

用社会化方法计算社会

孟小峰 余力
中国人民大学

关键词：社会计算 Web2.0 社会网络 概念框架

引言

随着互联网的迅猛发展，特别是Web2.0理念的逐渐深入，越来越多的虚拟社会网络出现了，如微博（Twitter、新浪微博）、社交网（Facebook、人人网）、社会标注系统（Delicious、Flickr）、论坛（BBS）、维基（Wiki）等，这些虚拟社会网络聚集了大量用户。据2010年2月资料显示，全球最大的社会网络Facebook注册的用户已达4亿多，成为排在中国和印度之后的全球人口第三大社会^[1]，如图1所示。虚拟社会网络已经成为一种新形态的数字社会^[2]。

虚拟社会网络不但聚集了大量的用户，而且用户参与网络活动的深度和广度都得到了空前的提高。网络用户不再仅是信息浏览和接受者，也是互联网信息资源的提供和传播者。虚拟网络已成为继报纸、广播、电视之后的“第四媒体”。这种由大众创造的社会媒体（social media）详细地记录了用户的思想和行为轨迹，这使得利用计算技术观察和研究社会成为可能。

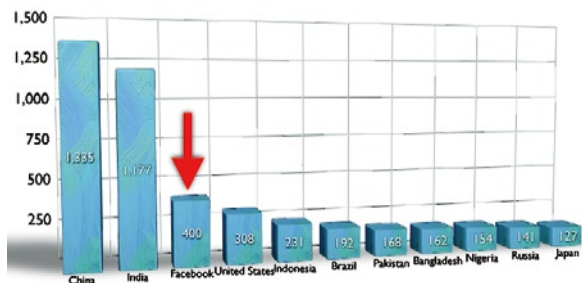


图1 Facebook成为全球第三大社会

同时，网络社会的虚拟性、用户匿名性、信息海量性、事件发展复杂性为虚拟数字社会的管理带来极大的挑战。特别是我国现处在经济转轨时期，各种事件频发，加强虚拟社会管理成为我国政府和社会管理的当务之急。

基于上述分析，社会计算出现的背景可以归纳为两方面，一是以用户为中心的Web2.0的思想得到广泛推广，逐渐产生了从个体行为到群体智能的社会化思维模式；二是许多虚拟社会网络的出现逐步形成了一种新型的数字化社会形态。

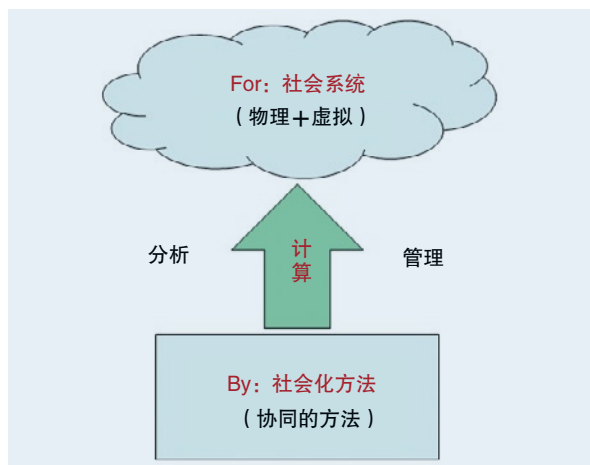
概念框架

社会计算作为一个新兴的跨学科研究领域，目前还没有一个公认的定义。但也许我们可以从社会计算出现的背景去剖析它的概念。基于这个认识，社会计算实际上可以简单地概括为“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”。如图2所示。

“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、正在发生、将要发生的事情，准确地了解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参考。

“社会化方法”是一种以草根用户为中心、

依靠草根用户的用户化方法，一种协同和群体智能的方法，也是一种从个体到整体，从微观到宏观的思维模式。许多事件都是由无数网民“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的工程。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，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。



下面分别从对象视角和方法视角两个方面简述社会计算的内涵。

从对象的视角：为社会计算

对象的视角表明，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。前者主要指传统意义上的社会，如某个国家或地区；后者主要指基于万维网的虚拟网络社区。从广义上来讲，整个互联网就是一个虚拟网络，但从狭义上来讲，虚拟网络主要指基于Web2.0的，强调以用户为中心的虚拟社区，如Facebook、Twitter等虚拟网络。但需要指出的是，尽管从社会形态可以分为现实物理社会和虚拟网络社会，但两者又是紧密相关的。虚拟网络社会是对现实物理社会的反映，研究虚拟网络社会的最终目的还是为现实物理社会的管理服务。

从本质上来讲，社会是人的社会，人是社会的人，社会是由若干个体人组成的一个系统。所以，从对象的视角来看，社会计算又包含两个层面的内容，即宏观社会层面和微观用户层面。

在宏观社会层面，社会计算的主要任务是通过对社会网络舆情的分析^[3]，回答“发生了什么事？大家对这些事观点是什么？将会怎么发展？如何引导和控制？”。在这个层面，研究的主要问题包括以下几点。

网络舆情搜集 网络舆情信息的收集是舆情分析的前提^[4]。目前舆情收集的方法主要分为两类，即基于内容分析的方法和基于网络计量法。前者主要通过对网页文本内容进行分析，具体实施上可以基于某一特定事件或话题上搜索公众的观点和看法，或基于某个人或组织搜索其对相关事件或话题的观点或看法。网络计量法就是将收集到的各种数据信息进行科学统计、分析、总结出一般规律，是一种基于对网络信息外在量的分析方法，包括利用链接分析法确定网络舆情的来源与走向；利用统计分析法收集各种网络舆情数据，如网站访问量、访问者的特征及其度量、网民的分布情况等；利用图论分析法研究网络链接以分析网络舆情的变化趋势。

舆情观点分析 舆情观点分析主要研究网民对各种事件的看法^[5]，包括观点分类及倾向性分析。观点分类是把观点文本信息进行分类的问题，倾向性分析是对用户观点所要表达话题事件的立场、观点和态度进行分析，分析出观点表达者对于话题事件的态度是积极的、赞同的，还是消极的、反对的，或者是中立的。观点倾向性分析常采用语料的方法，通过人工建立语料库，赋予语料库中的每种模式一个语义倾向性值，然后遍历语料库，分析其是否与待分析文本匹配，并累加所有匹配模式对应的倾向性值，把累加的倾向性值作为整个文本的倾向性。倾向性分析的核心是判断词汇的倾向性，一般先要抽取文本中能够体现主观色彩的情感词，然后对抽取出来的情感词进行倾向性判断。

舆情演化分析 舆情演化分析主要研究话题或事件是如何演化发展的^[6]，包括话题热度和事情演

化。话题热度主要指话题受社会公众的关注程度。从社会学角度来看,舆情话题的产生与群体行为的选择、社会大众的关注有密切关系。因此,网络舆情话题的热度要重点考虑用户的关注度。最简单的方法就是基于话题报道频率来反映话题关注度。舆情的演化过程相当复杂,既是话题的动态变化过程,又是话题多维属性的动态变化过程。许多研究在语言模型中嵌入时间变量来研究话题的演变。实质上,舆情演化是一个复杂的动态网络,基于动态网络模型来研究互联网舆情的演化,不仅能够帮助人们观察和理解舆情演化的动态过程,还可以揭示出舆情演化的微观结构和演化机理。

从微观用户层面来讲,社会计算主要关注的是如何促进用户与用户的交互,以及通过用户交互表现出来的用户社会影响分析。

用户交互研究 无论是Web2.0还是Facebook, Twitter等虚拟社会网络系统,其最大的特点就是强调用户与用户间的交互,实现的是人与人的互联。如何促进人与人的交互是社会计算研究的另一重要内容。笔者认为,随着Web2.0理念的深入,交互的重点已经从传统的人-机交互(HCI)转化为人-人交互(HHI)。传统人-机交互强调的是通过设计人员对系统形式和功能来控制来优化软件应用及界面以增加系统的友好性。而人-人交互更注重如何实现人与人互联,信息交换与知识共享。在不同的应用领域,人-人交互的模式不同,如在微博中,交互方式包括跟贴、回复、粉丝等,在人际关系网中,人-人交互一般显性表现为加某某为好友。目前有少数学者从信息系统行为角度对社会网络信息交互模式^[7]的基于web2.0的信息生成模式、Web2.0环境下知识共享问题进行研究。

用户影响分析 通过用户间的交互(回复、跟贴、加为好友等),用户与用户间形成一定的影响关系,并会逐步形成社会网络中的影响力^[8]。用户影响力分析主要研究如何基于用户的交互活动水平(activity level)来研究用户与用户是如何影响的,以及用户在社会网络中的影响力大小。具体来讲,影响分析包括影响关系分析和影响力分析。影响关

系是如何刻画两用户间的影响关系,影响力分析主要度量用户在社会网络中的重要程度。有不少学者基于博客论坛行为(提交、评论等),采用Web挖掘的方法和语义分析的方法研究博客论坛用户的影响力。从模型来讲,度量用户影响力通常表现为寻找最重要的节点,目前主要有两类:第一类是基于最小路径(geodesic path)的方法,如距离中心(closeness centrality)、图中心(graph centrality)、中介中心度(betweenness centrality)等;第二类是基于拓扑结构的方法,包括基于马尔科夫的方法(如PageRank算法、HITS模型)、度中心(degree centrality)的方法、基于路径的方法(如 α -centrality, SenderRank等)。

从方法的视角:用社会化方法计算

所谓“社会化方法”就是用户协同工作的方法,就是Web2.0的模式。从这个角度来讲,社会计算是一种计算模式。随着大量社会网络的产生,以Web2.0思想为核心的社会计算模式正逐步应用到诸多领域,如在个性化推荐、资源分类、电子商务、市场营销等,并相应地产生了许多全新的标志性研究方向,如社会(或协同)推荐、大众分类、社会商务、病毒营销,如表1所示。

协同推荐 协同推荐(social/collaborative recommendation)是目前个性化推荐中最重要的方法^[9],其基本思想是,具有相似兴趣爱好的用户会对同一商品表现出相似的偏好。根据此思想,要对用户进行推荐,先要确定其邻居(相似)用户,然后再综述其邻居的偏好。所以对某一用户的推荐,实际是取决于其他用户的偏好,用户与用户之间形成一种自助式、协同式的社会推荐模式。相反,传统推荐方法大多是基于内容的,首先需要对用户和候选推荐对象分别进行建模表示,然后把用户与产品进行匹配。也就是说,对某一用户的推荐,不需要参考其他用户的偏好信息。

大众分类 随着知识与信息的交叉,传统的基于固定分类目录的方法难以满足爆炸式信息增长的

表1 社会计算应用模式

社会计算模式应用	应用领域	应用特点	传统计算模式的特点
协同推荐	个性化推荐	推荐的产生由全体用户决定；用户自助式推荐；	寻找与用户偏好匹配的对象；
大众分类	资源分类	没有事先的分类标准；通过用户协同标注，形成资源信息的自动分类	分类标准及类别事先确定；
社会商务	电子商务	强调以用户为中心的商务模式	强调用户与商家的交互与便利；
病毒营销	市场营销	营销靠用户驱动；用户不但是购买者，更是营销者	用户仅是购买者；营销的成功取决于营销者的个人能力以及信息的不对称；

需要。大众分类（folksonomy）打破了传统目录分类的界限，资源目录的分类不再固定，而是由广大用户决定^[10]。大量用户使用标签（tag）对资源对象（如网页、文献、图片、视频、音乐等）进行标注（tagging），且由于用户标注时可以参考其他用户的标注，所以这种标注是一种协同的方式进行。用户的协同标注形成一种对资源对象的自动分类，这种分类模式需要大量用户的参与（标注），具有很强的社会性，是社会计算模式在资源分类中的应用。

社会商务 社会商务（social commerce）是近年来出现的一种商务模式^[11]，是社会计算模式在电子商务中的应用。尽管目前还没有形成对社会商务公认的概念，但对社会商务有一个基本共识，即社会商务是更加突出和强调以用户为中心的理论，用户不仅是一个购买者，还需要参与商务活动，如产品设计、产品定价、产品购买、售后服务等各环节。可以认为，在某种意义上凡是能体现用户参与商务活动的都属于社会商务。典型的社会商务模式如团购、拍卖（Penny Auction）等。传统的电子商务仅体现在基于网络平台的商家与用户的交互上，包括商务信息浏览、网上购买、网上支付等，缺乏用户与用户的交流互动。

病毒营销 病毒营销（viral marketing）实质上是社会计算模式在市场营销中的具体应用^[12]。在传统的营销模式中，营销员为主体，营销的成功与否在很大程度上取决于商品信息的透明度，以及营销员个人的能力。但在信息越来越透明的今天，用

户与用户可以非常方便地进行交互，信息越来越透明，新的营销模式必须适应这种新的环境要求。病毒营销基于社会网络的n-to-n的网络营销模式，其核心理念是依靠用户去营销用户，用户不仅是消费者，更是企业营销的主体，传统营销员的概念被拓展到用户，并通过用户的“口口相传”（word-of-mouth）进行营销，是一种社会营销的模式。

除上述社会计算应用模式外，典型的应用如维基百科，可看作社会计算模式在百科词典编辑方面的应用。

中国人民大学社会计算平台

2010年，中国人民大学信息学院成立社会计算研究小组，参与和组织了多种形式的社会计算研讨会。针对社会计算跨学科的特点，提出了社会计算通用实验平台框架，并对平台的各个层次的关键核心问题进行研究。社会计算实验平台是一个集成数据分析与处理、建模与仿真、预测与监控功能为一体的通用社会计算实验平台，平台由对象层、数据层、模型层、分析层、应用层五部分组成，如图3所示。

对象层 对象层是社会计算的研究对象。根据社会计算的概念，从对象视角来看，社会计算是要为社会进行计算。从根本上看，社会计算要为现实社会服务；从逻辑上看，主要包括信息网络社会和复杂经济社会，以及从中抽取出来的人工社会。

数据层 随着信息技术的发展，特别是Web2.0技术的发展，用户不但可以从互联网中获取数据信



图3 社会计算实验平台框架

息，参与互联网的程度也得到空前的提高，留下了大量社会“足迹”。收集这些社会足迹进行集成，

可为社会计算提供基础的数据信息。研究小组重点围绕海量多源数据集成、多源数据融合以及数据质量等海量数据信息处理的核心关键问题进行研究。

模型层 对系统的微观建模是社会计算的前提和基础。通过建模，对社会计算系统中涌现的宏观现象或情景进行生动、形象的展示或可视化，对实验参数的可控化进行不同的情景再现，并按研究需求进行适时调整以方便决策者及时讨论形成有效的决策方案。研究小组重点研究基于主体的人工社会经济系统建模问题。

分析层 社会系统行为是由个体交互而成，通过对个体行为分析研究系统内部结构是社会计算的理论基础。分析层将从基础层获取数据信息，并利用模型层建立的系统模型，对系统结构进行微观分析，为社会计算应用研究提供基础。研究小组重点从两方面结合社会系统的结构特征进行分析，包括社会网络社区结构分析和社会群体行为分析。

应用层 作为社会计算平台的最高层，应用层是把社会计算的理论与方法应用到实际的社会、经济系统中，为政府、企业等相关部门决策提供依据。研究小组重点对社会媒体传播、公共管理、社会商务、知识管理等领域进行研究。

CCF@U-92：沈阳工业大学

2011年10月28日，“CCF走进高校”系列活动在沈阳工业大学举办，CCF常务理事、清华大学林闯教授作了精彩演讲。沈阳工业大学信息科学与工程学院党总支书记王溪波教授主持。

林闯是CCF常务理事，奖励委员会成员，清华大学计算机科学与技术系教授，系学术委员会主任，2003年10月至2007年4月任系主任。主要研究领域为计算机网络，系统性能评价，安全分析，随机Petri网。2005年度获教育部提名国家科学技术奖自然科学一等奖，2007年度获北京市科学技术奖二等奖，2010年获中国电子学会信息科学技术奖一等奖。

沈阳工业大学是一所以工为主，涵盖工、理、经、管、文、法、哲等七大学科门类的多科性教学研究型大学。学校始建于1949年。1985年由沈阳机电学院更名为沈阳工业大学，原为国家机械工业部所属院校，1998年起由中央和地方共建，以辽宁省管理为主。学校由中央校区、兴顺校区、国家大学科技园和辽阳校区组成。中央校区（主校区）位于沈阳市经济技术开发区。学校具有学士、硕士、博士三级学位授予权。

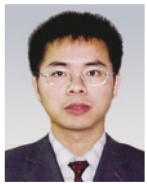
结语

社会计算是近年来随着Web2.0和虚拟社会网络发展而出现的一个新兴研究领域,本文基于社会计算出现的背景,提出了社会计算的概念模型,即社会计算是用社会的方法计算社会,该概念模型较好地统一了对社会计算的各种理解。中国人民大学社会计算研究小组从对象层、数据层、模型层、分析层和应用层构建了社会计算实验平台,并对不同层的关键问题进行研究,最终将形成一个集社会数据信息抽取、集成和分析为一体的实验平台,支持和服务社会科学。■



孟小峰

CCF理事,副秘书长,本刊编委,2009年CCF王选奖一等奖获得者。中国人民大学教授。主要研究方向为数据库技术,Web数据管理,XML数据管理,云数据管理等。xfineng@ruc.edu.cn



余力

中国人民大学副教授。主要研究方向为网络营销、电子商务、Web2.0与社会计算等。buaayuli@ruc.edu.cn

参考文献

- [1] Irwin King, Introduction to Social Computing, DASFAA2010 Tutorial, Tsukuba, Japan, April 1, 2010
- [2] 王飞跃, 社会计算与数字网络化社会的动态分析, 科技导报, 2005年9期
- [3] 耿骞, 刘颖. 网络舆情的形成、传播及监测方法[A]. 科学发展: 文化软实力与民族复兴——纪念中华人民共和国成立60周年论文集(下卷)[C], 2009
- [4] 许鑫, 章成志. 互联网舆情分析及应用研究[J]. 情报科学, 2008, (08)
- [5] 姜胜洪. 网络舆情热点的形成与发展现状与舆论引导[J]. 理论月刊, 2008 (4): 34~36
- [6] 朱恒民, 苏新宁, 张相斌, 互联网舆情演化的动态网络模型研究, 情报理论与实践, 2010-10
- [7] 李玉梅. 社会性网络信息交互模式系统分析[J]. 情报科学, 2010, (08)
- [8] R. Ghosh and K. Lerman. Predicting influential users in online social networks. In the fourth SNA-KDD Workshop Held in conjunction with ACM SIGKDD2010 (SNAKDD'10), May 2010
- [9] Resnick and Varian. Recommender systems. Communications of the ACM [J], 40(3):1997, 56~58
- [10] 苏中, 张小洵, 吴贤, 张俐, Web 2.0下的社会标注, Communications of CCF, 2008/10
- [11] Amir Afrasiabi Rad, Morad Benyoucef, A Model for Understanding Social Commerce, 2010 Conference on Information Systems Applied Research, Nashville Tennessee, USA
- [12] A.D. Bruyna and G.L. Lilien. A multi-stage model of word-of-mouth influence through viral marketing. International Journal of Research in Marketing, 25(3) : Sep. 2008, 151~163