,大数据时代的社会计算为科学研究提供了不同视角,蕴藏了大量有重大价值的研究课题,为社会预测创造了条件,为大数据在国民经济发展和社会安全中的应用提供了理论思路. 本世纪的第 2 个 10 年是复杂性科学深入各个领域、不断有新的突破且面临更大挑战的时代,许多重要复杂系统迫切需要运用社会计算研究方法,通过实证统计和理论模型分析相结合的手段进行全新的认识和探索.

5.2 理论谨慎派

社会科学界对大数据时代的社会计算所持的态度比较谨慎. 他们认为当前大数据的概念大而空洞,就数据论数据,忽视人性,带来了很大挑战. 对大数据分析大多处于商业应用层面,没有上升到社会科学层面. 如何运用社会科学理论来指导社会计算,特别需要提出新的理论来指导大数据对社会动态的分析,这是学术界亟需突破的问题.

5.3 工业界亟待解决的问题

社会媒体在互联网上迅猛发展,计算广告、人脸识别等技术在工业领域的广泛应用,虚拟社会如何良性运行与协调发展,已成为当前社会计算应用的主要挑战. 例如,电子商务企业在线交易中,虚假数据带来虚拟社会严重的信任危机,但当前反作弊业务流程主要关注算法的精确性和结果的可解释性,而作弊行为有突发性的特点,单规则认定作弊行为具有不确定性特点,基于这种方法解决虚假交易只能"堵"无法"疏",需要学术界探索解决这一问题. 工业界还提出很多亟需解决的问题,例如在线广告中展示广告与搜索行为之间的关系问题、大规模用户群的行为预测、同一个用户在不同网络终端个体识别、欺诈检测等问题.

6 社会计算面临的挑战

6.1 学科壁垒的挑战

"小世界"、"结构洞"等一系列影响深远的研究成果被发现之后,社会网络分析工作引起许多研究人员的关注,不仅社会学家跟踪研究,大量物理学、计算机等学科研究人员也开始关注社会网络分析. 近年来,用数学模型和各种算法对在线社会网络进行分析,发表了海量的论文,但这些研究工作大多是不断重复许多早期社会网络研究成果,用非常简单的数据模型解释非常复杂的社会现象,虽然取得了一些惊人的结果,例如"四度分隔"[45]等,但都无法超越社会学家早期在社会网络分析方面的成果. 正如 Watts 指出的:"物理学家可能是非凡的技术专家,但他们只能是二流的社会学家"[46].

社会学所研究的人类动态社会网络是具有社交性的网络[47],社会网络基于社会结构,社会结构包含社会地位和社会关系 2 方面,社会关系是制约社会结构的重要因素之一. 社会网络具有个体性、非正式性、持续性等特点,个体性是指具有私人层次上的交往和交流. 以社会媒体在线用户群为主要研究对象的社会网络分析中,研究者将社会关系简化处理,忽略社会地位的影响,将用户作为节点,用户间简单的"关注"关系作为边构成网络,以节点间边的紧密程度划分"社区",这种研究方法忽略了社会网络的本质,节点与节点之间无法体现出社会关系中的个体性特征,因此很难在社会性上有跨越性的研究突破.

社会科学与自然科学本质区别在于思维方式的不同[48],社会科学是总体逻辑思维,自然科学是类

型逻辑思维.类型逻辑思维认为应该重点关注典型现象,只要理解了典型现象的规律,就可以将其概括并推广到个体和具体问题.总体逻辑思维关注独立各异个案的整体分布,社会科学认为变异是社会现实的本质,社会学家的工作就是从变异中寻求规律,以经验为基础、以量化为导向地去概括总体变异的系统模式,社会科学的量化无法挖掘出普适规律来描述和解释所有个体行为[49].

正是因为思维方式的不同,社会科学与自然科学之间的壁垒仍然难以逾越,社会学家批评技术学派所作的社会计算研究缺乏理论指导,技术学派认为社会科学研究所用数据规模太小不可信任.为社会计算提出跨学科的协作与训练、提出学科间统一的理论指导是当前最大的挑战.

6.2 大数据带来的挑战

长期以来社会科学的定量研究都是通过问卷调查的方式收集数据,这种方式收集的数据量小且真实性难以确定,这些数据往往是在一个时间点或相隔很久的不同时间点获得,对连续的、动态的社会过程只能推断.大数据时代的到来,基于互联网的各种应用正以前所未有的方式生成和保留各种值得研究的大规模数据,这些具有空前宽度、深度和规模的数据对社会科学研究人员来说是宝藏和机遇,同时也是挑战.

一方面,绝大多数通信领域和社会媒体领域的企业都拒绝或限制研究人员获取其数据,跨组织的数据共享对于科学研究至关重要,但是社会计算研究正受到数据获取的限制所带来的严重挑战[50].另一方面,由于社会各系统相对独立,数据之间存在封闭性或关系的断裂性特点,大数据可能带来规律的丧失或失真,导致错误发现的风险增加.当前计算的速度越来越快,但分析的程度却越来越低,每个研究平台都有自己的数据,这些数据在表面上看起来很全面,但实际上都是各个领域信息的片段描述,而且数据背后看不见人性因素,失去社会意义.在这种社会背景下,看似人类拥有所有数据,但同时又什么数据都缺,大数据挖掘必须发挥人的主动性,体现人类对自身价值的终极关怀.要从大数据中采集到足够准确、系统而有代表性的社会个体特征,面临着伦理、法规和技术等多个方面的困难,这些难题已经构成大数据时代社会计算的严重挑战.

6.3 隐私保护

文献[51]通过手机轨迹数据分析了15个月150万人的移动规律,发现每个人的移动踪迹高度独立,在这个数据集中每个人的位置在每小时是特定的,4个时空点就足以唯一识别95%的个人.收集和发布移动数据、社会媒体数据、个人医疗数据等直接给个人隐私带来威胁,如果数据拥有者直接发布隐含着敏感信息的数据,而不采取适当数据保护技术,将可能造成个人敏感信息的泄露.在多样化的社会中隐私是基本的需要,是个人权利的基础,互联网和移动通信放大了个体的唯一识别性,进一步增强了传统的隐私挑战.

如何既能公开发布更多的数据,为社会计算提供更大的数据支持,同时又能保证个人的隐私不会泄露,作到数据开放与隐私保护的两者平衡,这是学术界需要关注的问题.

7 总 结

工业化时代的学科分类为推动社会进步作出了巨大贡献,随着后工业化时代的到来,这种学科划界越来越成为人们思想上的羁绊.计算机科学的基础虽然基于电子学等自然科学,但集成电路、操作系统等都不是自然界客观存在的,是人类智慧的产物.因此哲学家Popper称其为"第三世界科学"以别于自然科学.计算机科学既依赖于自然科学的发现又依赖于人的创造,既独立于人的意志之外又受人的意志驱动.正是由于学科的特殊性,将社会科学与计算机科学结合起来进行研究,越来越成为人们的共识.社会计算为人类发现自身价值提供了更多的机会,计算语言学、计算人类学、计算广告学、城市计算等跨学科研究领域不断产生.社会计算的研究还处于起步阶段,本文作了一些初步的探索,分析了社会计算产生的历史背景、学科分类,给出了确切的定义,对现有的研究工作及方法进行了归纳总结,最后指出了大数据时代社会计算面临的挑战性问题.希望能为研究人员提供参考.

参 考 文 献

[1] Zhu Hongwen. Study of the nature of social science [J]. Academic Monthly, 1999, 31(11): 33-40 (in Chinese)
(朱红文. 社会科学性质再探[J]. 学术月刊, 1999, 31(11): 33-40)

[2] Zheng Hangsheng. Introduction to Sociology [M]. Beijing: China Renmin University Press, 2003: 1-3 (in Chinese)
(郑杭生. 社会学概论新修[M]. 北京: 中国人民大学出版社, 2003: 1-3)

[3] Castellani B, Hafferty F W. Sociology and Complexity Science: A New Area of Inquiry [M]. Berlin: Springer, 2009: 246

[4] Zhu Hongwen. The nature of social science and its relationship with the humanities [J]. Philosophical Researches, 1998(12): 29-36 (in Chinese)
(朱红文. 社会科学的性质及其与人文科学的关系[J]. 哲学研究, 1998(12): 29-36)

[5] Castellani B, Hafferty F W. Sociology and Complexity Science: A New Area of Inquiry [M]. Berlin: Springer, 2009: 14

[6] Licklider J C R. Man-computer symbiosis [J]. IRE Trans on Human Factors in Electronics, 1960, HFE-1: 4-11

[7] Licklider J C R, Taylor R W. The computer as a communication device [J]. Science and Technology, 1968, 76(4): 21-31

[8] Hiltz S R, Turoff M. The Network Nation: Human Communication Via Computer [M]. New York: Addison Wesley, 1978

[9] Subramanian R. Starr Roxanne Hiltz: Pioneer digital sociologist [J]. IEEE Annals of the History of Computing, 2013, 35(1): 78-85

[10] Schuler D. Social computing [J]. Communications of the ACM, 1994, 37(1): 28-29

[11] Dryer D C, Eisbach C, Ark W S. At what cost pervasive? a social computing view of mobile computing systems [J]. IBM Systems Journal, 1999, 38(4): 652-676

[12] Wang Feiyue, Carley K M, et al. Social computing: From social informatics to social intelligence [J]. IEEE Intelligent Systems, 2007, 22(22): 79-83

[13] Musser D, Wedman J, Laffey J. Social computing and collaborative learning environments [C] //Proc of the 3rd IEEE Int Conf on Advanced Learning Technologies, Piscataway, NJ: IEEE, 2003: 520-521

[14] Li Jianhui. Toward computationalism [J]. Journal of Dialectics of Nature, 2003, 25(3): 31-36 (in Chinese)
(李建会. 走向计算主义[J]. 自然辩证法通讯, 2003, 25(3): 31-36)

[15] Liu Xiaoli. Querying computationalism [J]. Philosophical Researches, 2003 (4): 88-94 (in Chinese)
(刘晓力. 计算机主义质疑[J]. 哲学研究, 2003 (4): 88-94)

[16] Peter R M, Noshir S C. Theories of Communication Networks [M]. Translated by Chen Yu, Liu Ying. Beijing: China Renmin University Press, 2009: 276 (in Chinese)
(彼得·R·芒戈, 诺什·S·康特拉克特. 传播网理论[M]. 陈禹, 刘颖, 译. 北京: 中国人民大学出版社, 2009: 276)

[17] Faust K, Skvoretz J. Comparing networks across space and time, size and species [J]. Sociological Methodology, 2002, 32(1): 267-299

[18] Lin Freeman. What is social network analysis? [EB/OL]. [2013-05-23]. http://www.insna.org/what_is_sna.html

[19] Wasserman S, Faust K. Social Network Analysis Methods and Applications [M]. Translated by Chen Yu, Sun Caihong. Beijing: China Renmin University Press, 2011 (in Chinese)
(斯坦利·沃瑟曼, 凯瑟琳·福斯特. 社会网络分析: 方法与应用[M]. 陈禹, 孙彩虹, 译. 北京: 中国人民大学出版社, 2012)

[20] Milgram S. The small-world problem [J]. Psychology Today, 1967, 1(1): 61-67

[21] Watts D J, Strogatz S H. Collective dynamics of small-world' networks [J]. Nature, 1998, 393(4): 440-442

[22] Barabási A L, Bonabeau E. Scale-free networks [J]. Scientific American, 2003, 288(5): 50-59

[23] Granovetter M. The strength of weak ties [J]. American Journal of Sociology, 1973, 78(6): 1360-1380

[24] Burt R. Structural holes and good ideas [J]. American Journal of Sociology, 2004, 110(2): 349-99

[25] Easley D, Kleinberg J. Networks, Crowds, and Markets: Reasoning About a Highly Connected World [M]. Translated by Li Xiaoming, Wang Weihong, Yang Yunde. Beijing: Tsinghua University Press, 2011: 299-314 (in Chinese)
(大卫·伊斯利, 乔恩·克莱因伯格. 网络、群体与市场——揭示高度互联世界的行为原理与效应机制[M]. 李晓明, 王卫红, 杨韫得, 译. 北京: 清华大学出版社, 2011: 299-314)

[26] Castellani B, Hafferty F W. Sociology and Complexity Science: A New Area of Inquiry [M]. Berlin: Springer, 2009: 154-159

[27] Halpin B. Simulation in sociology [J]. American Behavioral Scientist, 1999, 42(10): 1488-1508

[28] Lazer D, et al. Computational social science [J]. Science, 2009, 323(5915): 721-723

[29] Han Jiawei, et al. Data Mining: Concepts and Techniques [M]. Translated by Fan Ming, Meng Xiaofeng. 3rd ed. Beijing: China Machine Press, 2012 (in Chinese)
(韩家炜, 等. 数据挖掘: 概念与技术[M]. 范明, 孟小峰, 译. 3 版. 北京: 机械工业出版社, 2012)

[30] Lugano G. Social computing: A classification of existing paradigms [C] //Proc of the 4th ASE/IEEE Int Conf on Social Computing, Piscataway, NJ: IEEE , 2012: 377-382

[31] O'Reilly T. What is Web2. 0: Design patterns and business models for the next generation of software [EB/ OL]. (2005-09-30) [2013-05-26]. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

[32] Ali-Hassan H, Nevo D, et al. Organizational social computing and employee job performance: The knowledge access route [C] //Proc of the 44th Hawaii Int Conf on System Sciences, Piscataway, NJ: IEEE, 2011: 1-10

[33] Tang Lei, Liu Huan. Community Detection and Mining in Social Media [M]. Translated by Wen Yimin, Bi Yingzhou. Beijing: China Machine Press, 2013 (in Chinese)
(唐磊, 刘欢. 社会计算: 社区发现和社会媒体挖掘[M]. 文益民, 闭应洲, 译. 北京: 机械工业出版社, 2013)

[34] Robles C, Benner J. A tale of three cities: Looking at the trending feature on foursquare [C] // Proc of the 4th ASE/IEEE Int Conf on Social Computing. Piscataway, NJ: IEEE, 2012: 566-571

[35] Golder S A, Macy M W. Diurnal and seasonal mood vary with work, sleep, and day length across diverse cultures [J]. Science, 2011, 333(6051): 1878-1881

[36] Wang Feiyue, Zeng Dajun, et al. A study of the human flesh search engine: Crowd-powered expansion of online knowledge [J]. IEEE Computer Society, 2010, 43(8): 45-53

[37] Zhang Qingpen, Wang Feiyue, et al. Understanding crowd-powered search groups: A social network perspective [J]. PLoS ONE, 2012, 7(6): 1-16

[38] IBM. IBM, Caltrans, and UC Berkeley aim to help commuters avoid congested roadways before their trip begins [EB/OL]. (2011-04-13) [2013-05-26]. http://www-03.ibm.com/press/us/en/pressrelease/34261.wss

[39] Ling K, Beenen G, et al. Using social psychology to motivate contributions to online communities [C] //Proc of CSCW2004, New York: ACM, 2004: 212-221

[40] Adam N R, Shafiq B, Staffin R. Spatial computing and social media in the context of disaster management [J]. IEEE Intelligent Systems, 2012, 27(6): 90-97

[41] Meng Xiaofeng, Ci Xiang. Big data management: Concepts, techniques and challenges [J]. Journal of Computer Research and Development, 2013, 50(1): 146-169 (in Chinese)
(孟小峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(1): 146-169)

[42] Hey T, Tansley S, Tolle K. The fourth paradigm: Data-intensive scientific discovery [M/OL]. Redmond, Washington: Microsoft Research, 2009. [2013-06-04]. http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[43] Schonberger V M, Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think [M]. Translated by Sheng Yangyan, Zhou Tao. Hangzhou: Zhejiang People's Publishing House Press, 2012 (in Chinese)
(维克托·迈尔·舍恩伯格, 肯尼思·库克耶. 大数据时代: 生活、工作与思维的变革[M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2012)

[44] Watts D J. A twenty-first century science [J]. Nature, 2007, 445(7127): 489

[45] Backstrom L, Boldi P, et al. Four degrees of separation [C] //Proc of WebSci12. New York: ACM, 2012: 33-42

[46] Watts D J. The "new" science of networks [J]. Annual Review of Sociology, 2004, 30(8): 243-270

[47] BianYanjie. The ten lectures of social networks theory [EB/OL]. (2007-11-09) [2013-06-04]. http://www.sociologyol.org/shehuixuedongtai/xinwengonggao/xinwengonggaoliebiao/2007-11-09/3907.html

[48] Xie Yu. Otis Dudley Duncan's Legacy: The demographic approach to quantitative reasoning in social science [J]. Research in Social Stratification and Mobility, 2007, 25(2): 141-156

[49] Xie Yu. Sociological Methodology and Quantitative Research [M]. Beijing: Social Sciences Academic Press, 2012: 1-31 (in Chinese)
(谢宇. 社会学方法定量研究[M]. 北京: 社会科学文献出版社, 2012: 1-31)

[50] Morstatter F, Huan L, Zeng D. Opening doors to sharing social media data [J]. IEEE Intelligent Systems, 2012, 27(1): 47-51

[51] de Montjoy Y A, Hidalgo C A, et al. Unique in the crowd: The privacy bounds of human mobility [J]. Scientific Reports, 2013, 3(1376): 1-5

Meng Xiaofeng, born in 1964. Professor and PhD supervisor at Renmin University of China. Executive member of China Computer Federation. His main research interests include cloud data management, Web data management, flash-based databases, privacy protection, etc.

Li Yong, born in 1979. PhD candidate at Renmin University of China. Student member of China Computer Federation. He is lecturer in Northwest Normal University. His main research interests include social computing, big data, etc.

Jonathan J. H. Zhu. Professor in the Department of Media and Communication at the City University of Hong Kong. His current research interests include the structure, content, use, and impact of social media and e-social science(j.zhu@cityu.edu.hk)