

CCF 推荐 B 类国际学术会议介绍

第24届国际万维网大会精彩纷呈

张永锋 刘奕群 唐杰
清华大学

关键词：互联网 国际万维网大会

第24届国际万维网大会(International World Wide Web Conference, WWW)于2015年5月18~22日在意大利佛罗伦萨召开,本次大会共吸引了来自学术界和企业界逾1400人参加,受到了极大的关注。此次会议精彩纷呈、亮点不断,如首届“十年最佳论文奖”授予了谷歌的联合创始人谢尔盖·布林(Sergey Brin)和劳伦斯·佩奇(Lawrence Page)的论文,会议投稿量创历史新高,中国作者的论文录用率位居前列等。

会议概况

本次大会共收到来自70多个国家和地区的929篇论文投稿。来自全球的554位研究人员组成了程序委员会,对投稿进行了严格的评审,共产生超过3000条评审意见。以论文第一作者所属机构统计,美国、中国、德国为录用论文最多的三个国家。其中来自中国大陆的论文有14篇,作者分别来自清华大学、北京大学、南京大学、中国科学技术大学、中国科学院自动化研究所、首都师范大学、东南大学、苏州大学、百度、微软亚洲研究院等,其中第一作者为清华大学的论文有3篇。此外,国防科技大学、北京邮电大学、天津大学等单位的论文和工作进行了海报展示。

本次大会还吸引了7个合办会议,并同时举办了26场研讨会、14场辅导讲座。会议还特设工业专题(industry track)和网络科学专题(Web sci-

ence track)两个环节。工业专题环节着重报告和展示来自工业界研究人员的最新研究成果和产业应用,网络科学专题则侧重报告和展示互联网与其他领域的交叉学科研究成果。这不仅能让我们了解当前互联网产业应用的最新进展,也让我们能够密切追踪当前互联网科学研究新的发力点。

研究论文

国际万维网大会涉及互联网研究、应用与标准化的诸多方面,是互联网学术界和产业界的研究人员进行交流、思想碰撞与合作的重要平台。为了确保走在互联网研究和应用的最前沿,每年都会根据当前互联网相关领域研究现状对会议的领域分布进行适当的调整,以反映当前研究的重点、热点和难点。本次会议设立了8个研究领域,分别为社交网络与图分析、用户行为分析与个性化、内容分析、语义网络、网络信息挖掘、网络检索与应用、网络安全与隐私、互联网经济与商业化。

在社交网络分析方面,随着微博、推特等开放式社交网络,脸谱等半开放社交网络以及近年来兴起的微信等私密社交网络的迅速发展和向日常生活的渗透,该领域的研究也不断涌现出新的重要课题。例如,在网络结构分析方面,如何应对大数据环境下的超大规模动态网络成为一个重要问题,并进一步衍生出超大规模动态网络下的小世界挖掘、密集子图挖掘、搜索、推荐等研究课题,既在理论上

难度，又在实践中有很高的重要性。此外，大规模网络中的采样、近似、剪枝、压缩等面向效率的相关研究也占据了重要位置。

用户行为分析领域的研究内容广泛分布并应用于社交网络、推荐系统、主题发现与聚类、个性化搜索等多个研究课题中，在本届会议中继续受到关注。除了传统的搜索引擎、社交网络、电子商务等领域的用户行为分析相关研究之外，本次会议论文还对研究较少的新领域和新场景中的用户行为分析、建模和预测进行了尝试，包括电子邮件系统、网络论坛、在线音乐等。

与往年不同的是，本届会议新增了众包与社交媒体、普适网络与移动计算、网络系统结构三个领域。其中，众包与社交媒体领域从今年开始独立出来，其论文录用量达到20篇，该领域的研究热情和质量由此可见一斑。本届的网络系统结构领域蜕变自往届的“软件架构、性能、扩展性与可用性”领域，不同之处在于本届会议更为明确地将其描述为“网络系统结构：数据中心、内容传输网与云计算”。这表明相关领域研究的概念和问题变得越来越清晰和具体，与互联网研究的关系也越来越明确，值得引起研究人员更多的关注和思考。

主题报告

本届大会邀请了安德烈·布罗德 (Andrei Broder)、狄波拉·艾斯汀 (Deborah Estrin) 和珍妮特·霍夫曼 (Jeanette Hofmann) 三位在学术界和工业界有较高声誉与影响力的学者为大会作主题报告。

安德烈·布罗德是谷歌的杰出科学家。他在报告中分享了过去二十年来关于互联网的主要预测中哪些已经成为现实，而哪些被证明是错的，从而帮助我们看清未来互联网的发展趋势。

狄波拉·艾斯汀是康奈尔大学纽约分校教授、雅各比健康技术研究中心的创办者。与当前火热的大数据相关研究不同，艾斯汀着重介绍了基于单个用户的小数据在描绘用户个性化时间线上的重要作用，并以可穿戴设备收集的健康数据为例，

介绍了健康小数据分析为用户和开发者双方带来的便利。艾斯汀指出，不同于大数据分析相关技术，特定领域的小数据分析需要较强的专业知识辅助才能更好地得出具有参考意义和可理解的结论。因此，从小数据分析技术入手，构建平台并将用户和开发者连接起来，具有重要的意义，也会带来潜在的机会。

珍妮特·霍夫曼是柏林社会科学研究中心的政策科学家、伦敦政治商学院研究员。其报告题目为“信息化的困境”，探讨了在数字信息时代互联网服务快速扩张带来的新难题，而这些难题更多地隐含着互联网信息的全球化所带来的逻辑和伦理上的悖论。例如，大众自由获取信息想法与希望互联网忘掉和移除个人信息之间的矛盾。该矛盾近年来为许多起案件在法律诉讼和判决上带来了困难。这些困境的破解或将成为引导互联网伦理和互联网逻辑在未来数十年发展的重要风向标。

最佳论文

本届会议的最佳论文为“HypTrails: A Bayesian Approach for Comparing Hypotheses about Human Trails”，作者分别是德国莱布尼兹社会科学研究所的飞利浦·辛格 (Philipp Singer) 和马克思·斯特罗迈尔 (Markus Strohmaier)、奥地利格拉茨技术大学的丹尼斯·海利克 (Denis Helic) 以及德国维尔茨堡大学的安德里斯·霍托 (Andreas Hotho)。用户在使用互联网应用的交互过程中会积累大量的序列信息，例如网页浏览序列、商品点击序列、在线音乐收听序列等，正确理解驱动这些序列产生的内在机制有助于我们理解网络结构、预测用户行为、提升个性化推荐效果等。本项工作在马尔科夫过程的背景下研究了驱动用户交互过程中状态变化的几种常见的假设，包括均一概率假设、自我循环假设、相似性假设、结构性假设、地理位置假设以及时间相关性假设等。该论文提出的框架可以将先验假设表示为状态转移的概率，从而使借助统一的框架来表述先验假设成为可能。

十年最佳论文奖



本次大会的一大亮点是颁发了 WWW 首届十年最佳论文奖 (The Seoul Test of time Award), 以表彰 WWW 历届论文中那些经得起时间检验、被认为最有价值的论文。

首届十年最佳论文奖授予了谷歌的联合创始人、美国工程院院士谢尔盖·布林和劳伦斯·佩奇, 用以表彰他们在 WWW 1998 上发表的论文 “The Anatomy of a Large-Scale Hypertextual Web Search Engine” 所带来的开创性成果。该论文描述了一种大规模的网页检索系统, 即当今搜索引擎系统的前身。今天, 搜索引擎已经深入到互联网用户的日常生活中, 甚至改变了人们的生活方式, 成为互联网历史上最重要的发明之一, 也成就了谷歌、百度等互联网公司巨头, 产生了巨大的社会影响。

未来展望

随着传统互联网应用的进一步发展成熟和互联网新应用的不断涌现, 围绕互联网技术的研究也呈现百花争放的态势。一方面, 社交网络、用户行为分析、网络内容分析等领域继续得到深入的研究与关注; 另一方面, 许多新的研究子领域正逐渐在互联网研究领域变热并具备更强的独立性, 如众包、可穿戴互联网、内容传输网等。值得注意的是, 互联网理论相关的研究变得更为 “社会化”, 即关注

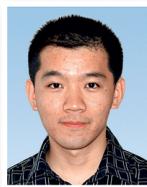
用户互联网行为的社会学意义, 这由本次大会的主题报告和本次最佳论文可见一斑。另外, 互联网安全与隐私领域的投稿数量不多, 而录用数量却占较大比例, 由此看出领域内专家对相关事件和日益凸显的网络安全问题十分重视。

目前互联网的发展已不仅仅局限于传统的桌面端联网形式, 基于移动设备的互联网正以飞快的速度融入人们的日常生活中。移动互联网的发展为网络科学的研究带来了新的机遇和挑战: 如何从网络结构方面对超大规模动态的移动网络进行建模和分析, 是区别于传统桌面终端网的新问题。另外, 随着移动互联网的发展, 逐渐积累了大规模的用户行为数据。这些信息往往是结构化或半结构化的, 其个性化特点更突出, 也具有更强的隐私性。这对移动端搜索、个性化推荐、网络内容变现等关乎互联网生态圈平衡的重要问题提出了挑战, 如何对这些数据进行有效的检索、整理、分析、建模和挖掘是其中的关键。在本届大会上, 这方面的研究在以可穿戴设备、移动网络数据挖掘等为主题的论文中有较强体现, 对于以智能化为核心特征的下一代互联网的研究建设与发展具有重要意义。■



张永锋

CCF 学生会员。清华大学博士生。主要研究方向为个性化推荐、情感分析、计算经济学。
yongfeng14@mail.tsinghua.edu.cn



刘奕群

CCF 高级会员。清华大学副教授。主要研究方向为互联网信息检索、自然语言处理、智能信息处理。
yiqunliu@tsinghua.edu.cn



唐杰

CCF 杰出会员、CCF 青年科学家奖获得者、本刊编委。清华大学副教授。主要研究方向为 Web 社会网络挖掘, 语义 Web。
jiatang@mail.tsinghua.edu.cn