

海外学人

2012年第3期 ISSN: 1556-861X

信息技术与安全

专刊



U.S.A.

中國旅美科技協會

Chinese Association for Science and Technology, USA (CAST-USA)

www.cast-usa.net

中国旅美科技协会名誉顾问

- | | | | |
|-----|----------------------|-----|----------------------|
| 杨振宁 | 清华大学教授、诺贝尔奖获得者 | 路甬祥 | 中国科学院前院长、全国人大常委会副委员长 |
| 周光召 | 中国科协名誉主席、中国科学院前院长 | 宋健 | 中国工程院前院长、全国政协前副主席 |
| 朱丽兰 | 中国科学技术部前部长 | 陈香梅 | 美国国际合作委员会主席 |
| 邓文中 | 美国工程院院士、美国林同琮国际公司董事长 | | |

中国旅美科技协会机构与主要负责人

总会会长
盛晓明

董事会主席
方 彤

理事会主席
曾大军

预备会长
蔡逸强

总会副会长

- | | | | |
|-----|------------------|-----|-------------------|
| 于 浩 | 协助会长处理科协日常事务 | 庞 涛 | 负责学术活动（学术协调委员会主席） |
| 郭 光 | 负责新闻通讯及杂志（新闻简报） | 陈志雄 | 负责组织学术会议 |
| 秦学文 | 负责企业交流与活动 | 沈 琦 | 负责对外宣传 |
| 汪 洋 | 负责会员发展 | 林馨思 | 负责战略规划 |
| 蒋为民 | 负责公共关系-美国事务 | 贾鹏程 | 负责公共关系-中国事务 |
| 谢长春 | 负责东部各分会工作的协调 | 彭 伟 | 负责中西部各分会工作的协调 |
| 潘星华 | 负责总会东部赞助事宜 | 程晓红 | 负责总会中西部赞助事宜 |
| 范 军 | 负责总会在国内的发展与活动 | 高梅梅 | 负责财务 |
| 陆 强 | 《海外学人》主编，顾问委员会主席 | 张 宣 | 总会法律顾问 |

海外学人

Overseas Scholars
OVERSEAS SCHOLARS

2012年第3期
ISSN: 1556-861X

本期责任编辑：陈志雄 曾大军 汪洋
发 行 人：曾大军

《海外学人》编辑部

名誉主编：王飞跃
主 编：陆 强
执行主编：郭 光
编 辑：盛晓明 蔡逸强 陈志雄
王志梁 胡增建
美术/摄影：徐 珺
公共关系：蒋为民
专刊审稿人：方 彤 贾鹏程 陆 强
庞 涛 宋德强 宋云明
谢长春 张永志 张良杰
期刊排版：汤淑芳

杂志主办

中国旅美科技协会
Chinese Association for Science and
Technology, USA (CAST-USA)
www.cast-usa.net



中国旅美科技协会简介

中国旅美科技协会（旅美科协）是1992年夏在纽约成立的非政治性、非盈利性的民间团体。旅美科协的宗旨是促进旅美学人的信息交流，事业发展；加强团员友谊；发扬中华传统文化；增进中美两国人民之间的了解；推动中美在科技、经济、文化方面的交流与合作。科协会员主要来自科技、文教、工程、法律、金融、人文等各个领域的中国旅美专业人士组成。现有注册会员近五千人，非注册会员三千余人。协会由地区分会和专业学会组成，总部设在纽约。目前在全美有十多个分会及学会，会员分布在几十州。总会设执行委员会负责日常工作，旅美科协还设有董事会、理事会、学术活动协调委员会和顾问委员会。

Table of Contents

1	《信息技术和安全》专刊前言	陈志雄 曾大军 汪洋
3	社交媒体分析与智能研究进展	崔凯楠 郑晓龙 曾大军 王飞跃
14	云管理：新型信息技术创新企业的管理模式.....	张良杰
26	信息技术在另类资产管理中的应用	杨 波
34	从服务质量的角度来看待社交网的挑战和机会	周念军
42	中国旅美科技协会大纽约分会简介	方 彤
44	第三个重要信息空间：车联网技术与应用分析	程久军 周孟初
58	微型飞行机器人在多障碍物复杂环境的自主导航方法研究.....	肖继忠
73	ACE and IEEE 简介.....	陈志雄
74	破茧成蝶：商务智能简史.....	涂子沛
92	提升金融 IT 技术对加快中国国际金融中心建设及人民币国际化的影响.....	邹云峰
96	预测市场：一种新兴的预测方法	陈伟运 李 昕 曾大军
112	面向服务的科学：以癌症网格为例	谭 伟
117	An overview on Pervasive Service Computing	Jiehan Zhou
126	哥伦比亚大学简介	沈 琦
127	支援移动保密与隐私实施模型.....	Patrick C. K. Hung
136	云计算时代的信息保证以及政府的作用	陈志雄

《信息技术和安全》专刊前言

陈志雄, 曾大军, 汪洋¹

旅美科协的《海外学人专刊：信息技术和安全》终于在旅美科协二十周年这个值得庆贺的纪念日发表了。这个专刊收集的近 15 篇文章在移动技术 (Mobile Technology)、社交媒体 (Social Media) 和信息安全与保障 (Information Security and Assurance) 等几个方面展现了现代信息技术和信息保障研究、应用的现状和发展趋势。不断进步的计算资源虚拟化技术和大数据处理技术为云计算提供了计算环境和数据处理方面的支撑, 使得许多大数据采集和处理在诸多重要领域, 如公共教育、公共卫生和金融交易和安全, 应用成为可能, 也直接促进了无线移动、网络社交和网络安全与保障技术的迅猛发展和应用。

首先, 智能终端, 尤其是智能手机, 已演变为一个让人们难舍难分的私人助手, 远远超出了一般意义下的通话呼叫功能。用户利用它来发短信, 规划每天的行程, 还用它来作为随身护士, 针对测试到的生命指标, 提供建议, 比如食物选择、体能锻炼和最佳的作息时间, 并将这些数据安全地传送到自己的家庭医生或专科医师那里以供远程诊断。在层出不穷的智能手机应用中, 最具有挑战的问题之一是身份和身份鉴定以及移动安全付款功能。智能手

机最终可以替代我们的身份证, 而有效和实时的身份鉴定功能是这种应用的前提。并且, 它可以携带充分多的个人信息, 根据不同的应用要求以显示不同级别的个人信息, 从而提供进一步的安全保障。

安全的付款功能可以替代我们每天携带的信用卡。当然, 它是建立在有效的电子身份及身份鉴定的基础之上。这个应用使日常生活如旅游和银行业务更加便利。比如, 我们可以不必经过营业员就可以购买货物, 并通过无线网银进行付款。

许多与实时地理位置有关系的应用 (Location-Aware) 问题, 也可以通过智能手机得到解决。比如, 在驾驶时, 我们可以及时知道交通的通行状况并进行实时线路调整。许多客户应用也可以根据实时地理位置的变化提供相应的服务。

越来越多的企业应用也被移到了智能手机上。比如, 证券市场的抄手们可以利用智能手机在任何地点和任何时间进行交易。著名的 Bloomberg 交易终端就已经被移植到蓝莓手机上, 尽管其使用年费高达七万美元以上, 抄手们还是愿意支付这样昂贵的费用以便进行移动交易。大型的博物馆已经开始用智能手机进行门票扫描。尽管门票类型愈趋多样化, 比如网上购买的电子票、VIP 赠票、现场购买的门票等, 智能手机可以在扫描这些门票的同时很容易地区别其类型, 并提供相应的服务, 比如给予 VIP 票的持有者以应有的特殊服务。越来越多的机场也开始了智能手机机票扫描机检。

¹陈志雄博士(旅美科协总会副会长, 大纽约分会理事会主席, zxchen@ieee.org), 曾大军博士(旅美科协总会理事会主席, zeng@email.arizona.edu)和汪洋博士(旅美科协总会副会长, 匹兹堡分会前主席, ywg@psc.edu)受旅美科协总会和海外学人学刊编委的委托, 一起组织和编辑了这期信息技术和信息专刊。同时, 这期专刊也是庆祝旅美科协成立 20 周年各项活动的一部分(<http://convention.castusa-gny.org/>)

除了智能手机，移动智能技术也体现在无人驾驶飞行器和机器人应用中。利用现代人工智能以及快速处理器，这些移动机器人可以完成许多人们所无法完成的任务。

其次，网络社交，或社交网、社会媒体，也颠覆了我们传统的截取信息和与人交往的模式。通过社交媒体，我们可以与人共享新闻包括个人的心得，并比传统的媒体更快地得到最新的信息。这是一种主动的信息获取而不是被动地被灌输。但是，社交网上的信息在很大程度上是可以随意发布而不必负责任，使得信息质量不能保障。许多谣言、蛊惑性的信息甚至可能影响到社会的稳定。网络社交诚信也是一个具有极大挑战的研究课题。

广义地讲，车联网也是一种社交网的表现和应用。撇开它的技术层面，从车联网得到的信息可以进行动态自治系统的分析和研究。

许多企业也开始进入以前主要为私人服务的社交网。利用社交网的特定群体，企业可以进行有效的广告宣传，推出特定的服务，与顾客进行互动。比如，花旗银行推出了它的顾客可以将自己积累的感谢点 (thank you point) 贡献给某些社会团体来推动社会公益活动。

第三，我们对于数据的认识又上升到了一个新的台阶。现在，每天产生的信息量在 2.5 个艾未数 (exabyte)。而最近两年的数据量比以前所有的数据量的总和都要多出几倍。伴随着大数据的产生是对大数据的解析和处理的要求。实时、实用是对任何解析处理方式的基本要求。Hadoop 并行处理技术和相应的 mapReduce 在这

个大数据时代适时成长，伴随着数据管理技术不断地向商业决策方向发展。同时，以服务为架构的中间件以及以松散型结合的应用软件也使得计算资源共享，存储共享和开发共享的云计算的抽象表述有了可靠的基石。

第四，信息安全和信息保障的重要性已经越来越为人们所理解和接受。不再是纸上谈兵或是可有可无。我们看到，在国家层面，美国政府已经动用了大量的行政资源来教育国民和企业理解信息安全和信息保障的重要性，培养专业技术人才和制定信息保障的各项标准。在企业层面，信息安全和保障也成为企业文化的一部分，不再是侧重发展安全的新技术，而是要把已有的安全技术植入到企业的方方面面。在研究领域，移动安全和云安全已是人们关注的重点。

最后，我们看到信息技术在医学、公共卫生、证券交易和教育的应用也越来越复杂。可以预见，将会有许多与目前不同的模式出现在大规模的公民普及教育中。Udacity 大学就是在这个领域一个令人瞩目的尝试。为了降低政府在医疗方面的支出和增强国民健康，美国政府已经把医疗信息电子化作为医疗改革的重点。而关系到国家金融安全的金融界更是信息技术应用的中心。几乎所有的交易已是机器化和智能化。不断衍生的金融新产品、高频率交易和就近 (colocation) 交易等等对信息技术提出了相当高的要求。

信息技术和信息安全已经深入到我们生活和工作的方方面面，有形和无形的计算无处不在，希望这部专辑能抛砖引玉。

社交媒体分析与智能研究进展

崔凯楠^{1,2}, 郑晓龙², 曾大军², 王飞跃²

(1. 西安交通大学电子与信息工程学院, 西安, 710049;

2. 中国科学院自动化研究所 复杂系统管理与控制国家重点实验室, 北京, 100190;)

摘要

随着互联网与移动通信技术的飞速发展, 社交媒体对信息的生成、传播与消费方式产生了巨大的影响。这种深刻的变革为学术研究与技术应用带来了无限机遇与巨大挑战。近年来, 社交媒体分析与智能已受到越来越多人的关注, 目前正成为一个新兴且极富吸引力的研究领域。基于此, 本文将主要分析并阐述社交媒体分析与智能这一领域近年来的研究工作。首先, 我们将介绍观点挖掘/情感分析、社会网络分析等社交媒体信息分析工具, 并描述影响力扩散、在线社区等社交媒体分析中的建模方法与预测研究。接着我们将介绍社交媒体智能的基本概念与应用情况。文章最后, 我们将总结并讨论该领域所面临的若干挑战。

1 引言

互联网与移动通信技术的发展推动了社交媒体 (Social Media) 的崛起, 新的技术平台使得内容的产生与分享能够以交互的方式进行。新鲜的应用功能, 良好的用户体验和含有社会属性的服务内容加速了用户生成内容 (User Generated Content, UGC) 的爆发。如果从平台的角度出发, 各式各样的网络应用服务勾画出社交媒体

的基本功能轮廓, 如表 1 所示列出了社交媒体中的主要服务类别。例如: 博客、微博、社交网络、维基百科、兴趣图谱、地理签到站点、图片/视频分享站点、社会标签站点、网络社区等等。事实上, 社交媒体中产生并流动着的海量用户生成内容正是 Web2.0 时代的主要特性。在 Web2.0 时代, 网民被赋予了更多的权利, 他们可以发表自己的观点, 或者发起某项集体行动, 这使得网络中的行为愈加密集, 更具风险性, 同时也更为复杂。

在维基百科中, 社交媒体被定义为“人们彼此之间用来分享意见、见解、经验和观点的工具和平台”^[1]。广义来说, 社交媒体代表了一种新的沟通方式, 这是一种基于社区的分布交互式的内容产生和传播机制^[2]。区别于传统媒体 (例如: 报纸、广播、电视) 的广播式传播机制, 社交媒体模糊了信息生产者与信息消费者之间的界限。这使得信息的生产、传递和消费模式发生了根本性的转变^[3]。各种各样的用户生成内容 (例如博客、评论、微博等) 在信息的生产者和消费者的复杂网络间实时地流动着。

表 1 社交媒体主要类别

类别	国际重要站点	中国重要站点
博客	LiveJournal, WordPress	搜狐博客, 新浪博客
微博客	Twitter	新浪微博, 腾讯微博
商业口碑	Yelp, Epinions	大众点评
图片视频分享	Instagram, Youtube, Flickr	优酷网, 土豆网, 网易相册
社交网站	Facebook, LinkedIn	人人网, 开心网
维基百科	Wikipedia	百度百科, 互动百科
地理位置服务	Foursquare	街旁
兴趣图谱	Pinterest	花瓣网, 堆糖网

据统计^[4], 2011 年全世界共有社会网络账户 24 亿个, Twitter 的用户数量在 2011 年已经达到 2.25 亿, 每天发送的 tweet 数量为 2.5 亿条。平台中最火热的话题每秒钟可吸引 8868 条 tweet, 平台最受关注的用户拥有 180 万名关注者; 相比于 Twitter, Facebook 拥有更广泛的用户

群体, 自 2004 年创建以来, Facebook 的用户数量持续增长, 如图 1 所示。Facebook 公布的最新数字显示的用户总数已经达到 9.01 亿人, 每日接受到的评论为 32 亿条, 每日新增图片 3 亿张。网站上的好友数量达到 1250 亿对^[5]。

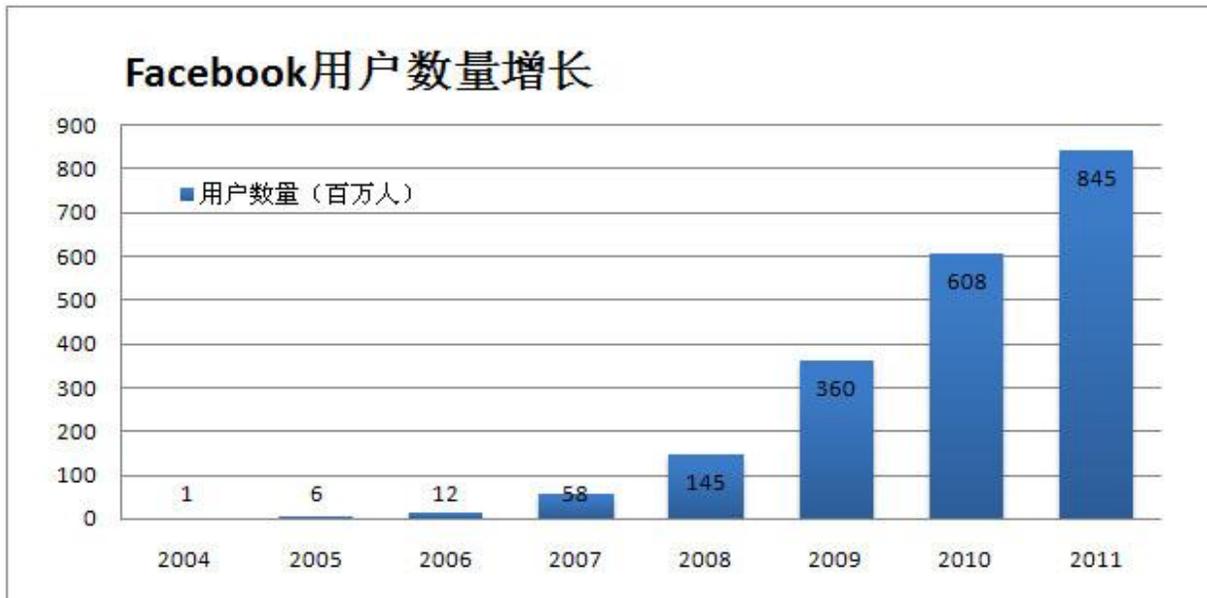


图 1 Facebook 用户数量增长情况 (根据新闻报道重绘^[6])

2 社会媒体的机遇与挑战

当下，社会媒体中自由涌动的海量异构数据已成为一种重要的资源，它以文本、图像、多媒体等形式呈现，其潜在价值将超越以实物呈现各种资源，例如：石油、矿藏等。因此对社会媒体的挖掘与利用引发了社会各界的广泛关注，消费者、选民、商业组织、政府机构、非营利组织都可以从社会媒体中找到关联的应用方向，这使得关于社会媒体的应用和研究在短时间内得到了爆发式的增长。以公共卫生领域为例，社会媒体中实时、海量的信息可以用于感知大规模人群健康状态，其将直接服务于公共卫生事件的实时监测中，如传染病事件暴发的早期检测与预警。传统的流行病学调查方法获得的资料需要巨大的时间与人力成本，这使得实证分析的结果不能及时反映传染病的真实传播过程。而社会媒体中的任何用户都是网络信息传感器的一个细小终端，可以实时感应并报道重大疫情的发生和发展状况，这一特点可以用于弥补流行病学调查数据之不足。最引人注目的应用是 Google 公司利用搜索关键词预估流感趋势的创新性研究方法——Google Flu Trends^[7]，它主要通过统计与流感症状相关的各类关键词出现频度与相关的空间信息进行预测。其预测结果与美国疾病预防控制中心的监测结果非常接近，而且实时性更强，可比传统方法提前一到两周得到监测结果。这种创新方法展现了传统手段难以企及的预测效果。Nelson 等人^[8]将世界范围中与传染病相关的疫情暴发新闻通过电子地图进行实时展示，这种可视化的方式可帮助人们分析疾病暴发的时空风险。

对于政府组织而言，社会媒体在防灾应急中也展示了巨大的应用价值。最近几年，每当突发事件出现时，社会媒体中便会以极快的速度大量涌现出从各个视角对突发事件的描述、评论与应对建议。网络用户不仅将突发事件现场的各类一手资料信息通过社会媒体分享，甚至会深入挖掘出突发事件的各种相关信息。突发事件出现后，社会媒体背后各种网络群体展现出的热情与高效所产生的重大价值是传统方法所无法想象的。例如，2010 年海地和智利特大地震中，基于对 Twitter 等微博内容的内容聚合平台——Ushahidi.com^[9]实现了对事件现场各类信息的即时反应、快速更新与全方面覆盖，为灾情实时感知、救援进展实时播报、搜寻幸存者等提供了具有重要价值的信息。事实上，早在 2004 年美国地质调查局 (United States Geological Survey, USGS) 开通了“Did You Feel It”网站，以便从网络民众直接获取地震活动相关的各类信息。2007 年，美国政府综合论坛、博客、掘客、标签、RSS、社交等社会媒体技术构建了专业政府防灾应急网“911.org”。社会媒体给非常规突发事件应急管理的创新研究提供了难得机遇。

通过追踪社会媒体中的信息，商业公司可以获得有助于推广和改进产品的反馈和信息。对于消费者，海量的信息和各种观点可以帮助他们走进群体智慧的大门，帮助他们制定更加理智的决定。社会媒体的出现极大促进了社会科学与信息科学的融合，为科学研究带来了新的机遇。社会计算采用信息与计算技术方法研究社会相关问题，例如：社会群体的形态及其演化

规律，人们的交互模式等。为了解决社会计算中复杂系统的管理与控制问题，王飞跃等人扩展了社会计算的概念^[11]，并提出了面向社会计算研究的 ACP 方法^[12,13]。ACP 方法通过建立人工社会，利用计算实验和平行执行分析评估影响系统的各种因素，进而实现实际系统和人工系统的相互借鉴和互动调节。这些研究可以帮助我们解答很多实际问题，如某个组织是达到了一个稳定的状态还是经常发生剧烈变化，具有创造力的团队应具有什么样的交互方式等^[10]。2009 年，David Lazer 等^[14]在 *Science* 杂志上提出计算社会学的概念，指出社交媒体中的大量信息，如博客、论坛、聊天、消费记录、电子邮件等，都是对现实社会的人及组织行为的映射，社交媒体中的数据可用来分析个人和群体的行为模式，从而深化我们对生活、组织和社会的理解。

然而社交媒体在带来巨大机遇的同时也面临着一些挑战，如海量数据的处理问题、网络隐私保护问题等。社交媒体中产生的海量数据促进了新学科的出现与发展，数据科学 (Data science) 是一门研究数据获取、数据存储与管理、数据安全、数据分析、可视化等内容的新兴学科。在 2008 年《*Nature*》杂志就出版了名为“Big Data”的专刊，从多个方面着重介绍了人们面对海量数据所遇到的困难。2011 年 1 月《*Science*》杂志推出专刊“Dealing with data”，同年 12 月又推出专刊“Data Replication & Reproducibility”，这两期专刊主要讨论如何应对海量数据处理中的各种问题。

不同于海量数据处理所带来的技术挑战，网络隐私保护问题需要考虑更多层面的因素。出于对日益严重的隐私保护方面问题的考虑，世界各国纷纷酝酿出台新的法律法规。2012 年 1 月 25 日，欧盟提交了新的数据保护法修订案，新数据保护法中规定泄露用户数据的公司将面临高额罚金。2012 年 2 月 23 日，美国向议会提交了新的隐私权利草案。由于法律环境的变化，一些社交媒体站点管理者开始抵制公众和第三方机构对于数据的大规模获取，例如：Foursquare (地理信息服务站点) 的服务条款中明确提及“禁止自动获取站点所提供内容”，Twitter 要求斯坦福大学移除先前公布的用于研究的数据集。在数据分享方面，Fred Morstatter 等^[15]总结了目前环境下社交媒体数据分享中面临的问题，并提出了一种基于开放准则的分布式数据分享框架。此外，研究人员已经开始尝试通过各种分割与合并算法降低社会网络数据公开后的隐私担忧。

3 社交媒体分析

社会媒体的研究具有天然的多学科特性，在过去的几年中，来自各主流学科的研究团体均开始关注社交媒体带来的科学机遇与挑战。如果从信息科学的角度出发，社交媒体分析和社会媒体智能是该领域中最重要研究方向。由于社会媒体的动态性与复杂性，信息传播的全面性和快速性、网上群体行为演化的动态性、网上以及网下行为虚实交互性，对于社会媒体的分析迫切需要各种分析工具与模型的支撑。具体说来，社交媒体分析的研究包括社交媒体数据的获取与追踪方法、数据挖掘方法与分析工具、社交媒体建模方法、分析结

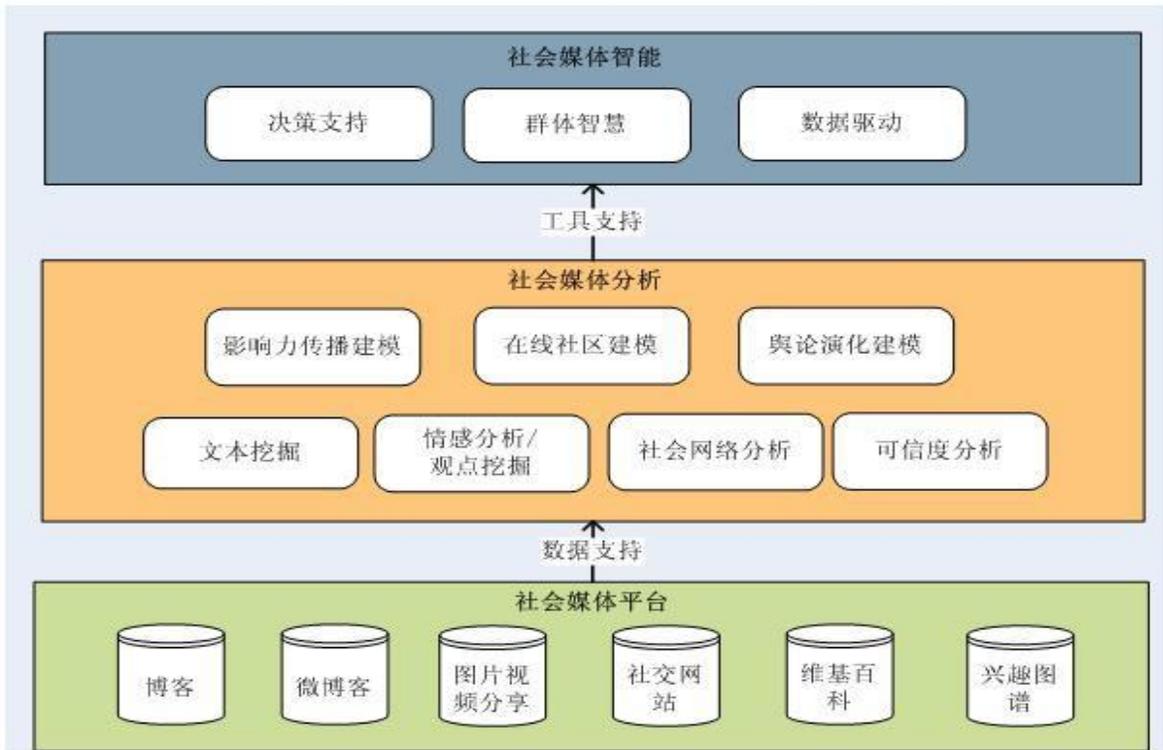


图2 社交媒体分析与社交媒体智能

果的可视化方案等。本文将主要介绍社交媒体分析中的数据分析技术与社交媒体建模方法，如图2所示。

3.1 社交媒体分析的工具与方法

目前，常用社交媒体分析工具有观点挖掘^[16]、社会网络分析^[17]、文本挖掘、可信度分析等。随着复杂网络研究的兴起^[18,19]，社会网络分析受到越来越多的关注，社交媒体中丰富的社会关系为使得开展大规模社会网络分析成为可能。此外由于社交媒体中含有大量用户生成的主观评论信息，通过情感分析和观点挖掘技术可以获取大众对于某一事件或产品的看法。情感分析/观点挖掘是对带有主观性的文本信息(以及语音识别后转化成的文本)进行分析、处理、归纳和推理的过程。简而言之就是获取意见的倾向性(正面、负面与中

立)与强度信息。进行情感分析的另一个重要目的是要寻找意见和情感的主体和客体。其中情感的主体就是观点和情感的持有人或者组织。观点的客体就是观点针对的目标，例如，某事件、事件的起因、事件的影响、事件的结局等等，上述目标也可以是产品、个体、公司、组织等等。如果从粒度这一维度进行划分，可分为文档级别分析与句子级别的分析。目前进行情感分析的主要方法为基于统计/学习的方法与基于规则的方法，更多的技术是两者结合使用。

3.2 社交媒体的建模与预测

在社交媒体中，网络中的节点通过信息传播与各种交互行为直接或间接影响其它节点行为。影响力传播建模主要分析研究信息和影响力在社交媒体中的传播扩散

的规律与机制^[20]。例如：Sun 等人^[21]通过研究 Facebook 中的扩散现象，发现不同于“扩散是由少量初始节点发起”这一传统假设，喜爱 Facebook 中公共页面的行为通常是由大量节点同时发起，并通过很短的传播链迅速结成大集群。类似的工作中，Lerman^[22]等人通过分析活跃用户分享新闻的行为，研究新闻在 Digg 和 Twitter 等站点中的传播特征。作者认为尽管两个站点在功能和界面上差异很大，但是信息传播的方式却很相似。Yang 等人^[23]则通过聚类算法研究信息扩散的时序特征，通过分析 Twitter 和 blog 中的数据共得到 6 种时序曲线的主要形状，这些信息可用于进一步的预测工作。

影响力传播建模通常以某种产品、技术、话题在社会媒体中演化与发展作为研究背景，通过各种模型研究这些信息在社会媒体中的传播与扩散过程。很多研究问题均基于这些影响力传播模型，如影响力最大化、爆发检测、意见领袖识别等。Bhatt 等人^[24]通过分析在线社会中产品购买行为的传播扩散过程，建立了相应的预测模型。该预测模型可以用于确定不同营销策略的目标人群，从而使得产品的市场份额最大化。Goyal 等人^[25]通过研究 Flickr 中图片的传播行为，提出机器学习算法，通过挖掘历史传播记录获取影响力传播模型的参数。根据模型种类不同，影响力传播模型大致可分为线性阈值模型，独立级联模型与博弈演化模型三种。

独立级联模型 (Independent Cascade Model) 认为网络中的感染节点 i 以概率 P_{ij} 来感染其邻居节点 j ，但每个节点只有一

次机会来感染同一节点，直到网络中所有节点的结束传染尝试为止。

线性阈值模型认为网络中的节点均存在在一个阈值 $h_i \in [0, 1]$ ，网络中的每条边 e_{ij} 都有一个权值 w_{ij} 并且满足 $\sum_{j \in N_i} w_{ij} = 1$ ，其中 N_i 表示节点 i 的邻居节点。当节点 i 满足条件 $\sum_{j \in N_i, j \in S} w_{ji} \geq h_i$ 时被感染，其中 S 为感染节点集合。

与上述两个模型不同，博弈演化模型中并不将状态分为感染或者未感染，而使用 x_i 表示一个节点的选择，节点会根据当前状态改变自己的选择。 $u(x_i, x_j)$ 表示收益函数。对于两个相连的节点而言，选择相同收益值大于选择不同的收益。网络中的每条边 e_{ij} 都有一个权值 w_{ij} ，每一时刻节点 i 都会调整自己的选择使得自己的总收益 $\sum_{j \in N_i} w_{ij} u(x_i, x_j)$ 最大，其中 N_i 表示节点 i 的邻居节点。

在线社区的建模与分析主要研究群体行为的建模理论，探索在线社区的形成、演化与发展规律。例如：在线社区识别主要通过网络节点间的链接关系发现潜在的社会群体。随着信息技术的发展，恐怖组织和仇恨群体越来越多的使用社会媒体进行传播思想、招募成员等活动。为了应对这些新的挑战，Chau 等人^[26]提出了一种识别博客中的仇恨团体的半自动方法，这种方法不仅可以应用于实际的信息监控中，也可以帮助研究仇恨团体的结构特征。群体行为分析的主要方法包括社会文化建模

分析和组织行为预测，此外，规划识别方法可以根据观察到的群体行为识别当前行为的目标和意图等深层信息。

4 社交媒体智能

越来越多的事实表明“群体智慧”在网络时代拥有巨大能量^[27]。以 2006 年的人肉搜索事件——“网络虐猫女”为例，网友们仅凭借网络中发布的一组照片，在短短六天时间里，找到了踩猫人、光碟拍摄人等关键人物，完成了几乎不可能完成的任务。同样，作为“群体智慧”，另一种形式，众包也因其强大的影响力渗透到生活中的各个方面。亚马逊还推出了专门提供众包服务的平台 Mechanical Turk。

如果说各种人肉搜索、众包现象充分展现了“群体的智慧”的力量与魅力，社交媒体智能则是建立在社交媒体分析之上，通过整合社交媒体分析的技术、方法与模型，旨在为各种应用问题提供有效的决策支持，帮助决策者利用“群体智慧”。社交媒体智能更倾向于如何激励、引导、收集、利用这些智慧。由国防高级研究计划局 (DARPA) 举办的网络挑战 (Network Challenge) 展示了社交媒体智能的巨大潜力^[28]。在该竞赛中，组织者秘密的将 10 只红色的气象气球放置到美国各地，在最短时间定位所有气球正确位置的参赛团队将获得 4 万美金的奖励。由于任务的跨地域性、时间敏感性，单单依靠个人或少数人的力量寻找所有气球几乎是不可能的。幸运的是，气球放置的位置多为各种公共场所，例如：公路旁，棒球场内等，普通民众均可以看到并提供信息。因此如何在短时间内动员人们进行大规模协作是参赛

队伍面临的主要挑战。由于需要实时、动态的信息交互方式，在线资源特别是各类社交媒体站点显现出明显的优势。在社交媒体的帮助下，获胜的团队通过“递归激励”策略开展动员，利用适当的信息确认方法收集整理各种结果与信息，最终在不到 9 个小时的时间内便找到所有 10 个气象气球，这一结果远低于组织者的预计，展示了社交媒体巨大的影响力。

在对获胜策略的具体分析中，Pickard 等人展示了社交媒体在竞赛过程中发挥的重要作用^[29]。通过分析 Twitter 中的数据可以发现，有些队伍主要依赖于初期密集的活动吸引参与者及注意力，但是这种方式所造成的影响力很快消失了。而获胜团队的“递归激励”策略可以使社交媒体中的影响力更加持续。该问题的实质是如何在有限的时间内开展大范围情报收集工作，因此竞赛结果将有助于思考如何开展大规模灾害，商业情报的收集。该活动也为网络社区传播理论、众包激励机制在社交媒体智能中的应用提供了良好的示例。

社交媒体分析的快速发展与广泛应用为社交媒体智能提供了坚实的基础，极大的促进了社交媒体智能在各领域中的应用。2012 年 3 月，美国红十字会正式启用了新的“数字化运营中心”，作为第一个基于社交媒体的人道主义救援部门，该中心的建成反映了社交媒体智能在应急管理日益增长的重要性^[30]。具体来说，数字化运营中心将帮助红十字会组织在紧急情况下收集各种受灾地区的信息，以确定需要救助的人们，并为他们提供更好的服务，例如帮助人们找到所需的食物、水、避难

所或是情感支持。在金融领域，2011年5月，世界首家基于社交媒体的对冲基金 Derwent Capital Markets 上线，它将利用 Twitter 中的公众情绪实时分析结果指导投资^[31]。这些应用的背后是相关研究的支持，目前尝试使用社交媒体预测股票市场可以分为两种，一种方法主要比较 Twitter 中的情绪信息与股票指数之间的关联性。如 Zhang^[32]等发现无论是“希望”等正面情绪或是“害怕”、“担心”等负面情绪在总推文数中的比例，都可以预示道琼斯指数、纳斯达克指数、标准普尔指数的变化。另一个研究中通过对 Twitter 进行更深入的分析，研究者发现推文的总数、交易量、看涨程度和标准普尔 100 指数之间密切相关^[33]。另一种方法则比较品牌在社交媒体中的受欢迎程度与公司股价之间的关联。通过分析星巴克、可口可乐和耐克等三个品牌在社交媒体中的数据，研究人员发现公司的股价与社交媒体中的社会网络信息密切相关，这些指标包括 Facebook 上的粉丝数、Twitter 上的听众数和 Youtube 上的观看人数等^[34]。

然而，与社交媒体分析相比，社交媒体智能的研究仍然处于起始阶段，尽管目前存在着许多概念层面的讨论，例如相关的技术挑战，相关的学科等，然而专注于社交媒体智能的系统性研究与细致而可信的研究结论依然非常稀少。在这里我们尝试讨论社交媒体智能的研究面临的主要挑战，以期该领域的研究能够得到更广泛的关注。首先，社交媒体智能的研究迫切需要不同学科之间的交互与整合。尽管以商业为代表的各种领域已经对社交媒体智能的研究表现出极大的兴趣，但是学科之间

的整合与交流程度仍然很低，信息科学的研究方法与研究问题仍然占据着最重要的位置。相较而言，社交媒体智能在心理学、社会心理学、传媒学、政治学等领域的研究比较分散，缺乏相应的跨学科项目支持。其次，对于社交媒体智能的研究需要制定客观和简明的性能评价方法。由于社交媒体智能的应用领域广泛，建立一套有效的决策支持评价机制是非常困难的。这种方法上的困难使得我们很难评价社交媒体智能的投入产出比，因而阻碍着相关研究的开展。最后，如果单纯从模型建立和决策支持的角度出发，随着社交媒体智能研究的不断成熟与发展，新的应用也将需要更先进的社交媒体分析方法与计算框架的支持。

5 总结

综上所述，在信息科学的研究中，社交媒体的分析与智能引起了广泛的关注并取得了丰硕的成果。社交媒体分析主要关注社交媒体数据的跟踪与获取、社交媒体数据的处理与整合、社交媒体的建模与预测等问题。在社交媒体分析的研究中，情感分析/观点挖掘、社会网络分析等方法与技术的成熟推动着社交媒体建模与预测的研究，诸如影响力传播建模、在线社区建模等研究均取得了一定的进展。社交媒体智能的研究建立在社交媒体分析之上，通过整合社交媒体分析的技术、方法和模型，为各种应用问题提供有效的决策支持，帮助决策者利用“群体的智慧”。与社交媒体智能的巨大应用价值相比，其研究仍然处于早期阶段，因此需要广泛的关注与深入的讨论。

致谢

本文作者感谢以下项目支持, 国家自然科学基金委(NNSFC)项目: 71103180, 91124001, 60921061, 90924302, 70890084, 91024030 and 40901219; 以及中国科学院项目(CAS)2F11N06。

参考文献

[1] 社交媒体-维基百科. Available: <http://zh.wikipedia.org/wiki/%E7%A4%BE%E4%BC%9A%E5%AA%92%E4%BD%93>

[2] D. Zeng, et al., "Social media analytics and intelligence," *Intelligent Systems, IEEE*, vol. 25, pp. 13-16, 2010.

[3] J. Leskovec, "Social media analytics," presented at the Proceedings of the 17th ACM SIGKDD International Conference Tutorials, San Diego, California, 2011.

[4] Internet in Numbers 2011. Available: <http://royal.pingdom.com/2012/01/17/internet-2011-in-numbers/>

[5] Facebook passes 900 million users. Available: <http://www.techweekeurope.co.uk/news/facebook-passes-900-million-users-74646>

[6] 增长:Facebook 最惊人的力量. Available: <http://www.36kr.com/p/80741.html>

[7] J. Ginsberg, et al., "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, pp. 1012-1014, 2009.

[8] R. Nelson, "HealthMap: the future of infectious diseases surveillance," *Lancet Infectious Diseases*, vol. 8, pp. 596-596, 2008.

[9] O. Okolloh, "Ushahidi, or'testimony': Web 2.0 tools for crowdsourcing crisis information," *Participatory Learning and Action*, vol. 59, pp. 65-70, 2009.

[10] X. Zheng, et al., "Next-Generation Team-Science Platform for Scientific Collaboration," *Intelligent Systems, IEEE*, vol. 26, pp. 72-76, 2011.

[11] F.-Y. Wang, et al., "Social Computing: From Social Informatics to Social Intelligence," *Intelligent Systems, IEEE*, vol. 22, pp. 79-83, 2007.

[12] F.-Y. Wang, "Toward a Paradigm Shift in Social Computing: The ACP Approach," *Intelligent Systems, IEEE*, vol. 22, pp. 65-67, 2007.

[13] F.-Y. Wang, "Computational experiments for behavior analysis and decision evaluation of complex systems," *Journal of system simulation*, vol. 16, p. 5, 2004.

[14] D. Lazer, et al., "Life in the network: the coming age of computational social science," *Science (New York, NY)*, vol. 323, p. 721, 2009.

[15] F. Morstatter, et al., "Opening Doors to Sharing Social Media Data," *IEEE Intelligent Systems*, pp. 47-51, 2012.

[16] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations*

and Trends in Information Retrieval, vol. 2, pp. 1-135, 2008.

[17] P. J. Carrington, et al., Models and methods in social network analysis: Cambridge Univ Pr, 2005.

[18] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," Science, vol. 286, pp. 509-512, October 15, 1999 1999.

[19] X. Zheng, et al., "Analyzing open-source software systems as complex networks," Physica A: Statistical Mechanics and its Applications, vol. 387, pp. 6190-6200, 2008.

[20] D. Kempe, et al., "Maximizing the spread of influence through a social network," presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C., 2003.

[21] E. Sun, et al., "Gesundheit! modeling contagion through facebook news feed," 2009.

[22] K. Lerman and R. Ghosh, Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks, 2010.

[23] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," presented at the Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, China, 2011.

[24] R. Bhatt, et al., "Predicting product adoption in large-scale social

networks," presented at the Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada, 2010.

[25] A. Goyal, et al., "Learning influence probabilities in social networks," presented at the Proceedings of the third ACM international conference on Web search and data mining, New York, New York, USA, 2010.

[26] M. Chau and J. Xu, "Mining communities and their relationships in blogs: A study of online hate groups," International Journal of Human-Computer Studies, vol. 65, pp. 57-70, 2007.

[27] F.-Y. Wang, et al., "A Study of the Human Flesh Search Engine: Crowd-Powered Expansion of Online Knowledge," Computer, vol. 43, pp. 45-53, 2010.

[28] DARPA Network Challenge. Available: http://en.wikipedia.org/wiki/DARPA_Network_Challenge

[29] G. Pickard, et al., "Time-critical social mobilization," Science, vol. 334, pp. 509-512, 2011.

[30] Social Media Digital Operations Center for Humanitarian Relief. Available: <http://content.dell.com/hk/en/corp/d/corp-comm/red-cross-digital-operations>

[31] The World's First Twitter-Based Hedge Fund Is Finally Open for Business. Available: <http://www.theatlantic.com/business/archive/2011/05/the-worlds-first-twitter-based->

hedge-fund-is-finally-open-for-business/239097/

[32] X. Zhang, et al., "Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear"," Procedia-Social and Behavioral Sciences, vol. 26, pp. 55-62, 2011.

[33] T. Sprenger and I. Welp, "Tweets and trades: The information content of stock microblogs," 2010.

[34] Study Claims Stock Price, Social Media Are Linked. Available: <http://www.pcmag.com/article2/0,2817,2382391,00.asp>

作者简介

崔凯楠：西安交通大学电信学院在读博士研究生。2007 年获西安交通大学学士学位，2009 年起赴中国科学院自动化研究所复杂系统管理与控制国家重点实验室进行联合培养至今。主要研究方向为社会媒体分析。

郑晓龙：中国科学院自动化所研究所助理研究员。2003 年获得中国计量学院学士学位，2006 年获得北京交通大学硕士学位，2009 年获得中国科学院博士学位，2009 年起，

在中国科学院自动化所复杂系统管理与控制国家重点实验室工作至今。主要从事社会网络与数据挖掘方面的研究与相关技术研发工作。是 IEEE、ACM Member 以及，Secretary of ACM social and economic chapter，获得国家省部级奖项多项。

曾大军：中国旅美科技协会总会 2012 年理事会主席，2010 年度会长，中国科学院自动化研究所研究员、博士生导师及 Web 科学及服务计算实验室主任。1990 年毕业于中国科学技术大学少年班、经济管理与系统科学系及计算机系，获学士学位；1998 年获美国卡内基梅隆大学工业管理博士学位。他主要研究领域为时空数据分析、传染病信息学、情报与安全信息学、多智能体系统及社会计算。

王飞跃：中国旅美科技协会总会 2005 年度会长，中国科学院自动化研究所研究员、博士生导师及复杂系统管理与控制国家重点实验室主任。1982 年获山东化工学院学士学位，1984 年获浙江大学硕士学位，1990 年获美国伦塞利尔理工学院计算机与系统工程博士学位。他主要研究领域为智能系统和复杂系统的建模、分析和控制，是智能控制方面的国际知名学者，也是该领域的早期开拓者之一。

云管理：新型信息技术创新企业的管理模式

张良杰

金蝶国际软件集团有限公司

中国深圳, 518057

Email: zhanglj@ieee.org

摘要:

在后经济危机时代, 企业所处的环境更复杂、更加不确定。传统经济优势面临巨大挑战, 在日益复杂多变的环境下, 企业如何应对挑战? 尤其是正处于产业结构调整与转型升级大背景下的中国企业, 如何寻找转型之路, 基业长青? 云管理就是应用社交网络、移动互联、云计算和大数据分析等新兴技术所催生的创新型管理模式。应用云管理, 企业可以整合产业链资源, 进行产业模式创新; 可以重塑企业与员工、供应商、客户、合作伙伴之间的关系进行企业管理创新; 可以整合资源, 创新协同价值链, 提供新的产品与服务, 打造新的商业模式。

1 引言

从技术发展历史来看, 每一次革命性发明和技术应用都会推动经济迅速增长, 如电话、铁路和汽车, 每一轮经济的增长无不受益于这些发明。不同的是, 互联网的全球化属性使得发达经济体和发展中经济体都会从中受益。未来关键新兴技术是什么? 一项调查研究显示^[1], 多数受访者(57%)认为, 未来五年中, 移动技术是对企业带来最大的正面影响因素, 其次是商业智能(37%)、云计算(36%)和社交网络。

1.1 社交网络

社交网络(Social Networking Services, SNS)最先源自于美国, 专指旨在帮助人们建立社会性网络的互联网应用服务。当前社会中, 社交网站正逐渐成为以人际关系为联接的互联网互动应用的平台, 应用覆盖到了商务、休闲、娱乐、教育, 其价值和地位也在日益突显。Forrester 研究公司的报告显示, 65%的受访公司正在部署至少一个社会软件工具。这是因为一方面, 利用社交网络可以有效地处理突发异常事件。另一方面, 社交网络可以让员工实现跨边界沟通, 这大大的提高了公司的运营效率。2009年, 社交网络用户首次超过电子邮件用户, 同时互联网使用习惯也发生了变化, 互联网用户22%的时间花在了社交网络上, 如 Facebook, YouTube 等。社交网络技术主要包括:

(1) Web 2.0: 指的是一个利用 Web 的平台, 由使用者主导来创造、协同合作、分享各种资讯与内容的一个分布式的网络现象。Web 2.0 是一种商业革命, 起因于开始把互联网当成交易平台, 并企图去理解在新的平台上通往成功的规则, 它是一个架构在知识上的环境, 人与人之间互动而产生内容, 然后在这个环境中被发布、管理和使用。Web 2.0 是网络运用的新时代, 网络成为了新的平台, 内容因为每位使用者的参与(Participation)而产生, 参与所产生的个人化(Personalization)内容, 借

由人与人(P2P)的分享(Share),形成了现在 Web 2.0 的世界。

(2) P2P: peer-to-peer 的简称,点对点技术或“对等”技术,是一种网络技术和网络拓扑结构,依赖网络中参与者的计算能力和带宽,而不是把依赖都集在较少的几台服务器上。

(3) 社交应用:主要作用是为一群拥有相同兴趣与活动的人建在线社区。这类服务往往是基于互联网,为用户提供各种联系、交流的交互渠道,如电子邮件、即时消息服务等。此类网站通常通过朋友,一传十传百地把网络延展开去,极其类似树叶的脉络,中国习称之“脉络(人面)网站”。多数社交网络会提供多种让使用者互动起来的方式,并为信息的交流与分享提供了新的途径。

(4) 六度分隔理论:小世界现象(又称小世界效应)假设世界上所有互不相识的人只需要很少中间人就能建立起联系。哈佛大学的心理学教授斯坦利·米尔格拉姆(Stanley Milgram)根据六度分隔理论(Six Degrees of Separation)做过一次连锁信实验,尝试证明平均只需六个人就可以联系任何两个互不相识的美国人。

1.2 移动互联网

移动互联网技术就是将移动通信和互联网结合起来的技术,运用无线通信技术,将智能终端设备接入互联网,实现互联网访问及相关业务,进而带动手机终端与PC、其它电子消费终端等设备的融合,并提供移动办公、信息化、多媒体、娱乐、广告内容于一体的综合信息服务。

普华永道认为,移动互联应用分为五个级别,第一个是普通智能便携服务,如文本,蓝牙,拍照等;第二个是大众消费应用,如内容应用包括新闻,图书和视频等;第三个是企业应用,如 ERP, CRM, 以云为基础的办公工具等;第四个是竞争性应用,如移动支付、视觉分析等;第五个是设备相关应用,如游戏,娱乐以及化学分析工具等。

移动互联网技术主要包括:

(1) HTML5:

HTML5 的目标是创建更易用的 Web 程序,书写更简洁的 HTML 代码。HTML5 设计原则包括内部功能非革命性的改进,只做发展性的升级;封装 HTML4 的复杂特性,语法简单易用;兼容老版本浏览器。HTML5 的 Canvas 支持复杂图形绘制、文本编辑。在多媒体处理领域,HTML5 可将音频、视频直接嵌入到网页,无需插件支持。另外,HTML5 支持 Web 存储、离线 Web 应用、跨文档消息传输和 Web Socket。在处理线程方面,HTML5 和前端分离,实现浏览器的多线程处理,且定义了一个并行执行脚本的 API,将 UI 和后台分离处理。另外,HTML5 可获取用户地理位置,省略 GPS 嵌入式设备集成以及外部地图服务商的选择。

选择 HTML5 的技术理由包括但不限于移动设备应用开发、跨平台,无需针对个别浏览器进行特殊开发、世界知名浏览器厂商的支持(如 IE9.0 支持 HTML5)、发展性升级方案,向后兼容、更简洁的语

法，开发效率高、GPU 技术可以在客户端支持复杂图形的绘制技术等。

(2) 移动通讯网络：包括电信通讯网络和移动互联网如 4G，主要解决业务流程复杂、带宽不足、通讯效率低、信息管理不足等问题。

(3) 智能终端：主要包括智能手机、平板电脑、便携式计算机等，主要指在移动通信设备中无线传输网络，并将终端设备的能力适配到无线传输的部分。

(4) 公众互联网服务：WIP(世界互联网项目)2010 年 最新调查报告显示，在其重点研究的 10 个国家和地区中，仅 5 个国家和地区网民数量比率超过了总人口的 50%。无论是发展中国家或发达国家，目前其网民比率仍相对较低。

1.3 云计算

云计算 (Cloud Computing) 是一种能够通过网络以便利的、按需付费的方式获取计算资源(包括网络、服务器处理能力、存储、应用和服务等)并提高其可用性的模式，这些资源来自一个共享的、可配置的资源池，并能够以最省力和无人干预的方式获取和释放^[2]。云计算最重要的特点就是资源共享，这种资源共享能力就是云计算所带给企业的活力。从以前的主机时代到分布式的计算机时代，以及到我们今天我们叫做相互连接在一起的主机，这个时候的主机就是数据中心。也就是说，现在的一个数据中心，实际上就是一个大的主机，只不过数据中心和数据中心之间发生了连接，这也是当今云计算时代和几十年主机时代的巨大差别。

当然，不同的人对云计算都会有不同的理解，大规模的数据中心可以是云计算、海量的存储也可能是云计算，虚拟化也是云计算。云计算是新的服务交付模式。云计算的兴起更多源于虚拟化技术和以服务为导向架构(SOA)的不断成熟。但当前云计算更多重概念而轻实施，重表象而轻本质、重基础而轻应用，使得云计算停留在概念炒作阶段，企业应用难以落地，很多人对于云计算只知其然却不知其所以然，云计算的发展成果并未落到实处。作为云计算的另一个技术基石，SOA 不仅可以将资源的描述规范化，而且能将得到的资源加以整合。

本文主要的工作是围绕着在后经济危机时代，企业所处的环境更复杂、更加不确定情况下，帮助正处于产业结构调整与转型升级大背景下的中国企业寻找转型与基业长青之路。本文介绍的云管理就是应用社交网络、移动互联、云计算和大数据分析等新兴技术所催生的创新型管理模式。应用云管理，企业可以整合产业链资源，进行产业模式创新；可以重塑企业与员工、供应商、客户、合作伙伴之间的关系进行企业管理创新；可以整合资源，创新协同价值链，提供新的产品与服务，打造新的商业模式。

2 新兴技术为企业带来的革新

如图 1 所示，随着社交化网络技术、移动互联网技术、云计算技术等新兴技术的成熟与广泛应用，企业管理模式有了根本性革新，新兴技术带来的新的计算模式、新的应用模式和新的商业模式，将推动企业进行颠覆性的管理变革及转型。企业需

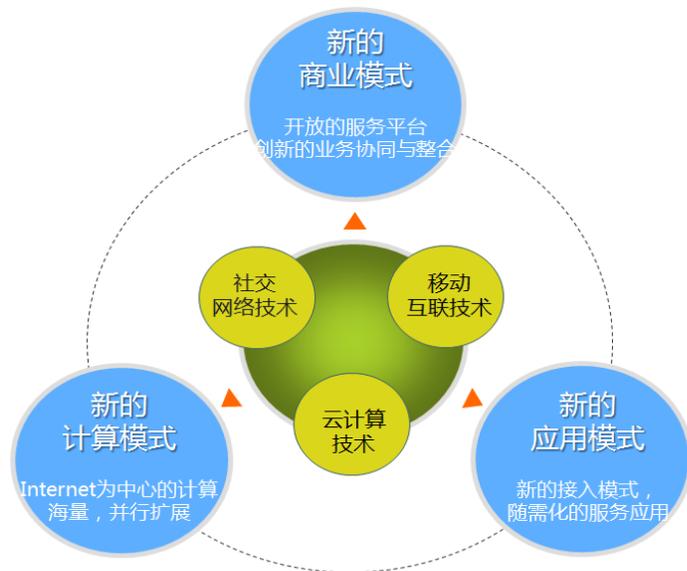


图 1 新技术催生企业创新

要利用 CIO 的智慧，挖掘企业当中不同种类数据带来的潜在商业机会，来有效地帮助企业决策。

新的计算模式：云计算技术带来了新的计算模式，而大数据具有大量化 (Volume)、多样化 (Variety) 和快速化 (Velocity) 和高价值 (Value) 等特征，能够从结构化和非结构化的海量数据中挖掘高价值数据。同时有具备可扩展性，即可横向扩展，也可纵向扩展，更容易地升级和提升处理更大规模数据能力。新的计算模式主要包括：

(1) 虚拟化技术：虚拟化是指计算单元在虚拟的环境里运行。常用的虚拟化方法使用 hypervisor 软件，在虚拟服务器和底层硬件之间建立一个中间层。一般来讲，完全虚拟化技术能让任何操作系统不用改动就可以安装到虚拟机上。缺点是 hypervisor 带来额外的处理器负担。另外一种准虚拟化 (para-virtualization) 技术，它支持在虚拟环境下与 hypervisor 协同。

其中，开源虚拟机技术 Xen 是准虚拟化的示例。但操作系统的核心层面必须进行某些改变，才能在作为虚拟服务器在 Xen hypervisor 上运行。目前 Xen 比较适用于开源操作系统，如 BSD、Linux、Solaris 等，对商用操作系统 Windows 等则无法适用，除非能改动 Windows 操作系统的内部代码。

(2) 大数据处理：政府、企业会创造大量非结构化和半结构化数据，海量数据分析常和云计算联系到一起，因为实时的大型数据集分析需要像 MapReduce 一样的框架来向数十、数百或甚数千的电脑分配工作。海量数据处理可以协助发现可重复的商业模式，并提高政府、企业管理的智能化。

(3) 资源弹性配置：根据管理和计算的需要，动态配置服务器的负载、数据源等资源，有效提高系统的水平伸缩性，提高资源的利用效率，降低资源浪费。

新的应用模式：运用移动互联网技术，将智能终端设备 (手机、PC、电子消费等终端) 接入互联网，移动通信和互联网二者成为一体，并为用户提供随需化的服务应用，使得计算机应用可被移植或被展示在不同的设备、系统、或平台上，为应用的普及化开辟了新的模式。消费者可以单方面地从云服务市场按需自动获取计算能力，如服务器的计算时间和网络存储，从而免去了与每个服务提供者进行交互的过程。微博作为社交网络平台正在改变每个人的生活，企业用户和个人用户都会意识到它的改变，现在的企业应用的用户不

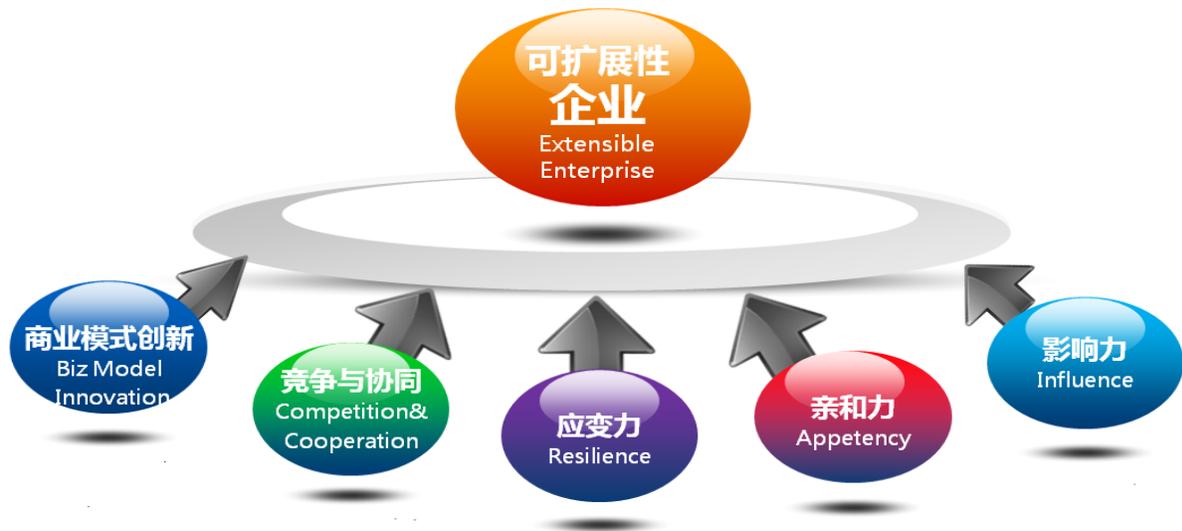


图 2 可扩展型企业的特征

仅仅是专业的管理者，也是企业中的每个人(企业人)，企业应用的开发者也在把用户对象从企业内部转向了企业人的个人社交网络，即整个社会的生态系统。例如，在大型企业的应用领域的软件，以前只有管理人员才能用，或者是很少部分的专业人士如财务人员才用。但是到现在，这样的企业管理软件可以被企业内部的每个员工使用了。他们每次要请假、出差，都需要报批，如今只要利用 iPhone 或 Andriod 手机上的 workflow 应用，点进去就可以随时批复员工提交的任何业务请求，这就是所谓的移动办公。

新的商业模式：云计算与社交网络等技术提供了一个开放的服务平台，通过业务协同与整合，企业可以利用该平台整合其生态链资源，并进行其商业模式的创新。例如，将软件应用以云服务的形式(SaaS)展现的商业模式，可逐渐的把大家平常用得比较多的个人商务管理软件，或者是管理应用，或者是 ERP 系统中还没有的管理功能，逐渐地碎片化，放到一个

公共的云上。当你想要用它的时候就去取，用多少就付多少费用。

3 云管理：创新型管理模式

云管理是应用社交化网络、移动互联网、云计算等新兴技术所带来的创新型管理模式。云管理为企业带来了如下的创新与突破。

突破了组织局限：云管理打破了传统的组织架构，让全员参与企业管理。基于社交化网络技术，提供各种联系和交流的渠道，实现管理的社交网络化。信息通过社交化方式沟通传达，扩大影响力，增强客户粘性，促进团队沟通与协同，提升管理效率，迅速灵活地推进工作，促进商业模式、产业模式创新。

突破了时空局限：伴随移动技术的发展，企业可以通过使用智能终端和云计算，以从未想象的方式持续改进业务流程，突破地域分散、时间协同、人员国际化导致的管理瓶颈，促进管理模式的持续优化。帮助企业由传统的集中式管理向网络化、

移动化过渡，从而在任何时间、任何地点都可以通过移动互联网进行日常工作的处理与审批，利用任何移动终端来实现日常工作与管理的移动化。

突破了资源局限：企业通过互联网按需访问虚拟化信息技术资源，对内整合硬件、软件资源，使资源利用率最大化；对外整合价值链和生态系统资源，实现海量数据的挖掘与处理，让最终用户可以通过任何终端最大化地使用与分享内外资源和辅助决策。同时，企业还可以最大限度地整合基础设施、服务共享、人才资产、智慧资产等各方面的资源，避免资本性投资，降低成本、提高营运效率及灵活性、提升市场响应能力，以绿色信息化的方式实现卓越绩效。

3.1 云管理时代的企业特征

在新兴技术成熟和应用之前，企业是单体运作，以自身为核心，提供产品与服务，受到资源、技术和时空等约束，企业无法进行各种资源整合。

如所示，进入云管理时代，置身于整个生态系统中的企业已经突破了资源、时空和组织局限，可以充分整合生态系统资源，企业具备了可扩展性，能够为客户提供更为丰富的产品与服务。

云管理时代的企业具备了以下五个特征：

商业模式创新：开拓新的创收模式，例如 Web2.0 在企业中已得到广泛应用，企业通过使用 Web 2.0 技术与客户进行交互来激发潜在销售机遇；企业可以利用

Web2.0 技术用来开拓新市场，产生新想法，建立新的业务模式等。

竞争与协同：建立协同产品研发机制，例如利用全球资源进行研发协同创新，选择新的想法和产品解决方案，最终形成差异化的竞争力。

应变力：充分社交网络媒体等，企业更敏锐地洞察客户，灵敏地捕捉市场的瞬息变化，灵活应对内外部变化，响应速度更快。

亲和力：利用社交化网络技术和移动互联网技术突破组织和时空局限，实现全员参与管理，使得人与人、企业与企业、人与企业之间更易接近和沟通交流。社交媒体不仅为客户提供了新的沟通渠道，也改变了企业与客户、雇主与雇员，合作伙伴间应用技术进行互动的方式。

影响力：企业利用新兴技术充分整合各种社会资源，打造属于企业自己的生态链，走正道，行王道，更加注重社会责任，企业品牌影响力和美誉度提升。

3.2 云管理推动企业转型与升级

从 2006 年，金蝶在企业界首次提出“中国管理模式”^[5]之后，我们一直致力于探索研究中国从“制造大国”走向“创新大国”的转型，中国企业如何获得持续的发展动力和最优的内在机制。针对中国企业的管理现状，金蝶联合著名商学院、管理学院，组织专家和学者，通过对众多的中国企业转型案例的研究与分析，得出如图 3 所示的初步结果——中国企业在管

理上迫切地需要从五个维度(战略、研发、运营、文化及企业家)出发进行转型。

今天，云管理成为了助力企业转型的核心驱动力，不但让企业的 IT 基础架构管理焕然一新，还为企业带来了更加智慧的管理模式。

第一，战略转型。云管理帮助企业进行商业模式创新，例如利用互联网技术整合生态链资源，改造和创新企业传统业务，建立新的业务模式，从而推动企业完成战略转型。作为战略转型的支撑手段，顶层设计原则上是一种至上而下、螺旋迭代式的主动型战略落地方法。将企业的业务、数据、应用、技术等资源通过一种结构化的逻辑组织起来就形成了企业架构。成功企业的业务能力、运营模式都是建立在企业架构之上。企业架构框架(TOGAF)方法论有助于企业通过架构化的方法，对热点技术创新领域进行系统化的设计与布局，

帮助企业打造新的商业模式、营运模式和管理模式。顶层设计在世界一流企业和政府得到实践。如美国行政管理与预算局就是通过企业架构规划、治理，确保联邦机构的业务和 IT 效率达到最大化。而中国企业与世界一流企业在基础管理、管理创新、投资并购、管理信息化、国际化经营管理方面，还存在较大的差距。中国企业，特别是转型任务最为艰巨的大型央企、上市公司，要缩短与世界一流企业的差距、具备国际竞争力，就应该向国际领先企业学习，强化企业架构。TOGAF 企业架构框架是国际领先企业建设企业架构实践经验的总结。企业或组织中的任何一位专业人士，无论是做人力资源、财务、技术，还是管理者，都要了解如何利用自己的专长来推动企业或组织业务发展的方法、框架；从“顶层设计”的角度规划企业的未来与自身的职业发展。



图 3 基于企业生态系统的转型支撑

第二，研发转型。云管理让企业实现研发协同，整合内外部资源，从原来中心的独立研发到外部的协同研发创新，最终实现研发转型。

第三，运营转型。云管理使得企业的管理流程更为精细化，更为顺畅和敏捷，对市场的洞察与客户反馈更为灵活和迅速，帮助企业实现运营转型。

第四，文化转型。云管理不仅实现了全员参与管理，而且让企业更具亲和力，实现了人与人、人与企业以及企业与企业之间更容易接近和沟通，打造亲和力文化，实现文化转型。一个企业的文化实际上是一个企业的命脉，如果没有好的企业命脉，是不可能健康地推动企业发展的。转型的欲望来自于思想上、心灵上的净化，企业的创始人和主要的领导者，必须要有转型的能力和勇气，同时也需要通过这样的方式，跟员工交流、通心，或者说交换心灵来推动企业的发展。

第五，企业家转型。云管理帮助企业建立和整合自己的生态链系统，提升在整个社会系统中的品牌影响力和美誉度，企业的社会责任感更为强烈，帮助企业完成从创业家到思想家的转型。例如，企业家的诚信非常重要，特别是目前很多企业追求的实际上是一种霸道型的资源侵占，或者是资源私有化的一种手段。这种商业的诚信问题，实际上为企业现在带来了非常大的挑战，为了进化目前的商业环境，企业必须利用各种技术的手段，来推动信息沟通与交流的渠道。

实际上，云管理为企业带来的改变绝不仅仅局限于 IT 系统。面对变幻莫测的

业务需求、国内外的竞争压力，云管理不但为企业注入更多潜能和活力，还在运营、安全、创新和服务层面提供前所未有的革新方案。

4 云管理实践

企业发展模式通常包括内生式增长和外延式扩张两种模式。内生增长 (endogenous growth) 是指完全通过公司现有资产和业务，而非通过兼并收购方式实现的销售收入和利润的增长。内生增长理论核心思想是认为经济能够不依赖外力推动实现持续增长，内生的技术进步是保证经济持续增长的决定因素。企业自身内生性的成长往往是一种线性成长模式。外延式增长是企业通过重组并购实现扩张，从而实现跳跃式发展。这类发展模式在迅速做大规模和提升发展速度的同时，也蕴含着极大的风险，如整合风险、不同企业文化的融合，管理风格的差异等。

在云管理模式下，企业通过资源共享、服务整合打造新的商业模式，用最小的成本提升业务，打造核心竞争力。云管理不仅为管理创新、文化变革提供新方法，她更具整合企业内外部资源、为企业注入新活力及竞争力的能力，显著提高企业效益、为企业创造突破性增长，最终实现企业的指数性增长。

云计算时代有非常多的应用场景，如可以随时随地批同事的出差申请、费用报销等，完全是一种游牧式的管理。随处可见的各种应用，比如通讯录的应用、社交网络像企业微博^[4]，可以让管理变得非常和谐，让同事之间的交流没有鸿沟，这种触手可及的云管理服务的确将改变企

业当中的每个人，以及每个人的家庭、每个人所拥有的社交网络的平台，以及我们拥有的社交网络的圈子。对企业应用软件的开发者和服务商来讲，这里面的机会非常巨大，因为它给了一种把企业应用的用户从企业内的专业人士扩展为整个社会大众的一条渠道。为了实现这样一个目标，我们需要把这些高价值的应用和高价值的云管理服务打造在一个智慧的后台上，这个智慧的后台要有智慧运算的能力，能够处理复杂的数据，处理不断增加的大数据，同时还能够在可拓展性方面打下坚实的基础。

4.1 社交化管理应用案例

集团 A 的总会计师负责集团的财务工作，她面临着两大管理困惑：一是如何把握好集团财务管控中人的因素，增强分子公司财务线员工的归属感，充分理解和贯彻总部财务管控的意图；二是加班加点是财务人员的家常便饭，如何鼓舞员工士气，增进沟通和创建和谐的工作氛围。因此，公司需要所有同事(包括跨地域)可以零距离交流，提出要求和建议，获得支持，加深了解；不同部门同事之间需要深度分享资讯和文档，但是使用 QQ^[3]和邮件的交流和分享过于分散，不利于知识的管理和共享；不同的部门和项目之间需要协同，缺少合适的工具，容易导致项目小组成员之间的信息失衡。

通过金蝶微博^[4]，集团 A 零时间、零成本的构建了一个私密的如图 4 所示的企业内部社交网络，把全集团财务线同事汇聚在一起；免费使用

金蝶微博可节约 10 万购置企业软件的预算；分子公司员工需要总部同事的帮助，直接@一下，总部的回复更加及时，增强了分子公司财务线员工的归属感，减少 60%的邮件负担；基于话题和小组的讨论应用，更好的解读和贯彻总部的财务政策；关注/粉丝/赞/照片分享/话题等社交行为的融入，拉近了管理者与员工的距离，创建更和谐的工作氛围。这个就是在云计算时代下的管理领域有多少的梦想正在被实现的一个非常成功的案例

4.2 移动化管理应用案例

云计算时代要想充分展现它的价值或者它的威力,需要利用智能终端，手机、iPad 等移动设备在我们手中无时不在。我们以一家以供应链运营和房地产开发为主业的现代服务型企业 B 为例，来介绍一个移动化管理的案例。企业 B 的供应链服务涉及采购、深度物流、销售、结算等在内的“一站式”服务，目前公司在供应链运营领域已形成浆纸、钢铁、矿产品、化工产品等十大核心业务品牌。房地产综



图 4 集团总部与子公司之间的沟通

合服务已在中国 15 个城市进行房地产开发,初步形成全国性战略布局。随着业务的增加和全国布局的形成,如何克服集中管理的瓶颈,突破管理的物理局限,实现随时随地的移动化、随需化的管理,是企业 B 面临的管理挑战。

为此,我们为企业 B 量身定制了移动化的管理方案,在 30 天内就让 80% 以上人开始使用金蝶移动应用,51 条业务流程中有 44 条能通过金蝶的 EAS^[6] workflow 移动应用审批,占比 86%,3 月开始进行房地产决策移动应用的验证。同时,通过企业通讯录服务^[6] 方便地将和勤 HR 系统中 3000 多员工的信息方便的供大家查询。企业 B 的移动化管理应用的总投资回报率达到了 300%。企业 B 的 CTO 评价道:这是第一个不用培训就让全员都会使用的应用,也是第一个让董事长和管理层觉得好用的 IT 应用。

4.3 云管理的应用部署的思考

打造云管理的美妙世界,除了上面介绍的社交网络技术 with 移动互联技术外,还需要非常强大的后台,那就是云计算的基础设施或者计算模式,只有这种模式才可以让所有资源(硬件资源、软件资源、流程资源等)都可以得到有效的分享,并做到系统化的集成。“云管理”作为一种全新的管理模式,它可以为企业带来方便易用、资源可伸缩的企业价值创造的新模式。这种新模式的创造需要智慧的后台、处理强大的计算能力、强大的存储空间、以及非常灵活的资源弹性调配机制。

以前所有的应用都是部署在一个企业里面,在云计算时代下就衍生出私有云环

境。私有云逐渐走向公有云再走向混合云的时候,应用的部署和管理将变得非常复杂,具有很大的挑战性。在这种情况下,作为一个企业的应用,它需要有交付模式的改变和创新才能适应现在技术时代的变化。以前交付一个应用相对简单,除了开发,然后做应用的设计,硬件、软件的评估,然后到交付,通过软硬件的集成来部署,花的时间非常多。

但是在云计算时代,我们可以利用一种新技术来简化企业应用部署时间、加快应用部署的速度。在这种情形下,新的云计算时代的应用部署是有几个环节的,第一从创建应用开始,可以在云计算环境中或者云计算社交网络平台上,在开放平台的基础上开发你的应用,应用的开发、测试、运维已经不再有边界了,一定是共同设计的模式。当开发好了应用以后,需要知道把这样的应用放在什么平台上,需要去申请你获得的资源来运行你的程序,同时应该想到,这样的程序如果部署在这个平台上,这个平台未来能够支撑巨大用户量的增长吗?所以要考虑到未来的基础架构需求。在这种情况下的确需要非常多的具有专家技能集成的系统环境,这种环境是我们以后每个企业的应用或者每个政府的应用当中都会用到或者说正在面临的挑战。

作为云计算时代下,我们所有的资源都已经被虚拟化,这种情况下我们不仅仅只有基础设施的共享,同时要让我们的数据库、应用可以共享,同时还有非常得多企业的应用变成了一种云服务,这种云服务也要在这个平台上共享。如何用同一个系统打造适合于不同场景或者不同规模环

境下的数据中心，这也是我们以后要加以追求的一个方向。

4.4 云管理应用落地的挑战与机会

面对当前公众对于云计算认识的差异，世界可以通过不同行业来划分云计算，包括物流交通运、食品安全云以及智慧城市云，现今很多企业有云计算需要，但是云计算在企业的实施中面临六大挑战，企业的领导者则要审视以下六大挑战，即安全性、可用性、低能高率、可伸缩性、可移植性和互操作性。

此外，企业要想打造安全可信的云，得不到用户的信赖将是最大的障碍，无论既有技术是否真正可以保证安全，但始终是用户关心的问题。另外，安全可信的云还需要灵活的安全机制支撑。当前没有绝对安全并且稳定可靠的云服务，不稳定的因素会时时威胁云的可用性。虽然没有100%稳定的机器，但需要建立100%有效的机制，进而打造稳定可用、智能化的弹性云，让互操作性、可移植性等得到最大的保证。只有深入研究关键技术，才能最大限度保证云计算的落地。而这些关键技术则要立足于云计算的参考框架，从操作系统、运营以及应用平台多角度出发，整合SOA与虚拟化技术，融合国际标准和行业最佳实践，打造开放的云生态系统。

5 结语

本文介绍了在移动互联、云计算、社交网络技术以及大数据分析时代，以及这些新技术为企业管理或者为企业的价值创造打造了一个全新的模式：“云管理”的模式。中国企业家要培养远见卓识和对未

来的洞察力，带领企业进行商业模式和企业战略、创新、运营、文化以及组织领导力等管理领域全面转型。中国必须努力把握这个千载难逢的机会，利用新一代信息技术驱动下的创新管理模式来实现中国企业实现质的飞跃，推动经济跨越到一个新的发展水平。

云管理作为一种新兴的革命性管理模式，其必将快速推动企业在管理模式、商业模式、产业模式上的实现新的转型、升级。中国企业也必须面对云管理时代的到来。在云管理时代，信息化在企业运营中的角色和定位必将发生根本的变化，从传统对业务的支撑和适应转向战略牵引和持续使能，促使企业的管理信息化进入一个新的境界，使越来越多的企业成为云端管理的企业。

致谢

本文的部分资料由金蝶研究院原商业价值研究所的同事们收集、整理，在此向他们从事的《金蝶云管理白皮书》的提炼工作表示感谢。

参考文献

[1] The New Digital Economy - How it will transform business, Oxford Economics, June 7, 2011

[2] National Institute of Standards and Technology (NIST, 美国国家标准和技术研究院)

[3] QQ, <http://qq.com>

[4] 金蝶微博,
<http://www.kdweibo.com>

[5] 中国管理模式,
<http://www.ccmr.org.cn>

[6] 金蝶EAS企业管理软件,
<http://www.kingdee.com/products/eas/>

作者简介

张良杰：金蝶国际软件集团高级副总裁、首席科学家兼金蝶研究院院长。清华大学工学博士，毕业后入选并完成IBM "Micro MBA"和IBM"现代服务业领袖"培训计划，担任过IEEE服务计算专业委员会首任主任。现被聘为武汉大学软件工程国家重点实验室、北京邮电大学网络技术研究院客座教授，同时担任中国旅美科技协会大纽约分会会长，IEEE服务计算汇刊(IEEE Transactions on Services Computing)创刊总编，IEEE国际服务计算大会、云计算大会、WEB服务大会等学术会议的执行委员会主席，以及The Open Group全球董事会董事兼中国分会会长。他是IEEE Fellow、ACM杰出科学家和“千人计划”国家特聘专家。

信息技术在另类资产管理中的应用

杨波

摘要

另类投资是指除传统股票、债券和现金以外的金融衍生产品投资。另类投资常用来增强投资组合的整体回报并通过多元投资来降低投资风险。另类投资管理要求更为强化和具有特定功能的信息技术平台来处理日益复杂的日常交易和运营及满足投资者对高回报高增长的需求。管理另类投资的信息技术平台要足够灵活以处理各种另类投资产品的特殊需求，要易于扩容以处理不断增长的资产规模，要易于扩展以处理新生产品和迎合多变的市场需求和地区性差异。本文在简介另类投资的基础上探讨以对冲基金为主的另类投资管理各个环节的信息技术解决方案。

关键字

另类投资，对冲基金，信息技术

1. 另类投资介绍

在另类资产投资领域，从开始到结束的整个生命周期的投资交易管理集成软件平台存在着巨大的市场需求。另类投资定义为除传统股票，债券和现金以外的金融投资，常用来增强投资组合的整体回报，通过多元投资降低投资风险。对冲基金是另类投资领域里的典型代表。另类投资在欧美国家已经占据巨大的投资份额，已不再是传统的投资。投资交易的生命周期主要包括交易模型的建立，交易订单管理，投资组合分析，交易后处理，投资仓位管

理，抵押品管理以及支付管理和处理，基金会计，投资者服务报表及向监管机构提供的各类报表。为了管理一个投资交易的全生命周期，软件平台需要和市场中的各个相关机构进行通讯连接，这些相关机构包括交易机构，如纽交所、上交所；市场数据供应商，如彭博、路透社；交易执行经纪商，如高盛、摩根士丹利；交易中央配对服务提供商，如 Omgeo、MarkitSERV；资产托管商，如巴克莱资本、摩根大通；现金支付银行，如美洲银行、中国银行等。开发这样一个复杂的集成软件平台需要了解业务知识的专家，并采用先进合理的技术来建立这样的平台。

2008 年金融危机后，为了了解另类投资的发展前途 KPMG 对 26 个国家的 200 个组织于 2010 年 2 月至 6 月进行投资方向的调查^[1]。调查结果发现 52% 的受访者将会增加另类投资的资产配置，另有 33% 的受访者将保持现有的另类投资资产配置。这些数据表明大多数投资者都对另

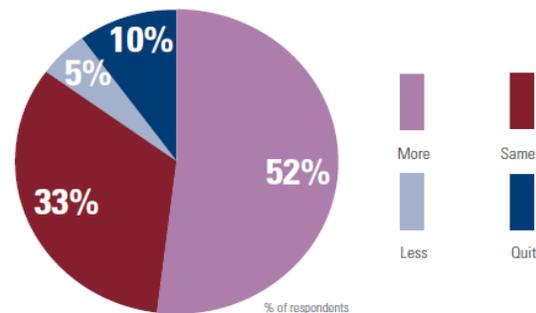


图1，愿意将资产配置于另类投资的受访者比例（根据KPMG于2010年2月至6月组织的全球调查）^[1]

类投资充满信心从而愿意将资产配置于另类投资以获得更多回报。图表 1 显示了被调查人对将来配置资产于另类投资的选项的百分比。

资成为防御型投资组合的主要部分，它可使投资在不可预知的市场中受益多样化。图 2 显示 2004 年至 2011 年另类投资对冲基金变化。中国正快速改变着投资管理

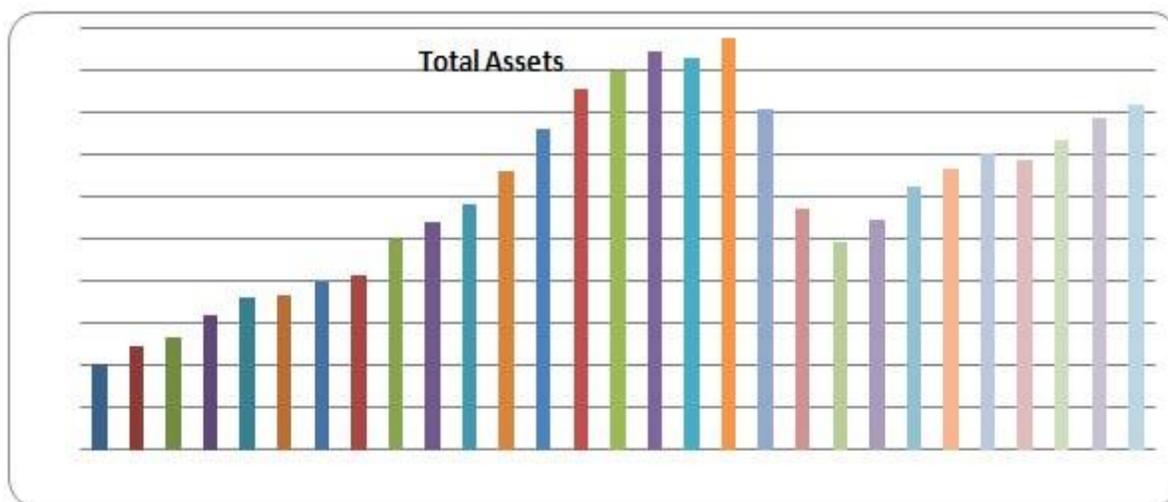


图 2, 对冲基金从 2004 年 1 季度至 2011 年 1 季度管理的总资产 (以 10 亿美元为单位)

由 HedgeNet.com 发布的对冲基金行业 2011 年 3 月 31 日的资产流及业绩报告^[2]表明从 2004 年 1 季度至 2008 年 2 季度期间，对冲基金管理的总资产一直保持平稳增长，在 2008 年金融危机时突然迅速下降，在 2009 年 2 季度触底后到 2011 年 1 季度统计时，对冲基金管理的总资产一直处于上升之势。这也证实了 KPMG 预测的大部分投资者会增加资产配置于另类投资这一结论。2011 年，另类投资份额占高净值个人投资组合的 24% (出自 2011 年世界财富报告)。另类投资在机构投资需求的驱动下将继续快速增长，尤其在养老金投资于另类资产方面。此外，基于近年来的投资趋势，另类投资已从许多传统管理人眼中的投资 " 特例 " 转变为一种核心资产配置。越来越高的认可度使另类投

策略，市场已开始创建和交易更多的金融产品，如基于外汇、汇率、股票和信用的金融衍生品^[3]。对于中国金融市场来说，拥有能够管理高度复杂的另类投资的全生命周期的高自动化高集成度的大型信息处理系统至关重要。本文将着重介绍另类投资管理生命周期中所要求的信息处理系统。

2. 另类投资的交易生命周期

交易的生命周期经历前台、中台和后台交易处理系统。每个阶段都有他们从开始到结束不同的管理职能。图 3 描述了一个典型另类投资管理公司里完整的交易生命周期和系统的主要功能以及各系统与其他系统之间的连接。

2.1 交易订单生成管理和前台技术

一个交易的生命周期从使用订单管理系统 (OMS) 或执行管理系统 (EMS) 下订单或合同开始。一个订单定义为客户端给交易商的指令，用于在一定条件下买入或卖出证券。合同是经过双方协商的买入或卖出证券的条款，这类交易通常称为场外交易 (OTC)。大部分的掉期如利率掉期，信用违约掉期，股权掉期等，都是通过场外交易。交易执行管理系统是管理证券订单执行的软件平台，其指令通常是通过金融信息交换协议 (FIX) 下达。执行管理系统有 4 个关键特性：交易记账簿 (blotter)，与交易所的连通 (Connectivity)，多目的

地指令下达和实时市场数据获取。订单记账簿主要有一个订单输入界面，用户可以根据市场条件生成订单和监控交易活动。该平台能兼容单个证券或投资组合交易。通过执行管理系统，用户可以连接到主要的证券交易所，电子通讯网络 (ECNs)，另类交易系统，跨越多个网络和交易商进行互联。实时市场数据由交易所或数据提供商供给，有时是从执行管理程序供应商提供的内部数据。两者比较，交易订单管理系统是设计为管理投资组合，合规管理 (compliance)，和交易活动监控。交易员使用订单管理系统为各种证券填写订单并

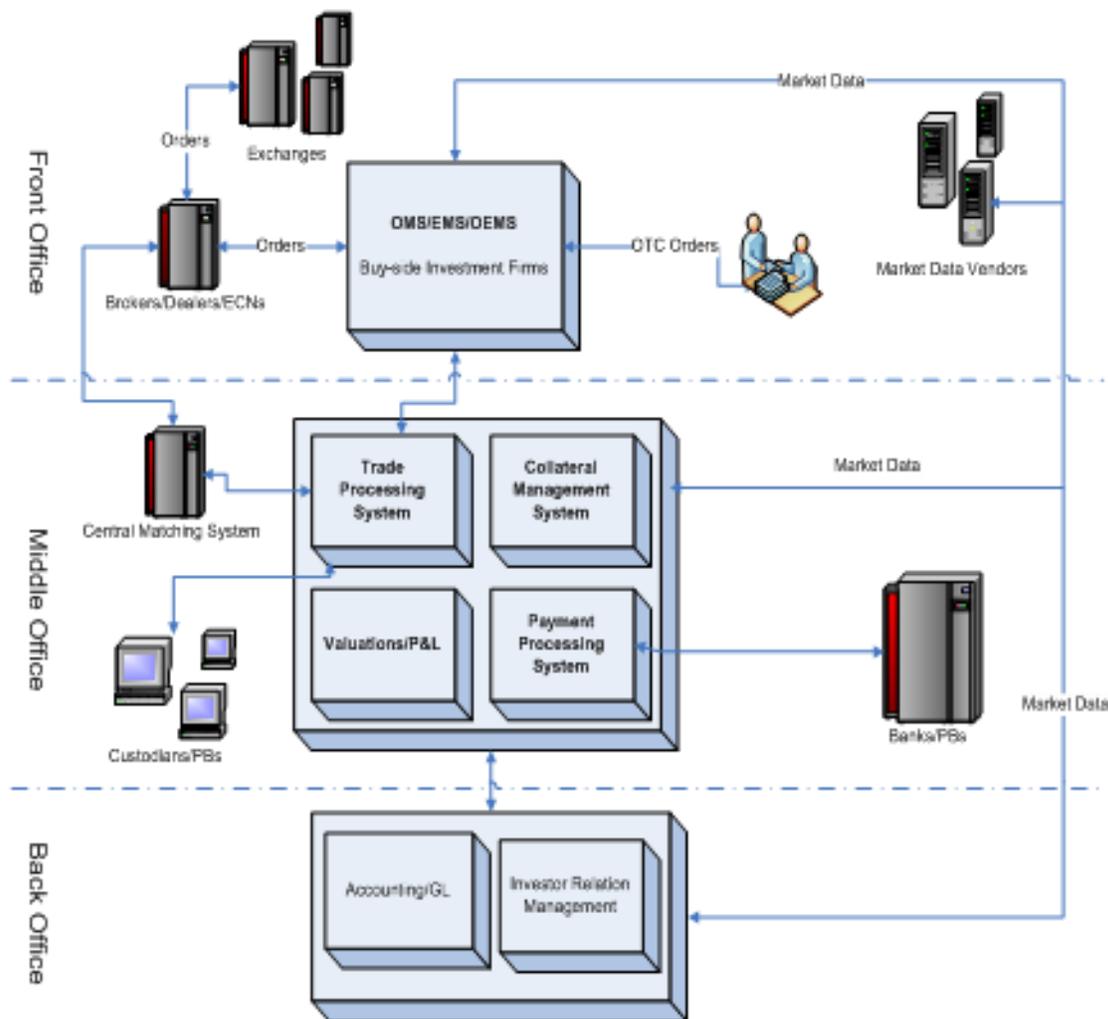


图 3，管理另类投资的信息系统功能和互联示意图

跟踪每个订单的交易过程。通常只有交易所会员能直接连接到交易所，交易中的卖方 (sell side) 通常有交易所的连接，买方 (buy side, 像对冲基金公司这样的投资公司) 是通过卖方公司的连接与交易所连接的。当前越来越多的交易系统设计为在一个集成平台同时提供执行管理系统和订单管理系统，称为订单执行管理系统 (OEMS)。

2.2 交易后活动和中台技术

交易下完订单后，由前端至中端的信息传输器将该交易传递到中台交易处理系统以对该交易进行中处理。图 4 描述了一个典型的交易中处理系统的功能和数据流。把订单输入交易处理系统有两种方式：从订单管理系统自动导入或在交易管理系统界面中手工输入。订单交易的产品由投资

品主信息 (investment master, IM) 来描述和定义。投资品主信息是前中后台共享的信息，其最初是从数据提供商处获得的。如果交易证券的投资主信息在交易处理系统中找不到，交易处理系统将自动发起请求向数据供应商获取信息并储存在系统中以备将来使用。在订单被输入交易处理系统后，交易处理系统与交易商交易系统确认交易细节。通常一个交易确认信息会从交易商的订单管理系统发到买方基金经理，然后买方返回一个确认信息确认交易细节。如果出现交易细节不符，双方需要审查原因并达成一致。由于基金管理公司会与大量交易商交易，与每个交易商建立自动确认/再确认连接是非常耗时的高成本工作。集中匹配系统 (Central Matching System) 用来桥接交易商和买方投资人直接的自动确认。交易商和基金管理公司都向集中匹

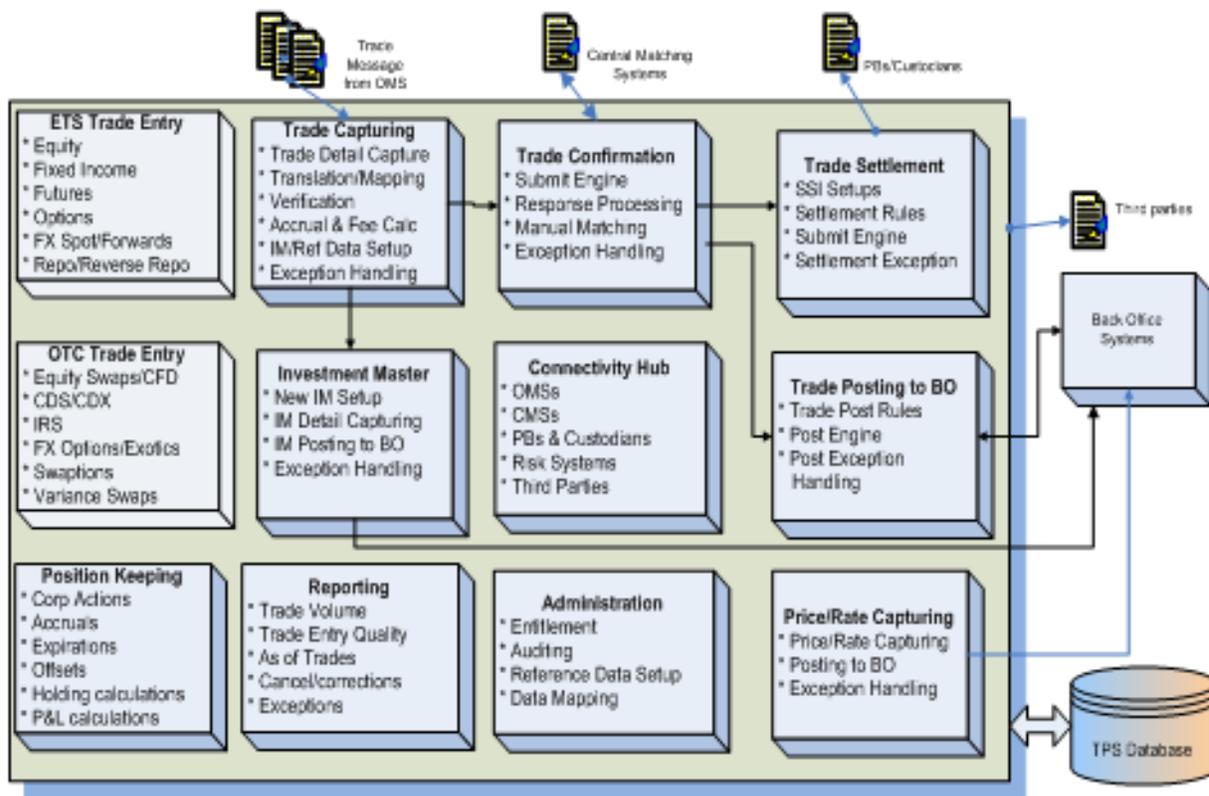


图 4，中台交易处理系统的功能和数据流

配系统提交交易细节，然后由该系统比较交易细节并发送结果报告。

交易匹配完成后，状态被更新为可以结算。一个交易结算提交引擎把交易信息和银行账户发送给负责结算的主经纪商(Prime Broker, PB)或托管机构(Custodian)。SWIFT(环球银行金融电信协会)通信协议通常被用来发送交易结算指令以使系统间接口自动化。SWIFT是向207个国家和地区将近8100家金融机构提供安全、标准化交易信息服务和相应应用软件的行业协会。在第二天，主经纪商或托管机构返回请求结算过的交易信息供对账使用。如果发现不符，投资经理修正交易清单并给主经纪商发送修正的交易信息。所以，一般交易不会在下单的第二天或第三天之前结算。在交易向主经纪商发送结算请求的同时，也可以向后台系统发送信息用于基金财务相关的活动。

在投资产品的持有阶段，为了使持有资产与市场的变动保持一致基金管理人需要进行日常性的行为。这些行为称为持有保障(position keeping)。根据特定投资品主信息的具体特征，这些行为包括处理公司行动如证券股息获取及计算,并购,重组,证券保期和合同重置管理,ISDA文档管理和合同执行,企业组合银行贷款服务等等。

为了管理基于合同交易如掉期,回购,和外汇交易等的交易对方的信用风险,需要有一个抵押管理系统(Collateral Management System)管理主合约风险和头寸,计算抵押品需求,发出保证金和抵押请求,验证交易对方的抵押请求,连接支

付管理系统以结算抵押,以及计算息票和利息等。

为了管理如抵押支付、投入和清算场外交易,终止和额外费用支付过程,需要有支付管理系统管理支付过程。这些支付管理系统(Payment System)需通过接口从网络向银行和主经纪人发送SWIFT格式的支付指令。

另一个中台范围的重要功能是计算投资损益(profit & loss)。系统的主要任务包括执行定价政策,验证价格,对交换和衍生品估值从而产生精准的基于策略,交易经理和基金级别上的日、月和年损益数据。系统将损益报表自动分发给投资人和其他相关各方。

2.3 后台服务和技术

后台的任务包括基金会计、基金行政和投资人关系管理。基金会计系统(Fund accounting),亦称通用分类账管理系统,可为同一个投资公司管理的多个基金机构及其合并主体提供合计运算,对包括潜在公司实体,特定目标交易机构,多方投资人的母基金,专项账户的基金会计及行政管理。计算服务范围包括计算每种证券的每日净资产值,管理费和预提费,维护全套财务信息,管理公司预提税账户,在依据申报当地管辖及投资人要求制定的公司会计制度之上,准备会计报表。基金管理要解决现金和资产持有中的不匹配情况,核准基金估值,处理购买,赎回,汇兑,资产转换,记录所有交易和非交易利息,计算管理和绩效费用,理顺和核准平衡表,在普通合伙人和投资人中分配收入,审核基金资产净值并审批通过。投资人关系管

理系统 (Investor Relationship Management) 提供股东服务, 注册和转换代理功能, 反洗钱和客户信息登录等。为提供这些服务, 需要一个综合的全面的基金会计系统和投资人关系管理系统。

2.4 另类投资对 IT 系统的整体要求

显而易见的是, 另类投资管理的直通式处理过程 (straight through processing) 对信息技术有非常高的要求。许多传统资产管理进入了另类资产领域^[4]。投资经理需要用多种交易方式管理资产, 如做多/做空, 套利 (固定收入、信用、波动), 商品期货, 全球宏观基金, 结构化基金, 对冲基金中的基金等等, 来满足不同投资客户的需求。订单管理系统将不再只管理做多做空证券的交易, 还需要管理信用, 场外衍生品, 债务和商品, 所以现代订单管理系统将更复杂, 并能够应对所有类型投资资产全部功能的综合性和风险健壮性。现在很多另类投资经理并没有技术产品来支持管理另类 and 结构化投资的多样化功能需求。数据经常分散存储在各种数据库和文件中, 一个公司整体报表视图需要大量的手工处理和计算才能产生投资历史和策略。这将会给运营人员带来较重的工作负担, 他们需要手工复制数据, 然后在离线状态下处理数据并把数据送到下游系统。另外, 商业或自主开发的系统与其他系统对接还需要不同程度的额外开发。这些系统分别为不同的投资需求设计。例如对冲基金, 基金的基金, 私募基金等。由于各系统技术还在不断变化, 与其他相关系统的集成形成了对另类投资公司的严峻挑战。

在根本上说, 另类投资技术必须满足以下需求才能满足客户和基金经理的期望:

- 客户想能够容易了解到他们投资组合在当前市场环境下的表现, 也就是与另类投资的历史表现比较, 还可以与传统类别投资资产比较。

- 公司必须能够进行如果-将会怎样的预测分析比较成本和绩效, 有工具包可以为客户定制模型、产品、 workflow、用户界面、横跨任何组合进行风险希腊 (Greeks) 参数和聚集计算。

- 信息必须经过裁剪以满足不同投资人的需求和偏好, 而基金经理将可以简单快速的从多个数据源抓取数据, 构建准确可靠的报表, 来展现一定范围类产品的整体的可定制的示图。

- 理想情况下, 信息始终是实时有效的, 通常通过互联网给客户交互体验。还可以通过智能手机, 移动设备接入互联网后获取信息。

- 另类投资平台应当无缝连接, 只需要每天很少的维护就可以处理多个变化的客户。

- 另类投资平台应当有完整的风险管理工具如压力测试、对冲情景、风险价值 (VaR)、极限监控等。

- 分析系统必须全面集成, 有可靠的估值计算和 24 小时监控。

3. 新一代另类投资 IT 系统的技术解决方案

对于新一代管理另类投资的IT信息系统需具有以下设计理念。

- 系统的高集成化及完整性。系统产品线功能完整覆盖交易完整生命周期，为用户提供整体化解决方案，避免不同公司产品数据格式不一致，无法及时同步而造成的信息孤岛现象，保证业务数据和报表的准确性。

- 面向服务的架构SOA设计。系统所有产品尽量基于统一的开发语言和服务器平台以及统一的后台数据库，系统内部基于SOA架构，通过ESB将各业务模块连接起来，可根据客户需求灵活配置业务模块和业务流程。

- 基于云计算的解决方案。现在另类投资公司可以通过SaaS的云服务平台所提供的一整套应用来满足IT服务上的需求。许多IT服务如基金公司财务，投资人关系，客户关系管理，网络门户等等都可以通过云服务方案提供，从而节省在IT基础设施和支持上的大量投入。对不同级别的客户，根据QoS配置自动分配处理请求所需带宽和运算资源，服务器可以根据需要扩充运算能力和数据存储，保证用户的数据安全和使用体验。

- 覆盖投资产品的多样性。系统可处理的金融产品种类，除了常见的证券、各种产品期货、国债和基金等，还包括更多欧美市场上的金融衍生品，例如国债、汇率的期货、掉期、期权、互换等。随着中

国金融市场的开放，能够处理计算这些复杂产品的系统将必不可少。

- 灵活多样的安装选择。系统可灵活部署运行，即可以独立安装在客户的数据中心提供定制服务；也可以运行于公司内部云计算中心，通过网络为用户提供服务，从而满足不同客户的需求。

4. 总结和结论

本文在简介另类投资的基础上，探讨以对冲基金为代表的另类投资管理每个环节的信息技术解决方案。着重介绍了另类投资日常运营中所运用的前、中、后台信息系统及相关技术。然后介绍了新一代另类投资 IT 系统的技术解决方案以对新系统的设计及开发提供参考。另类投资 IT 系统可归结为以下几点：

- 以对冲基金为代表的另类投资管理要求更为强化的信息技术平台来处理日益复杂的交易和运营管理。

- 管理另类投资的信息技术平台要足够灵活以处理多种多样另类投资产品的特殊需求。

- 信息技术平台要易于扩容以处理不断增长的资产规模。

- 信息技术平台要易于扩展以处理新产品和迎合多变的市场需求和地区性差异。

- 在设计新型系统时要充分了解最新信息技术并加以合理运用以满足现实及未来的需求。

参考索引

[1] Anthony Cowell , Andrew Stepaniuk , 2010 , Transformation The Future of Alternative Investments, KPMG International.

[2] HFN Hedge Fund Industry Asset Flow/Performance Report, 2011

[3] Noriyuki Oharazawa, Kyla Roberts, 2011, Fundamental Indexes: An alternative view of the Asia Pacific Region, Global Markets Exploration, Russell Investments.

[4] S. K. Rao , 2012 , Alternative Investments: Technology Solutions for Improving Client Experience , Tata Consultancy Services。

作者简介



杨波现担任美国证券软件与咨询技术控股集团(SS&C Technologies Holdings, Inc. (SSNC))公司的副总裁, 管理该公司基金服务及另类资产的产品研发及企业架构的整体实施。SS&C的金融软件及服务管理着世界上160万亿美元的金融资本。在SS&C之前, 杨波担任Citco Fund Services公司的运营技术总监。在Citco之前, 杨波担任OpHedge Investment Services公司的运营技术总监, 管理产品研发部 (Research & Development), 金融数据部 (Market Data), 金融信通部 (Connectivity), 及质量监控部(Quality Control)组成的全方位地支撑另类投资日常运作及长远发展的集技术与业务知识于一体的运营部门提供为Soros Fund Services之类大型对冲基金的日常服务。在OpHedge之前杨波在美国Pitney Bowes公司从事技术和管理工作。杨波于2000年获新泽西理工大学机械工程博士和计算机科学硕士学位。1994至1996攻读华中科技大学CAD中心博士。于1994年获哈尔滨工业大学机械工程硕士。杨波在校期间发表16篇科学技术期刊论文, 荣获2000年IEEE/ISEE最佳学生论文一等奖第一作者, 一项美国专利的第一发明人。

从服务质量的角度来看待社交网的挑战和机会

周念军 (jzhou@us.ibm.com)

(Research Staff Member, IBM T. J. Watson Research Center, Hawthorne, NY 10532)

摘要

随着社交网的兴起，产生了新的人际交流方式。今天，我们正在见证这一方式对社会的各个方面正在发挥着的巨大影响，并在一步步地改变我们的生活方式。在社交网和社交媒体蓬勃发展的同时，也带来了新的管理方面的挑战。在这一新的媒体上，传统的基于长期认识的和相互信任的小范围内的人际关系的被大大地扩展了到了相对缺少互信的巨大的社交网上。今天，除了传统的权威性的媒体信息发布者以外，存在着大量的自发的非专业的业余信息提供者。传统的媒体空间被大大地压缩。同样，传统的零售和服务受到了社交网上的提供的跨区域的和面向所有网上用户群的服务商的激烈竞争。以前自上而下的互动，被横向互动的方式而代替和补充。新的互动，带来了人际交往的随意性和非持续性。并随之而来的带来了很大的服务质量和服务管理问题，提出了对社交网和媒体服务管理的挑战。在本文中，我们要从服务质量的角度来探讨社交网的现状和发展趋势。并在进而进一步讨论社交网可能会带给我们的新机会。

关键词

社交网；社交媒体；众包；博弈；服务平台；知名度；实名认证

1.前言

近年来，社交网和社交媒体在全球兴起和发展。在中国，以新浪微博、人人网和腾讯为代表的社交媒体的也跟这一趋势同步地高速发展。信息的传播，从垄断性的大众传播开始向互动为主导的格局转变。年轻一代，使用社交媒体，活跃在面书 (facebook)、推特 (Twitter)、博客、微博、淘宝和其他网络平台上。随着新的智能手机的普及，和通信带宽增加和价格的大幅降低，使用社交网络人数将在全球范围内进一步地持续增加^[1]。用户从年轻人向各个年龄层发展，并由个人使用变为个人和企业维持信誉和知名度的重要场所^[2]。社交网上的功能将由开始的交友，信息的共享和网上交易，进一步发展为全面的服务的整合平台。为整个社会的各种信息的创造、加工、获取和传播方式引发深刻的变革。进而影响我们的生活方式。并进一步地改变我们的彼此间的互动，带来文化、经济和政治的重大改变^[3]。举例而言，美国现任总统奥巴马就利用面书在 2008 年的总统选举中得到了很多年轻人的支持^[4]。

由于社交网的出现以及和服务业的整合，正在把传统的无法进行精确定量分析的服务行业慢慢地带入可以进行更优化的管理的行业中。在传统的服务中，提高服务质量的最大的障碍是无法对服务进行定量的优化。为了提高服务质量，我们必须能够能够在潜在的服务对象和服务提供者

之间建立一个可持续的联系。而社交网提供了恰恰提供了一个低成本的社会联系的机制和一种可以进行分析和预测的网络结构^[18]。这一变化为新的经济形态与模式在提供了一种可能。在下一阶段的发展中, 社交网正从纯粹的社交场所演化为一个可能和现代服务相结合的一个交易平台。在当前, 这一发展还刚刚体现在广告商在社交网单向的信息发布的广告上。经济和简便的信息发布机制给很多的服务商以及各种政府机构、大学提供了直接面对它们的潜在的服务对象的机制^{[6][7]}。同时, 在政府的层面, 利用社交网来提高公共服务才刚刚开始, 公共服务要讲究公平性和民主性, 因此如何结合社交网提高市政服务将是一个很大的题目。对于众多的非盈利组织, 社交网的出现也为他们提供了很多的机会。一般而言, 非盈利组织的生存和发展依赖于大量的志愿者的参与。而如何低成本地找到这些志愿者一直是困扰非盈利组织的大问题。在不久的将来, 如何利用社交网来寻找高质量的志愿将是一个很有意义的一个发展方向。今天, 这一想法已首先在软件开发上得到应用, 也就是众包的概念(crowdsourcing)^[5]。众包利用大量社会上的计算机的人才的创意和能力来合作和完成对社会大众有益的软件应用开发。毫无疑问, 这一种新的组织劳动力的全新方式, 同样适合其他需要降低开发成本和降低服务费用的领域。而对于非盈利组织来说, 这是一个能少用费用而完成高质量的服务的最佳途径。

2. 社交网发展面临的挑战

在传统的社交体系中, 我们的信誉是以互相认识和彼此推荐为依托, 并由整个社交圈作为一体来维护这个信誉的。我们在这个社交圈寻求生活上方方面面的支持。任何个人或企业对社交圈内的个体的侵犯(帮助), 就会得到整个社交圈所有人的惩罚(鼓励)。而在今天的社交网上, 由于技术的支持, 我们开始拥有和更广泛的人群的接触的能力, 而他们的大多数可能并非在我们的传统的社交圈之中, 并完全可能是匿名的。这样的话, 传统的社交圈是无法保护我们在社交网上的行为和利益。今天, 在社交网上, 由于缺乏类似于传统社交圈的信誉和信用机制, 我们可以看到许多在社交网肆无忌惮地侵犯他人的利益的例子。这一缺陷, 正影响着社交网的健康发展。而如何建立新的社交的信誉机制, 是今天社交网发展面临的巨大挑战。

任何社交网的三方: 1) 平台的提供者; 2) 交流的发起者; 3) 交流的跟随者三者形成的是一种博弈关系。参加的各方具有不同的目标或利益。为了达到各自的目标和利益, 各方必须考虑对手的各种可能的行动方案, 并力图选取对自己最为有利或最为合理的方案。一个好的平台要求他们三方之间有明确的互相的牵制关系。好的社交网游戏规则的核心是在于相互发生作用的三方之间应有明确的互相的牵制的关系和具有互相有强制力约束力的能力, 使任何对规则的破坏都有可能得到抑制。如果没有的话, 就会产生非合作的博弈。这会为商业欺诈和谣言的散布提供可能, 进而导致整个社交平台的总体信誉的大幅度地降低。

如何来进一步分析这个问题呢？我们可以由易而难地在以下三个层面来讨论社交网的管理方面的挑战——1)发现违规行为了的容易度；2)违规行为的惩罚的可能性；3)游戏规则的公平性。首先，让我们看一看今天在社交网上违规行为的检测的容易度。在信息互享的社交网站总有一些黄色和对他人名誉损害的和网络垃圾的消息，而出现这一问题原因在于社交服务平台首先缺乏对这一类行为的检测机制。由于处理成本太高，或由于加害者是匿名的。这一问题很难在用户个人层面加以解决。在社交媒体中，除了实名制的邻客(LinkedIn)、面书(facebook)和人人网以外，在大多数的情况下，社交网上的用户是匿名的。这一匿名的方式，对于信息共享的服务为信息的发布者提供了宝贵的保护。但由于这一匿名性，在要求社交网提供高质量的服务时，平台的提供者自己必须扮演一个强力的仲裁者和执行者的角色。对这一类有违公德的行为予以合适的惩罚和抑制。

一种可行的解决的方法是在在技术的层面来预防。比如利用图形和内容的分析来提前发现黄色照片和网络垃圾^{[11][12]}，和此社交网的用户关系网，来对用户的内容和行为加以预测，包括对用社交网来进行对社会有害的联系^{[13][14]}。另一方面，我们可以用用户的分级来减少这一类问题的危害。根据用户的信用程度和是否被实名认证的进行分类。对不同的用户提供不同的权限，对那些可以被证明是可靠的用户，提供更大的自由度。从技术的层面上讲，用户的信用可以通过对用户的使用记录的分析来完成。在今天的社交网上，尽

管对实名制的鼓励正成为一种趋势，但大多数的用户的活动仍然是匿名的。当然，必须指出，在社交网刚刚发展的时候，匿名的方式为社交网的发展提供了动力，大大地保护了个体的自由和保护了个人的隐私。但当社交网全面发展以后，应成为我们获取信息的主要和提供社交的主要的方式以后，仅仅靠匿名的方式在社交网网上进行交流是不够的。因为，匿名某种程度保护了那些社交网的散布谣言和欺诈的破坏者。未来的社交网的管理必须在匿名和实名中间取得平衡。我们应逐步加强社交网的实名认证制度(比如在新浪微博的实名认证)，特别是对那些社会知名人士的实名认证。同时，我们在提高实名认证的用户的信用度同时，可以降低匿名用户对社交网的信用度，进而减少匿名用户滥用社交网的危害。

第二，我们来分析一下如何在社交网上对违规行为的进行惩罚。今天，在社交网上有很多的诈骗行为。这类行为对社会和对社交网的危害都很大，影响整个社交网的声誉。由于社交网大大地提高了信息的传播的效力，这一特征特别有利于把社交网发展成为现代化服务的平台，而这类诈骗破坏了商业行为中最重要的信用和信誉基础。如果没有在技术上提高对这一类行为的设别，和进一步在法律上加以制裁，那么就可能窒息很有前景的利用社交网来提升服务现代化的发展。

今天在以社交网为基础来发展服务业，仍有许多商业上的挑战要解决：例如 1)信任、信用、信誉的缺乏；2)服务和产品信息不对称；3)缺失适当的反馈机制等一系列的问题。这些都大大高于在网上发展

服务业的风险。我们必须充分地认识到这些问题^[9]，并努力在技术的层面上找出解决的方法问题^{[10][12]}。而其中，建立可靠的社会网信用体系^[8]应是最重要的第一步(包括商家实名认证和实行经由认证的评价系统)。这样，我们就能把社交网对经济的发展，从简单的提供商业广告，进一步全面扩展到一个服务平台，进而提升全民的服务业的质量和效率。

举例而言，今天，在社交网上有很多的诈骗行为。由于诈骗的发起者的很多的时候都是匿名的和生活在境外，抓到他们的可能性不是很高。解决这一问题的方法可以在技术上加以预防。对于有商业性的社交网，在现阶段可以用第三者介入而提高服务的可逆性。如何由第三方的介入解决商业欺诈的问题？对于随着社交网兴起的网络交易(eBay 和淘宝网)，正成为新的网上商业活动的重要组成部分。由于这类交易都是在小商户和小批量采购的个人间进行的，商业信用度很低。如何防止欺诈就变得很重要^[21]。在今天无法直接在两者之间建立直接的约束力的情况下，通过加入仲裁第三方的介入，在双方的博弈的利益函数上加入一个时间延时因子。提供受伤害的一方有机会发现问题，进而通过第三方取消交易或制裁对方的机会。这样就可以很快在社交网上网络交易的平台中，把规则的破坏者赶出去。

最后，保证社交网游戏规则的公平性也是我们面临的一大挑战。今天，参与制定规则的主要是社交平台的提供者和政府。但是，必须指出的是提供商和政府本身也是(见下节的讨论)社交网发展的主要利益的相关方。如果，我们希望有个长远和

稳定的社交网，进而发展为新的服务平台，我们必须定出适合于所有参与者的游戏规则。而这方面的工作才刚刚开始起步，各国法规的制定仅仅集中在网络的安全性方面。而如何对用户的私权的保护¹，在公私权的之间如何鉴定和管理，如何在社交网上解决出现的各种名誉和商业纠纷，都将是我们要面对的问题。

所有的一切，都提出了在社交网上建立一个信用体系提出了迫切的要求。今天，在多个相同或类似社交网彼此竞争的商业环境下，一个有意义的题目是如何建立一个独立于任何社交网的社交信誉系统。为了体现公平性，可以由独立的非盈利组织来提供这一独立的认证和信用的系统，并成为公共服务的一部分。通过政策和技术的调控，使实际起作用的社交网的激励和惩罚机制和长远的社会目标一致起来。提高社交网提供服务的可信度，进而达到提高社会总体效益和减小社会总体成本的目的。

3.社交网和服务经济的发展

随着社交网的成熟，除了大量的个人用户外，也加入了很多的企业、机构和政府。社交网已超越信息交换的功能，渐渐成为整合社会服务的平台。社交网的服务内涵正向全面的生活需求服务发展。当然，这是一个长期的发展的过程。但从国际和国内社交网站的发展来看，这一判断

¹ 在 2011 年下半年炒得沸沸扬扬的关于“丈夫离世，妻子有无权力继承 QQ”的新闻明显反映了世界上社交网立法迟后的问题。这个继承的要求有腾讯直接来回答，反映了这一问题的困惑。以为鉴定用户私权的问题，本应由司法来加以回答。

<http://news.sohu.com/20111013/n321995474.shtml>

是适合的。我们开始看到以社交网为基础,以信息服务为核心的,整合新闻服务、旅游和餐饮服务(包括和移动服务的结合)、广告服务和银行服务的平台在一天天成熟起来。由于社交网的综合整合力和大量地用户群,社交网的用户可以通过社交网获得更好的服务。由于社交网信息同享的特征,使得服务的提供方必须对服务价格和质量加以认真的对待,以求其商业信誉,而求得长期和稳定的客户群。

在所有的商业活动中,首先进入社交网的是广告业。按照美国的市场调研公司 eMarketer (<http://www.emarketer.com/>) 对社交网络广告的估计,全球社交网络广告客 2012 投入 77.2 亿美元,而到 2014 年,总投入可近 120 亿美元。国内的多家公司也在积极开展微博营销。在新浪和博讯微博上我们能看到各种大的品牌的官方微博。而各种中小企业也不时地发条广告。这一新的广告模式对于中小企业和要求特殊服务对象的企业尤其重要。

实名社交网站以其与社会极高的相关性而拥有其他互联网媒体不可比拟的“真实”优势。今天,国际上的邻客和面书以及中国国内的人人网采用了实名注册方式,用户提供了真实的社交圈子。由于采用实名的方式,大家可以更方便和精确地找到想要的服务和信息,服务商可以精确选择服务对象,从而造就了一个极为方便于精准营销的平台。在实名社交网站营销层面上,由于提供了便捷分享平台,用户可以很高兴和朋友分享购物或其他服务的体验,而服务信息可以为服务商提供的可靠的用户行为的监测报告,从而为用户提供更加

精准和高质量的服务,而形成一种在社交网上的良性循环。

毫无疑问,公共服务是政府的责能。和其他的商业服务一样,在社交网上,公共服务的质量受到了大众的监督。在强调民主和法制的现代社会里,社交网的普及不仅为各级政府的管理带来了新的挑战,同时也为和公众互动和建立公信力带来了新的机会^[20]。在美国的一些小城镇,已经利用社交网这一手段,来提高重大公共项目决策过程的公众参与程度。事实上,开放对社区有重大利益的项目的讨论,实际上是公众和政府同时分担公共项目的开发的责任和风险。这样可以减少以后的争议,使重大和长期的项目的发展能够避免政府换届的影响。毫无疑问,利用社交网来提高公共服务的质量和公众参与度,是一个很有前景的发展方向。当然,我们也要注意事情的另一面,在社交网提供民众越来越多话语权的时候,如何提高民众在社交网上社会责任心,让社交网的讨论更理性和规范也是迫切和重要的题目。今天,许多国家政府近年都在加紧在社交网上的立法,为保护公众利益和公众参与找到新的平衡点。

社交网拥有了大量的用户的信息。今天,由于在系统结构和分布计算方面的成功^[16]。社交网已经能可同时支持上亿的用户。而大量的科学研究也开始集中在大量有网上搜集的大数据研究^[18]。这一类研究是随着互联网的巨大的用户群而发展而兴起的,并随着各种各样的信息和服务在社交网的整合,大量针对优化的数据挖掘^[19]和实时服务正在兴起,并一步一步地整合到社交网中。对实行实名认证的用

户和社交网(像邻客和人人网),还可以提供更准确的年龄、性别、学历、工作单位等社会信息,以及用户动态的地理位置信息(这一动态的信息可以由用户的登录点的IP地址或手机的地址获取)。就基础信息而言,从社交网上搜集的数据不仅有用户的数量和使用数据、个人兴趣,个人的社会关系网,还包括地区间联系的强度、社会的动态、产品的欢迎程度、甚至能可挖掘到季节性的流行感冒的发展趋势^[22]。

随着信息格式定义(XML)的普及,及信息交换(RESTful)的应用,社交网和其他服务的结合在技术上成为非常容易的事情^[17],大量的商品和服务信息和浏览和评价信息正在整合到同一的数据分析中。这一信息为政府的公共服务和企业的私人服务提高定量化和优化提供了大量的数据。其中的一个重要的应用就是对广告点击率加以分析。今天,广告的费用是以点击率来定价的,但很多的点击是无效的。如果我们在技术层面对社交网上的用户的特性加以分析,找出潜在的有价值的用户,进而提高广告点击率转换为购买力的能力。

最后,在这一新的社交网成为新的信息的和整合的门户时,要注意的可能形成的新的商业和信息的垄断的可能性。由于社交网成为信息和整合的平台,扮演者承上启下的服务经销商的角色。而有着一般用户(即便是明星用户)和一般传统的经销商没有的巨大权力。如何在社交网上维护一般用户和商家的利益不受社交网本身的侵犯也可能是我们要关注和立法的问题。

4.小结

在本文中,我们回顾了社交网的发展,展望了以后商业发展的前景。今天,不管是个人、企业、机构、和政府,我们在社交网的存在和社交网上对我们的评价,都会影响我们的在现实生活中的生存和发展^[2]。而随着社交网和其他系统的结合,社交网正在为我们提供一个极好的机会来促进我们的服务业的整合和更快地走向服务业的现代化转型。

为了加快社交网对社会变化的正面影响。我们需要在三个方面加以努力和改进。首先,我们要尽快地建立社交网的社会信用体系。第二,我们要加强利用社交网的综合能力,对社会的动态、服务的优化进行大数据方面的挖掘和利用,提高公共和商业服务的水平和效率。第三,要加强对社交网的立法,使我们的公权和私权有一个很好的保护。为社交网的健康发展提供一个可靠和坚实的法律基础。

参考文献

[1]Dube J., *Why smartphones are the future of social networking*; Forbes, May 08, 2012 (web article – www.forbes.com).

[2]Madden, M. & Smith, A. (2010). *Reputation Management and Social Media*. Washington, DC: Pew Internet & American Life Project. Retrieved May 26, 2010 from http://pewinternet.org/~media/Files/Reports/2010/PIP_Reputation_Management.pdf.

[3]Rory O'Connor R., *Friends, Followers and the Future: How Social Media are Changing Politics, Threatening Big Brands, and Killing Traditional Media*, Publisher City Lights Publishers,ISBN-10

0872865568 <<Book>>

[4]Williams, C.B. and Gulati, G. J., *Explaining Facebook Support in the 2008 Congressional Election Cycle*,2009, Working Paper. http://opensiuc.lib.siu.edu/pn_wp/26

[5]Leimeister J., Huber M., Bretschneider U.,Krcmar H.,*Leveraging Crowdsourcing: Activation Supporting Components for IT-Based Ideas*, Journal of Management Information Systems, Vol. 26 Issue 1, Number 1 / Summer 2009, pp197-224.

[6]Weber L., *Marketing to the social web: how digital customer communities build your business*, John Wiley & Sons, Inc. New York, NY, USA 2007, ISBN: 9780470124178 <<Book>>.

[7]Evans D.,*Social Media Marketing: The Next Generation of Business Engagement*, John Wiley and Sons, Sep 16, 2010 - 400 pages <<Book>>

[8]Dwyer C., Hiltz S., Passerini K. *Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace*,Americas Conference on Information Systems (AMCIS) 2007 Proceedings, 2007.

[9]Mutz D.C.,*Social Trust and E-Commerce, Experimental Evidence for the Effects of Social Trust on Individuals' Economic Behavior* , Public Opinion Quarterly, 2005 - AAPOR

[10]Ma H., King I.,Lyu M.; *Learning to recommend with social trust ensemble*; Proceeding SIGIR '09 Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, pp 203-210.

[11]Chan Y., Harvey R.,Smith D., *Building systems to block pornography*, Challenge of Image Retrieval, Newcastle, 1999

[12]Markines B., Cattuto C., Menczer F., *Social Spam Detection*; Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web, Pages 41-48, 2009.

[13]Otte E., Rousseau R., *Social network analysis: a powerful strategy, also for the information sciences*; Journal of Information Science 2002; 28; 441.

[14]Yang C.C.,Liu N.,Sageman M., *Analyzing the Terrorist Social Networks with Visualization Tools*, Intelligence and Security Informatics, Lecture Notes in Computer Science, 2006, Volume 3975/2006, 331-342, DOI: 10.1007/11760146_29.

[15]Kirkpatrick D., *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*, Simon and Schuster, Jun 8, 2010 - Business & Economics - 372 pages

[16]Thusoo A., et. al., *Data warehousing and analytics infrastructure at facebook*, SIGMOD '10 Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, pp 1013-1020.

[17]Feiler J., *How to Do Everything: Facebook Applications*, McGraw-Hill Professional, Feb 11, 2008 - 330 pages

[18]Asur S., Huberman B.A., *Predicting the Future with Social Media*, Proceeding , WI-IAT '10 Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and

Intelligent Agent Technology - Volume 01,
Pages 492-499.

[19] Bakshy E., Hofman J.M., Mason W.A., Watts D.J., *Quantifying influence on twitter*, Proceedings of the fourth ACM international conference on Web search and data mining, pages 65-74.

[20] *Social Networks and Government*,
<http://www.howto.gov/social-media/social-networks/>

[21] Boyd J., *In Community We Trust: Online Security Communication at eBay*, Journal of Computer-Mediated Communication Volume 7, Issue 3, page 0, April 2002.

[22] J Ginsberg, et.al., RS Patel, L Brammer Detecting influenza epidemics using search engine query data, Nature, Vol 457, Letters, 19 February 2009.

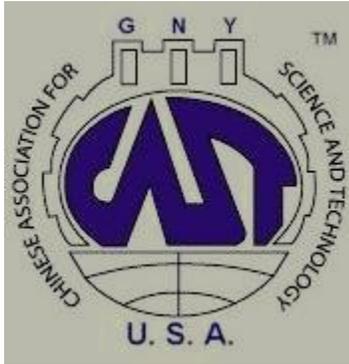
作者简介

Dr. Nianjun Zhou (周念军) is a research staff member (RSM) of IBM T.J. Watson Research Center. Currently, his research is in the area of services research including services economics, services computing, cloud computing, SOA architecture, and project estimation. Currently, he is the Chair of Professional Interest Community of Services Computing at T.J. Watson Center. Before joining IBM, he has been a research scientist of New York State Government. Dr. Zhou received his Ph.D from RPI, focused on analysis of Ad Hoc network routing overhead for variable topology network.

中国旅美科技协会大纽约分会简介

方彤

中国旅美科技协会董事会主席



中国旅美科技协会是于 1992 年 8 月在美国纽约注册成立的一个非政治和非盈利的华人专业人士社团组织。来自大纽约地区的历任会长是周华康博士、张群博士、徐震春博士、陆重庆博士、马启元博士、周孟初博士、谢家叶博士、石宏博士和左力。2002 年，为适应旅美科技协会在全美发展以及各地区分会成立的需要，基于纽约州和新泽西州成立了旅美科协大纽约分会。分会历任会长包括谢家叶博士、张宽博士、左力、陈志雄博士和方彤博士。张良杰博士是现任会长，陈志雄博士科协大纽约分会理事会主席，方彤博士是科协总会董事会主席。

目前科协大纽约分会拥有来自美国纽约州和新泽西州 2000 多名会员，会员大多数拥有科技、教育、商业、法律、医学、艺术和其他领域的高级学位。科协大纽约分会是大纽约地区华人专业人士最知名和

最活跃的华人科技社团组织之一。一直致力于促进中美之间的学术和人才交流，以及会员的职业发展。科协大纽约分会还通过举办多种形式的研讨会、学术会议、IT 高峰论坛、海外人才交流论坛等，努力打造科技、经济、教育和贸易等领域的中美合作与交流平台。此外大纽约分会还组织很多高层次代表团参加在中国的各种学术和人才交流会议。通过和其他专业协会的合作，旅美大纽约分会在促进当地社区发展、加强会员间交流和友谊等方面起到日益重要的作用。

作为旅美科协创始分会，旅美科协大纽约分会将与其他分会，为打造中美专业人士沟通的桥梁做出更大的贡献。同时，旅美科协大纽约分会也将继续致力于帮助会员们在美国或中国的进一步的职业发展。



电子邮件: info@castusa-gny.org

网址: www.castusa-gny.org



举办 IT 高峰论坛



海归创业 -- 旅美科协成功协办 2011 年浙江省纽约大型人才洽谈会

第三个重要信息空间：车联网技术与应用分析

程久军¹ 周孟初^{1,2}

(同济大学嵌入式系统与服务计算教育部重点实验室, 上海, 201804)¹

(Department of Electrical and Computer Engineering, New Jersey Institute of Technology,
Newark, USA, 07102)²

摘要

车联网作为新兴产业物联网最能够率先取得突破的应用领域, 发展车联网已经成为一项具有国家战略意义的重要决策。然而, 随着车联网热潮的升温, 在技术、标准以及社会资源的整合层面上, 车联网的发展仍会遭遇诸多瓶颈。本文从发展车联网的必要性入手, 剖析了目前国内外车联网在国家政策、市场趋势、技术与行业标准等现状, 并对车联网目前的研究现状进行了分析, 希望对促进车联网健康快速发展、满足国家经济和科技发展需求具有一定的应用价值和现实意义。

关键词

车联网; 专用短程通信技术 (DSRC); IEEE 802.11p; IEEE 1609

1. 发展车联网的必要性

随着汽车保有量的持续增长, 汽车已经变成人们除“家与办公室”之外的第三重要空间, 城市交通的发展已经无法满足汽车保有量的增长, 堵车成为社会性问题, 常发生于世界各大市区, 如北京、上海、纽约、华盛顿、伦敦、巴黎、韩国、新加坡等。汽车使用率增加是导致交通堵塞的主要原因。由于汽车的方便, 导致市区内车流日益升高。汽车的一大缺点就是十分

浪费空间, 随着数量的不断增加, 导致现有道路无法负荷如此大的车流量, 造成交通拥堵。据预测, 车联网的应用可以使交通拥堵减少约 60%, 使短途运输效率提高将近 70%, 使现有道路网的能力提高 2~3 倍。以动态交通信息服务为基础的车联网服务, 成为缓解该问题的重要方向之一。

交通事故已经成为全球公共的交通安全问题。当车主身处浓雾或者超视距范围的地方时, 能见度低, 驾驶汽车往往面临着严重的危险。中国每 5 分钟有一人因车祸死亡, 每一分钟有一人因车祸伤残, 每天死亡 280 多人, 每年死亡 10 万多人, 每年都会因交通事故而致死、伤残大量人员, 财产损失更是一个天文数字。如果能获得邻近车辆的实时信息, 包括车速、行驶方向、位置等, 车主就能在第一时间内做出相应的操作以避免交通事故的发生。

在停车场泊车方面, 一份来自通用汽车的调查显示, 目前, 在超大型城市, 30% 的石油浪费在寻找停车位的过程中, 造成车主每月多支出 336 元, 七成车主每天至少碰到一次停车困难。传统汽车需要超过 10 平方米的车位, 90% 的时间处于停车状态。车联网的应用可以在很大程度上能够解决停车困难问题。

如同因特网中的电脑、移动互联网中的手机，车载终端是车主获取车联网最终价值的媒介，用户长时间在公路上驾驶和旅行的人们，需要方便、及时地得到各种消息，并得到必要的视听娱乐服务，这些需求在催生并推动车联网的发展方面，将会发挥重要作用。

2. 车联网国内外发展现状分析

2.1 车联网国内发展现状

2.1.1 国内市场现状

咨询机构研究表明，中国汽车保有量和销量还在快速增长，在 2011 年中国汽车保有量将达 9100 万辆，较 2010 年增长 21%。中国车联网产业各环节厂商均在积极运作，为即将到来的产业爆发布局。2009 年车联网用户规模达 300 万，至 2011 年用户规模将突破 450 万，预测 2014 年车联网用户增长速度将首度超过汽车保有量的增长速度，用户规模将达到 1310 万，车联网用户初具规模效应，车联网市场迎来爆发点。如图 1 所示。

自 2010 年底，车联网系统已开始应用和普及，安吉星系统几乎覆盖了上海通用旗下所有车型，而东风日产、荣威等品牌也都拥有了属于自己的车联网系统，车联网已逐渐由概念走向应用。目前我国已经有超过 20 万用户正在体验车载信息服务，有业内人士预测，随着车联网相关技术的逐渐成熟，3G、CallCenter、云平台与语音识别技术等均已开始与车联网产业深度融合，在市场需求带动下，车联网终端设备有望迎来爆发式

增长。预计未来 5 年车联网产业的产值将有望超过 1000 亿元。专家预测，未来 10 年将迎来国内车联网市场的高速发展期，汽车“后市场”即将迈向“车联网”时代，预期 2020 年实现可控车辆规模达 2 亿。

2.1.2 国内体制现状

当前，汽车信息化已成为我国物联网产业“十二五”规划的重要融合推广和示范应用领域。早在去年 7 月，交通运输部就已经明确提出要推动车联网、船联网的建设；国家汽车移动物联网工程是国务院决策，并且得到高度重视和多次批示的国家战略工程；去年 10 月，国务院在 863 高科技研究发展计划中，提出两项涉及车联网关键技术的项目，就是智能车路协同关键技术研究、大城市区域交通协同联动控制关键技术；今年 4 月，交通运输部、公安部、国家安监总局、工信部四部委联合下发《关于加强道路运输车辆动态监管工作的通知》，明确提出，从今年 8 月 1 日起，“两客一危”车辆出厂前应安装符合规定的卫星定位装置，这意味着至少在客运车辆和危险品运输车辆中，车联网产

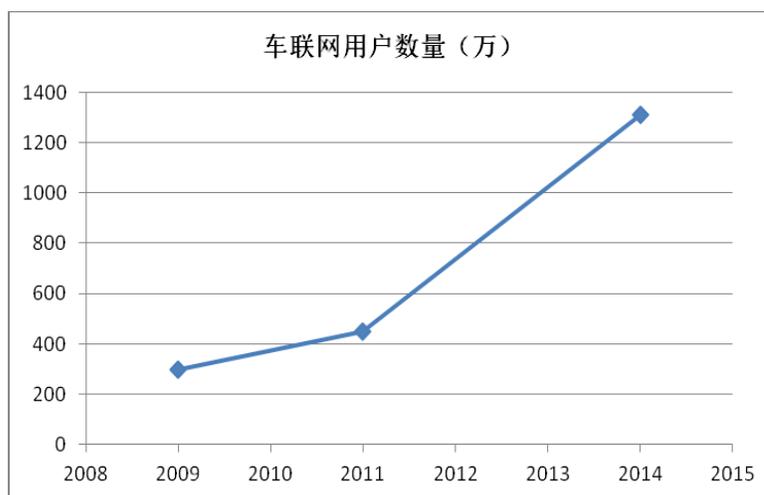


图 1 车联网用户数量增长趋势图

品将成为标配；国家工业和信息化部在其主导的《物联网产业“十二五”规划》草案中，也明确将从产业规划、技术标准等方面着手，加大对车载信息服务的支持力度，以推进汽车物联网产业的全面铺开。汽车物联网项目已被列为我国重大专项，将获财政扶持资金，扶持资金将集中在汽车电子、信息通信及软件解决方案上，车联网平台投资需求或超过百亿元。

世界各科技强国都将车联网放到未来科技经济发展战略中的重要位置，大力发展车联网产业已经成为我国今后一项具有国家战略意义的重要决策。

2.1.3 国内科技现状

2009年，长安汽车对国内外智能交通和主动安全技术的发展现状、产业化前景以及国内基础等进行充分调研和论证，在此基础上，制定实施了重点发展基于智能交通的汽车主动安全技术的战略规划，并于2010年与清华大学开展基于机器视觉的车道偏离和前方障碍物预警系统的研究，已经完成样车开发。该样车具备车道偏离报警、自适应巡航、前撞预警功能。

2011年4月11日，国内首个“智能驾驶与车联网实验室”在重庆科技研究院揭牌。今后，这里将作为国内研究“智能交通”的重要“练兵场”。其中，造价20余万元的“操作平台”按照1:12的比例制作成为的道路交通模型，可完成路标感应、路况分析等十几项检测功能，目前

参与其中的有力帆、长安汽车以及湖南大学、重庆邮电大学等多个企业或高校。

2011年3月15日，大唐电信与启明信息携手共建联合实验室，研究下一代通信服务与汽车电子产品的融合，开发有自主知识产权的高性能、低功耗汽车电子产品。奇瑞公司和韩国SK电讯联合组建“车联网技术联合实验室”，并宣布研制成功MIV(手机远程操控汽车服务)；陕汽重卡同期也推出了重卡行业车联网系统——天行健。从汽车企业、通信运营企业、软件服务企业到通信设备企业，产业链上下游的强大力量正在蓄积。

目前，国内企业高校联盟联合共建车联网研究实验室如图2所示。

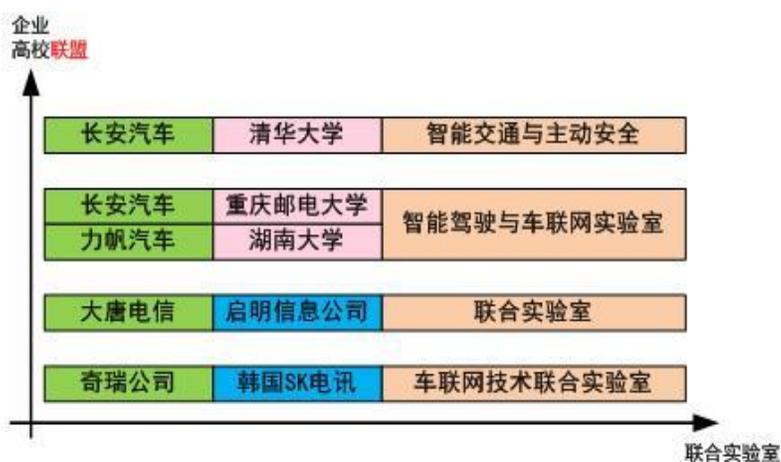


图2 企业高校联盟联合共建车联网研究实验室

2.2 车联网国外发展现状

2.2.1 市场现状

“车一路”信息系统一直是智能交通发展的重点领域。在国际上，欧洲CVIS、美国的IVHS、日本的SmartWay等系统通过车辆和道路之间建立有效的信息通信，实现智能交通的管理和信息服务。

美国和欧洲的汽车公司较早应用车联网技术管理车队，主要公司有如曼、沃尔沃、斯堪尼亚、奔驰等，同时欧美的客运公司也积极使用相关技术。除了欧美，日本和台湾在车辆管理系统的开发上也做了很多尝试，取得了不错的效果。为解决城市交通拥堵、提高车辆的运行效率等问题，国外企业也开发了很基于实时道路信息及诱导方式的车辆信息服务系统，日本的道路交通信息通信系统 VICS 就是其中的例子之一。1996 年，VICS 在全日本范围内投放市场，到 2008 年销量达到 300 多万台，至今累计销售达 2119 万台。目前，日本 80% 的车辆导航系统集成了 VICS 系统。日本丰田公司的试验数据表明，在同一路段上，使用 VICS 系统的车辆行驶时间最大减少了 21.86%；平均减少了 9.52%；平均时速提高了约 5%。

中国是全球汽车最大的生产国和消费国，车联网的成立，不仅会给我国智能汽车业带来广阔的前景，也将会带动其他相关产业链的发展。与国外车联网产业相比，我国虽然在系统性应用方面存在一定弱势，但是在发展潜在规模上我国已经具备了一定的布局优势。在企业层面来看，部分车企的确已经具备了一定的技术融合能力。

2.2.2 政策现状

2004 年，美国政府与工业界积极参与到车载物联网的研发中，出台“车路一体化”（Vehicle Infrastructure Integration, VII）^[1]计划作为国家目标，致力于利用无线通信技术使行驶中的车辆更紧密地与周围的环境相联系，从而全面提高整个交通

系统的安全性。2009 年，美国总统奥巴马对“智慧地球”给予了积极的回应，并很快将支持落实在行动上。美国奥巴马政府表示，将最高支出 60 亿美元推动电动汽车发展，未来电动汽车的发展将肯定伴随车联网技术的发展。已经走出破产困境的通用汽车已成为电动汽车及车联网坚定的拥护者。美国政府与工业界也积极参与到车载物联网的研发中，车辆基础设施集成计划主要参与者包括美国交通部、加州交通部以及戴姆勒、福特、通用等汽车公司，致力于利用无线通信技术使行驶中的车辆更紧密地与周围的环境相联系。由美国交通部主导的 IntelliDrive 项目致力于在个人移动设备、车辆以及路边基站之间建立安全、灵活的无线通信。美国交通部目标在 2013 年前对现有无线通信技术进行测试和评估。

欧洲车联网研究组织 Car 2 Car Communication Consortium 目标是为车车通信建立一个公开的欧洲标准，使得不同制造商的汽车能够互相通信^[2]。欧洲委员会启动若干与车联网相关的项目，如 FleetNet^[3]是一个由欧洲多个汽车公司、电子公司和大学的合作项目，合作者包括 NEC 公司、DaimlerChrysler 公司、Siemens 公司和 Mannheim 大学，该项目利用无线多跳自组织网络技术实现无线车载通信，能够有效提高司机和乘客的安全性和舒适性。另外还包括 Network on Wheels (NoW)^[4]，SAFESPOT^[5]，COOPERative networks for intelligent road Safety (COOPERS)^[6]，Secure Vehicular Communication (SeVeCOM)^[7]等相关项目。以西欧、日本为代表的发达国家更是投入

了大量的人力、物力和财力进行智能交通系统(ITS)相关技术的研究、试验和开发,许多新的技术和产品已走出实验室进入工程应用。

2.2.3 技术与标准现状

1992年,DSRC由美国材料与试验协会(ASTM)最早提出。2010年7月正式颁布的IEEE 802.11p是IEEE 802.11协议在车辆领域的推广与扩充,主要规定了DSRC协议的媒体访问控制层(MAC)和物理层(PHY)标准。IEEE 1609标准则是以802.11p通信协议为基础的高层标准,是IEEE针对无线通信技术应用于车辆在通用公司展示车载安全系统设计的,车辆通过DSRC无线技术实时监控周围车辆的位置、速度与方向,一旦发生紧急情况,车辆通过声、光信号警告司机。1992年,ASTM主要针对ETC业务的开发而最先提出DSRC技术的概念,其采用915MHz频段。2002和2003年ASTM分别通过了DSRC标准E2213-02及其改进版本E2213-03,工作频率为5.9GHz。DSRC能在短距离通信范围内进行高速数据传输,而且成本低廉。

2004年,IEEE专门成立了车辆通信环境下的无线接入(WAVE)工作组,DSRC标准化工作也转入IEEE802.11p与IEEE 1609工作组^[8]进行,主要还是针对应用于高速移动环境中的IEEE 802.11标准的相关内容进行一些修改。IEEE 802.11p是美国交通部针对欧洲的车辆通信网络,特别是电子道路收费系统、车辆安全服务与车载商业交易系统等应用而设计的一种中长距离通信的空中接口

标准,它计划将被用在车辆通信系统中,能够提供高速的车到车和车到路侧单元的数据传输。在WAVE系统中,一个路侧单元大约可以覆盖方圆1km。IEEE 802.11p协议主要描述车辆无线通信网络的物理层和MAC层协议:MAC层采用增强分布式信道接入机制(EDCA),物理层仍采用802.11a标准使用的正交频分复用(OFDM)技术,只是其物理层参数在802.11a的基础上进行了一些调整。比如,其带宽调整为10MHz,信息传输速率相应地降低为3~27Mbit/s,工作频率由美国联邦通讯委员会分配,为5.850~5.925GHz。

IEEE 1609系列标准主要目的为制订V2V和V2I的标准无线通信协议,并提供行车环境下,包括汽车安全性、自动收费、增强导航、交通管理等广泛应用环境的通信标准。802.11p协议主要负责物理层及MAC层,而IEEE 1609则主要规定了相应的高层协议。IEEE 1609.1主要用做资源管理,为DSRC设备提供额外的管理机制,让具有控制能力的节点能够远端控制一个区域内的所有节点。IEEE 1609.2主要负责制订IEEE 1609标准中的安全机制。IEEE 1609.3主要用来制订WAVE系统中网络层通信协议及管理机制,并为降低WAVE设备间传输时所需的时间而专门制订了短信通信协议(WSMP)。IEEE 1609.4负责制订信道间切换方式。WAVE中包含1个控制信道和6个服务信道。因为必须定期或者不定期对控制信道与服务信道之间进行切换,所以不同WAVE设备之间必须通过共通的控制信道交换信息。车用通信长

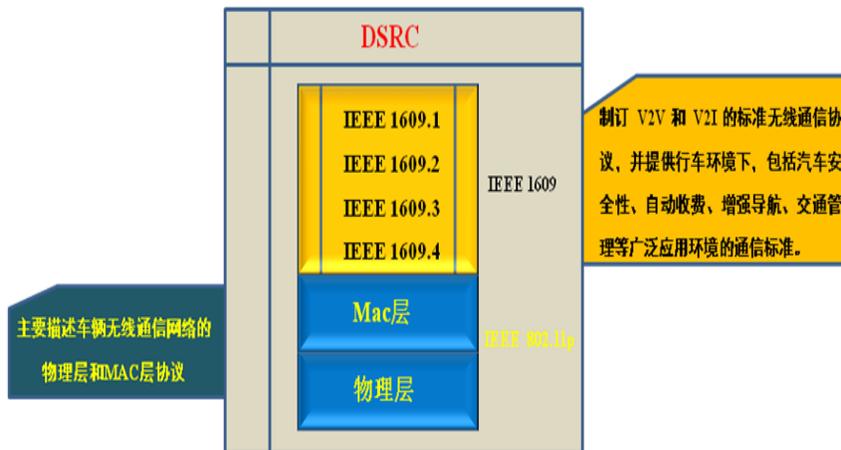


图3 车联网无线通信技术与标准

久以来受限于无线通信技术难以应用在高速移动环境中；不同车商之间缺乏统一的通信接口，使得车辆与外界的沟通受到相当大的限制。IEEE 1609 协议族在这方面有了突破。所以，世界各国都以制定车辆通信标准为其重要交通政策，如美国交通部提出 ITS 相关计划便是一例。IEEE 结合了政府、车商及运营商共同制订 IEEE 1609 系列技术标准，为行车环境中各项应用情境提供了解决方案。如图 3 所示。

3.车联网研究现状分析

近年来，车辆的爆发式增长和无处不在的信息需求已日益将通信网络和车辆紧密结合起来。人们在车辆移动过程中的通信服务需求日益增大，车载移动网络的研究已成为世界瞩目的焦点，同时也促进了车辆向智能化、网络化方向的发展。车联网作为物联网产业中最容易形成系统标准以及最具备产业潜力的应用，同样也逐步引起了学术界和工业界越来越多的关注和研究，特别是 2009 年以来已经成为国内外新的研究热点。

3.1 车联网网络进程演化理论研究

到目前为止，国内外关于车联网的研究都只处于起步阶段，大多数车联网应用仅停留在一个小尺度的闭环系统中进行，并取得一定的成效，例如，在智能公交定位管理和信号优先、智能停车场管理、车辆类型及流量采集、路桥电子不停车收费及车辆速度计算分析等。然而距离实现车与车、车与路之间互联互通还相差甚远，制约车联网发展的最大原因是目前车联网严重缺乏理论基础，从而造成现有的车联网应用大多局限于局部环境下的简单应用，从整体上来说缺乏一套规范化的、系统的、科学的理论模型。车联网的科学研究和设计开发通常基于传统的、适用于小尺度闭环系统中的理论和方法。这些理论难以有效解决大规模开环跨域系统的互联互通问题，不能支持动态自组织系统的弹性可伸缩，也不能指导大规模开环跨域系统性能指标的可测性。而解决车联网大规模开环跨域系统的上述问题恰恰是目前车联网发展所迫切需要的。针对车联网的核心科学问题，开展车联网网络进程演化理论模型的研究，形成规范有效的理论和方法体系，这对未来车联网的发展有着重要的理论指导意义。

3.2 车联网网络通信的研究

近几年来，研究者致力于把先进的通信技术应用到车辆交通系统中，使其更加安全、智能和高效。2004 年，IEEE 成立了 IEEE 802.11p 工作组以制定 IEEE

802.11 在 WAVE 的版本, 并以 IEEE 1609 系列协议作为上层协议, 从而形成车辆无线通信的基本协议构架。美国伊利诺伊大学 Urbana Champaign 分校 Nitin Vaidya 教授为首的团队开发了多信道测试的无线 Mesh 网络测试台^[9]。UCLA 教授 G.Pau 提出了车辆间特殊路由协议 (PVRP), 搭建了系统测试平台进行了验证^[10]。密歇根大学郭锦华和向卫东教授开发了基于 5.9GHz 的 WAVE 系统信道测试平台^[11]。

现阶段, 车载移动网络的研究热点主要集中在基于 WAVE 协议 (IEEE 802.11p) 的车辆通信多信道协调应用、组播路由管理, 以及基于 WiMAX 协议 (IEEE 802.16) 的固定中继技术的切换、资源调度方面。但 802.11 具有覆盖范围小、车辆移动过程中需要频繁切换连接路边单元、服务质量 (QoS) 支持弱、无法对多媒体信息提供高质量支持的弱点^[12,13]。为此, 文献^[14,15]提出采用基于 WiMAX 的技术来为车辆及其内部所属用户进行车载移动宽带无线接入, 首次将 WiMAX 技术应用于车辆通信网络。为了解决车内用户终端在高速移动情况下的宽带无线接入问题, IEEE 802.16 标准制定组 2006 年 3 月成立基于 IEEE 802.16j 的移动中继 (MRS) 工作小组, 以研究采用 MRS 的可行性, 想采用车载 MRS 站点为车内的群体用户终端提供宽带无线接入服务^[16]。

在基于 WAVE 协议的车辆与车辆之间自组织通信网络中, 整个车辆网络的安全和非安全应用都在一个信道上完成, 难以保证安全应用的 QoS。文献^[17]采用多信道的媒体访问控制 (MAC) 机制来解决

上述问题。采用多个信道后, 节点间可以使用不同的信道进行通信, 可以获得优于单信道的网络吞吐量和时延特性。文献^[18,19]采用时隙间隔方法把时间交替分为控制间隔和数据交换间隔, 在控制间隔所有节点跳到控制信道进行信道协商, 在数据交换间隔再跳到不同的信道进行数据传输。WAVE 协议中原有路由机制也不完全适合拓扑动态变化的车载通信网络, 这些路由协议随着通信跳数增加和车辆移动速度加快, 建立路由的延迟相应增大, 难以满足需要低延迟的安全应用, 因此出现基于位置的组播路由^[20,21]。组播路由的目标是将报文从源节点传递到位于关联区域 (ZOR) 内的所有节点。针对组播路由机制, 提出了簇的概念, 将车辆网络组织成多个对等的单元 (簇), 从而提高移动环境下的可扩展性^[22]。

车辆与路边基础设施之间的通信仅仅适应车辆在低速行驶或者相对静止的环境下, 车辆在高速行驶过程中无法提供与路边单元的基础设施进行长时间的信息交互。文献^[23]提出了一种基于固定中继的两级资源调度机制, 提高了系统吞吐量, 降低了业务的丢包率和延时时间。文献^[24]提出了多跳蜂窝网络中继辅助切换的技术, 移动终端通过中继节点进行信息的传输, 利用这种技术保证了信道的 QoS 指标, 降低了掉话率。文献^[25]首次提出了基于 MRS 的群组切换, 移动中继站辅助车内用户终端完成接入目标基站的切换, 并通过切换过程中资源的重新分配来提高切换成功率, 降低切换阻塞率和延时。

文献^[26]提出了 WiMAX 与 WAVE 新型异构网络融合的车载移动网络架构, 在

该移动网络架构中,车车之间的通信通过基于 WAVE 来实现,车辆与路边基站的通信通过 WiMAX 实现。在车内用户终端与路边基站的两层结构中引入 MRS 概念,车内用户通过 MRS 站进行网络通信,从而支持车辆之间的应急通信,支持智能交通控制中心之间实现实时、可靠的信息交互,以及进行多媒体数据业务的传输。文献^[27]根据通信距离划分,车辆通信网络主要包括车域网(VAN)和车辆自组网(VANET)两大类。车域网在移动车辆内部各部件之间构建无线局域网,并通过车载网关接入周边的无线广域网。移动自组织网络(MANET)在交通道路上的应用被称之为车辆自组网,它为高速运动的车辆之间及运行在高速公路上的车辆提供了一种可能的高速率无线通信接入方式。文献^[28]提出较为详细的车联网的系统架构和协议架构,分为 in-vehicle、ad hoc 和 infrastructure 域。

以上车联网网络通信通过协议转换的方式实现网元间的连接,这会带来效率和性能上的严重不足,从根本上制约车联网大规模信息的交换。

3.3 车联网环境下动态自组织技术的研究

随着车联网技术的不断发展和成功应用,目前已取得一些关于动态自组织技术方面的研究成果。文献^[29-30]中提出了一种基于集群的自组织协议 CSP,这是一种主动自组织协议,它通过将道路分成相邻的段,将处在同一段中的汽车构成一个集群的方法智能地构造车联网。文献^[31]在对车载自组织网络的特点和研究现状分

析的基础上,给出了车载自组织网络(VANET)中多维多层的理论模型和网络体系结构,讨论了物理层技术及其相关标准以及 MAC 层、网络层协议设计的重点和难点,阐述了广播协议的设计思路。文献^[32]提出了分布式构造基本拓扑算法,这种自组织算法意在通过最小化网络参与者间的干扰来提供高效的可靠的分层拓扑。通过将密集连续节点分成集群产生一个网络的地理最优集群,允许集群成员间的强连接和减少集群间的干扰。文献^[33]提出了一种基于内容路由和传播技术的新网络架构,比较适用于汽车 Ad Hoc 网络。文献^[34]提出了一种基于路由算法的簇方法用于规范车流量。这种方法是结合移动路点和组合移动模型的一种混合模型。为了传送实时更新信息,并维持长的联络时间以提高数据传输率,提出了基于位置的多径洪泛算法,此算法在多变移动性上适用于车辆。文献^[35]认为利用汽车移动的信息(例如位置,方向,速度和道路的数据地图)在一个可能的连接断开事件发生之前来预测它。汽车根据他们的适量速度来分组,这种分组确保同一组中的汽车更可能建立稳定的单跳或多跳路径。文献^[36]提出了一个可行的解决方案和工具以确保在没有中央管制的情况下 VANET 的运作,该方法通过有关用户的合作可以侦测和警告异常交通状况。文献^[37]提出了一个针对 VANET 的基于发布订阅的新型动态分布感知数据传播系统,在 VANET 中基于汽车移动性的特点,提出通知令牌机制,只有令牌的持有者向周围的订阅用户广播信息。文献^[38]提出了一种新的路由方案 BAS,每条道路通过多重副本传播来缓存数据包,目的是在路口给分组交换提供更

多的机会。与传统的 VANET 协议不同，BAS 中副本的传播在路由路径上是双向的。文献^[39]通过使用渗透理论来分析 VANET 的连接。经过推导，发现在网络连接中，车辆密度和传输距离之间存在定量关系。当车辆密度或者传输范围足够大时，存在一个网络连接的跳跃。这一结果对于现实世界中 VANET 的发展具有重要意义，根据给出的车辆密度，这一定理可以用来计算达到良好的网络连接的最小传输范围。文献^[40]通过协作来协调数据收集和处理，识别交通现象和改变结构。文献^[41]提出了一个对于集中车辆网络组织的新框架 LTE4V2X，运用集中式架构来优化集群管理和提供更好的性能，同时研究了它的性能和一个分散式组织协议的比较，这个分散式组织协议是一个知名的城市传感应用。文献^[42]提出了一种架构，意在整合两种或多种协议来发挥它们各自的优势。定义了车辆集群和网管管理的相关设计准则。前者旨在加强 VANET 中连接的稳定性，而后者则维持 VANET 和 3G 以及其他网络间的连接性。文献^[43]对车载自组网的安全需求及目标进行分析，然后从安全路由、密钥管理和入侵检测三个方面对车载自组网安全研究的现状进行介绍。在这三个方面中，安全路由研究最为广泛，而密钥管理则是保障车载自组网安全的关键。文献^[44]利用基于车载自组网的交通信息采集和交通事件协同检测算法、大规模海量交通信息的分布式组织和实时处理方法、基于位置的交通信息分发协议等技术，构建一个动态交通信息系统，实现改进交通信息系统中的事件检测的准确性和实时性，提高大规模交通信息处理的实时性，改善交通信息分发的针对性。

文献^[45]仿真软件 SUMO 生成的跟踪轨迹文件，在仿真软件 ONE 下做网络参数的对比实验。通过模拟道路网络的动态密度变化，使得最佳发送时机随网络参数不同而变化。文献^[46]介绍了车载自组网的发展历史、特点和应用领域。使用分析和比较的方法，讨论各种无线通信技术用于车载自组网的优缺点，并针对车载自组网的应用及特性提出搭建车间通信系统的设计思想和突破方向。

现有的研究成果基本上都是针对车联网自组织网络路由算法相关的技术，但大都停留在小尺度闭环系统中，并没有考虑大规模新型车联网网络环境下随着需求和环境的变化弹性构建满足当前任务的需求网络问题。因此，必须提出新型的适用于未来车联网智能服务的需要动态自组织网络构建方法。

4. 车联网发展建议

4.1 建议目标

对于车联网的打造，一方面需要解决一系列的技术问题，另一方面涉及的部门、行业较多。因此，政府需要提出发展规划、发展步骤，规范化管理和稳定有序的经营，政策指导要硬，需要相关政策去磨合和解决，大力配合和支持企业实现自主创新。具体如下：

- 发展车联网，应向智能方向深入推进，并实施中央提出的智能发展战略；
- 发展车联网，应实施安全发展战略，努力破解交通安全难题，以贯彻落实党中央、国务院关于“以人为本，安全发展”的重大战略方针；

- 发展车联网，实施新型现代化战略，建立自主创新和“中国制造”的重大物联网产业技术平台和产业技术标准，建设技术先进、结构高效、自主开放、体系完整的中国车联网产业体系；
- 实施“产业主导，需求牵引，科技服务”的新型科技经济协同体制，从而推进车联网的研发和实践，进而推动我国车联网产业的快速良性可持续发展。

4.2 技术建议

对于车联网，事关车联网行业的健康发展的技术标准等，目前为止还是一片空白。尤其是缺乏全局性的行业标准，使车联网的概念没有真正为消费者所接受，难以区分互联网、车联网和车载通信平台之间的区别。在技术标准缺失的情况下，车联网的发展实际上处于相关企业各行其是、自行发展的阶段。因此，在技术层面建议如下：

- 联网网络通信通过协议转换的方式实现网元间的连接，这会带来效率和性能上的严重不足，从根本上制约车联网大规模信息的交换。因此有必要建立车联网可互联互通的通信协议；
- 建立全国乃至全球范围内统一的车联网技术、服务以及质量标准。标准化是促进一个产业健康发展的基础，车联网是物联网分支中最容易形成系统标准，最具备产业潜力的应用。车联网标准化的建立，也将会推动物联网的发展。

4.3 政策体制建议

中国现代化的初步实践和世界大国现代化的实践，从正反两个方面告诫我们，发展车联网等新兴支柱产业，没有国家强大意志和国家重大培育和投入，难以取得根本成功。我们认为，中央财政应大幅增加国家汽车移动物联网工程的财政预算和政策扶持力度。因此，在政策体制建议方面如下：

- 建立完整的车联网产业链。车联网背后是汽车制造商、车载终端企业、电信运营商、IT企业、硬件供应商、交通信息内容运营商及服务商等组成的一长串产业链条。车联网的成立，不仅会给我国智能汽车业带来广阔的前景，也将会带动其他相关产业链的发展，只有彼此积极配合，才能更好地推动车联网的发展；
- 政府部门与企业要分工明确。政府需要提出发展规划、发展步骤，并完成生产和采购环节的落实，优先提供科研人员攻克技术难关，保障投资资金到位。企业则应该克服行业准入的困难，对相关技术、产品、商业模式不断创新。

5. 小结

本文结合国内外同行的研究报道与我们的研究结果，不难推断，近两年来，尽管对车联网的研究已有了一定的研究成果，但基本还处于初期的探索阶段，其政策、资金、技术及支撑的产业链仍不完善成熟。因此，政府要构建一个博弈环境，引导汽车业、电子业、通讯业等企业协同，共同打造未来车联网价值的增长点，才能吸引

更多的人来参与和开发车联网市场。车联网要取得最终形成成熟的应用技术，还需要更具有创新性的研究。在欧盟 IST 将启动新的研究与创新框架计划 (FP8)——展望 2020 (2014-2020) 项目研究，以及我国将车联网列入“十二五”国家科技重大专项的重要背景下，适时启动关于车联网技术标准 and 行业的发展以及国家对车联网政策的大力扶持，是必要而且迫切的。

参考文献

- [1]. CVIS. www.cvisproject.org.
- [2]. Car-2-Car Communication Consortium. <http://www.car-to-car.org>, 2003.
- [3]. Hartenstein H. Position-Aware Ad Hoc Wireless Networks for Inter-Vehicle Communications: The FleetNet Project, Mobile Ad Hoc Networking & Computing, ACM Press, 2001, 259-262.
- [4]. Network on Wheels (NoW). www.network-on-wheels.de.
- [5]. SAFESPOT. www.safespot-eu.org.
- [6]. CO-OPERative networks for intelligent road Safety (COOPERS). www.coopers-ip.eu.
- [7]. Secure Vehicular Communication (SeVeCOM). www.sevecom.org.
- [8]. ASTM E2213-03. Standard specification for telecommunications and information exchange between roadside and vehicle systems-5 GHz band dedicated short range communication (DSRC) medium access control (MAC) and physical layer (PHY) specifications[S], 2003.
- [9]. Shen T. Experiments on a Multichannel Multi-Interface Wireless Mesh Network, Master's Thesis, 2008.
- [10]. Pau G. and Rowstron A. PVRP: Efficient disruption tolerant discovery in vehicular ad hoc networks, ACM MobiComm, 2009.
- [11]. Xiang W. Richardson P. Guo J. Performance Evaluation of Wireless Access in Vehicular Environments (WAVE) Systems for Dedicated Short Range Communications (DSRC), IEEE Wireless Communication Magazine, 2009.
- [12]. Yin J. Elbatt T. Yeung G. Ryu B. Habermas S. Krishnan H. Talty T. Performance evaluation of safety applications over DSRC vehicular ad hoc networks, Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks, 2004.
- [13]. IEEE Std 802.16e. IEEE standard for local and metropolitan area networks, Part 16: Air interface for fixed and mobile broadband wireless access systems, 2006.
- [14]. Yang K. Ou S. Chen H. He J. A multihop peer communication protocol with fairness guarantee for IEEE 802.16 based vehicular networks, IEEE Transactions on Vehicular Technology, 2007, 56(6):3358-3370.
- [15]. Shan L. Liu F. Yang K. Performance analysis of group handover scheme for IEEE 802.16j-enabled vehicular networks, Proceedings of the Joint International Conferences on Asia-Pacific Web Conference and Web-Age Information Management, 2009.
- [16]. IEEE C802.16j-D5. Draft

amendment for local and metropolitan area networks, Part 16: Air interface for mobile broadband wireless systems multihop relay specification[S], 2008.

[17].Crichigno J. Wu M. Shu W. Protocols and architectures for channel assignment in wireless mesh networks, *Ad Hoc Networks*, 2008, 6(7): 1051-1077.

[18].Mo J. So H.W. Walrand J. Comparison of multichannel MAC protocols, *IEEE Transactions on Mobile Computing*, 2008, 7 (1): 755-765.

[19].So H.W. Walrand J. Mo J. A multi-channel MAC proposal for ad hoc wireless networks, *Proceedings of IEEE Wireless Communications and Networking Conference*, Mar 11-15, 2007, 255-260.

[20].Lochert C. Hartenstein H. Tian J. Fussler H. Hermann D. Mauve M. A routing strategy for vehicular ad hoc networks in city environments, *IEEE Intelligent Vehicles Symposium*, 2003.

[21].Korkmaz G. Ekici E. Ozguner F. Umit O. Urban multi-hop broadcast protocol for inter-vehicle communication systems, *Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks*, 2004.

[22].Li F. Wang Y. Routing in vehicular ad hoc networks: A survey, *IEEE Vehicular Technology Magazine*, 2007, 2(2):12-22.

[23].Wang L. Ji Y. Liu F. Adaptive subframe partitioning and efficient packet scheduling in OFDMA cellular systems with fixed decode-and-forward relays, *IEICE Transactions on Communications*, 2009, E92-B(3):755-765.

[24].Jeon S and Lee S. A relay-assisted handover technique with network coding over multi-hop cellular networks, *IEEE Communications Letters*, 2007, 11(3):252-254.

[25].Shan L. Liu F. Wang L. Ji Y. Predictive group handover scheme with channel borrowing for mobile relay systems, *Proceedings of the IEEE Wireless Communications and Mobile Computing*, 2008.

[26].刘富强, 单联海. 车载移动异构无线网络架构及关键技术, *中兴通讯技术* [J], 2010:16(3).

[27].罗涛, 王昊. 车辆无线通信网络及其应用, *中兴通讯技术* [J], 2011, 6(3):01-07.

[28].Car 2 Car Communication Consortium. Car 2 Car Communication Consortium Manifesto Overview of the C2C-CC System, <http://www.car-to-car.org/index.php?id=31&L=oksjfr>.

[29].Cherif M.O. Senouci S.M. Ducourthial B. CSP: Cluster-based Self-organizing Protocol for vehicular networks, *Information Infrastructure Symposium*, 2009.

[30].Cherif M.O. Senouci S.M. Ducourthial B. A new framework of self-organization of vehicular networks, *Information Infrastructure Symposium*, 2009.

[31].Cherif M.O. Senouci S.M. Ducourthial B. Vehicular network self-organizing architectures, *IEEE GCC Conference & Exhibition*, 2009.

[32].Yair A. Segal M. Near-optimal, Reliable and Self-organizing Hierarchical Topology in VANET, *VANET '11*

Proceedings of the Eighth ACM international workshop on Vehicular inter-networking, 2011.

[33].Arnould G. Khadraoui D. Habbas Z. A Self-Organizing Content Centric Network Model for Hybrid Vehicular Ad-Hoc Networks, Proceedings of the first ACM international symposium on Design and analysis of intelligent vehicular networks and applications, 2011.

[34].D.Rajini Girinath, S.Selvan. A novel cluster based routing algorithm for hybrid mobility model in VANET, 2010 International Journal of Computer Applications, 2010, 1(15).

[35].Taleb T. Sakhace E. Jamalipour A. Hashimoto K. Kato N. Nemoto Y. A Stable Routing Protocol to Support ITS Services in VANET Networks, IEEE Transactions on Vehicular Technology. 2007, 56(6):3337-3347.

[36].Caballero C. Molina J. Caballero P. Mart ́n-Fern ́ndez F. Yanes-Garc ́a D. Introducing secure and self-organized vehicular ad-hoc networks, Proceedings of the 12th International Conference on Computer Systems and Technologies, 2011.

[37].Wu L. Liu M. Wang X. Gong H. Dynamic distribution-aware data dissemination for Vehicular Ad Hoc Networks, International Conference on Future Computer and Communication, 2010, 2:353-360.

[38].Song C. Liu M. Wen Y. Cao J. Chen G. Buffer and Switch: An Efficient Road-to-Road Routing Scheme for VANETs. International Conference on Mobile Ad-hoc and Sensor Networks, 2011.

[39].Jin X. Su W. Wei Y. A study of the VANET connectivity by percolation theory. IEEE Consumer Communications and Networking Conference, 2011.

[40].Fekete S.P. Schmidt C. Wegener A. Hellbrück H. Fischer S. Empowered by wireless communication: Distributed methods for self-organizing traffic collectives. ACM Transactions on Autonomous and Adaptive Systems, 2010, 5(3).

[41].Rény G. Senouci SM. Jan F.Gourhant Y. LTE4V2X: LTE for a Centralized VANET Organization, IEEE Globecom proceedings, 2011.

[42].Taleb T. Benslimane A. Design Guidelines for a Network Architecture Integrating VANET with 3G & beyond Networks, IEEE Global Telecommunications Conference, 2010.

[43].高永康, 郝建军. 车载自组织网络中的网络与信息安全, 中兴通讯技术[J], 2011, 17(3):21-23.

[44].向勇. 基于车载自组织网络的动态交通信息的挖掘和利用, 中兴通讯技术[J], 2011, 17(3):29-34.

[45].王琳, 徐俊, 江昊. 基于车载自组织网络的消息发送时机研究. 中兴通讯技术[J], 2011, 17(3):40-43.

[46].常促宇, 向勇, 史美林. 车载自组网的现状与发展, 通信学报[J], 2007, 28(11):116-126.

作者简介



Jiujun Cheng (程久军) received his Ph.D. degree from Beijing University of Posts and Telecommunications in 2006. He is presently an associate professor of Tongji University. He is the author of more than 30 conference and journal papers. His research interests include P2P Overlay and Social networks with a focus on IN/Internet interworking, and Internet of Things.



MengChu Zhou (周孟初) received his B.S. degree in Electrical Engineering from Nanjing University of Science and Technology, Nanjing, China in 1983, M.S. degree in Automatic Control from

Beijing Institute of Technology, Beijing, China in 1986, and Ph. D. degree in Computer and Systems Engineering from Rensselaer Polytechnic Institute, Troy, NY in 1990. He joined New Jersey Institute of Technology (NJIT), Newark, NJ in 1990, and is a Professor of Electrical and Computer Engineering. He is presently a Professor of The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China. His research interests are in Petri nets, wireless ad hoc and sensor networks, semiconductor manufacturing, transportation and energy systems. He has over 460 publications including 10 books, 200+ journal papers (majority in IEEE Transactions), and 18 book-chapters. He is the founding Editor of IEEE Press Book Series on Systems Science and Engineering, Editor of IEEE Transactions on Automation Science and Engineering, and Associate Editor of IEEE Transactions on Systems, Man and Cybernetics: Part A and IEEE Transactions on Industrial Informatics. He is a life member of Chinese Association for Science

and Technology-USA and served as its President in 1999. He is Fellow of American Association for the Advancement of Science (AAAS) and Fellow of the Institute of Electrical and Electronics Engineers (IEEE)

微型旋翼飞行机器人在多障碍物复杂环境的自主导航方法研究

肖继忠

纽约城市大学城市学院电气工程系

Jxiao@ccny.cuny.edu

摘要

近年来,微型旋翼飞行机器人在复杂环境的自主导航和应用研究是机器人学最前沿的热门研究方向之一。由于具有垂直起降、高机动性、体积小、噪音低、不易被发现等特性,它特别适合在低空复杂环境执行侦察监视,证据搜集等任务,在反恐执法、军事侦察、城市巷战、治安和民用等领域具有广阔的应用前景。本文首先介绍课题的研究意义和国内外研究现状;然后概括介绍肖继忠博士领导的纽约城市大学机器人学研究团队在该领域的主要创

新成果,包括微型四旋翼机器人的实时感知,自主导航和三维地图创建;最后指出未来的主要研究方向。

关键词

无人驾驶飞行器,自主导航,路径规划,同步自定位与地图创建,全方位立体视觉

1 引言

1.1 研究意义

近年来,结构紧凑、外形小巧的微型旋翼飞行机器人(MAV, micro aerial vehicle)成为无人驾驶飞行器(UAV)研究领域的热门平台。对于大/中型无人驾驶直升机或高速飞行的固定翼无人机来说,它主要依靠全球定位系统(GPS)在开阔的高/中层空间实现遥控或基于航点规划的自主导航以执行各种军事和民用任务(如图1所示)。而对于微型旋翼机器人来说,由于具有垂

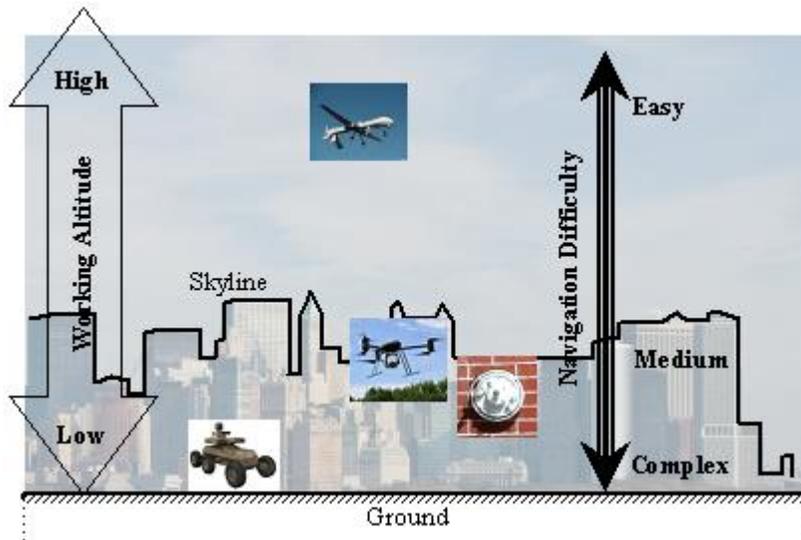


图 1: 导航区域与在此区域中地面机器人、微型无人直升机、爬壁机器人,以及固定翼 UAVs 或大/中型无人直升机导航困难对照。高度越高,运动空间的障碍物越少,导航难度越低。微型无人直升机在导航难度适中的地面和地平线之间的低空环境运动。

直起降、高机动性、体积小、噪音低、不易被发现等特性，它特别适合在近地面的低空环境(例如，室内、市区、山洞和丛林中)执行侦察，监视，证据搜集，洞穴搜索，交通监控，火灾搜救及航空摄影等任务。在反恐执法、军事侦察、城市巷战、治安和民用等领域具有广阔的应用前景。(如图2所示)。例如，装备有摄像头的警用微型旋翼机器人可以低空飞入危险小巷，近距离监视和拍摄新疆暴乱分子的罪行；也可以从窗户飞入高层建筑物内部以帮助救火队员的搜救工作。

难进行复杂的信号处理和视觉运算。这些关键技术的挑战，激发了世界各地的科学工作者在这一领域进行探索。

1.2 国内外研究概况

微型旋翼机器人的自主导航控制是一项涉及多门交叉学科的高、精、尖技术，在军事、反恐和民用领域都有广阔的应用前景，极具研究价值。美国国防部为了推动微型机器人关键技术的突破，促进系统集成与实际应用研究以满足国防建设的迫切需求，于2008年10月由陆军科研局牵头



图 2:微型旋翼飞行机器人在 a) 洞穴搜索, b) 丛林探秘, c) 反恐侦察等领域具有广阔的应用前景。

然而，要使微型旋翼机器人真正走向成熟与实用，还面临着诸多关键技术的挑战。与高/中层空间不同，低空环境非常复杂，存在大量的障碍物(例如，树杈、电线、室外标志等)令微型旋翼机器人的自主导航控制困难重重。这些困难来自多方面因素，首先，微型旋翼机器人是在三维空间运动，其飞行姿态和运动轨迹比地面机器人更加难以控制；其二，不确定性干扰(例如，阵风和载荷的变化)对飞行器控制稳定性的影响很大；其三，微型旋翼机器人的载荷有限，大型传感器无法安装，必须特别设计轻便的小型传感器系统以感知观测外部环境；其四，由于负载和电源的限制，机载计算机的运算能力有限，很

设立“微型自主系统与科技协作研究联盟”(MAST-CTA, *Micro Autonomous Systems and Technologies Collaborative Technology Alliance*, <http://www.mast-cta.org/>), 资助成立了三个以大学为主的研究中心: 微机械中心, 微电子中心, 和自主系统信息处理研究中心。协作研究联盟的远景目标是开发全自主, 多功能, 敏捷机动, 并具有多个体协同工作能力的微系统以增强在城市和复杂地形中对战术情报的收集能力。以宾夕法尼亚大学领导的自主系统信息处理研究中心 (*processing for autonomous operation center*) 包括麻省理工大学, 佐治亚理工大学, 加州大学伯克利分校, 新墨西哥大学等成员, 主要以德国 Ascending

Technologies公司设计生产的Asctec微型四旋翼飞行器为平台,从事飞行控制,自主导航,三维地图创建,多机协调控制等方面的基础理论及工程应用研究。陆军科研局的项目指南明确提出了要实现的任务目标,由易到难依次是:建筑物内部的搜索,洞穴搜索,多个体协同进行周边防卫(small unit perimeter defense)。

在国际学术界,微型旋翼机器人的研究工作主要集中于飞行器控制算法设计、路径规划与自主导航、三维空间的同步自定位与地图创建、人机互动以及多机协调控制技术。宾夕法尼亚大学(Upenn),麻省理工大学(MIT),斯坦福大学,纽约城市大学(CUNY City College),瑞士苏黎世联邦科技学院(ETH Zurich),瑞士洛桑联邦科技学院(EPFL, Lausanne),澳洲昆士兰大学等机构在微型旋翼飞行机器人自主导航和应用研究领域处于领先地位。宾夕法尼亚大学,瑞士苏黎世联邦科技学院实现了微型旋翼机器人的高机动飞行控制,所拍录像展示了微型四旋翼直升机高速穿越窗户和随音乐起舞的机动能力,然而,它们的飞行运动轨迹是事先规划好的。麻省理工大学,纽约城市大学,宾夕法尼亚大学实现了微型四旋翼机器人在结构化的室内环境自主飞行并绘制三维地图的能力。如今,微型旋翼机器人已经初步具备在结构化环境,以及开阔地带实现航点追踪和基于视觉的自主导航能力。但是,如何在遍布障碍物的环境,如洞穴、市区、林地空间中实现自主导航仍是一个极具挑战性并亟待突破的难题。

在工业界,加拿大的Draganfly Innovations Inc.设计生产了四旋翼,六旋

翼,八旋翼飞行器,可以通过操纵杆遥控实现航空摄影,警用侦察等任务。德国Ascending Technologies公司设计生产的Asctec四旋翼飞行器由于结构开放,负载较大,被很多大学和研究机构选为实验平台。Ascending Technologies公司与LaserMotive公司合作,利用激光远程充电,创造了四旋翼飞行器在空中连续飞行十二个半小时的世界纪录。

国内在微型旋翼机器人方面的研究主要集中于硬件系统的设计,控制算法的实现。国防科学技术大学、北航、南航、西北工业大学等单位在这一方面有相关研究工作的报道,实现了四旋翼飞行器的姿态稳定控制和遥控。国内很多厂家也有能力生产制造各种各样的遥控直升机玩具。然而,微型旋翼直升机不可能完全依靠操作人员的现场遥控以实现复杂环境中的避障飞行。在实际应用中,微型旋翼机器人必须智能化,必须具备足够的自主导航和视觉信息处理能力以完成各种任务。国内在这一方面的相关研究尚处于起步阶段。

2. 研究进展

微型旋翼机器人在复杂环境的自主导航和应用研究是机器人学最前沿的热门研究方向之一。近年来,肖继忠博士领导的纽约城市大学机器人学研究团队在复杂环境中微型四旋翼机器人的实时感知,自主导航和三维地图创建领域已经取得了许多研究成果,在国际上处于领先地位,赢得了德国Ascending Technologies公司的赞助*。本节概括介绍该项研究的主要创新成果。*<http://www.asctec.de/sponsorship>

2.1 飞行控制基本原理

四旋翼飞行机器人是由完全对称的刚性十字交叉框架和安装于尾端的四个螺旋桨组成，其滚转轴(roll)，倾斜轴(pitch)，偏航轴(yaw)，力(F_i)，叶片旋转产生的力矩(τ_i)由图3所示。为了避免由反作用力矩引起的偏航漂移，四旋翼飞行机器人的一对叶片(2-4)被设置为逆时针方向旋转，另一对叶片(1-3)被设置为顺时针方向旋转。它的基本工作原理由简化了的动力学方程表示，其总合升力为： $F_{tot}=F_1+F_2+F_3+F_4$ ，无偏航移动的情况下其总合力矩应该保持为零，即， $\tau_{tot}=\tau_1+\tau_3-\tau_2-\tau_4\approx 0$

为了实现上(下)运动，四个叶片的旋转速率要保持相同，通过同时改变四个叶片速率大小而增加(或减少)总合升力。四

旋翼飞行机器人的前/后，左/右和偏航运动是由控制四个叶片的差分速率而实现。偏航运动可以通过增加一对叶片的速率并相应减小另一对叶片的速率以保持总合升力，但产生水平反作用力矩来实现在固定飞行高度的偏航旋转。前向运动可以通过减小前叶片的速率并相应增加后叶片的速率以保持总合升力并产生倾斜角以向前飞行。其它方向的运动控制遵循同一原理。

2.2 实时状态估计和自主飞行控制

微型旋翼飞行机器人在三维复杂空间的自主导航是非常困难的任务。由于惯性导航参数的漂移很大，微型旋翼机器人运动位置不可能通过惯量参数的二次积分准确获得，在没有GPS信号的情况下，准确进行自身运动估计是首要克服的问题。我们采用的解决方案是利用图像/激光扫描

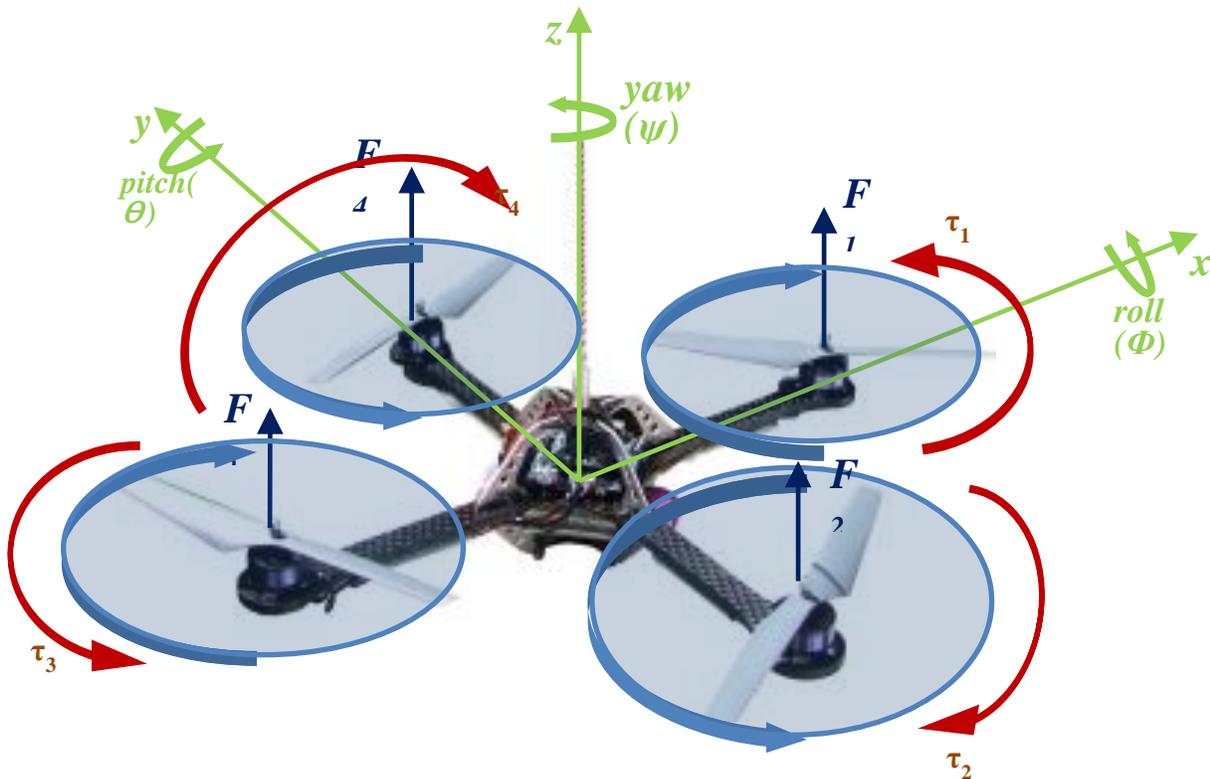


图 3：简化了的四旋翼飞行器自由刚体模型。

匹配技术来估计自身运动(ego-motion)，并综合利用小型激光测距仪提供的距离信息，设计信息融合算法，以实现避障，轨迹规划和自主飞行控制。

图 4 显示的是纽约城市大学 CityFlyer 四旋翼飞行机器人试验平台，它是以德国 Asctec Pelican 微型四旋翼飞行器为基础，安装 Hokuyo 线扫描激光测距仪，并利用一个小平面镜分离一束激光射向地面以获取精确的高度信息。



图 4: CityFlyer 四旋翼飞行机器人试验平台，配置有 Hokuyo 激光测距仪做为主要的外部环境感知器件。

2.2.1 体系结构

CityFlyer 四旋翼飞行机器人体系结构由图 5 所示，它由地面工作站，机载计算机和自动驾驶仪(autopilot board)构成。地面工作站主要用于场景显示和通过操纵杆或控制软件界面进行遥控操作。机载计算机选用的是 1.6G 赫兹阿童木 (Atom) 单板计算机，软件模块主要用于完成视觉里程计算，自定位，轨迹规划与地图创建算法。自动驾驶仪由两个 ARM7 机载控制器组成，下位机负责处理惯性导航器件(IMU)，马

达控制和基本的飞行姿态控制(升力控制，滚转角控制，倾斜角控制和偏航速度控制)。上位机负责卡尔曼滤波和信息融合算法，以实现飞行机器人的位置/速度控制和自主飞行控制。机载计算机和自动驾驶仪通过串联通讯口传递数据，地面工作站和机载计算机以无线通讯网络实现连结。我们设计的软件模块和工具包都是基于机器人操作系统(ROS——Robot Operating System)的开源程序。这些软件程序包发布在 ROS 网站

<http://www.ros.org/wiki/ccny-ros-pkg> 与广大机器人研究社群共享^[Dryal1]。

2.2.2 状态估计

惯性导航器件(IMU)提供了滚转角、倾斜角、偏航角(RPY)，和x-y-z各轴的线加速度，但这些读数噪声很大，必须经过卡尔曼滤波处理，融合通过激光测距仪(laser scanner)获得的测量信息，来估计飞行机器人运动状态。首先，我们利用室内环境的直线投影特性和惯性导航器件提供的滚转角和倾斜角读数，把激光测距仪的线扫描点投影到x-y平面(如图6所示)。其次，我们运用激光扫描匹配算法(scan matcher)，处理一序列的激光扫描投影，并从中估算出飞行机器人的位置(x, y)和偏航角度。此外，向下投射的激光高度计(laser altimeter)也可以利用激光测距读数，滚转角和倾斜角读数计算出高度值(z)。这些测量参数通过简单快速的 $\alpha\beta$ 滤波器($\alpha\beta$ -Filter)以减低噪音，做为卡尔曼滤波器的测量参数。

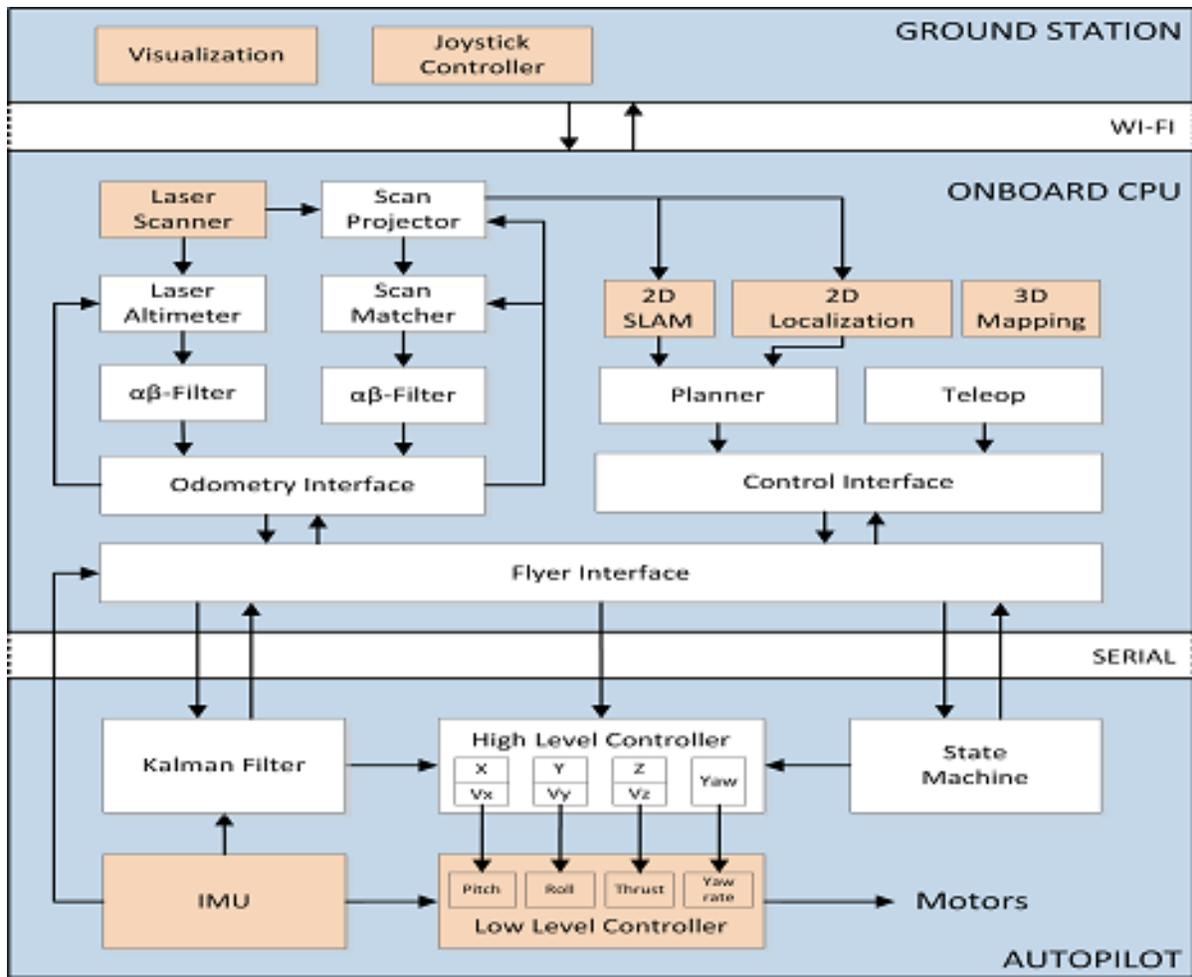


图 5:CityFlyer 四旋翼飞行机器人体系结构

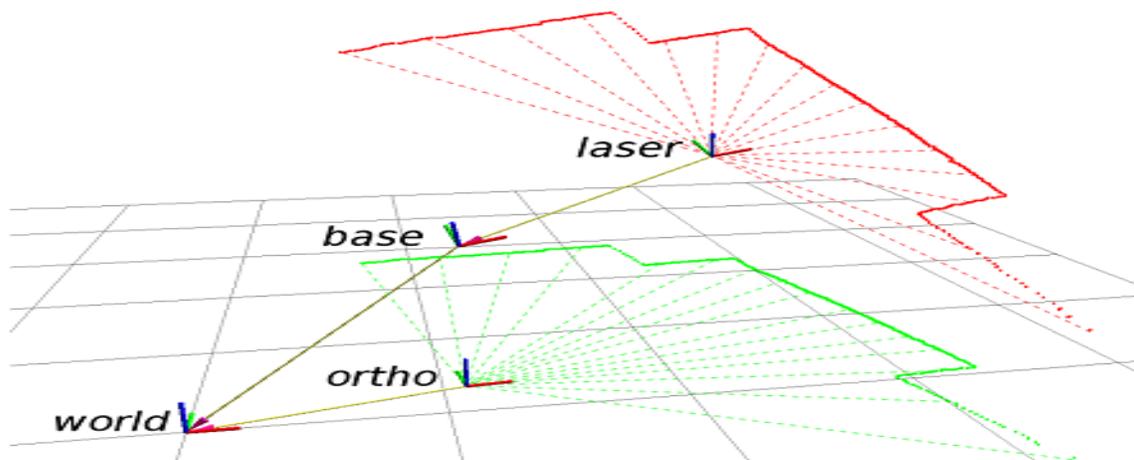


图6: 激光扫描投影步骤, 真实的激光扫描点(红色), 直角投影到x-y平面(绿色)。激光扫描仪坐标系(Laser), 机器人基坐标(base), 固定的世界坐标系(world), 直角投影坐标系(ortho)。

我们采用如图7所示的去耦卡尔曼滤波器，以x轴为例，其位置(x)、速度(v_x)和加速度(a_x)的关系由以下系统模型方程式表示，观测模型中的x_L, v_{xL}是从激光扫描匹配算法中获得的测量参数。

系统模型(固定加速度模型)

$$x_{k+1} = x_k + v_x \cdot dt + a_x \cdot \frac{dt^2}{2}$$

$$v_{k+1} = v_k + a_x \cdot dt$$

$$X_{k+1} = A \cdot X_k + B \cdot U_k$$

$$\begin{bmatrix} x_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & dt \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ v_k \end{bmatrix} + \begin{bmatrix} \frac{dt^2}{2} \\ dt \end{bmatrix} [a_x]$$

测量模型

$$Z_k = H \cdot X_k$$

$$H = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow Z_k = X_k$$

$$\text{修正: } \hat{Z} = \begin{bmatrix} x_L \\ v_{xL} \end{bmatrix}$$

运用如图8所示的标准卡尔曼滤波算法，我们可以得到稳定的飞行机器人运动状态估计值，包括x-y-z方向的位置，速度和偏航角。

2.2.3 位置控制

在此基础上，我们在上位机上设计了具有直接速度反馈的PI-D控制器，实现了基于激光扫描匹配的自主飞行位置控制(如图9所示)。我们用卡尔曼滤波器输出的速度估计值取代传统PID控制器采用的位置求导获得的速度值，极大地改善了控制性能，抑制了刺波。图10显示的是四旋翼机器人沿着规划航迹点飞行的实际轨迹。图11显示四旋翼机器人定点盘旋控制精度达到0.2米。

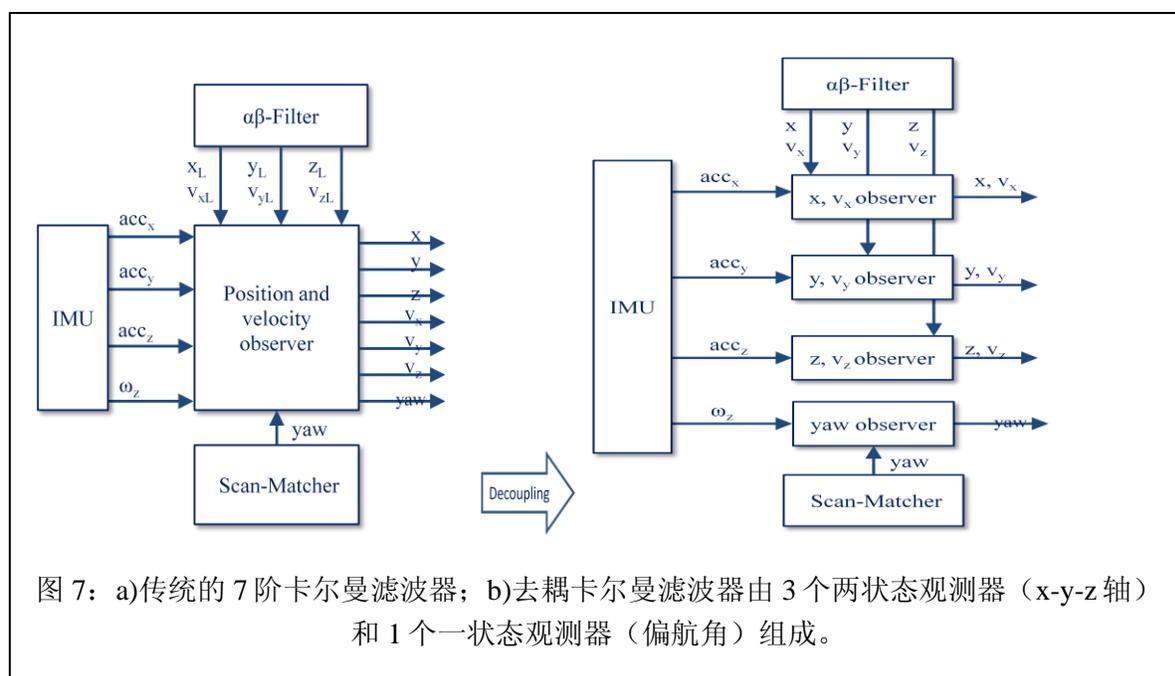


图 7: a)传统的 7 阶卡尔曼滤波器; b)去耦卡尔曼滤波器由 3 个两状态观测器 (x-y-z 轴) 和 1 个一状态观测器 (偏航角) 组成。

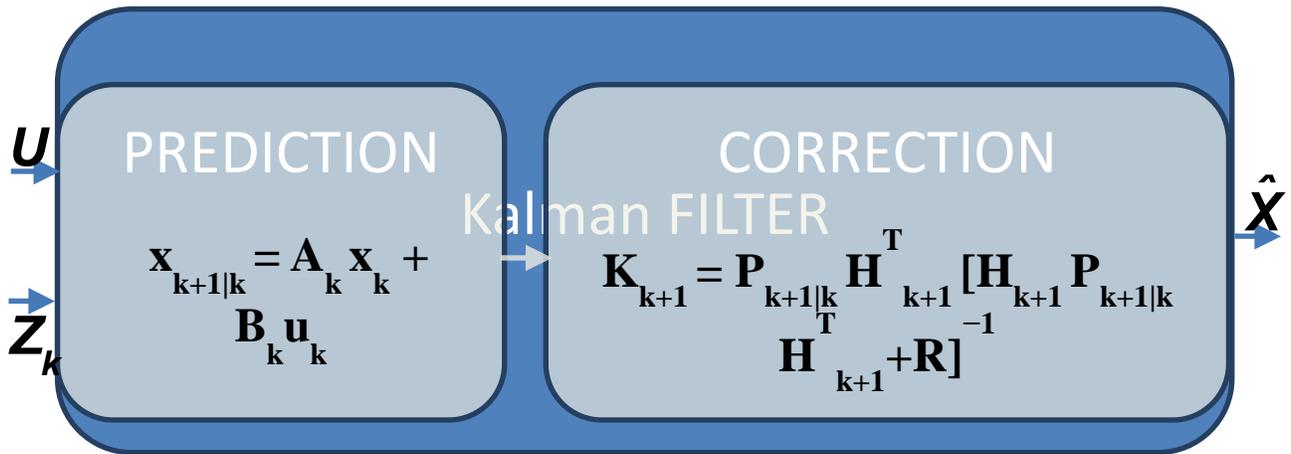


图8: 卡尔曼滤波器算法, K是卡尔曼矩阵, P是协方差矩阵, AB是系统模型矩阵, H是观测模型矩阵, Q, R 是符合高斯噪声分布的方差矩阵。

$$u(t) = K_p \cdot e(t) + K_i \cdot \int_0^t e(t) + K_d \cdot \frac{de}{dt} \quad \text{PID}$$

$$u(t) = K_p \cdot e(t) + K_i \cdot \int_0^t e(t) + K_d \cdot [-v(t)] \quad \text{PI - D}$$

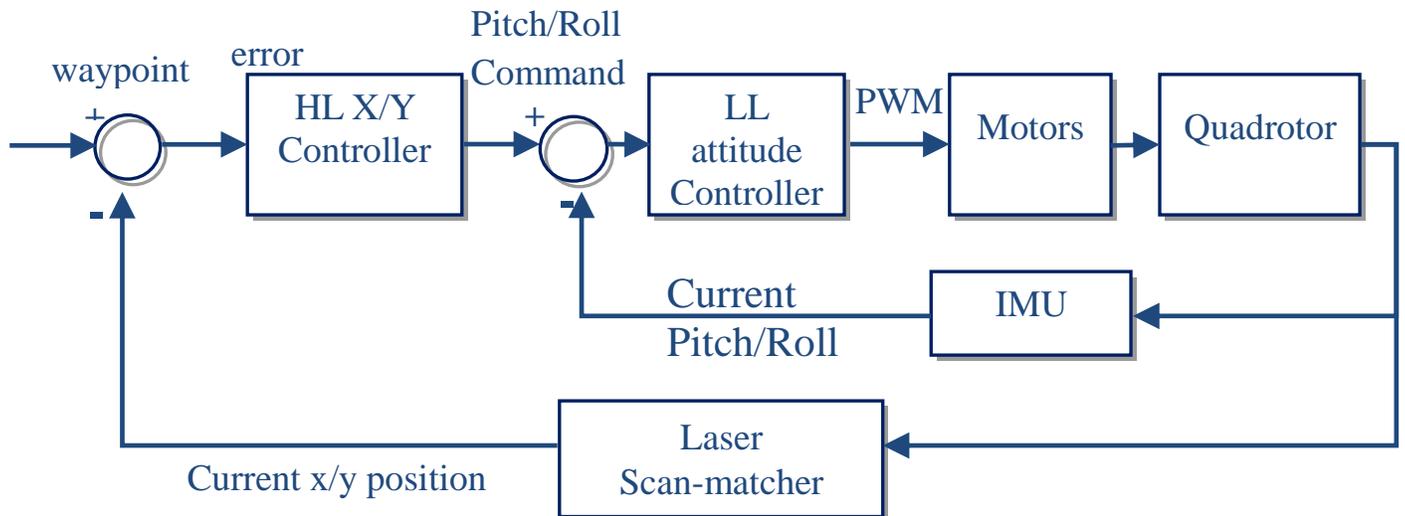


图 9: 具有直接速度反馈的 PI-D 控制器

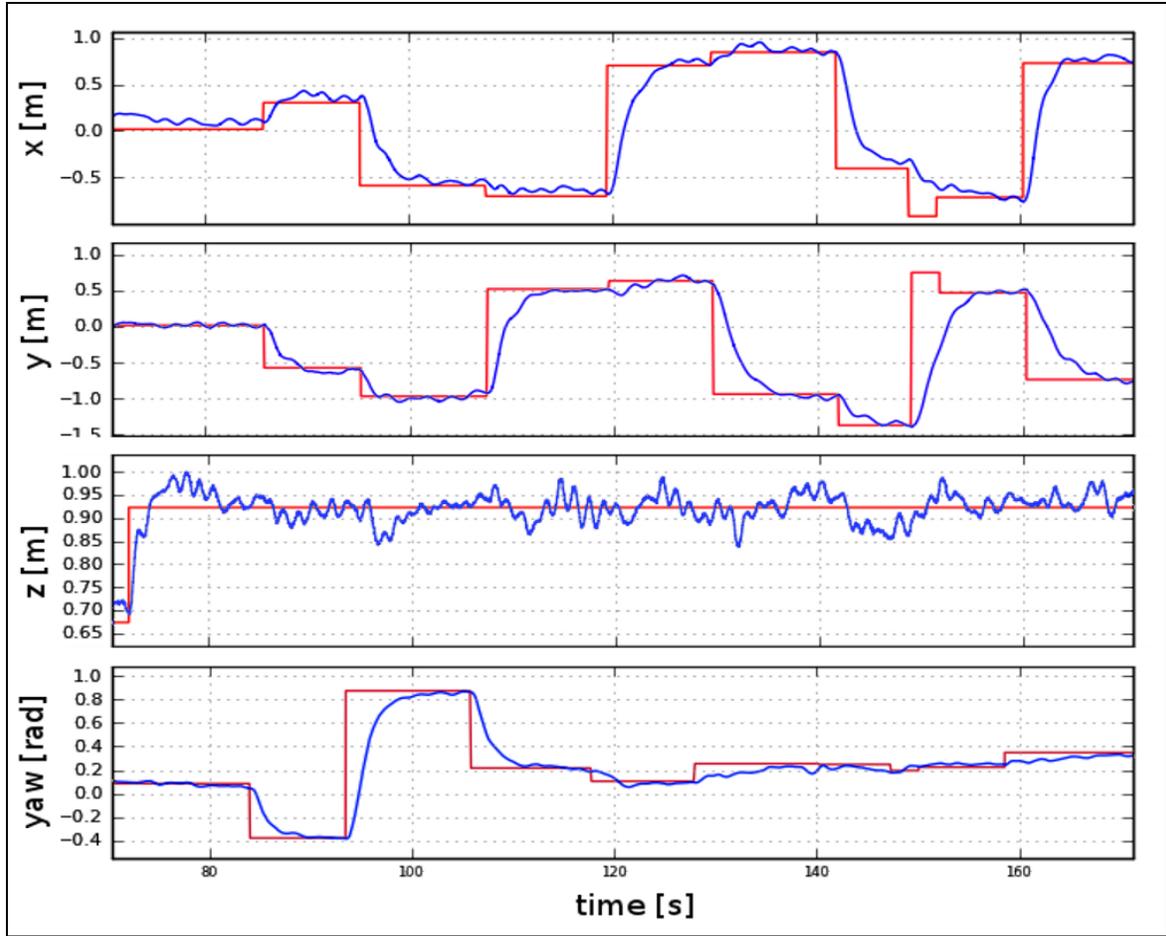


图 10: 四旋翼飞行机器人位置控制实验结果, 航点规划轨迹由红色表示, 蓝色线表示实际飞行轨迹。

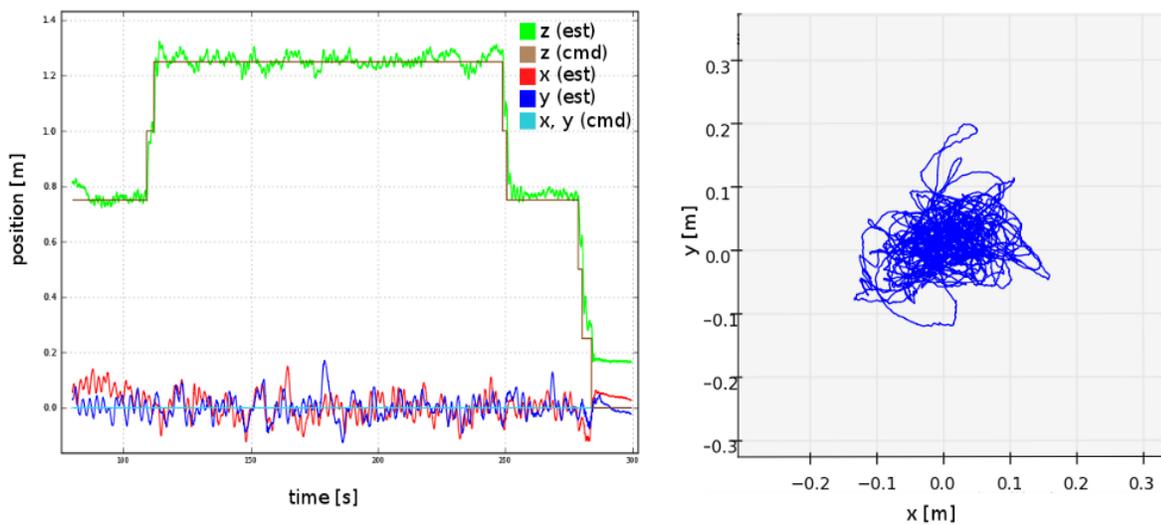


图 11: 四旋翼飞行机器人定点盘旋控制实验结果, 左图的绿色线表示高度控制轨迹, 右图蓝色线显示的是飞行机器人在 x-y 平面的实际轨迹 (飞行 4 分钟), 其定点盘旋控制精度达到 0.2 米。

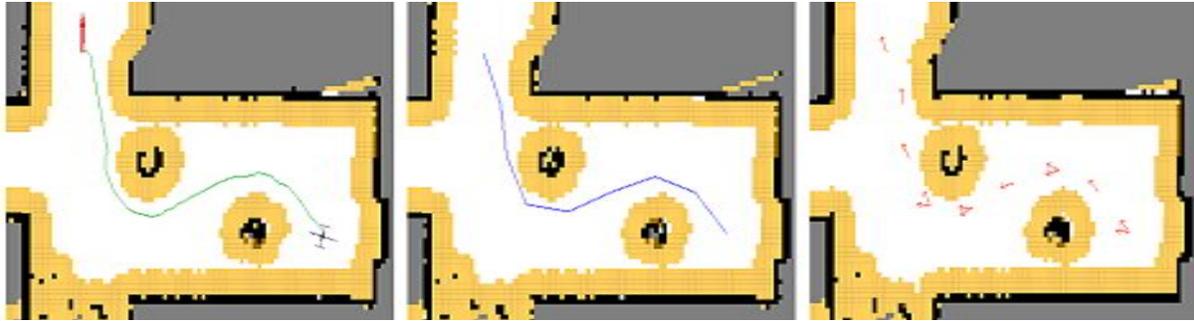


图 12: 轨迹规划步骤。左图: 用户在二维栅格图上点击选择飞行器目的地; 中图: 飞行轨迹分解为无碰撞的直线段; 右图: 产生一系列的规划航点, 包括定点转向航点。

2.2.4 轨迹规划

四旋翼飞行机器人的轨迹规划可分为三步。首先, 根据激光扫描点绘制二维栅格图, 每个栅格划分为已占用或空闲栅格。沿障碍物或围墙边界线扩张以飞行机器人的物理半径为宽度的已占用区域 (图12黄色区域)。这样, 飞行机器人就可以被认为是缩小为一个点, 用Dijkstra算法在空闲区域规划壁障轨迹以连接起始点到目标点。第二步是采用递归分解算法反复二次分割直线段轨迹, 直到所有的轨迹段都可以避开障碍物。第三步是沿着规划轨迹产

生一系列的航迹点, 每个航迹点包含 x, y 坐标和偏航角 $[x, y, yaw]$ 。如果一个航迹点到下一个航迹点之间的偏航夹角小于 20° , 允许四旋翼飞行机器人向前飞行并同时调整偏航角。如果两个航迹点之间的偏航夹角过大, 可以中间插入一个定点转向航迹点以强制飞行器先调整偏航角再向前飞行。飞行机器人不断检查两个航迹点之间是否有动态障碍物以随时重新规划轨迹。

2.3 三维环境建模

三维地图创建面临三大挑战, 首先, 它要求机器人有适当的传感器观测环境并自定位, 其次, 它需要设计合适的数据结构以简化地图的表达, 最后, 它需要高效的算法把多个局部地图融合起来。我们利用线扫描激光测距仪和室内结构化环境特有的直线投影(rectilinear)特性, 成功实现了室内同步自定位和创建三维地图的功能。我们提出了 Multi-Volume Occupancy Grid (MVOG) 方法 [Drya10, Drya11, Morr10, Morr11], 可以把激光点云图数据极大地压缩并以连结树的数据结构表达同一个二维点上

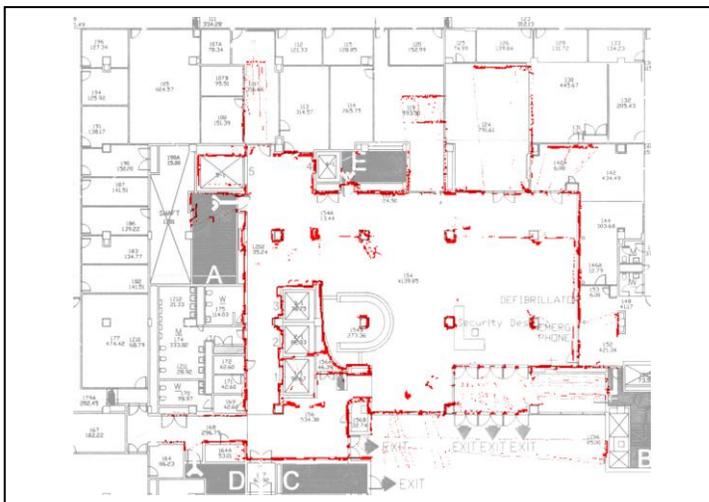


图 13: 四旋翼飞行器产生的二维图 (红色) 覆盖在纽约城市大学工学院大厅的实际建筑平面图上。

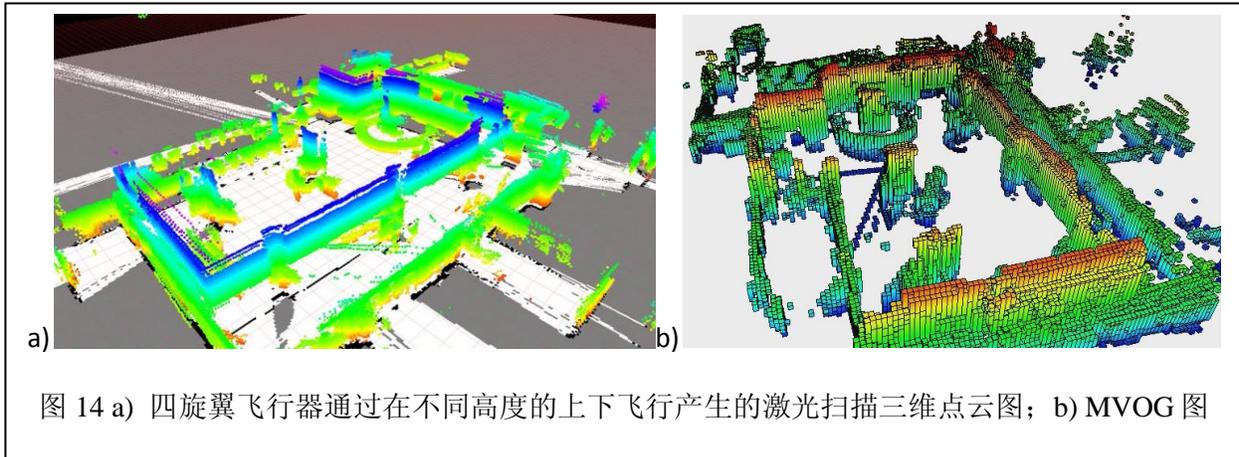


图 14 a) 四旋翼飞行器通过在不同高度的上下飞行产生的激光扫描三维点云图；b) MVOG 图

不同高度的地方是否被占据。如图 14 所示，MVOG 方法建立的地图只占很少的内存，可被直接用于三维空间的自主导航飞行，并在微型四旋翼飞行器的嵌入式控制系统上实现。我们制作的 Youtube 录像*展示了微型四旋翼飞行器在室内环境中自主飞行和创建三维地图的过程。
*<http://youtu.be/8TXxnGXIEjg>

然而，在洞穴搜索应用中，直线投影的假设不成立，获取三维地图的难度成倍增加。众所周知，机器人可以在二维地图上用点和线表示障碍物和走廊等物体。而

三维地图创建的基本元素是平面。机器人可以用激光点云图或分段平面来表示三维地图，其复杂度和对计算机运算能力的要求都大大提高。解决这一关键问题的创新点在于以下几方面：**A)** 设计简化的可在机载计算机平台上实时运行的基于图像匹配的自定位算法。**B)** 综合利用新型 RGB-D 传感器提供的图像和三维点云图，设计特征提取和机器人姿态估计算法。**C)** 设计高效的数据结构以表达立体信息，减少存储器用量。**D)** 设计快速高效的算法精炼并把多个局部地图融合。

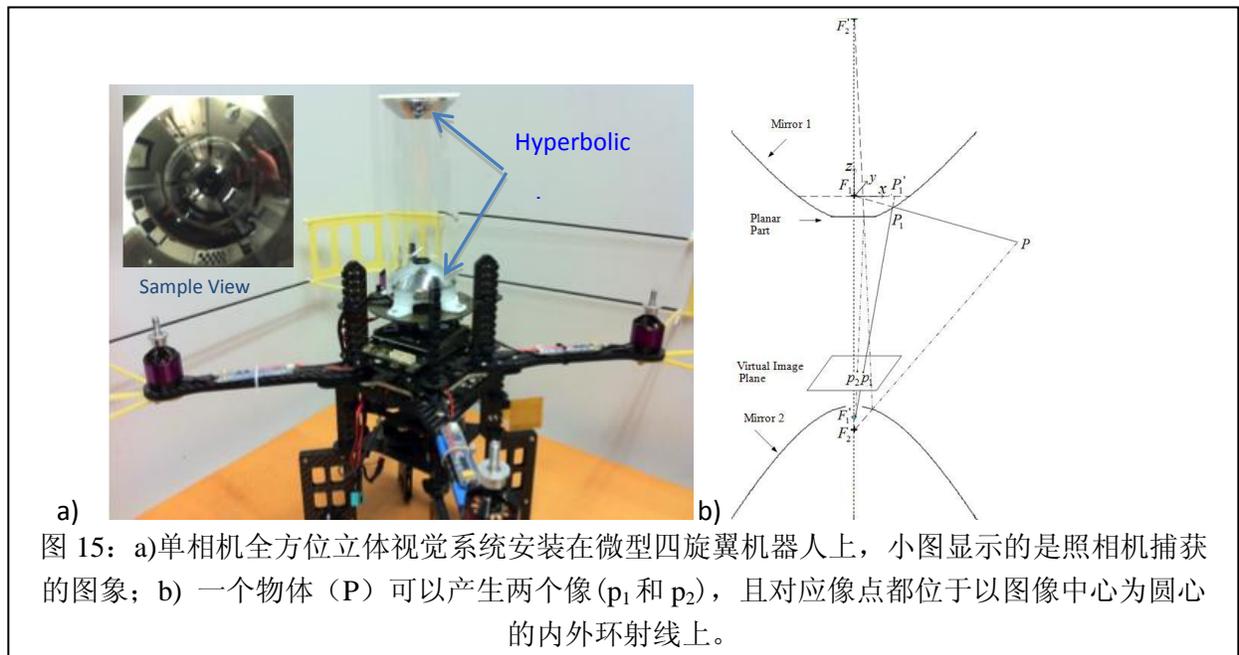


图 15: a) 单相机全方位立体视觉系统安装在微型四旋翼机器人上，小图显示的是照相机捕获的图象；b) 一个物体 (P) 可以产生两个像 (p_1 和 p_2)，且对应像点都位于以图像中心为圆心的内外环射线上。

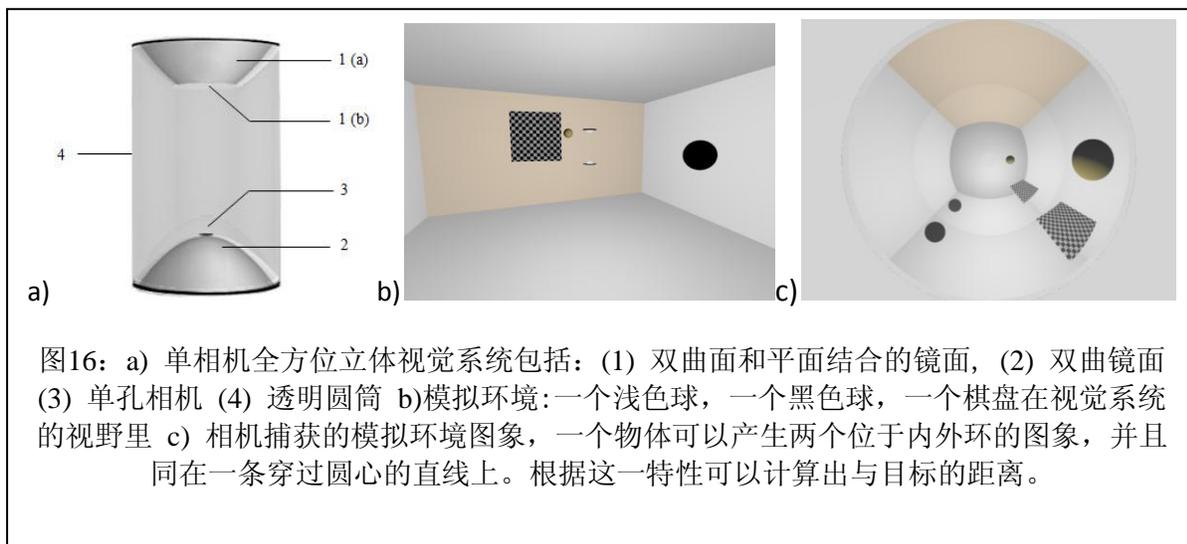
2.4 单相机全方位立体视觉系统

视觉系统能够通过成像硬件设备快速获取周围环境的图像，通过软件模块分析图像中目标的特征，是机器人和其他智能系统中不可或缺的感知环境、辅助目标识别、路径规划等任务的重要手段。我们设计了一种紧凑轻便的单相机全方位立体视觉系统以及相关的深度恢复算法和图像处理软件^[Guo10]，这一系统可以安装在四旋翼飞行器上(见图 15)以获取周围 360 度的图象信息并计算出与目标的距离。

单相机全方位立体视觉系统的设计原理由图 16 所示。它由一个透明圆筒，上下两个特定形状的反射镜面和一个透视相机组成。上侧镜面为双曲面和平面镜二者构成的组合镜面；下侧镜面为双曲面。透视相机安装于下侧镜面的内部，其主光轴与上下两个反射镜面的主轴共轴，镜头比下侧镜面顶部的通光孔略小。一个位于上下两个镜面视场相交迭处的物体可以在透视相机上产生两个图象，分别位于内外环，且对应像点都位于以图像中心为圆心的射线上。如图 15(b)所示，上侧组合镜面的

双曲面单独成像到外环；平面镜与下侧镜面构成折返系统成像到内环。该系统结构紧凑、轻便小巧，具有单视点 (single view point) 的特性，能形成水平 360 度、垂直 150 度的三维视场，便于利用深度恢复算法获取立体视觉信息。

图 17 中的圆形图像是照相机获取的原始全景图像，右边显示的是一对展开的经过表格查寻矫正 (look-up table) 的全景投影图像；上面的图像对应外环，下面的图像对应内环。我们采用了两种方法在两幅图像的交迭区域寻找同一物体相对应的匹配点。一种是传统的区块匹配方法，即用特定大小的图像窗口沿着投影线寻找匹配点，可以产生密集立体图；另一种是特征匹配方法，即搜索具有比例/旋转不变特性的匹配点，产生稀疏立体图。单相机全方位立体视觉系统的基线是上下两个双曲镜面原点之间的距离。根据三维空间目标点投影到上下两幅图像的匹配点的三角几何关系，我们可以计算出目标点到相机的距离，用于路径规划和自主导航控制。



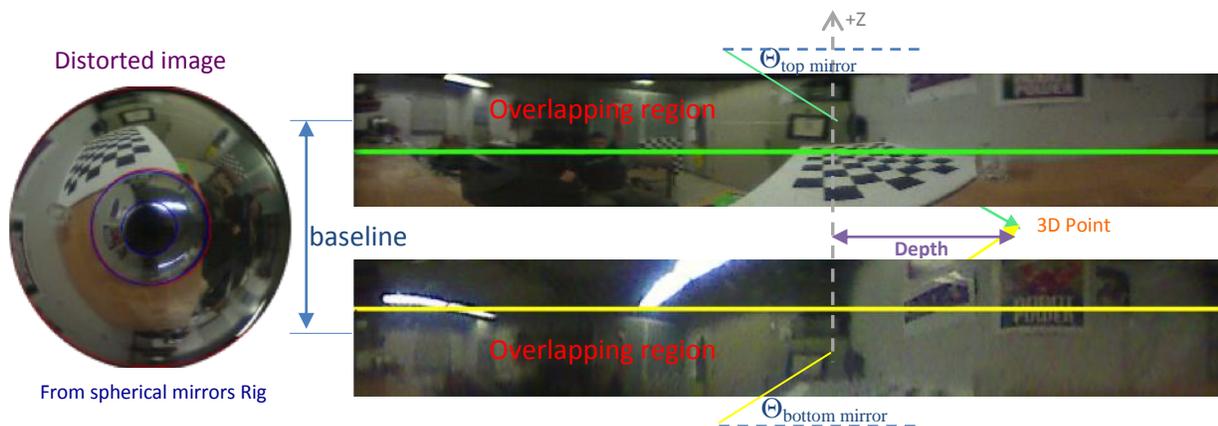


图17: 立体视觉信息的获取方法: 1) 经过表格查寻矫正算法, 把原始图像展开为上下两幅全景投影图像, 2) 沿着投影直线在两幅图像的交迭区域寻找同一物体相对应的匹配点, 3) 根据三角几何关系(基线, 仰角 Θ_{top} , Θ_{bottom})计算出目标点到相机的距离。

3. 研究方向

本节指出未来研究方向和拟解决的关键科学问题及创新点。

1) 飞行控制问题。 微型旋翼飞行器是一个具有六自由度(位置与姿态)和多个控制输入(旋翼转速)的多变量、非线性、强耦合的复杂系统。它对旋翼空气动力学, 陀螺效应以及阵风等干扰因素很敏感, 负载配置的变化也使飞行控制系统的设计变得非常困难。解决这一关键问题的创新点在于研究设计既能精确控制飞行器姿态, 又具有较强抗干扰和自适应能力的非线性控制算法, 使微型旋翼机器人能够在复杂环境下敏捷机动地全天候飞行。

2) 实时感知问题。 由于负载有限, 微型旋翼飞行器只能安装小型轻便的传感器和嵌入式计算机, 其对外部环境的感知能力和信息实时处理能力受到制约。目前国际上通常采用小型的视觉摄像机和线扫描激光测距仪作为主要的传感器, 并不能达到理想的效果。解决这一关键问题的创

新点在于设计新型三原色加距离(RGB-D)传感器, 单相机全方向立体视觉系统, 微型面扫描激光测距仪, 并研究多传感器信息融合技术。

3) 自主导航问题。 在没有GPS信号的情况下, 准确进行自身运动估计是自主导航首要克服的问题。这个问题的一个很自然的解决方案是利用图像匹配技术来估计自身运动(ego-motion)。然而, 复杂的视觉导航算法不易在低运算力的机载计算机平台上实现。解决这一关键问题的创新方案有以下几种: A) 采用基于面-线匹配的视觉里程推算法估计自身运动。在城市环境中, 建筑物的垂直边缘和侧平面是比较容易提取的环境特征。与点特征相比, 平面和边缘是更具鲁棒性的特征。旋翼机器人水平移动对垂直边缘的视场变化非常敏感。通过改进RANSAC方法, 可以明显降低特征空间维数, 提高算法的精确度, 使之适合在机载计算机平台上运行。B) 从昆虫飞行机理得到灵感, 研究基于视觉

流场(optical flow)的导航控制方法并推广到三维空间。

4) 环境建模问题。微型旋翼飞行机器人已经成功实现了室内同步自定位和创建三维地图的功能。然而,在洞穴搜索应用中,直线投影的假设不成立,获取三维地图的难度成倍增加。众所周知,机器人可以在二维地图上用点和线表示障碍物和走廊等物体。而三维地图创建的基本元素是平面。机器人可以用激光点云图或分段平面来表示三维地图,其复杂度和对计算机运算能力的要求都大大提高。解决这一关键问题的创新点在于以下几方面: A) 设计简化的可在机载计算机平台上实时运行的基于图像匹配的自定位算法。B) 综合利用新型RGB-D传感器提供的图像和三维点云图,设计特征提取和机器人姿态估计算法。C) 设计高效的数据结构以表达立体信息,减少存储器用量。D) 设计快速高效的算法精炼并把多个局部地图融合。

5) 多机协同问题。单个旋翼飞行机器人的功能和有效载荷毕竟有限,多机协同操作被认为可以更加有效地完成复杂任务(例如,大范围的搜索)。然而,要达到理想的协同作战效果,必须在以下几方面有创新: A) 研究多机协同控制的体系结构并设计任务管理软件模块以监控整个任务完成情况,合理调度、分配资源。B) 研究机器人之间的无线通讯协议和信息交换机制,以确保移动网络的连通性和鲁棒性。C) 研究人机互动机制,运用时序逻辑(temporal logic)实现操作人员对多个机器人的使命规划控制。D) 研究与三维地理信息系统结合的飞行轨迹自动规划。

4. 结语

本文概括介绍作者在探索智能化微型旋翼飞行机器人在多障碍物复杂环境下的感知,控制,自主导航,环境建模等领域的阶段性研究成果,指出未来的主要研究方向和拟解决的关键技术,希望能起到一个抛砖引玉的作用,为国家安全与公共安全提供能满足实际需求的高技术侦察监视工具。

REFERENCES

[Drya10] Ivan Dryanovski, William Morris, Xiao Jizhong, "Multi-Volume Occupancy Grids: an Efficient Probabilistic 3D Mapping Model for Micro Aerial Vehicles", 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2010), Teipei, Oct. 2010, pp1553~1559.

[Drya11] Ivan Dryanovski, William Morris, Jizhong Xiao, "An Open-Source Pose Estimation System for Micro-Air Vehicles", Proceedings of 2011 International Conference on Robotics and Automation (ICRA2011), pp4449-4454.

[Drya12] Ivan Dryanovski, Carlos Jaramillo, Jizhong Xiao, "Incremental Registration of RGB-D Images", 2012 International Conference on Robotics and Automation (ICRA2012), May 14~18, St Paul, Minnesota, USA, pp1685-1690.

[Guo10] Guo Ling, Igor Labutov, Xiao Jizhong, "Design and calibration of single-camera catadioptric omnistereo system for Miniature Aerial Vehicles (MAVs)", 2010 IEEE/RSJ Int. Conf. on Intelligent Robots

and Systems (IROS2010), Taipei, Oct. 2010, pp. 622-627.

[Kaus10] Ravi Kaushik, Xiao Jizhong, Samleo L. Joseph, William Morris, “Polygon-based Laser Scan Registration by Heterogeneous Robots”, 2010 IEEE International Conference on Robotics and Biomimetics (Robio2010), Tianjin, Dec. 2010, pp. 1618-1623.

[Kaus12] Ravi Kaushik, Jizhong Xiao, “Accelerated Patch-based Planar Clustering of Noisy Range Images in Structured Indoor Environments”, Robotics and Autonomous Systems, Elsevier, Vol. 60, Issue 4, 2012, pp. 584~598.

[Labu11a] Igor Labutov, Carlos Jaramillo, Jizhong Xiao, “Generating Near-Spherical Range Panoramas by Fusing Optical Flow and Stereo from a Single-Camera Folded Catadioptric Rig”, Machine Vision and Applications, Springer, 17 September 2011, pp. 1-12.

[Labu11b] Igor Labutov, Carlos Jaramillo, Jizhong Xiao, “Fusing Optical Flow and Stereo in a Spherical Depth Panorama Using a Single-Camera Folded Catadioptric Rig”, Proceedings of 2011 International Conference on Robotics and Automation (ICRA2011), pp3092-3097. (ICRA2011 Best Vision Paper Award Finalist)

[Morr11] William Morris, Ivan Dryanovski, Jizhong Xiao, “CityFlyer: Progress Toward Autonomous MAV Navigation and 3D Mapping”, Video, 2011

International Conference on Robotics and Automation (ICRA2011), pp2972-2973.

Biography:



Jizhong Xiao received his Ph.D. degree from Michigan State University in 2002; Master of Engineering degree from Nanyang Technological University, Singapore, and MS, BS degrees from East

China Institute of Technology in 1999, 1993, and 1990, respectively. He joined the EE Dept. of the City College of New York (CCNY, CUNY City College) in 2002 and currently is an associate professor and the EE Ph.D. Program advisor at CUNY City College, as well as a doctoral faculty member of the Ph.D. program in Computer Sciences at the Graduate Center of the City University of New York (CUNY Graduate Center). Dr. Xiao started the robotics research program at the City College and is the founding director of CCNY Robotics Lab (website: <http://robotics.cuny.cuny.edu>) and the Center for Perceptual Robotics, Intelligent Sensors and Machines (PRISM Center).

His research interests include robotics and control, cyber-physical systems, autonomous navigation and 3D SLAM, real-time and embedded computing, digital signal processing, assistive technology, multi-agent systems and swarm robotics. He is a recipient of National Science Foundation (NSF) CAREER award in 2007 and the CCNY Outstanding Mentor Award in 2011. Dr. Xiao has served the robotics community in various roles of many robotics conferences, and served as guest editor and reviewer for major robotics journals and conferences. He is a senior member of IEEE.

ACM 和 IEEE 简介

陈志雄

旅美科协总会副会长，大纽约分会理事会主席

zxchen@ieee.org



计算机协会 (Association for Computing Machinery, ACM, <http://www.acm.org/>) 是世界上最大的与计算、信息技术科研和教育相关的组织。它于 1947 年 9 月 15 日在美国纽约哥伦比亚大学召开的一个会议上正式创立，原名是东部计算机学会，一年后改为现在的名称。它的成立与当时计算领域的最新发展(如具有储存功能的计算机的发明、数字计算的快速进展)密切相关。今天，计算机协会通过举办学术会议、出版学术刊物和数字图书馆，设立专业委员会，颁发各类科技奖，组织全球编程比赛，帮助职业选择等来推动计算作为一门学科和一种职业的发展。它在全美各地都有分会，也在世界各地发展会员。

图灵奖是 ACM 最有影响的奖项 (<http://amturing.acm.org/byyear.cfm>)。今

年 6 月，ACM 在美国三藩市举行了图灵奖世纪庆祝大会。有三十多位图灵奖获得者参加了会议。

美国电气电子工程师学会 (the Institute of Electrical and Electronics Engineers, IEEE, <http://www.ieee.org>) 于 1963 年正式成立。它由成立于 1884 年的美国电气工程师协会 (AIEE) 和成立于 1912 年的无线电工程师协会 (IRE) 合并而成。



今天，IEEE 是全球最大的电气及电子工程、计算机、通信等领域的专业技术组织。它的主要目标是培养和培育技术创新和技术卓越。IEEE 计算机学会是 IEEE 的重要组成部分。于 ACM 一样，IEEE 计算机学会也举办众多学术会议，出版学术刊物及数字图书馆，设立各种专业分会，制定技术标准，颁发各类奖项，帮助职业选择。

破茧成蝶：商务智能简史

你可能不知道，商务智能，这个二十一世纪的时尚名词，其实有着几乎和计算机一样长的历史。

涂子沛

缘起：决策支持系统¹

信息消费了什么是很明显的：它消费的是信息接受者的注意力。信息越丰富，就会导致注意力越匮乏……信息并不匮乏，匮乏的是我们处理信息的能力。我们有限的注意力是组织活动的主要瓶颈。

——赫伯特·西蒙（1916-2001），美国经济学家、政治学家、人工智能的创始人之一，1973年²

1946年，人类历史上第一台电子计算机在美国费城问世³。来自匈牙利的移民冯·诺伊曼是这台计算机的主要设计者，他被后世称为“计算机”之父。

仅一年之后，卡内基梅隆大学的赫伯特·西蒙(Herbert Simon)教授出版了《行政组织的决策过程》一书。在这本被后世喻为经典的著作里，他指出，人类的理性是有限的，因此所有的决策都是基于有限理性(Bounded Rationality)的结果。这位天才科学家继而提出，如果能利用存贮在计算机里的信息来辅助决策，人类理性的

范围将会大大扩大，决策的质量就也能提高。

他进而预测：在后工业时代，也就是信息时代，人类社会面临的中心问题将从如何提高生产率转变为如何更好的利用信息来辅助决策。

西蒙毕业于芝加哥大学，1943年获得政治学博士学位，此后半个多世纪，他长期在卡内基梅隆大学任教。

卡内基梅隆大学，是美国信息技术研究的“火车头”，它以计算机科学和“交叉性研究”闻名于世。西蒙的整个学术生涯都浸润着卡内基梅隆的色彩，他从政治、经济出发，把毕生的精力都集中在对决策和信息的研究上，将不同学科之间的“交叉性”应用得炉火纯青，也硕果累累。1973年，因为对人工智能的贡献，他获



赫伯特·西蒙(1916-2001)，二十世纪最具影响力的科学家之一，他横跨多个学科和领域，曾获得1975年的图灵奖、1978年的诺贝尔经济奖、1993年的美国心理协会终身成就奖。

图片来源：卡内基梅隆大学图书馆

¹ 作者注意到，对于数据分析技术，在商业领域多被称为“商务智能”；在政府领域则多被称为决策支持系统。

² 英语原文为：What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention... The scarce resource is not information; it is processing capacity to attend to information. Attention is the chief bottleneck in organizational activity *Designing Organizations for an Information-Rich World* Simon, 1971

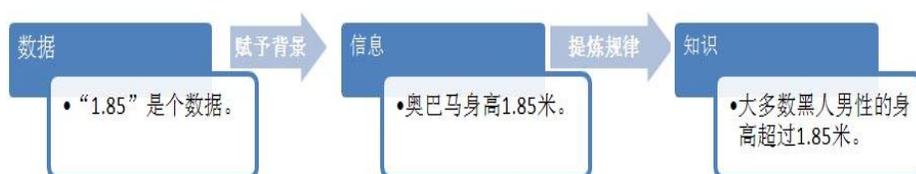
³ 人类第一台计算机到底在哪年发明的，近年来学界对此有所争议。本书以1946年2月在宾夕法尼亚大学发明的ENIAC计算机为准。

得了计算机界的最高奖项：图灵奖；1978年，他又因为对“商务决策过程”的出色研究戴上了诺贝尔经济学奖的桂冠。

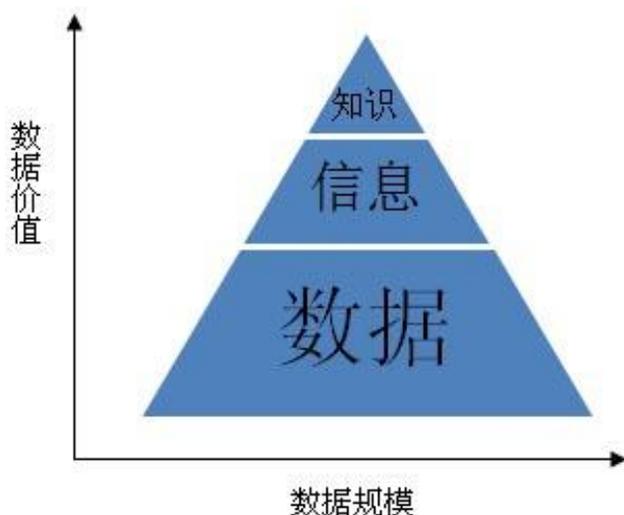
追本溯源，学界普遍认为，西蒙对决策支持系统的研究，是现代商务智能概念最早的源头和起点。但西蒙可能没有想到，他播下的“决策支持”的种子，在半个世纪以后，却结出了“商务智能”的果实，并成为信息时代的一道奇葩。

从决策支持系统到商务智能，名字变了，但新瓶装的还是旧酒。现代商务智能技术回答的还是决策支持系统面对的老问题：如何将数据、信息转化为知识，扩大人类的理性，辅助决策？

从数据、信息到知识的演变



说明：如果经过对现有数据的分析和统计，能得出“大多数黑人男性的身高超过 1.85 米”的结论，那这就是一则知识。



从数据到知识，这个跨越，人类用了半个多世纪。

数据、信息和知识的区别和联系

在半个多世纪的漫长过程中，决策支持系统曾经一度因为缺乏有效的数据组织方式而徘徊不前。直到上世纪 90 年代，由于若干新技术的出现，打破了瓶颈，“商务智能”的概念才横空出世。随后，其发展取得了前所未有的加速度，在本世纪第一个十年蓬勃向上。今天，回头考察这些新技术的一一出现，可以清楚的看到商务智能的产业链条不断向前延伸的轨迹。

这个轨迹的起点当然就是计算机。计算机，是硬件和软件相结合的产物。它的发明，是诸多不同领域的科学家共同努力的结果。冯·诺伊曼其实是一名数学家，他之所以被称为“计算机之父”，其最大的贡献之一，在于他明确了计算机内部的数据组织形式：二进制。

二进制的引进，解决了在没有“情感、智能和生命”的物理机器中表达、计算、传送数据的最大难题，有了二进制，软件的运行才有了支点。

如前文所述，软件是由程序和数据组成的。二进制的确定，解决了数据在计算机内部“理解”和“流动”的问题，但当数据在计算机内部累积得越来越多的时候，如何快速的组织、存贮和读取数据又成为了新的挑战。

计算机科学家一直在研究数据在软件内部的最佳组织方式。1970年，

IBM 的研究员埃德加·科德(Edgar Codd)发明了关系型数据库,成了软件发展历史上一个跨越性的里程碑。

此前,数据库的组织结构以网状、层级制为主,复杂多变,程序和数据之间你中有我、我中有你,彼此有很强的依赖性。科德提出的关系型数据库具有结构化高、冗余度低、独立性强等优点,彻底的把软件中的程序和数据分立开来。从此,软件的发展成了“两条腿”走路,程序和数据在各自的轨道上自由奔跑。

科德后来又总结出构建关系型数据库的“黄金十二定律”,把理论扎扎实实的推向了实践,关系型数据库开始得到大面积的推广,引发了一场软件领域的革命。科德也因此获得 1981 年的图灵奖。

此后,大型软件、即大型信息管理系统的应用一日千里、遍地开花。

这些信息系统的建立和运行,使人类从繁杂的重复性劳动当中解放出来,大大的提高了商业效率。但这些信息系统,都是针对特定的业务过程、处理离散事务的“运营式”信息系统。

所谓“运营式系统”,是指为提高日常工作的效率而设计的系统,数据在其中的作用,是一个个商务流程的记录,数据在这些系统内不断累积的结果,仅仅用于查询,而不是分析。

上个世纪 90 年代,面对信息管理系统普及、各行各业数据纪录的激增,管理大师彼得·德鲁克(Peter Drucker)曾发

出慨叹说:迄今为止,我们的系统产生还仅仅是数据,而不是信息、更不是知识⁴!

怎样从各个独立的信息系统中提取、整合有价值的信息,从而实现从数据到信息、从信息到知识、从知识到利润的转化?这个要求,随着信息管理系统的普及,变得越来越迫切。企业的规模越来越庞大、组织越来越复杂,市场更加多变、竞争更加激烈,信息是否及时准确、决策是否正确合理,对组织的兴衰存亡影响越来越大,一步走错,可能全盘皆输。

因为实业界的这些迫切需要,决策支持系统的旧问题又重新占据了顶尖科学家的大脑。

商务智能的“幽灵”开始徘徊……

结蛹：厚积薄发之数据仓库

岳不群叹了口气,缓缓的道:“三十多年前,咱们气宗是少数,剑宗中的师伯、师叔占了大多数。再者,剑宗功夫易于速成,见效极快。大家都练十年,定是剑宗占上风;各练二十年,那是各擅胜场,难分上下;要到二十年之后,练气宗功夫的才渐渐的越来越强;到三十年时,练剑宗功夫的便再也不能望气宗之项背了。然而要到二十余年之后,才真正分出高下,这二十余年中双方争斗之烈,可想而知。”

——金庸,《笑傲江湖》第九章,1967 年

决策支持系统面临的“瓶颈式”难题,是如何有机的聚集、整合多个不同运营信息系统产生的数据。对这个问题的关注起

⁴ Our systems are great at producing data but not at producing information. In our daily life, we deal with huge amount of data and information. Data and information is not knowledge until we know how to dig the value out of it. Forbes, 24 August 1998

源于美国计算机科学研究的另一所重镇：麻省理工学院。和卡内基梅隆大学一起，这两所大学先后为现代商务智能的发展奠定了主要的基石。

20世纪70年代，麻省理工学院的研究人员第一次提出，决策支持系统和运营信息系统截然不同，必须分开，这意味着要为前者设计独立的数据存储结构。但受限于当时的数据存储能力，该研究在确立了这一论点后便停滞不前。

但这个研究如灯塔般为实业界指明了方向。1979年，一家以决策支持系统为已任、致力于构建独立数据存储结构的公司 Teradata 诞生了。Tera，是太字节，其大小为 2^{40} ，Teradata 的命名表明了公司处理海量数据的决心。1983年，该公司利用并行处理技术为美国富国银行 (Wells Fargo Bank) 建立了第一个决策支持系统。这种先发优势令 Teradata 至今一直雄居在数据行业的龙头榜首。

另一家信息技术的巨头，国际商业机器公司 (IBM) 也在为集成企业内不同的运营系统大伤脑筋。越来越多的 IBM 客户要面对多个分立系统的数据整合问题，这些处理不同事务的系统，由于不同的编码方式和数据结构，象一个个信息孤岛，处于老死不相往来的状态。1988年，为解决企业的数据集成问题，IBM 公司的两名研究员创造性的提出了一个新的术语：数据仓库 (Data Warehouse)⁵。

一声惊雷，似乎宣告了数据仓库的诞生。可惜的是 IBM 在首创这个概念之后，

也停步不前，只把它当作一个花哨的新名词用于市场宣传，而没有趁胜追击、进一步提出实际的架构和设计。IBM 很快在这个领域丧失其领先地位，2008年，IBM 甚至通过兼并 Cognos 才使自己在商务智能的市场上重占一席之地，这是后话。

但这之后，更多的信息技术企业垂涎于数据仓库的“第一桶金”，纷纷开始尝试搭建实验性的数据仓库。

又是几年过去。1992年，尘埃终于落定。比尔·恩门 (Bill Inmon) 出版了《数据仓库之构建》 (*Building the Data Warehouse*) 一书，第一次给出了数据仓库的清晰定义和操作性很强的实战法则，真正拉开了数据仓库走向大规模应用的序幕。恩门不仅是长期活跃在这个理论领域的领军人物，还是一名企业家。此后，他的江湖地位也得以确定，被誉为“数据仓库”之父。

恩门所提出的定义至今仍被广泛的接受：

“数据仓库是一个面向主题的 (Subject Oriented)、集成的 (Integrated)、相对稳定的 (Non-Volatile)、反映历史变化 (Time Variant) 的数据集合，用于支持管理中的决策制定。”

数据仓库和数据库的最大差别，在于它是以分析、决策支持为目的来组织存贮数据，而数据库的主要目的则是为运营性信息系统保存数据。

江山代有才人出。

⁵ 该二人为 Barry Devlin 和 Paul Murphy。

恩门一统江湖没多久，风头又被拉尔夫·金博尔(Ralph Kimball)抢了去。金博尔是斯坦福大学毕业的博士，长期在决策支持系统的软件公司工作。1996年，他也出版了一本书：《数据仓库的工具》(The Data Warehouse Toolkit)，金博尔在书里认同了比尔·恩门对于数据仓库的定义，但却在具体的构建方法上和他分庭抗礼。



拉尔夫·金博尔(左)

他的数据仓库构建方法目前在市场上占据了主流。和普适计算的创始人马克·韦泽一样，他也曾经在施乐公司的帕罗奥多研究中心公司(PARC)长期工作过。

图片来源：datamgmt.com 网站



比尔·恩门：数据仓库之父(右)

2007年曾被《计算机世界》评为近40年计算机产业最具影响力的十大人物之一，目前还活跃在数据仓库领域，他的最新成果是将“非结构化的文本数据”通过特定的工具装入数据仓库。

图片来源：Computerworld 网站

恩门强调数据的一致性，主张由顶至底的构建方法。一上来，就要先创建企业级的数据仓库。金博尔却说：不！务实的数据仓库应该由下往上，从部门到企业，并把部门级的数据仓库叫做“数据集市”(Data Mart)。两人针锋相对，各自的追随者也唇舌相向，很快形成了明显对立的两派。

两派的异同，就好比华山剑法的气宗和剑宗。主张练“气”的着眼全面和长远，

耗资大，见效慢；主张练“剑”的强调短、平、快，效果可能立竿见影。

如金庸在《笑傲江湖》中描写的剑气之争一样，两派华山论剑的结果不难猜测，金博尔“从易到难”的架构迎合了人类的普遍心理，大受欢迎，商务界随即掀起了一阵创建数据集市的狂潮。“吃螃蟹”的结果，有大面积的企业碰壁撞墙、也有不

少企业尝到了甜头，攒了个盆满钵盈。

潮起潮落中，两派又有新的融合和纷争。油灯越拨越亮，道理越辩越明，数据仓库的理论和技术的争论中不断地得以丰富，到2000年，其理念和架构，已经完全成熟，并被业界所接受。

如蚕之蛹，数据仓库是商务智能的依托，是对海量数据进行分析的核心物理构架。它可以形象的理解为一种格式一致的多源数据存储中心，数据源可以是来自多

个不同的系统，如企业内部的财务系统、客户管理系统、人力资源系统，甚至是企业外部的系统，这些系统，即使运行的平台不同、编制的语言不同、所处的物理位置不同，但其数据可以按统一定义的格式被提取出来，再通过清洗、转换、集成，最后百川归海，加载进入数据仓库。这个提取、转换、装载的主要过程，可以通过专门的 ETL(Extraction, Transformation, Load)工具来实现，这种工具，如今已是数据仓库领域的主打产品。

ETL 工具和数据仓库理论的成熟，突破了决策支持系统的瓶颈。从此，商务智能的发展走上了顺风顺水的“快车道”，接下来，好戏连台上演。

蚕动：联机分析之惊艳

当越来越多的组织认识到联机分析的需要以及其带来的巨大收益的时候，分析型的用户就会增加。在人类的历史上，只有很少一部分运筹学专家曾经负有这样的责任：为企业开展如此高端的分析。

——埃德加·科德，关系型数据库之父，1993年⁶

数据仓库的物理结构出现以后，活跃在前沿的科学家一下子找到了自己的专属“阵地”，商务智能的下一个产业链：联机分析，如水到渠成般迅速形成。数据仓库开始散发真正的魅力。

联机分析 (Online Analytical Processing)，也称多维分析，本意是把分立的数据库“相联”，进行多维度的分析。

“维”是联机分析的核心概念，指的是人们观察事物、计算数据的特定角度，举个例子，跨国零售商沃尔玛如果要分析自己的销售量，它可以按时间序列分析、商品门类分析、地区国别分析；也可以按进货渠道分析、客户群体分析，这些不同的分析角度，就叫“维度”。

分析问题的任何角度，都可以视为一个或多个维度的交叉。例如：

沃尔玛 2011 年在美国纽约州的销售量是多少？这是个“地区”和“时间”两个维度交叉的问题。

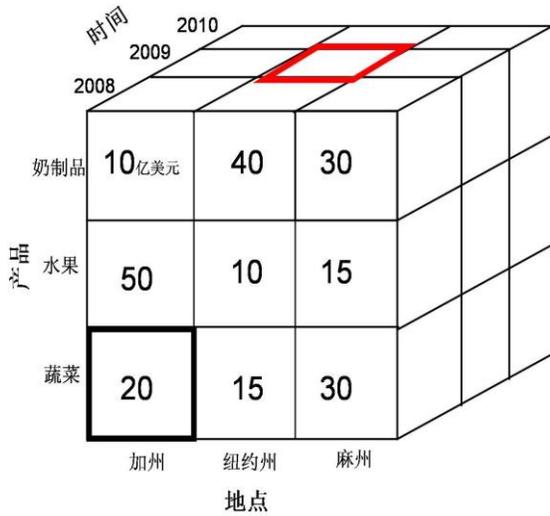
沃尔玛 2011 年在纽约州奶制品的销售量是多少？这是个“地区”、“时间”、“产品类别”三个维度交叉的问题。

沃尔玛 2011 年在纽约州进口奶制品的销售量是多少？这是个“地区”、“时间”、“产品类别”、“供货渠道”四个维度交叉的问题。

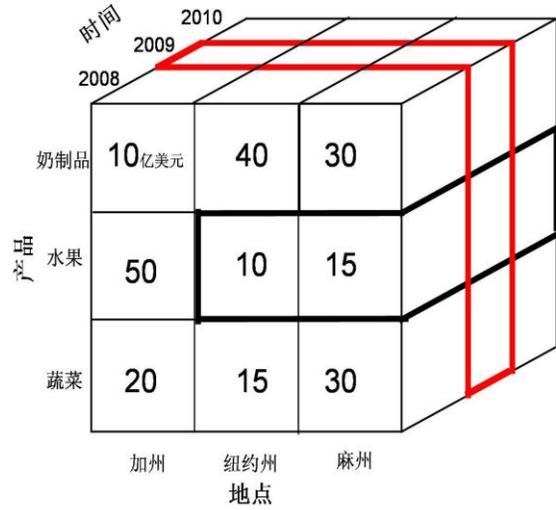
随着维度的不断增多，问题可能变得很复杂。三个维度就是三度空间，也可以想象成一个立方体。一旦超过了三个维度，人类的思维和想象能力就受到了很大的限制。

⁶ 英语原文为：As more and more organizations recognize the need and significant benefit of OLAP, the number of user analysts will increase. Historically, a small number of experts in operations research have been responsible for performing this type of sophisticated analysis for business enterprises. Providing OLAP to User-Analysts: An IT Mandate E.F. Codd, 1993

理解时间、产品和地点三个维度的交叉



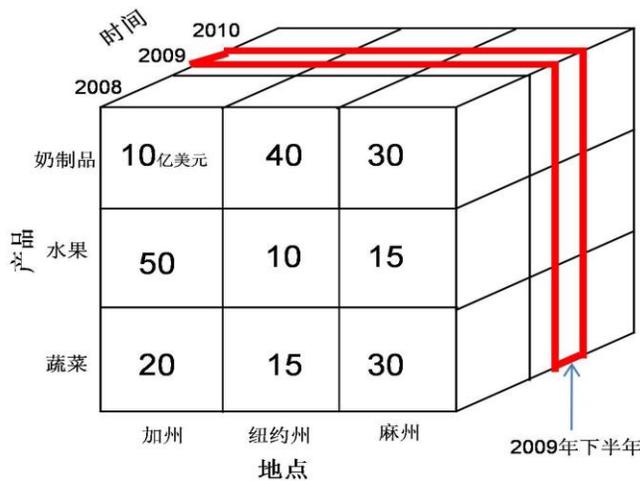
理解一个维度或两个维度的交叉



- 加粗方块（上）：表示2009年纽约州奶制品的销售量
- 加粗方块（下）：2008年加州蔬菜的销售量为20亿美元
- 横割（地点和产品两个维度的交叉）：所有年份（本图只有2008年至2010年）纽约州和麻州水果的总销量，为黑线加粗部分
- 竖切（时间一个维度）：所有州、所有产品（本图只有3个州、3种产品）在2009年的总销量，为红线加粗部分

说明：为了绘图的方便，这个例子每一个维度只取了3个值。事实上，每个维度的值都可以无限制增加，例如，年份可以增加2005、2006、2007年等，产品可以增加甜点、咖啡等，地点可以增加佛州等等。

理解下钻和上卷



- 下钻：可以按“年——半年——季度——月——日”的层级一直下钻到每一天
- 红色加粗：下钻到半年——2009年下半年3个州、3种产品的总销量
- 上卷：这个例子里只有3个州，如果有55个州，就可以上卷到全国

同时，一个维度，还可以下钻细分(drill down)，例如，就时间的维度而言，问完了一年的销量，分析人员可能会对半年的、一个季度的销量、甚至每个月、每一天的销量感兴趣；又如，就地点的维度而言，知道了纽约州的销量，分析人员可能又立刻想知道某个地区、某个城市、甚至某个小区的销量。和下钻相对应的，是上卷(roll up)，例如，从各个州的销量，加总到全美国的销量，就是一个典型的上卷。

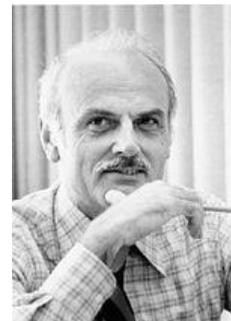
以关系型数据库为基础的运营式信息系统，事实上，也可以回答以上任何一个问题，但它回答问题的方式，是通过事先设计的报表，也就是说，根据用户指定的条件，由软件开发人员事先一一定制，通过“一对一”的查询，将结果通过报表的形式返回用户。报表，是关系数据库时代将数据转化为信息和知识最主要的手段。基于一、两个维度的分析，是简单报表；交叉的维度越多，报表就越复杂，而且不同维度的组合将产生不同的报表，对一个立足于决策的用户来说，他的需要是“动态”的：他可能问出任意维度交叉和细分的问题，但软件开发人员只能将最常见的问题定制在软件中。没有定制的问题，系统就无法回答。所以，在联机分析技术出现之前，这种静态的、固定的报表根本无法满足决策分析人员的全部需要。

早在 1960 年代，研究人员就意识到了这种“动态”决策需求和“静态”报表之间的矛盾，决策支持系统的先行者就开始探索联机分析的方法。1970 年，第一个联机分析的产品就已经问世。它通过建立一个复杂的、中介性的“数据综合引

擎”，把分布在不同系统的数据库人为联结起来，实现了联机分析。

1993年，发明关系数据库的科德再一次站到了创新的潮头。他发表了论文《信息技术的必然：给分析用户提供联机分析》(Providing OLAP to User-Analysts: An IT Mandate)，在这篇文章中，他详尽的阐述了联机分析的定义，并为如何构建联机分析提出了“黄金十二定律”。他形象的比喻说：“用关系型数据库来分析数据，是试图用“锤子”把一个“螺丝钉”硬生生的“敲”进墙，虽然最后可以勉强完成任务，但很费劲，为什么不用“螺丝刀”呢？”

科德认为，联机分析就是解决“数据分析”问题的“螺丝刀”。其惊艳之美在于用户可以根据自己的需要随时创建“万维”动态报表，也就是说，报表的定制权由后台的开发人员直接转移到了前端用户。



埃德加·科德(1923-2003)：
英国人，1948年移居美国，加盟IBM，因为提出关系型数据库，获得1981年图灵奖。1993
又率先定义了“联机分析”(OLAP)，

图片来源：维基百科

有了联机分析，用户可以自己创建自己所需要的报表。开发人员只需要预先为用户在后台构建多维的数据立方体

(Cube)。一旦多维立方体建模完成，用户可以在前端的各个维度之间自由切换、并可以从不同的维度、不同的粒度对数据进行分析，从而获得了全面的、动态的、可随时加总、细分的分析结果。在多维立方体的构建和运算方面，曾在 IBM 和微软工作过的吉姆·格雷 (Jim Gary) 多有贡献，他也于 1998 年获得了图灵奖。

因为有了数据仓库，不再需要不同数据库之间的人为“联机”，联机分析找到了真正的用武之地，如有源之水，活力四射。任何复杂的报表，都可以在鼠标的瞬间点击之内从用户的指尖弹出，数据尽在手中，如玲珑剔透的水晶体，任意横切竖割，流畅的美感令人叹为观止。

破茧：数据挖掘——智能生命的真正产生

每天早上一醒来，我就要问自己：如何才能让数据流动得更好、管理得更好、分析得更好？

——罗林·福特，沃尔玛首席信息官⁷

数据仓库、联机分析技术的发展和成熟，为商务智能奠定了框架，但真正给商务智能赋予“智能”生命的是它的下一个

数据挖掘 (Data Mining)

数据挖掘是指通过特定的计算机算法对大量的数据进行自动分析，从而揭示数据之间隐藏的关系、模式和趋势，为决策者提供新的知识。

之所以称之为“挖掘”，是比喻在海量数据中寻找知识，就象开矿凿金一样困难。

⁷ 英语原文为：Every day I wake up and ask, “How can I flow the data better, manage data better, analyze data better?” Rollin Ford, Chief Information Officer of Wal-Mart.

产业链：数据挖掘。

一开始，数据挖掘曾一度被称为“基于数据库的知识发现”(Knowledge discovery in database)。随着数据仓库的产生，“数据挖掘”的叫法开始被广泛接受。也正是因为有了数据仓库的依托，数据挖掘如虎添翼，如“巧妇”走进了“米仓”，在实业界不断创造点“数”成金的故事，其中，最为经典的例子当属啤酒和尿布。

这是一个关于零售帝国沃尔玛的故事。

沃尔玛，是全世界最大的零售商，拥有 8400 多家分店、两百多万雇员，它的人数，和美国联邦政府的雇员等量齐观，它的收入，2010 年突破了 4 千亿美元，超过了很多国家的 GDP 总值。

沃尔玛拥有世界上数一数二的数据仓库，是最早应用数据挖掘技术的企业之一，也是数据挖掘技术的集大成者。在一次例行的数据分析之后，研究人员突然发现：跟尿布一起搭配购买最多的商品竟是啤酒！尿布和啤酒，听起来风马牛不相及，但这是对历史数据进行挖掘的结果，反映的是数据层面的规律。这种关系令人费解，这是一个真正的知识吗？

经过跟踪调查，研究人员终于发现事出有因：一些年轻的爸爸经常要到超市去购买婴儿尿布，有 30% 到 40% 的新爸爸会顺便买点啤酒犒劳自己。

沃尔玛随后对啤酒和尿布进行了捆绑销售，不出意料之外，销售量双双增加。

沃尔玛还有很多利用数据挖掘、扩大销售的故事。2004 年，分析人员发现，

每次飓风来临，一种袋装小食品“Pop-Tarts”的销售量都会明显上升。手电筒、电池、水，这些商品的销量会随着飓风的到来而上升，很容易理解，但 Pop-Tarts 的上升是不是必然的呢？

研究人员后来发现，这也是一个有用的规律：Pop-Tarts 的销量上升，一是因为美国人喜欢甜食，二是因为它在停电时吃起来非常方便。此后，飓风来袭之前，沃尔玛也会提高 Pop-Tarts 的仓储量，以防脱销，并把它和水捆绑销售。

如果没有数据挖掘，Pop-Tarts 和飓风的微妙关系就难以被发现。

1989 年，可谓数据挖掘技术兴起的元年。这一年，图灵奖的主办单位计算机协会 (Association of Computing Machinery) 下属的知识发现和数据挖掘小组 (SIGKDD) 举办了第一届数据挖掘的学术年会、出版了专门期刊。此后，数据挖掘被一直热捧，其发展如火如荼，甚至成为一门独立的课目走进了大学课堂，在美国的不少大学，还先后设立了专门的数据挖掘硕士学位。也正是 1989 年，高德纳咨询公司 (Gartner Group) 的德斯纳 (Howard Dresner) 在商业界为“商务智能”给出了一个正式的定义：

“商务智能 (Business Intelligence)，指的是一系列以事实为支持、辅助商业决策的技术和方法。”

这个定义，强调了商务智能是一系列技术的集合，获得了业界的广泛认同。

商务智能在 1989 年完全破茧而出，并不是历史的巧合，而是因为数据挖掘这种新技术的出现，使商务智能真正具备了“智能”的内涵，也标志着商务智能完整产业链的形成。

如果说联机分析是对数据的一种透视性的探测，数据挖掘则是对数据进行挖山凿矿式的开采。它的主要目的，一是要发现潜藏在数据表面之下的历史规律，二是对未来进行预测，前者称为描述性分析，后者称为预测性分析。沃尔玛发现的啤酒和尿布的销售关联性就是一种典型的描述性分析；考察所有历史数据，以特定的算法对下个月啤酒的销售量进行估计以确定进货量，则是一种预测性分析。

数据挖掘的两个侧重点



数据挖掘把数据分析的范围从“已知”扩大到了“未知”，从“过去”推向了“将来”，是商务智能真正的生命力和“灵魂”所在。它的发展和成熟，最终推动了商务智能在各行各业的广泛应用。

近过十多年的发展，数据挖掘的范围正在不断扩大。传统的数据挖掘是指在结构化的数据当中发现潜在的关系和规律，但随着商业竞争的白热化，更加高端的数据挖掘也开始初现端倪。例如，通过网络

结构化数据和非结构化数据

按结构，数据可以划分为两类。

结构化的数据是指存储在数据库当中，有统一结构和格式的数据，这种数据，比较容易分析和处理；非结构化的数据是指无法用数字或统一的结构来表示的信息，包括各种文档、图像、音频和视频等等。这种信息，没有统一的大小和格式，给分析和挖掘带来了更大的挑战。

从结构化数据到非结构化数据的推进，也代表着可供挖掘的数据在大幅增加。

留言挖掘顾客的意见。顾客在博客、论坛、社交网站和微博上用文字记录的消费体验、对商品和服务发表的看法和评价，是一种非结构化的数据。如何把在散布在网络上的这些资源整合起来，并从中自动挖掘有价值的信息和知识，正是当前数据挖掘面临的巨大挑战之一。数据仓库之父比尔·恩门近年来就在这个领域多有建树。

化蝶：数据可视化的华丽上演

图形是解决逻辑问题的视觉方法。

——杰克·伯廷(1918-2010)，法国统计学家，1977年⁸

随着数据仓库、联机分析和数据挖掘技术的不断完善，业界曾一度认为，商业智能系统已经功德圆满，很好的完成了智能分析的使命，因此早期商务智能的产业链条只含有这三块。

但技术无止境。进入 21 世纪之后，风生水起，新的技术浪潮又使商务智能的产业链条向前延伸了一大步：数据可视化。

⁸ 英语原文为：Graphic is the visual means of resolving logical problems. Graphics and graphic information processing P.16, Jacques Bertins. 1977

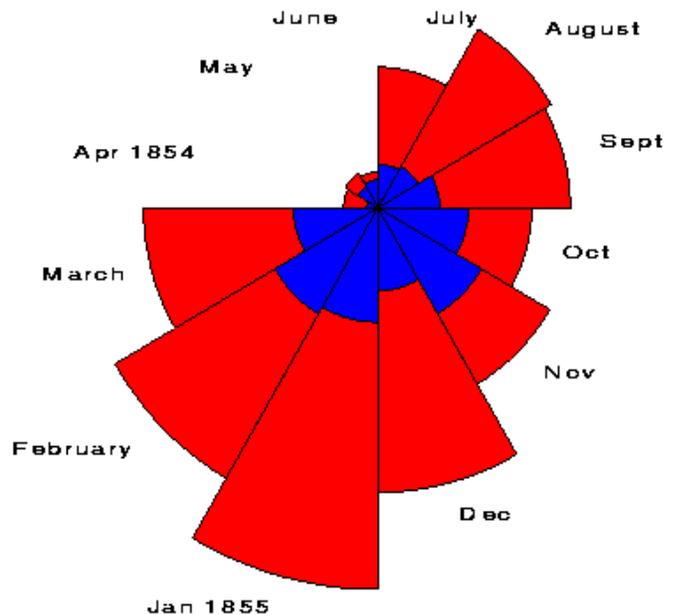
数据可视化 (Data Visualization)

数据可视化是指以图形、图像、地图、动画等更为生动、易为理解的方式来展现数据的大小、诠释数据之间的关系和发展的趋势，以期更好地理解、使用数据分析的结果。

数据可视化也是几代统计学家上百年的梦想。

故事可以追溯到十九世纪中期。1855年，英国和俄罗斯之间爆发了克里米亚战争。这场战争共死亡 50 多万人，异常惨烈。佛罗伦斯·南丁格尔 (Florence Nightingale) 是英国的一名战地护士、也是一名自学成才的统计专家，她在考察了英国士兵的死亡情况之后，发现由于医疗卫生条件恶劣导致的死亡人数，大大超出了一线的阵亡人数。南丁格尔将她的统计

1854 年 4 月至 1855 年 3 月，英国军队士兵的死亡原因



图形说明：每月的死亡人数以 30 度的扇形面积表示，内环蓝色代表因战斗死亡的人数，外环红色代表非战斗死亡的人数，也就是可以预防、改善的医疗卫生原因

图片来源：SAS 公司

结果制成了一个图表，该图表清晰的反映了“战斗死亡”和“非战斗死亡”两种原因死亡人数的悬殊对比，强烈的视觉效果引起了英国社会的极大反响，最后直接促成了英国政府出台建立野战医院的决定。

南丁格尔后来被誉为现代护理业之母，她的这份图形，是历史上第一份“极区图”(Polar area diagram)，也是统计学家对利用图形来展示数据进行的早期探索。

一份图表催生了一座医院、改变了一个制度。南丁格尔的贡献，充分的证明了数据可视化的价值，特别是在公共领域的价值。官僚们麻木的神经尤其需要强烈的

视觉效果来冲击、来刺激。生理学也证明，人的大脑皮层当中，有 40% 是视觉反应区，人类的神经系统天生就对图像化的信息最为敏感。通过图象，信息的表达和传递将更加直观、快捷、有效。

更重要的原因在于：人的创造力不仅仅取决于逻辑思维，还取决于形象思维。数据可视化的技术，可以通过图象在逻辑思维的基础上进一步激发人的形象思维和空间想象能力，吸引、帮助用户洞察数据之间隐藏的关系和规律。

到了二十世纪 70 年代，由于计算机技术的兴起，美国有一批远见卓实的学者

第一组数据

	1	2	3	4	5	6	7	8	9	10	11
X1	10.0	8.0	13.0	9.0	11.0	14.0	6.0	4.0	12.0	7.0	5.0
Y1	8.04	6.95	7.58	8.81	8.33	9.96	7.24	4.26	10.84	4.82	5.68

第二组数据

	1	2	3	4	5	6	7	8	9	10	11
X2	10.0	8.0	13.0	9.0	11.0	14.0	6.0	4.0	12.0	7.0	5.0
Y2	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74

第三组数据

	1	2	3	4	5	6	7	8	9	10	11
X3	10.0	8.0	13.0	9.0	11.0	14.0	6.0	4.0	12.0	7.0	5.0
Y3	7.46	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73

第四组数据

	1	2	3	4	5	6	7	8	9	10	11
X4	8.0	8.0	8.0	8.0	8.0	8.0	8.0	19.0	8.0	8.0	8.0
Y4	6.58	5.76	7.71	8.84	8.47	7.04	5.25	12.50	5.56	7.91	6.89

都看到了这个领域巨大的潜力。耶鲁大学的统计学教授弗朗西斯·安斯科姆 (F.J. Anscombe) 就是其中的先驱人物。1973 年, 他发表论文《统计分析中的图形》, 专门阐述了图形在统计研究当中不可替代的作用, 他认为:

“未来的计算机不仅要能计算, 还要能将计算结果转变为直观的图形。我们应该研究这两种结果, 因为每一种都有助于我们理解问题。”⁹

在这篇文章中, 安斯科姆教授提出“安斯科姆四重奏”, 通过这个例子, 他强调, 在研究数据、使用数据的时候, 图形和计算同等重要, 有的时候, 图形甚至是解决逻辑问题更为直接有效的方法。这个著名的“四重奏”, 是 4 组同时呈现在你面前的数据 (X, Y)。

当你粗略浏览这 4 组数据之后, 你会感觉其数值大多在 5 到 11 之间, 比较杂乱。稍做对比, 你会发现:

$$X_1 = X_2 = X_3$$

X₄ 的值, 除一个之外, 全部都等于 8

$$Y_1 \neq Y_2 \neq Y_3 \neq Y_4$$

如果再进行简单的统计学计算, 很容易得到以下结果:

X₁、X₂、X₃ 和 X₄ 的平均值都等于 9, 其方差等于 10

Y₁、Y₂、Y₃ 和 Y₄ 的平均值都等于 7.50, 其方差等于 3.75

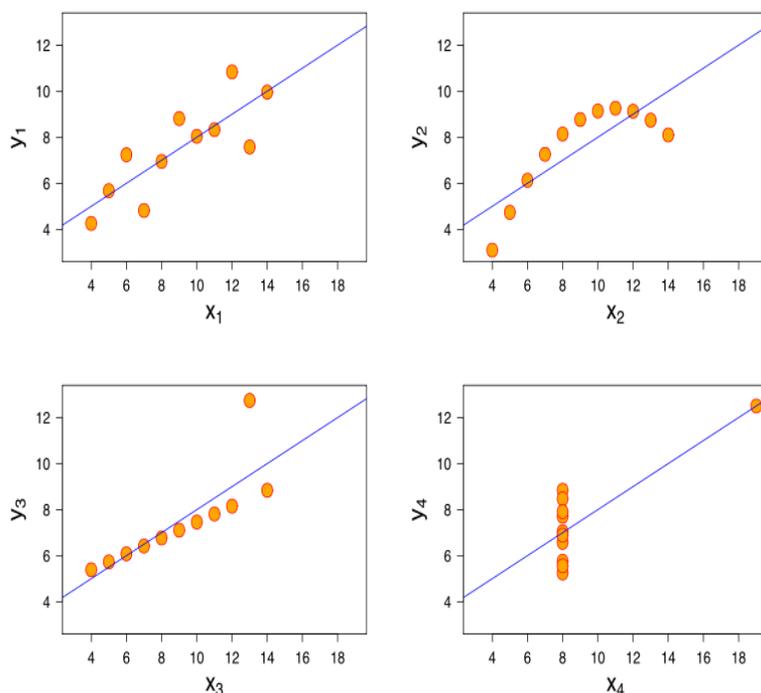
4 组数据都符合线性回归: $y = 3 + 0.5x$

也就是说:

4 组数据当中, X 和 Y 之间的关系都是相同的, 个别数据的偏离, 可以视为随机产生的干扰。

但当我们用散点图把它们在坐标中标出来之后, 面对图形, 就会立刻发现, 统计学“欺骗”了我们:

图形是解决逻辑问题的视觉方法: 安斯科姆四重奏的真实分布



⁹ 英语原文为: A computer should make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding. "Graphs in Statistical Analysis, F.J. Anscombe, 1973

4 组数据当中，仅仅只有第一组数据严格符合利用统计学作出的线性回归结论。

第二组数据存在某种规律，但显然不是线性的；

第三组数据大部分符合线性回归的模型，但有一对数据明显异常，它是第三对数据(13.0, 12.74)；

第四组数据则呈垂直分布，其之所以貌似符合线性回归的分布，是因为其第 8 对数据(19.0, 12.50)在其中起了很大的扭曲作用。

1983 年，耶鲁大学政治学教授爱德华·塔夫特(Edward Tufte)成了数据可视化这门学科的奠基人。塔夫特系统的考证了人类用“图形”表达“数据”和“思想”的渊源，整理了种种历史古籍中的图形瑰宝，并结合计算机的发展给统计领域带来的革命，出版了《定量信息的视觉展示》(The Visual Display of Quantitative Information)一书。这本书后来被公认为“数据可视化”作为一门学科的开山之作。

这本书的出版，也有一段曲折。因为塔夫特整理了从古到今很多优秀的图表，他坚持要在新书中使用高质量、高精度的彩色插图。几乎所有的出版商都认为这是赔本买卖，没人愿意出版。塔夫特最后无计可出，用自己的房子做了抵押，自费出版了这本书。

结果当然令出版商大跌眼镜：这本书最终获得了很大的商业成功，塔夫特教授也由“政治学”专家成功转型为“信息学”专家，近十多年来，他又先后出版了《视

觉解释》(Visual Explanations)、《美丽的证据》(Beautiful Evidence)等几本重量级的著作，本本都洛阳纸贵，造成了不小的轰动。他本人也成了数据可视化领域当仁不让的掌门人。2010 年 3 月，奥巴马刚刚上任，就任命塔夫特为顾问，要求他运用“数据可视化”的技术推进联邦政府的行政透明。

塔夫特教授强调数据可视化的关键在于“设计”，他认为：

“信息过载这回事并不存在，问题出在糟糕的设计，如果你用来表达数据的图形让人感觉杂乱不解，那就要修改你的设计。”¹⁰

进入 21 世纪之后，大数据的爆炸使人们更加需要展示数据、理解数据、演绎数据的工具。这种需求，刺激了数据可视化专业市场的形成，其产品迅速增多，现在的市场可谓绚丽多彩、百花齐放。从最早的点线图、直方图、饼图、网状图等简单图表，发展到以监控商务绩效为主的仪表盘(dashboard)、记分板(scorecard)，到以交互式的三维地图、动态模拟、动画技术等等更加直觉化、趣味化的表现方法，曾经冰冷坚硬、枯燥乏味的数据开始“动”了起来、“舞”了起来，变得“性感”！

数据可视化把美学的原素带进了商务智能，给它锦上添花。一幅好的数据图像不仅能有效地传达数据背后的知识和思想，而且华美精致，如一只只振动翅膀的彩蝶，

¹⁰ 英语原文为：There is no such thing as information overload, just bad design. If something is cluttered and/or confusing, fix your design. Edward Tufte

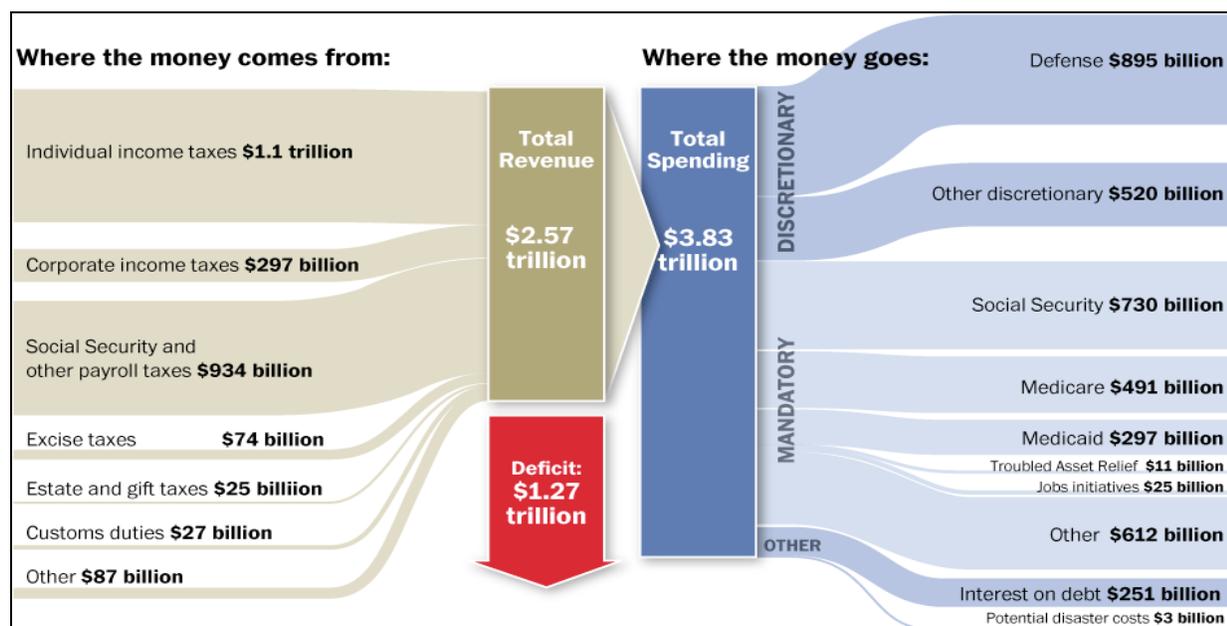
刺激视觉神经、调动美学意识，令人过目不忘，留下栩栩如生的印象。

2010年2月，奥巴马宣布了联邦政府新的年度预算。《华盛顿邮报》立即对这份新鲜出炉的预算进行了分析报道，它利用数据可视化的技术，抓住了读者的眼球。图形以线条的粗细表明各项收支金额的大小，左边是收，右边是支，中间的红色部分是赤字缺口，形象贴切，奥巴马收了多少钱，要办哪些事，各项收入支出的轻重大小，一目了然。

个以时间为横座标，以大奖得主的国籍为纵座标的散点图。不难看出，1940年以前，德国是世界科学和文化的中心，但二次世界大战之后，这个中心毫无疑问转移到了美国。还能看到，美国人的崛起首先在物理领域，其次是医学领域，再次是经济学领域。1970年以前，没有美国人获得过经济学奖，但1970年后，美国人几乎囊括了全部的经济学奖。

作为一个新兴的行业，数据可视化的发展潜力不容小觑。2010年，谷歌的首

奥巴马政府 2010 年度预算开支的可视化展示



图片来源：《华盛顿邮报》，2010年2月1日¹¹

每年的10月，诺贝尔奖花落谁家是全世界的热门话题。2011年10月，福布斯(Forbes)对一百多年来各项诺贝尔奖的获得情况做了一个可视化的展示。这是一

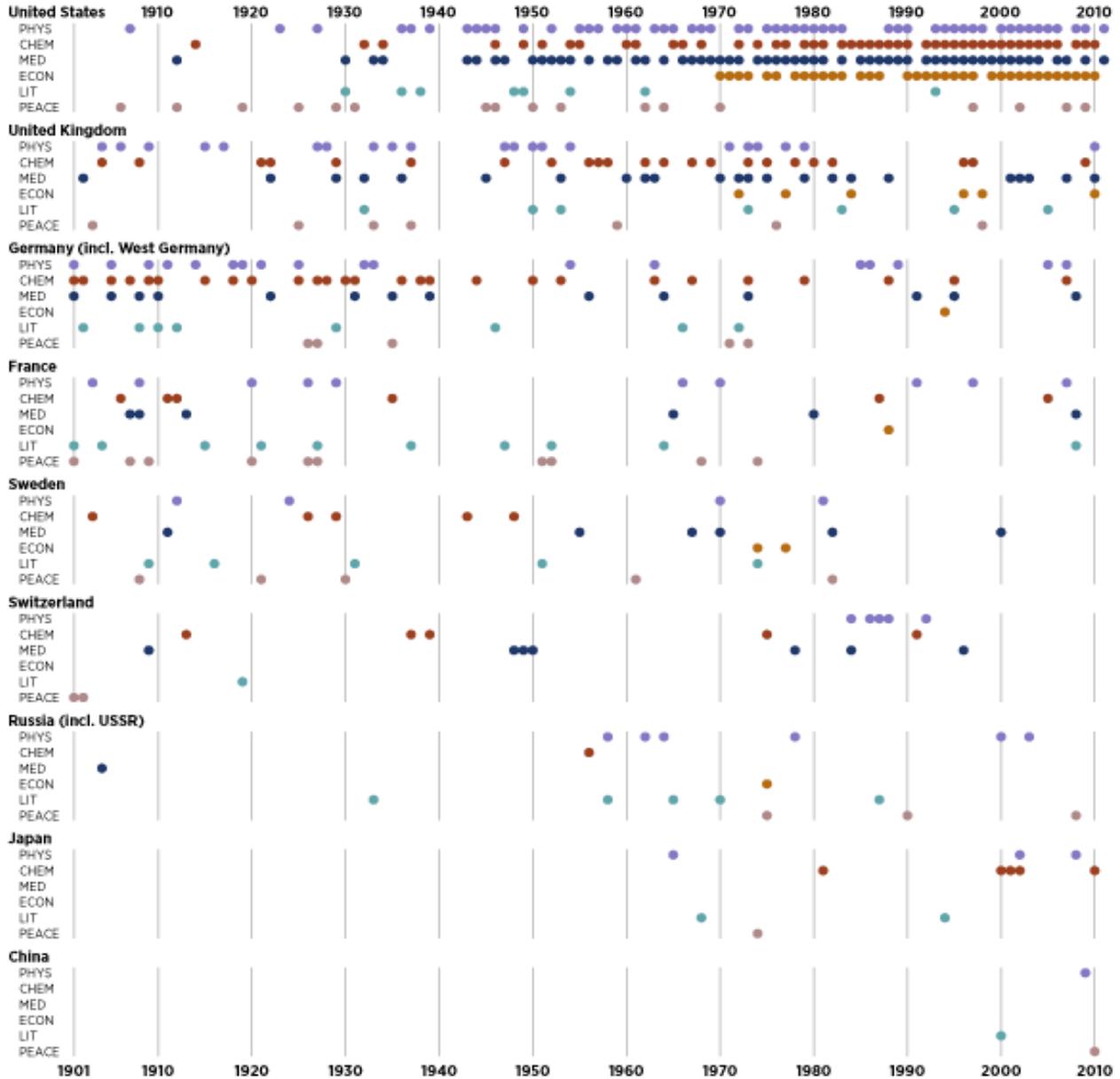
席经济学家范里安(Hal Varian)就一直在多种场合强调，下一个十年，将出现一类新的专业人才：数据科学家。其中一种，正是数据可视化工程师，这种人才既懂得数据分析、又精通构图的艺术，集故事讲述和艺术家的特质于一身，将是我们大数据时代的导航员。

¹¹ Taking apart the federal budget, Data Source: White House Office of Management and Budget; GRAPHIC: Wilson Andrews, Jacqueline Kazil, Laura Stanton, Karen Yourish - The Washington Post, Feb1, 2010

百年诺贝尔的分布

Nobel Laureates

By Country and Prize



SOURCE: NOBELPRIZE.ORG

Laureates are shown in the country that hosted their research at the time of award
Last updated on October 4, 2011

Forbes blogs.forbes.com/jonbruner

注：获奖人的国籍，有时候难以甄别，例如：历史上曾出现以难民身份的获奖者。又比如，2009年，高锟获物理奖时，为英国国籍，但持有香港身份证、并居住在香港，作者将他归入中国。作者还指出：在美国的314个获得者中，有102位（32%）是在美国本土之外出生的，这其中德裔15名、加拿大裔12名、英国裔10名、俄裔6名、华裔6名；而德国的65名获奖者中，只有11名出生在海外；日本的获奖者，则全是在本土出生的。

图片来源：《福布斯》，2011年10月5日¹²

¹² American Leadership in Science, Measured in Nobel Prizes (Infographic), Jon Bruner, Forbes, Oct 5th, 2010

数据可视化的这种“导航”作用也极大的推动了商务智能的大众化。通过把复杂的数据转化为直观的图形，并呈现给最普通的用户，商务智能已经不再是少部分高级分析人员的专利，而是贴近大众生活的、浅显易懂、人皆可用的工具和手段。

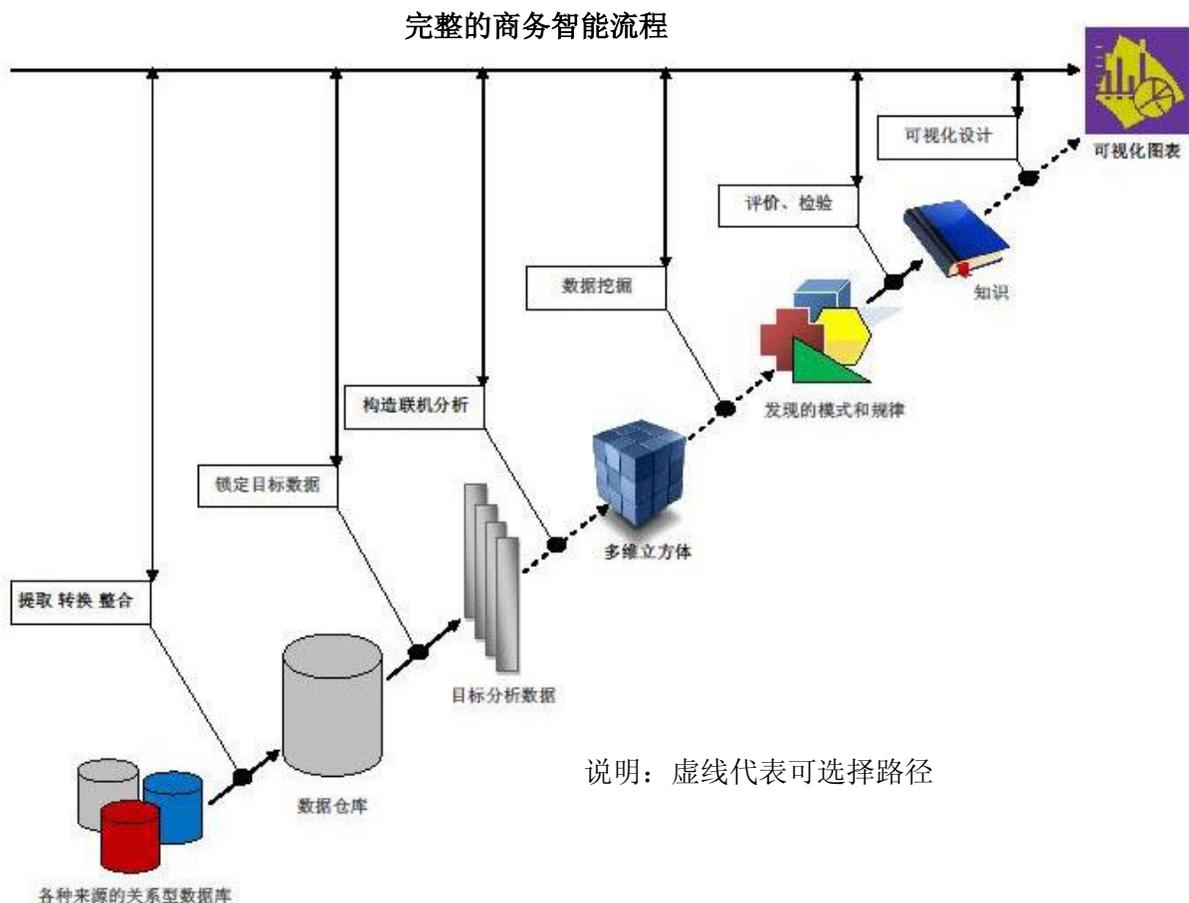
美国联邦政府也意识到“数据可视化”的战略意义。2004年，联邦政府在国土安全部成立了国家可视化分析中心(National Visualization and Analytics Center)，专门推动该项技术在政府部门的应用，特别是在情报分析领域的应用。

可视化技术的出现，使商务智能的产业链形成了一个从数据整合、分析、挖掘、到展示的完整闭环。它的起点是多个独立

的关系型数据库，经过数据整合之后形成统一的、多源的数据仓库，再根据用户的需要，重新取出若干数据子集，或构造多维立方体(Cube)、进行联机分析，或进行数据挖掘，发现潜藏规律和趋势，如果挖掘的结果经得起现实的检验，那就形成了新的知识，这种知识还可以通过数据可视化来表达、展示和传递。

商务智能的这四个产业链，每一块都相当的复杂，彼此的独立性也很强。一个好的商务智能产品，并不见得一定要面面俱到，时下不少公司，都专注在一个链条上、大做文章。

商务智能的历史，是一个渐进的、复杂的演进过程，至今为止，它的内涵和外



延，还处于动态的发展之中。它的各个产业链条，还有不断丰富扩大的趋势。特别是作为其“智能灵魂”的数据挖掘技术，潜力非常巨大，可以预见，将对人类社会的发展产生深远的影响。

大数据时代的竞争，将是知识生产率的竞争。以发现新知识为使命的商务智能，无疑是这个时代的最为瞩目的竞争利器。

作者简介

涂子沛：知名专栏作家、信息管理专家，华南理工大学公共政策研究院副教授，中国旅美科协匹兹堡分会主席。涂子沛先后毕业于华中科技大学、中山大学和卡内基梅隆大学，现担任美国一软件公司数据中心的主任。赴美留学之前，曾在省、市、县几级政府的不同部门磨砺10年，做过职业程序员，担任过公安边防巡逻艇的指挥官，也从事过政府统计工作。目前为《南方都市报》、《IT经理世界》、艾瑞网等国内多个报刊网站撰写专栏。著有《大数据：正在到来的数据革命》。

浅谈提升金融 IT 技术 对加快中国国际金融中心建设及人民币国际化的影响

邹云峰

美国赢透证券集团 (InteractiveBrokers LLC, Greenwich, CT, USA 06830)

摘要

本文讨论金融产品交易规模(交易量)对国际金融中心地位的影响,金融 IT 技术对美国金融业的提升作用,及促进中国金融业发展的一些建议。

国际金融中心与金融产品交易量

正如在传统行业的竞争中,同等技术条件下企业规模是竞争的关键,金融行业的发展竞争也不例外。国际金融中心的地位大体可用其金融产品交易品种、交易量及市场方向影响力来衡量。具体来说就是在市场缺乏方向时,交易中心完成主要的交易量。在市场出现方向性时,中心引导并决定方向。

进一步观察,市场交易量才是关键。理由是如果交易同样产品,交易量大的市场往往决定市场的最终走向。国家经济、政治、军事、文化、媒体等综合软硬实力因素对市场的方向影响不言而喻。这些实力的增长非一日之功,不在本文讨论范围。中国国际金融市场的发展也不可能等到具备哪些条件再加速展开。历史事实也表明,成为经济大国并不能自动成为真正的国际金融中心。日本成为世界经济第二大国的数十年间,其国际金融市场的影响力远不相称其经济实力。经济规模小得多的新加坡及香港却有不俗的影响力。可见促进发

展的政策及正确的方向选择也非常重要。如果能以促进中国金融市场交易量为中间目标,建立起一到两个全球金融交易中心,中国成为金融大国的目标就水到渠成了。在中国还缺乏超级大国的软硬实力以前,集中精力提高交易规模似乎是提升中国国际金融中心地位的合理选择。

如果我们假设提高交易量是实现我们最终目标的必要途径,那么我们的思路及要做的主要工作就是找到并促进有利于提高交易量的各项措施。建立和完善多种金融产品交易市场,降低交易成本(税费)及门槛,简化交易产品,对全球投资者放开中国金融市场,扩大参与度等都是容易想到的措施。从技术上讲,我们必须大力促进和提升 IT 技术在金融产品交易中的作用。

我们应该建立和扩大金融产品品种及市场。提供多种金融交易产品无疑促进提高交易总量。因此我们需要提供期货(Future)、期权(Options)、差价合约(CFD)、交易所交易基金(ETF)、外汇(Forex)、国债、企业债等交易产品。这些都是在国际市场大受欢迎的产品。我们不提供交易市场,如何能在交易量上超过别人?我们的市场应提供国际通行的交易产品。成功的交易市场应能提供全世界主要产品的交易,应当支持多种货币结算的

交易。如果国际投资者能通过投资澳元，加元，石油等产品来分享中国经济的发展成果，中国的投资者更应当拥有好的交易产品投资世界市场。

我们应该提供多种可相互对冲的交易产品。合理的市场有涨有跌。能卖空产品是合理市场的必要条件。如果不能卖空，企业和投资者如何对冲市场下跌风险？泡沫化的企业如何能有效地被市场挤压？此外，我们的金融服务业需要持续的交易量才能持续发展。目前的状况有如传统的农业产品靠天收。牛市时金融业日进斗金，熊市时市场门可落雀。这妨碍了金融产业的持续发展。市场提供的是金融工具及服务犹如田径场提供跑道，胜负/盈亏本应与市场及服务提供者无关。

我们应该降低交易税费，成本越低则交易量越高。目前的税费太高妨碍了市场规模的增长。具有全球竞争力的低税费是进一步发展国际金融中心的必要条件。

我们应该尽量简化产品交易标准，降低交易者准入门槛。让更多的交易者参与才能造就流动性更高，更稳定，更安全的市场，从而带来更高的交易量。

我们应该向各国投资者，甚至投机者开放市场。在我们的市场交易和管理他们，远比在国外交易场被他们蒙在鼓里强。另外，这是一个全球化的时代，我们的竞争对手和市场遍布世界。关闭市场不可能关闭世界市场对中国的影 响，隔离只能扩大我们与世界先进水平的差距。

金融 IT 技术在美国的现状 & 新进展

我们必须大力促进 IT 在金融产品交易中的作用。当今世界，各种金融产品 & 市场已高度关联和衍生化，没有 IT 技术的全面支持与计算机程序交易的广泛运用，提供多种金融交易产品及提高交易量就不可能实现。没有低的交易成本及标准简化的产品，缺乏多种相互关联的金融产品，计算机程序交易就难以起步及发展。IT 技术的广泛应用将提高金融业的生产效率，从而推进交易费用的降低。交易量的扩大反过来为国家贡献更多的税费。降低税费促发展，预取先与才是正确的策略。

金融 IT 技术在美国各大交易所，券商，机构投资者，和对冲基金中得到广泛运用。传统的由交易员配对的交易场所已逐步被电子交易系统所取代。越来越多的成交量来自于计算机程序交易。纽约证交所发布的程序交易量经常占到总量的 40% 以上。这还不包括由投资者完成的未标示为程序交易的成交量。做市商大多使用程序交易为市场提供流动性，经纪商依赖网络 IT 技术建立起连接世界各大交易场所和产品的交易平台，对冲基金大量使用程序交易进行统计套利，股指及 ETF 套利，组合套利和各种量化交易活动。金融 IT 技术是这一切活动的基石。

IT 技术已广泛应用于金融交易的各个环节。从下单、到多市价格寻优 (smart routing)、算法交易、清算交割、风险控制、税务服务、实时新闻、客户支持培训及公司和市场分析等全面采用计算机程序，使交易和服务成本大幅降低。

对许多金融机构来说，IT 相关的成本占据了相当大的比重。各金融机构不惜重金聘请优秀的 IT 研发人员，金融 IT 工程师成为 IT 同行中薪水最高的一族。金融机构财大气粗可采用最新的计算技术及设备。这一切被媒体称为金融技术的“军备竞赛”。为提高交易速度，交易电脑已被放置于交易所计算机的旁边 (colocation)。

最近几年，IT 及网络技术的全面进步使美国金融业出现了几个值得注意的新进展。

- 大额交易下单广泛运用程序交易，划整为零，以减小对市场的冲击和影响，同时可防止泄露投资意图。暗盘 (Dark Pool) 交易应运而生。最近美国证监会已批准纽约证交所 (NYSE) 可直接提供暗盘交易。当今的趋势是程序交易使得平均单位成交量不断下降。

- IT 技术的普及和使用成本的不断下降拉低了竞争平台。许多小型投资机构开发自己的程序交易与大机构分享市场。高频交易 (HFT) 的成功就是典型的例子。

- 大交易所开展跨国整合，提供多种交易产品，加大规模以提高市场竞争力。例如纽约证交所 (NYSE) 与德国 Eurex 交易所合并成为 NYSE Euronext，纳斯达克 (NASDAQ) 与瑞典/芬兰 OMX 合并成为 NASDAQ OMX 集团。

- 出现了提供单一账户可交易全球交易场所及产品的超级证券经纪商。这种券商提供的交易网络相当于一个跨国，跨产品的交易中心。这里要提到的是美国赢透证券集团 (InteractiveBrokers Group)。该公司提供一个账号可交易全球 19 个国家

100 多个交易市场的网络交易平台。交易产品包括股票、期货、期权、差价合约、国债、企业债、外汇等多种产品。

- 交易所交易基金 ETF 大量上市，为中小投资者提供了多种灵活的投资工具及杠杆。

- 自营交易及经纪业务共享交易系统，降低开发及维护成本以提高 IT 资源的使用效率。

通过以上几点可以看到，美欧金融公司已视全球金融市场为统一市场。正在以快速的步伐展开新的全球布局。提供多产品，跨区域和时区服务，一场以 IT 技术引领的新的金融革新正在快速展开。金融业实际上已转型为高技术产业。金融 IT 的整体发展水平在相当程度上代表金融业的水平。

促进中国金融业发展的一些建议

目前，中国金融行业与美欧先进国家比较，从理念、经验、知识、人才及技术上的差距还十分明显。然而，中国成为未来世界金融中心的潜力和条件十分优越。中国强大实体经济规模，众多投资者，以及高储蓄率，高财富积累率为金融业创造了天然雄厚的基础。在金融领域，中国人是最敢于冒险的投资者。乐观地讲，中国成为世界金融中心只是时间问题。为加速这一天的到来，国家应当从政策，技术，和人才等方面早做筹划。因此本文提出如下几点建议：

- 中国金融业的发展重点应以围绕提高金融交易总量和规模为近期目标。制定和实施有利于提高交易量的政策。不断降低交易税费，提供丰富的各种相互关联可

对冲的金融交易产品，延长金融产品的交易时间实为首要措施。

- 必须十分重视提升金融 IT 技术。正如本文所述，金融产业已成为高技术产业，金融 IT 技术是现代金融业的基石。中国金融业长期以来与外界隔离，技术已成为今后发展的短板。另一方面，金融 IT 技术的提高并非一朝一夕可以取得，更不可能从外部引进。所以必须从战略高度认识这一问题的紧迫性。

- 从政策上鼓励发展计算机程序交易。没有大量的程序交易，市场没有流动性和容量，市场就难以生存。在美国，一些交易市场为提高流动性，对增加流动性的交易方，交易所不仅不收交易费，反而付费给交易方。我们也应当采取同样的策略。

- 金融 IT 技术的提升需要一大批人多年的努力。国家应当马上有组织地从国外引进一批有丰富经验的金融 IT 专才，建立专门的联系通道。值得提出的是中国在国外已有非常好的金融 IT 人才储备。很多早期的留学国外的中国人现工作于这一领域。已具有丰富的经验。现今美欧国门大开，一大批新的留学生所学专业为金融、财务及管理专业，形成了很好的人才梯队。几年后大批青年将回国工作。没有金融业的大发展，他们中的很多人将学非所用，对国家和他们的家庭将是极大浪费。所以国家也因从这一角度未雨绸缪。

另一方面，人民币国际化与发展金融市场的关系密切。本文认为，没有强大的金融市场，人民币的国际化影响力也无从谈起。我们应该想到当外国投资者拥有大量人民币后何处配置和投资这些人民币等相关问题，必须考虑到他们需要一个规模大、安全、透明、涨跌都有去向的金融交易市场。

在金融全球化的今天，金融市场是一个世界性的关联市场。金融产品也高度关联和衍生化。封闭自己并不能隔绝世界市场对中国的冲击和影响。中国的金融安全必须靠向外发展成为世界金融中心来取得。这是一个以国家甚至以国家集团和文化为单位的竞争。金融 IT 技术是这场较量的基础和兵器。我们应当从战略高度上加以重视和提高，必须从今天就抓紧。

作者简介

Yunfeng Zou (邹云峰) is a senior manager and software engineer at InteractiveBrokers Group LLC since 2001. His expertise is on trading system research and development. He received his Ph.D. from State University of New York at Stony Brook and M.S. in Shanghai Jiaotong University. You can reach him at yunfengzou@hotmail.com.

预测市场：一种新兴的预测方法

陈伟运¹，李昕²，曾大军¹

¹Institute of Automation, Chinese Academy of Sciences, Beijing, China

{weiyun.chen, dajun.zeng}@ia.ac.cn

²Department of Information Systems, City University of Hong Kong, Hong Kong, China

xin.li@cityu.edu.hk

摘要

本文详细介绍了预测市场的基本概念、预测市场设计中需要考虑的几个主要方面、预测市场的理论基础以及与其它传统预测手段相比所具有的优缺点。在此基础上，我们还简单介绍了我们使用固定赔率机制作为预测市场交易机制的初衷和初步结果。我们相信，在网络和通信技术日益普及的今天，在线预测市场这种崭新的群体智慧方法将发挥越来越重要的作用。

1. 引言

预测未来不确定事件的结果对我们的生产和生活具有重要的意义。例如近期天气和交通状况、房价的涨跌和各行业发展趋势等的预测莫不与我们每个人的日常生活息息相关。不同的预测对象和预测任务所需要的信息量不同，有些简单的预测我们一般人就可以胜任，但是对于比较复杂的事件，我们个人很难具有相对完备的信息。例如火山地震等自然灾害、社会态势和大规模群体事件、突发传染性疾病的传播^[1-4]、世界区域政治形势等等的预测都涉及到诸多不确定的已知或者未知因素，特别是近二十年来互联网和移动通信技术的不断发展和普及所带来的革命性的影响、虚拟空间和实际物理空间的频繁互动所导

致的高度的社会动态性，更是对我们的预测和决策手段提出了新的要求和挑战。这些变化迫切要求我们能及时地就事态的发展做出准确的预测^[2,5-7]。然而，这些高度复杂或者动态的事件往往涉及到多方面的大量不确定因素，所有可利用的相关信息往往分布在群体之中，单靠个别专家很难迅速有效地收集到所有的事件相关信息^[8,9]。我们需要寻找更高效准确的群体信息聚集机制以探求分散在群体中的“智慧”来辅助决策。

许多新颖有效的挖掘和利用群体智慧的方法已经成功地被应用于一些相关领域^[10-12]。例如谷歌搜索通过利用网页链接之间的规律开发出强大的搜索引擎，取代了传统的雅虎百科目录；维基百科通过提供简单的网络编辑功能使其成为世界上首屈一指的“人民百科全书”；百度知道、谷歌地图等等通过提供简单的在线活动形式，极大地丰富和方便了我们的日常生活。这些看似简单的机制在互联网和通信技术飞速发展的背景下完成了我们以前无法想象的任务。

长久以来，人们就知道市场能及时有效地聚集、反映分散在群体之中的信息。简单的市场交换原理实现着任何模型或者计划无法替代的复杂的资源配置功能，市

场价格就仿佛一只“看不见的手”，反映并调控着整个群体的预期，各种与交易对象相关信息都能通过群体之间的交易活动不断地反映在市场价格中。由于市场价格能通过群体之间的市场交易反映群体对未来的预期，自由市场这一古老的机制正在被应用于未来事件的预测。一种新型的在线虚拟交易市场，即预测市场 (Prediction Markets, PMs) 也受到越来越多的关注。

2. 预测市场及其沿革

预测市场是一种专门用来预测未来不确定事件的投机市场，预测市场价格可直接解释为事件发生的概率。该市场中的合约可以在指定的时间以预先明确规定的“价格”兑现一定的现金或者积分等，该价格与事件的实际结果相关。由于预测市场中的交易者群体对事件可能结果持不同的预期，因此通过市场交易产生的价格能反映分散在群体中的与合约事件相关的信息^[13]。预测市场的价格与未来事件结果密切相关，能反映分散在群体之中的事件相关信息，所以预测市场也叫做信息市场 (information markets)、事件期货 (event futures) 和虚拟市场 (virtual markets) 等。

例如，在线预测市场 Intrade 的 2012 美国总统

选举结果预测中，参与者注册后能自由交易奥巴马当选的合约，如果最后的选举结果奥巴马当选，那么该合约就以 10 美元的价格出清；如果奥巴马没有当选，那么该合约价格为 0。如果某个交易者认为奥巴马会继续当选，就会愿意以低于 10 美元的价格买入该合约以期获利；如果同时存在另外的一个交易者认为奥巴马不会当选，那么该交易者愿意以高于 0 美元的价格出售他手头持有的奥巴马当选合约。如果市场上奥巴马当选合约的买入价格高于卖出价格，交易双方就可以根据一定的规则，比如双方喊价的均值作为成交价进行交易，该成交价格即为当时的市场价格。交易者的出价大致代表了交易者对合约所代表事件是否发生的主观概率，例如如果



图 1：在线预测市场 Intrade 中关于奥巴马在 2012 美国总统选举中获胜的预测结果。图例中当前的合约价格为 6.06 美元，表示奥巴马当选的概率大约为 60.6%。（图片来自 www.intrade.com）

交易者出价以低于 7 美元的价格收购奥巴马当选的合约，那么可认为该交易者大致认为奥巴马当选的概率不超过 70%。当市场中存在大量的交易者进行交易时，预测市场的该合约价格即代表了交易者群体对相应事件的预期，因此奥巴马当选合约的市场价格可解释为交易者群体对奥巴马能够继续当选的概率的预测。图 1 为 Intrade 中奥巴马再次当选美国总统的合约的交易界面，在该交易界面中，最近的市场成交价格为 6.06 美元，同时下面的绿色按钮还显示有 4 份以 6.06 美元出售的合约，而红色按钮则显示 3 份以 6.04 美元买入的合约。这表示整个预测市场对奥巴马 2012 再次当选美国总统概率的估计大约为 60.6%。

预测市场价格不仅能反映群体对事件的集体预期，还能及时反映最新发生的事件相关信息。很多事例表明，现代通讯技术条件下信息在社会网络中传播的速度往

往甚至超过传统的普通媒体正式报道的速度。很多传统媒体还没来得及报道的事件或者新闻在微博、Twitter 和 Facebook 等这样的社交媒体上往往能提前找到蛛丝马迹。因此，当预测市场中的一部分交易者获得影响预测事件走势的最新相关信息时，在市场激励条件下这些信息就会通过交易者之间的交易及时反映到市场价格中。图 2 为 Intrade 预测市场中 2011 年 1 月份至 2012 年 3 月期间的奥巴马 2012 美国总统选举获胜合约的价格变化趋势。从中我们可以清楚地看到，该合约在 2011 年 5 月 1 号突然由 6 美元猛涨到 7 美元，然后在短短两天之内就跌回到 6 美元附近。这是因为 2011 年 5 月 1 号这一天美国当局宣布了本·拉登的死亡消息，因而奥巴马的民意支持率空前提高，预测市场的价格变化反映了这一有利消息，因而迅速上涨。但是由于这一信息并不能长久影响奥巴马的获胜概率，因此不久之后预测市场价格回落到正常水平。

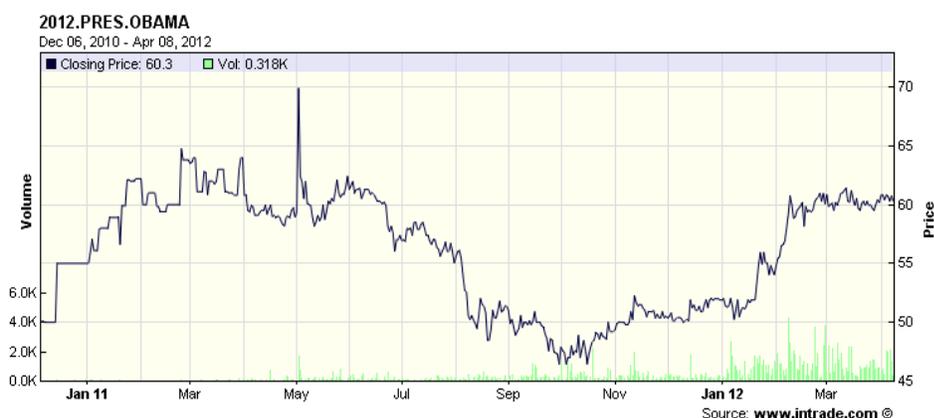


图 2：2011 年 1 月至 2012 年 3 月在线预测市场 Intrade 中关于奥巴马在 2012 美国总统选举中获胜的合约的价格波动情况。2011 年 5 月 1 号该合约的价格从 6 美元猛涨到 7 美元反映了美国当局公布的本·拉登死亡消息。（图片来自 www.intrade.com）

预测市场的历史可以一直追溯到 19 世纪晚期。据统计，从 1884 年到 1940 年期间有 15 个预测市场预测了这一时期的美国总统选举结果，那个时期还没有较科学的民意调查方法，而预测市场的

预测结果已经非常准确(15次选举预测中只有1次实际投票结果与市场价格所预测的结果相反,正确率为 $14/15=93.3\%$)^[14]。近年来,互联网和信息技术的不断发展普及为预测市场提供了更为便利的渠道,越来越多的在线应用关注于这一颇有潜力的预测方法,预测市场的组织形式也逐渐由‘线下’转变为‘在线’运行。在这一演变过程中,市场激励机制也由单一的现金交易变为现金、积分排名和奖品等混合的形式。比较著名的在线预测市场有位于美国爱荷华州的爱荷华电子市场(Iowa Electronic Markets, IEM)和位于爱尔兰的在线预测市场 Intrade 等等。其中 IEM 自 1988 年到现在一直用于预测美国总统选举结果的预测^[15],其预测精度明显超过传统的民意调查方法,因而引起了很多相关研究领域学者们的注意。而 Intrade 成立于 1999 年,其发行的合约种类涉及政治、金融、气候、科技、社会、娱乐等各个方面,其预测数据被美国全国广播公司下的消费者新闻与商业频道(Consumer News and Business Channel)、美国有线电视新闻网(Cable News Network)、

福克斯广播公司(FOX)、华尔街日报(The Wall Street Journal)、金融时报(Financial Times)、纽约时报(The New York Times)和超过 50 所美国知名大学以及一些著名的金融研究机构所引用。一些知名的在线预测市场有: Betfair(网址为: www.betfair.com), Foresight Exchange(网址为: www.ideosphere.com), Inkling Markets(网址为: home.inklingmarkets.com)等。

一些实证研究的数据表明,预测市场的价格准确地反映所预测事件的实际概率,而且市场价格能及时反映事件相关信息。比如,假如我们把所有价格为 0.3 的合约所对应的事件放在一起,其中最后的确发生的事件比例大约占总数的 3 成。而所有价格为 0.8 的合约所对应的事件发生比例为 8 成,以此类推。这个发现已经被来自许多领域的实践和实验数据所证实。例如好莱坞股票交易市场(Hollywood Stock

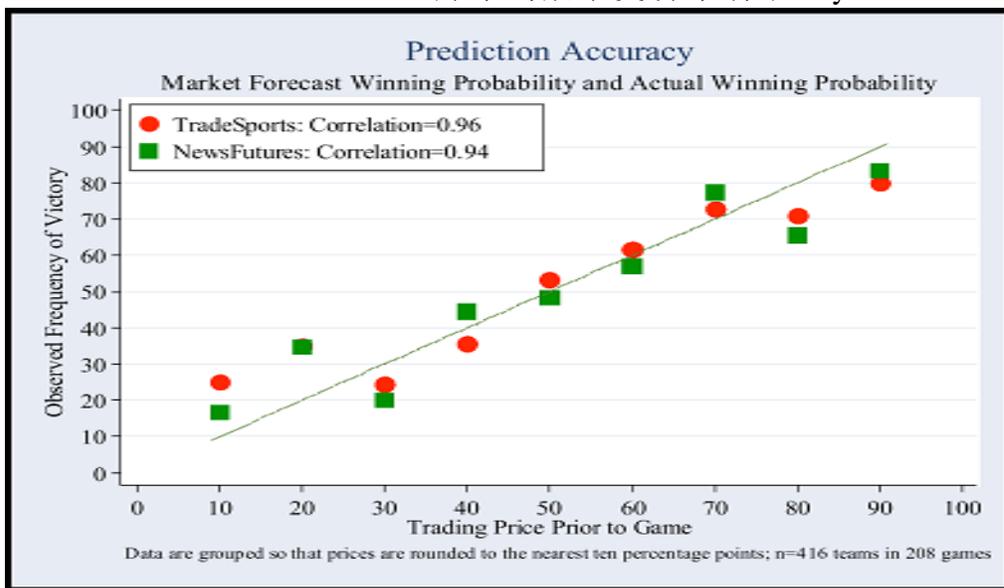


图 3: 预测市场价格反映相关事件概率

Exchange, HSK) 预测奥斯卡奖得主, TraderSports 和 NewsFutures 分别利用现金交易 (real money) 和点数交易 (play money) 的预测市场预测足球联赛的结果等等, 观测到的实际数据都表明市场价格能反映事件发生的概率。下图为 Schreiber 等报道的利用预测市场预测足球比赛结果的例子, 可以看到预测市场价格准确反映了事件发生的概率。

预测市场引起了一些大企业和大学、研究机构的实验室的关注。许多大企业 (例如微软、雅虎、英特尔、诺基亚和惠普等) 都在企业内部建立预测市场进行财务、销售额和公司策略方面的预测以提高公司内部的决策过程^[13,16]。预测市场的特点和性能同时也吸引了大学和一些研究机构的注意。除了前面提到的爱荷华大学的 IEM, 在维吉尼亚大学的维康实验室 (Vecon Lab)、麻省理工学院的集体智慧研究中心 (MIT Center for Collective Intelligence)、宾西法尼亚州立大学的经济管理与拍卖实验室 (Laboratory for Economics Management and Auctions, Pennsylvania State University) 等都建立起了用于群体智慧科学研究的预测市场^[17]。另外, 美国爱荷华大学商学院和医学院在 2007 年合作建立了“爱荷华健康预测市场” (Iowa Health Prediction Market), 该市场预测流行病疫情和疫苗效果的市场运作方式, 参与者限定为他们所邀请的医疗专业人士。此外, 若干新闻媒体也建立了公开的预测市场网站, 例如 CNN 的政治

市场 (CNN Political Market)、《金融时报》 (Financial Times) 的 FT PRED\$CT、《华尔街日报》 (Wall Street Journal) 的政治市场 (WSJ Political Market)^[17]。在公共部门运作方面, 美国国防部曾计划通过预测市场机制对地缘政治风险进行预测, 不幸的是最后由于两位民主党参议员把这个研究计划贴上了“恐怖市场”的标签而搁浅^[17-19]。有关预测市场的实证研究还在不断增加, 从现有的文献来看主要的发现有:

- 1) 预测市场价格能反应相关事件概率, 准确性优于传统的办法^[20-22]
- 2) 预测市场价格能及时反应新信息^[23-25];
- 3) 预测市场价格不可能长期被操纵^[26-28];
- 4) 预测市场价格存在高估低概率事件和低估高概率事件的偏差^[29-34];

3. 预测市场特点及其设计

预测市场和传统的期货市场和体育博彩非常相似。预测市场和期货市场的共同点在于, 这些市场合约的价格都反映了群体对交易对象的预期, 也反映了与交易对象的相关信息。与这些传统市场不同的是, 预测市场的合约价格是明确定义的, 即只要合约规定的事件在规定的时间内发生或者不发生, 那么合约就能或者不能兑现预先约定的现金; 而期货市场中的合约在交割日期的市场价格是不确定的。预测市场与期货市场的区别还在于它们的侧重点和目的有所不同。期货市场具有价格发现和规避风险的能力, 而预测市场主要用于反映未来事件

的相关信息。预测市场与体育博彩之间的联系在于，预测市场和体育博彩对

事件的定义与合约在未来某个日期的价格都非常明确，都是通过交易的形式就未来明确定义的事件进行预测。预测市场与体育博彩的区别在于，现有预测市场所采用的交易形式与期货市场类似，比如连续双向拍卖 (continuous double auction, CDA) 和类似于自动做市商 (automatic market maker, 即可以根据预先指定的规则与用户自动进行交易的计算机程序) 的对数市场得分规则 (logarithmic market scoring rules, LMSR) 等。虽然 LMSR 是一种专门针对预测市场提出的交易机制而且已经被应用于实际的预测市场，例如 Inkling Markets，但是该交易机制与自动做市商相近，而

与体育博彩中所流行的交易机制差别较大。体育博彩中流行的交易机制有三种：按注分彩 (pari-mutuel)、让分制 (point spread)、固定赔率 (fixed odds betting)，这些交易机制有异于常规的动态定价交易而更像是“赌博”。比如，固定赔率博彩中设定的赔率相当于一般股票交易中做市商设定的买卖差价 (bid-ask spread)，只不过这个价差在博彩过程中

保持不变，不像 CDA 等交易机制中价格是随着交易的进行而不断动态变化的。因此现有预测市场的交易机制与期货市场的交易机制相似，而交易合约与未来事件结果相关且具有明确指定的价格这一点又与体育博彩非常相近，兼有娱乐和现金套利的特点，因此预测市场可以看做是期货市场和体育博彩的混合形式^[35]。

与传统的期货市场或者体育博彩游戏一样，合理的市场设计是预测市场发挥作用的重要因素。预测市场的设计需要考虑多方面的因素，例如如何定义合约类型使之与真实的事件相联系、如何选择适当的交易机制以便于交易者理解和参与、市场激励机制是使用现金还是使用虚拟货币或者积分排名等形式、国家法律是否允许建立在线预测市场、预测市场是对所有人员开放还是只对选定的交易者开放、每次交易数量是否有上下限制以及是否收取交易手续费、预测市场的开市和闭市时间等等。其中比较重要的三方面是：合约类型、交易机制和市场激励机制。

根据合约和报酬计算方式的不同，可

表 1: 预测市场与期货市场、体育博彩的比较

	价格反映相关信息?	到期时合约市场价格是否明确	交易机制
期货市场	是	不明确	一般为连续双向拍卖
体育博彩	是	明确定义	按注分彩；让分制；固定赔率等
预测市场	是	明确定义	连续双向拍卖，动态按注分彩机制，对数市场得分规则等

将预测市场分为如下三类：

- 1) 合约价格由事件是否发生决定。例如，如果候选人当选，那么对应的合约就以预先规定好的价格清算。
- 2) 合约价格由事件发生的量级决定。例如，合约的最终价格以候选人最终的得票百分比进行清算。
- 3) 合约价格由事件是否以一定的量级发生决定。例如，如果候选人以超过 80% 的得票率当选，那么合约就以预先规定好的价格清算。

第一类合约的报酬由事件是否发生定义，事件结果揭晓后，以预先规定好的价格清算，因此市场价格可以被看做事件发生概率的期望值。第二类合约的报酬由事件发生的程度进行度量，市场价格向最有可能的程度逼近。因此市场价格可作为事件实际发生值的参考，所反映的信息比第一类合约更为详细。比如候选人最终以多大的得票百分比当选，不仅暗含了当选事件发生的概率，而且还能提供第 1 类合约提供更详细的信息。第三类合约主要关注事件发生的程度是否超过了一定的阈值，该阈值往往是关注的重点，因此可以看做是介于第一类和第三类合约的某种组合。比如，在足球比赛中，我们已经知道某个队的实力远远超出其对手，这时候仅仅预测输赢已经没有意义，我们感兴趣的可能是这个强队能超过对手多少分。从这三类合约的特点中可以看出具体选用哪种合约与预测对象的特征有关。仍以选举为例，如果两个相互竞争的候选者实力相当，那么使用第三类合约预测他们的得票率是否超过某一个阈值就没有太多意义，阈值也

不容易设定；而如果两个后选者实力悬殊，胜负判然，那么选票的差距就有更多的意义。第二类合约则比较适合于预测单个连续变量的取值，该变量一般不涉及到竞争对手的输赢和胜负，例如预测公司销售量、产量等等第二类合约就比较适合。

交易机制是用来促成买卖双方达成交易的方式，一旦确定了预测市场中的合约形式，这些合约将以交易机制规定的形式在市场中进行买卖。常用的预测市场交易机制有三种：CDA^[13]、LMSR 和动态按注分彩市场机制 (dynamic pari-mutuel markets, DPM)^[36]。CDA 按照交易者提交的交易价格和数量对交易者进行匹配，如果买家的出价高于卖家的喊价那么就可以成交。CDA 的一个主要问题是由市场流动性 (liquidity) 差导致的瘦市场 (thin market) 问题，即当市场中交易者较少而且买卖双方的出价没有重叠造成无法成交时容易导致预测市场的失效。LMSR 是 Hanson 根据适当得分规则 (proper scoring rules, PSR) 发明的一种自动定价算法，该算法的主要目的是给预测市场提供一种自动定价机制，能够保证预测市场在任何时刻接受交易者提出的交易请求。DMP 则是 Pennock 根据体育博彩中的按注分彩机制设计的一种新的自动定价机制，该机制与传统按注分彩机制的差别在于：传统的按注分彩机制中每单位货币的购买力与购买的先后顺序无关，因而不能鼓励交易者积极披露自己掌握的最新信息；动态按注分彩机制则保证先购买的交易者能以较

低价格购买合约，从而鼓励了交易者积极披露事件相关信息。CDA 被应用于大多数的预测市场，如 IEM、Intrade 等，LMSR 的实际应用相对较少，如 Inkling Markets 和一些实验场合，DPM 则只在一些实验性的预测市场中有所报道。Luckner (2008)^[37]比较了这三种交易机制的之间异同，如下表：

表 2：CDA、LMSR 和 DPM 三种交易机制的比较

	CDA	DPM	LMSR
是否连续反映信息	是	是	是
是否保证流动性	否	对买者：是 对卖者：否	是
是否对市场 资助者有风险	无	无	有限风险
应用实例	IEM, Foresight Exchange, Intrade, BetFair	Yahoo! Tech Buzz Game	Inkling Markets

除市场交易机制外，市场的激励机制对参与者影响也比较大。预测市场的激励机制分为现金交易和虚拟货币交易两种方式。现有的文献对于这两种激励方式的预测效果存在不同的看法，因此这是一个有待于继续研究的问题。一般认为，现金交易能使交易者更认真地对待输赢和结果，因此现金交易的准确性要高于虚拟货币交易。例如 Rosenbloom 和 Notz (2006) 发现使用现金交易的预测市场效果较好^[38]。

而来自实验经济学的一些证据则表明，现金交易容易歪曲人们参与预测市场的真实内在动机^[39]，一些本来因为个人兴趣而参与预测的人可能因为现金的使用而退出。同时，现金交易受到一些国家反赌博法律的限制，因为使用现金交易的市场容易滋生赌博行为^[40]。而来自该领域的另外一些实证研究则表明，虚拟货币的激励机制

并不逊于使用现金进行交易的性能。例如 Servan-Schreiber 等 (2004) 发现这两种报酬机制对预测精度没有大的分别^[41]。Luckner (2007) 以德国大学生为对象研究市场激励机制对预测准确度的影响^[42]。在他们的实验研究中，受试者被分为三组，每组总的期望收益相同，但是具体的激励机制有所差别。第一组的收益为固定的，与预测的准确性无关；第二组每个人的收益预测的准确度成正比，第三组则只对排

名靠前的几名进行奖励。结果发现,按照排名进行奖励的预测市场准确性高于其它两组。这些问题除了需要理论上的探索,更需要实证的研究。

4 预测市场理论基础

由于预测市场与传统的期货市场有着很大的关联。因此,在已有的微观经济理论中可以找到一些用来解释预测市场工作原理的理论和模型。这些理论模型主要集中在有效市场假设 (Efficient Market Hypothesis, EMH)^[43]、理性预期均衡 (Rational Expectation Equilibrium, REE)^[44,45] 和公共知识模型 (Common Knowledge)^[46] 三个方面。由于预测市场中明确地规定了合约价格与事件的关系,因而对于理性的交易者而言,他们只需要根据自己对于事件是否发生的主观信念和当前的市场价格就能作出是否交易以及交易数量的决定。所以,预测市场的价格解释直接与交易者对事件的信念分布有关。

Manski (2004)^[47] 和 Wolfers^[48] (2005) 基于期望效用理论 (Expected Utility Theory, EUT) 对预测市场的有效性进行了理论分析。他们回避了群体信念与真实事件概率之间关系的问题,一开始就假定群体信念均值就是事件概率。在这个前提下,他们通过利用不同的效用函数模型来验证市场价格是否能反映群体信念均值,并间接证明预测市场的有效性。

Robert Forsythe 等提出边际交易者假设试图对观察到的预测市场有效性进行解

释^[49]。该假设认为市场中的交易者可以分为两类,边际交易者 (marginal trader) 和非边际交易者。边际交易者的行为比较理性,不存在行为偏差,能根据当前已有的信息主动设定最合理的价格。当市场价格偏离预期时,边际交易者会主动交易修正该价格,并且边际交易者能根据最新事件相关信息及时修正自己主观预期进而更新市场价格。非边际交易者则是价格接受者,一般不会主动设定市场价格,而是根据市场价格和自己的偏好(存在偏差)决定买卖哪个合约,而且对事件相关的新信息不敏感。

除了以上这些依据市场供求理论所建立的模型, Feigenbaum 等 (2003)^[50] 和 Chen (2005)^[51] 将交易者通过市场价格交换信息的过程建模为一个分布式计算过程,具体研究了预测市场中的信息聚集过程。他们证明了在一定的条件下,市场价格在有限步内收敛并准确反映真实的世界状态。

5 预测市场与其它传统决策方法比较

相对于一些传统的群体决策方法,例如民意调查、专家座谈和 Delphi 方法等,预测市场需要参与者先将对事件的预期翻译为数字形式的价格,这造成了参与者的认知负担,同时市场价格传递给参与者的信号也不如语言文字那样直接,因此预测市场的机理并不像传统的群体决策方法那样直观。同时,由于单个参与者能够通过大量的交易暂时影响市场价格,因此潜在的市场操纵也会降低决策者对该方法的信

心。加上传统的组织决策环境中我们一般倾向于听取专家的意见，使用专家座谈、Delphi 等方法形成比较科学的决策，所以虽然预测市场具有一些传统群体决策手段难以比拟的优势，但仍然很难获得大规模

的应用成为一种常规的决策支持手段。

童振源等通过总结现有的预测市场研究，详细比较了预测市场和民意调查、专家意见或者座谈等常见的传统预测方法^[17]，如下表。

表 3: 预测市场与民意调查、专家座谈等群体决策方法的比较

	预测市场	民意调查	专家意见或座谈
参与对象	主动参与	随机抽样	遴选或推荐
意见表示时间	连续参与直到事件结束	一次性	一次性或周期性
意见表示方式	互动式	独立式	独立式或互动式
意见表示内容	事件发生几率 例如：谁将当选	个人偏好 例如：您支持谁	个人偏好或者事件发生几率
参与者权重	不平等 权重为投资比重	平等	不一定
参与诱因	经济报酬	一般没有	声望或者一次性报酬
说实话诱因	经济诱因	缺乏奖惩机制	声望 缺乏奖惩机制
意见整理结果	以价格信号反映参与者共识的变动 连续性趋势变化	静态一次性分析 结果只能反映“过去”的民意	一次性分析
预测准确度	准确	普通	略好
执行方式	电子市场	大规模访谈或问卷调查	慎重选择参与专家

Green、Armstrong 和 Graefe (2007) 针对预测市场与 Delphi 方法进行了比较 [52][17]，他们认为预测市场的优势在于能够连续定量地不断反映事件相关的最新消息，可以避免面对面座谈造成的团体压力等，而 Delphi 方法等则是一次性的意见，很难持续不断地反映最新的信息。相对于预测市场，Delphi 等方法的优势在于不需要将意见翻译成价格，专家有机会在每一回合用文字表达自己的意见，群体内部信息交换过程比预测市场的价格更有效和透明，也不用担心潜在的市场价格操纵。由于这些传统方法与我们的日常交流方式更加接近，因而适用的范围也更加广泛。

6 发展设计多样化的预测市场

一般意义上的预测市场是预测可验证的事件的虚拟市场，视事件结果是否可表达为连续变量这一类预测市场又可以分为预测离散事件的预测市场和预测连续变量的预测市场。例如，预测奥巴马是否能重新当选美国总统的市场只有当选或者不当选两个可验证的结果，因此该预测市场为预测离散事件的预测市场。而预测奥巴马得票率的预测市场则属于连续事件，可以设计适当的合约预测最有可能的得票率，如合约最后价格与奥巴马最后的得票率成正比，这在本文第 3 节已经讨论过。除了预测真实的可验证的事件之外，通过人为定义“事件结果”，预测市场还能用于政策结果分析和创意选择，这两种预测市场分别叫做

决策市场 (decision markets) 和创意市场 (idea markets)。Hanson 于 1999 年提出将预测市场的信息聚集机制用于政策分析，他认为预测市场能有效地在群体中共享信息因而能更好地预测某一政策的效果。例如，为了分析持枪法案 (hidden-gun bill) 对犯罪率影响可以建立两个不同的预测市场，一个市场预测持枪法案通过条件下的犯罪率，另外一个预测持枪法案没有通过条件下的犯罪率，通过比较这两个预测市场的预测结果就能知道持枪法案对犯罪率的影响程度，从而能够起到决策支持的作用。创意市场则主要用于评估新创意或者产品设计的受欢迎程度。在创意市场中，通过定义适当的事件结果，例如定义价格最高的合约对应的创意或者产品设计就是最后被选择的创意或者产品设计，市场价格一样能聚集反映分散在群体中的偏好信

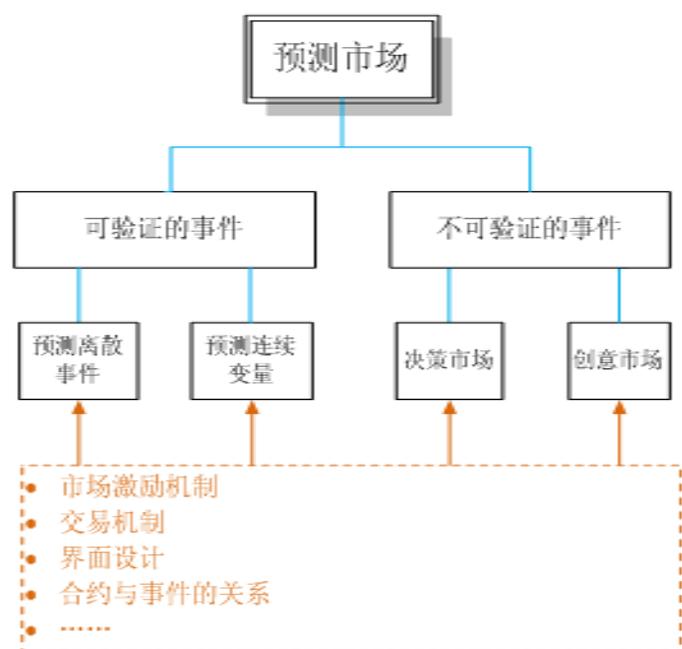


图 4: 预测市场的几种形式

息。这几种市场的关系如图 4 所示。

这四种预测市场的具体实现和运作都涉及到市场激励机制、交易机制、Web 人机交互界面和合约类型选择等等市场设计方面的问题。在实际的预测市场设计过程中，需要根据用户知识背景、决策者偏好和实际问题选择不同的交易机制、激励机制和合约类型，设计合适的人机交互界面。例如，现有的标准预测市场一般都采用 CDA 机制或者 LMSR 等作为交易机制，这些交易机制需要交易者显示地给出价格，给预测市场参与者造成了认知上的困难；单个交易者能通过交易暂时影响市场价格使得市场价格存在被操纵的可能，这降低了决策者对预测结果的信心，因而在某些情况下可考虑使用比较简单的按注分彩、让分制或者固定赔率等机制作为预测市场的交易机制。特别是当预测市场参与者大部分都是一般的交易者、没有太多的主动定价经验时，可选择固定赔率机制作为交易机制、积分排名作为市场激励手段，通过设计适当的事件概率估计方法从固定赔率市场的下注结果中估计事件的概率。

针对动态价格交易机制的缺点，固定赔率机制简单易用，不易被少数交易者操纵，可以避免 CDA、LMSR 等交易机制所带来的不利因素，因此我们详细研究了固定赔率作为一种轻量级预测市场交易机制的可能性^[53,54]。由于固定赔率市场中的赔率不能直接作为事件概率的预测结果，因此如果使用固定赔率机制作为预测市场

的交易机制，需要从赔率和下注结果估计的事件概率。基于现有的认知心理学成果对主观概率的解释，我们提出了事件概率—群体信念分布映射模型用来从固定赔率预测市场的下注结果中估计事件概率。实际数据对该模型的测试表明，事件概率—群体信念分布映射模型比专家设定的赔率预测精度要高。在平方适当得分准则 (quadratic proper scoring rule) 意义下，专家赔率所给出的事件概率预测值的平均得分为 8.56，得分方差为 6.54。而我们提出的事件概率—群体信念分布映射模型平均得分为 15.60，得分方差为 6.69^[55]。

7 小结

预测市场能及时聚集和反应分散在群体中的信息，是一种现代网络技术条件下先进的社会计算方法，其市场价格可用来预测未来不确定事件的可能性，有利于公共政策的制定和决策支持。同时，普及应用预测市场的社会意义在于它能唤起人们以更理性、更积极的态度关心和参与社会公共事务。因而，预测市场的良性发展和普及具有积极而重大的社会效用。

8 致谢

本文作者感谢如下科研支持：NNSF 60921061，90924302，71025001，91024030，40901219，71071152，70890084，NSFC 71102117，71103180；CAS 2F11N06；CityU SRG 7002625，CityU StUp 7200170。

参考文献:

- [1]. 曹志冬, et al., 2009年北京市甲型H1N1流行的气象因子与时空传播风险. 科技导报, 2010. 28(8): p. 26-32.
- [2]. 曾大军, et al., 传染病信息学和症状监测. 科技导报, 2007. 25(21): p. 17-23.
- [3]. Cao, Z.D., et al., An epidemiological analysis of the Beijing 2008 Hand-Foot-Mouth epidemic. Chinese Science Bulletin, 2010. 55(12): p. 1142-1149.
- [4]. Cao, Z.D., et al., Spatio-temporal evolution of Beijing 2003 SARS epidemic. Science China Earth Sciences, 2010. 53(7): p. 1017-1028.
- [5]. 曾大军, 王飞跃 and 曹志冬, 开源信息在突发事件应急管理中的应用. 科技导报, 2008. 26(16): p. 27-33.
- [6]. 王飞跃, et al., 基于平行系统的非常规突发事件计算实验平台研究. 复杂系统与复杂性科学, 2010. 07(4).
- [7]. 王飞跃, 曾大军 and 曹志冬, 网络虚拟社会中非常规安全问题与社会计算方法. 科技导报, 2011. 29(12): p. 27-38.
- [8]. Zeng, D., F.Y. Wang, and K.M. Carley, Guest Editors' Introduction: Social Computing. IEEE Intelligent Systems, 2007: p. 20-22.
- [9]. Zeng, D., et al., Social media analytics and intelligence. Intelligent Systems, IEEE, 2010. 25(6): p. 13-16.
- [10]. Peng, J. and D. Zeng. Topic-based web page recommendation using tags. 2009: IEEE.
- [11]. Zeng, D. and H. Li, How useful are tags?—an empirical analysis of collaborative tagging for Web page recommendation. Intelligence and Security Informatics, 2008: p. 320-330.
- [12]. Li, X., et al., Online communities: a social computing perspective. Intelligence and Security Informatics, 2008: p. 355-365.
- [13]. Wolfers, J. and E. Zitzewitz, Prediction Markets. Journal of Economic Perspectives, 2004. 18(2): p. 107-126.
- [14]. Rhode, P.W. and K.S. Strumpf, Manipulating political stock markets: A field experiment and a century of observational data. University of Arizona, mimeo, 2006.
- [15]. Berg, J.E. and T.A. Rietz, The Iowa Electronic Markets: Stylized Facts and Open Issues, in Information Markets: A New Way of Making Decisions, R.W. Hahn and P.C. Tetlock, Editors. 2006, AEI Press: Washington D.C. p. 142-169.
- [16]. Berg, J.E. and T.A. Rietz, Prediction Markets as Decision Support Systems. Information Systems Frontiers, 2003. 5(1): p. 79-93.
- [17]. 童振源, et al., 台湾选举预测: 预测市场的运用与实证分析. 选举研究, 2009. 16(2): p. 131-166.
- [18]. Hanson, R., Designing Real Terrorism Futures. 2005, Department of Economics, George Mason University.
- [19]. Hanson, R., Impolite Innovation: The Technology and Politics of 'Terrorism Futures' and Other Decision Markets, in Promoting the General Welfare: American Democracy and the Political Economy of Government Performance, E. Patashnik and A. Gerber, Editors. 2006, Brookings Institution Press: Washington D.C.
- [20]. Gruca, T.S., J.E. Berg, and M. Cipriano, Consensus and Differences of

Opinion in Electronic Prediction Markets. *Electronic Markets*, 2005. 15(1): p. 13-22.

[21]. Hahn, R.W. and P.C. Tetlock, Using Information Markets to Improve Public Decision Making. *Harvard Journal of Law & Public Policy*, 2005. 29(1): p. 214-289.

[22]. Gruca, T.S., J. Berg, and M. Cipriano, The Effect of Electronic Markets on Forecasts of New Product Success. *Information Systems Frontiers*, 2003. 5(1): p. 95-105.

[23]. Brüggelambert, G., Information and efficiency in political stock markets: using computerized markets to predict election results. *Applied Economics*, 2004. 36: p. 753-768.

[24]. Forsythe, R., et al., Using market prices to predict election results: the 1993 UBC election stock market. *Canadian Journal of Economics*, 1995. 28(4a): p. 770-794.

[25]. Smith, M.A., D. Paton, and L.V. Williams, Market Efficiency in Person-to-Person Betting. *Economica*, 2005.

[26]. Hansen, J., C. Schmidt, and M. Strobel, Manipulation in political stock markets - preconditions and evidence. *Applied Economics Letters*, 2004. 11(7): p. 459-463.

[27]. Hanson, R., Foul Play in Information Markets, in *Information Markets: A New Way of Making Decisions*, R.W. Hahn and P.C. Tetlock, Editors. 2006, AEI Press: Washington D.C. p. 126-141.

[28]. Sunstein, C.R., Deliberation and Information Markets, in *Information Markets: A New Way of Making Decisions*, R.W. Hahn and P.C. Tetlock, Editors. 2006, AEI Press: Washington D.C. p. 67-100.

[29]. Chen, K.-Y., L.R. Fine, and B.A. Huberman, Eliminating Public Knowledge Biases in Information-Aggregation Mechanisms. *Management Science*, 2004. 50(7): p. 983-994.

[30]. Oliven, K. and T.A. Rietz, Suckers Are Born but Markets Are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market. *Management Science*, 2004. 50(3): p. 336-351.

[31]. Spann, M. and B. Skiera, Internet-Based Virtual Stock Markets for Business Forecasting. *Management Science*, 2003. 49(10).

[32]. Berg, J., F. Nelson, and T. Rietz, Accuracy and Forecast Standard Error of Prediction Markets. 2003, Departments of Accounting, Economics and Finance, Henry B. Tippie College of Business Administration, University of Iowa.

[33]. Berg, J.E. and T.A. Rietz, Longshots, Overconfidence and Efficiency on the Iowa Electronic Market. 2002, Tippie College of Business, University of Iowa.

[34]. Rabin, M., Inference by Believers in the Law of Small Numbers. *The Quarterly Journal of Economics*, 2002. 117(3): p. 775-816.

[35]. Othman, A., Zero-intelligence agents in prediction markets, in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 2*. 2008, International Foundation for Autonomous Agents and Multiagent Systems: Estoril, Portugal. p. 879-886.

[36]. Pennock, D.M. A dynamic pari-mutuel market for hedging, wagering, and information aggregation. in *5th ACM conference on Electronic commerce*. 2004. New York, USA: ACM Press.

- [37]. Luckner, S., Prediction Markets: Fundamentals, Key Design Elements, and Applications. 21st Bled Econference Ecollaboration: Overcoming Boundaries through Multi-Channel Interaction, Conference Proceedings, ed. J.F. Hampe, et al. 2008, Kranj: Univ Maribor. 236-247.
- [38]. Rosenbloom, E.S. and W. Notz, Statistical Tests of Real-Money versus Play-Money Prediction Markets. *Electronic Markets*, 2006. 16(1): p. 63-69.
- [39]. Gneezy, U. and A. Rustichini, Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 2000. 115(3): p. 791-810.
- [40]. Arrow, K., et al., Statement on prediction markets. AEI-Brookings Joint Center Related Publication No. 07-11, 2007.
- [41]. Servan-Schreiber, E., et al., Prediction Markets: Does Money Matter? *Electronic Markets*, 2004. 14(3): p. 243-251.
- [42]. Luckner, S., Prediction Markets: How Do Incentive Schemes Affect Prediction Accuracy? 2007, Schloss Dagstuhl - Leibniz-Zentrum für Informatik.
- [43]. Fama, E.F., Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 1970. 25(2): p. 383-417.
- [44]. Muth, J.A., Rational Expectations and the Theory of Price Movements. *Econometrica*, 1961. 29(6): p. 315-335.
- [45]. Radner, R., Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices. *Econometrica*, 1979. 47(3): p. 655-78.
- [46]. Aumann, R., Agreeing to Disagree. *The Annals of Statistics*, 1976. 4(6): p. 1236-1239.
- [47]. Manski, C.F., Interpreting the Predictions of Prediction Markets. *Economic Letters*, 2006. 91: p. 425-429.
- [48]. Justin, W. and Z. Eric, Interpreting Prediction Market Prices as Probabilities. 2006, Institute for the Study of Labor (IZA).
- [49]. Forsythe, R., et al., Anatomy of an experimental political stock market. *American Economic Review*, 1992. 82(5): p. 1142--1161.
- [50]. Feigenbaum, J., et al., Computation in a distributed information market. *Theoretical Computer Science*, 2005. 343(1-2): p. 114-132.
- [51]. Chen, Y., T. Mullen, and C.-H. Chu, An in-depth analysis of information markets with aggregate uncertainty. *Electronic Commerce Research*, 2006. 6(2): p. 201-221.
- [52]. Green, K.C., J.S. Armstrong, and A. Graefe, Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared. *Foresight: The International Journal of Applied Forecasting*, 2007(8): p. 17-20.
- [53]. Chen, W., X. Li, and D. Zeng, Deciphering Wisdom of Crowds from Their Influenced Binary Decisions, To appear in the 2012 International Workshop on Social Computing (SoCo). 2012: Washington, DC. p. 1-12.
- [54]. Chen, W., X. Li, and D. Zeng, Estimating collective belief in fixed odds betting, in Proceedings of the 6th Pacific Asia conference on Intelligence and security informatics. 2011, Springer-Verlag: Beijing, China. p. 54-63.
- [55]. Chen, W., X. Li, and D. Zeng, Prediction based on Fixed Odd Betting Markets. Working paper, 2012.

作者简介

陈伟运：中国科学院自动化研究所复杂系统管理与控制国家重点实验室项目助理，主要研究方向为群体智慧和预测市场，在该领域先后发表相关论文 6 篇。分别于 2002 年和 2005 年获华中科技大学工学学士和硕士学位，2012 年获中国科学院工学博士学位。

李 昕：香港城市大学信息系统系助理教授。2000 年获清华大学工学学士学位，2003 年获清华大学工学硕士学位，2009 年获亚利桑那大学博士学位，IEEE, ACM 和 AIS 会员，主要研究领域包括 Business Intelligence & Knowledge Discovery, Social Network Analysis, Social Media, Scientometric Analysis, 在相关国际会议和期刊上发表论文多篇。

曾大军：中国旅美科技协会总会 2012 年理事会主席，2010 年度会长，中国科学院自

动化研究所研究员、博士生导师及 Web 科学及服务计算实验室主任。1990 年毕业于中国科学技术大学少年班、经济管理与系统科学系及计算机系，获学士学位；1998 年获美国卡内基梅隆大学工业管理博士学位。他主要研究领域为时空数据分析、传染病信息学、情报与安全信息学、多智能体系统和社会计算。

面向服务的科学：以癌症网格为例

谭伟

IBM T. J. Watson Research Center, Hawthorne, NY, USA

wtan@us.ibm.com

1. 引言

2007年, Jim Gray总结了历史上科学研究的四种模式^[1], 分别是基于经验的(empirical)、基于理论的(theoretical)、基于计算的(computational)和数据密集的(data-intensive)或者称为 e-Science。

数据密集的模式有其深刻的应用背景。科学家正面临来自诸如望远镜、传感器网络、加速器和超级计算机等处的海量的数据。这里我们给出一些例子, 例如, 2010年欧洲核子研究组织的 ATLAS 和其他三个 LHC 探测器产生了 13petabytes 的数据^[2]。生命科学和医学领域的 PubMed 数据库截止至 2011年7月, 收录了 2100万条数据^[3]。2011年 Nucleic Acids Research 杂志的 Database Issue and the online Molecular Biology Database Collection^[4]收录了 1330个分子生物学领域的数据库; 其中 GenBank 是美国 US NIH 维护的一个 DNA 序列数据库, 截止 2010年8月它收录了超过 2860亿条记录^[5]。这种数据泛滥(data deluge)^[6]的现象急需一种统一的、能够屏蔽底层细节的方法来访问数据和计算资源, 并进一步进行数据分析。

2 面向服务的科学(Service-oriented Science, SOS)

在 e-Science 的广阔背景下, 结合计算机领域面向服务的体系结构(Service

oriented Architecture, SOA)的兴起, 芝加哥大学和 Argonne 国家实验室的 Ian Foster 教授 2005年在 Science 杂志发表文章, 提出了面向服务的科学(Service-oriented Science, SOS)的概念^[7]。SOS 提供了一个集成数据(如 DNA 序列、天文图像、粒子观测器数据)、仪器(如电子显微镜、加速器、天文望远镜)、计算资源(如超级计算机、网格和云计算平台)的基础架构。在这个架构上, 用户使用资源, 就像我们使用电网的电力一样便利。这个基础架构提供的功能如下:

虚拟化: 所有数据、仪器和计算资源的访问都使用标准的 SOA 协议, 如 HTTP, SOAP 和 REST。从而用户可以访问这些异构的资源而不必关心其内部的实现。

安全: 隐密性(Privacy)、完整性(integrity)、认证(authentication)和授权(and authorization)是安全性的核心问题。这个架构支持安全策略的定义和运行时的监控。

元数据: SOS 鼓励在一个跨学科、跨机构的环境下科学家的共享和协作。因此一个元数据的框架对于科学家理解异构数据的语法和语义具有十分重要的意义。

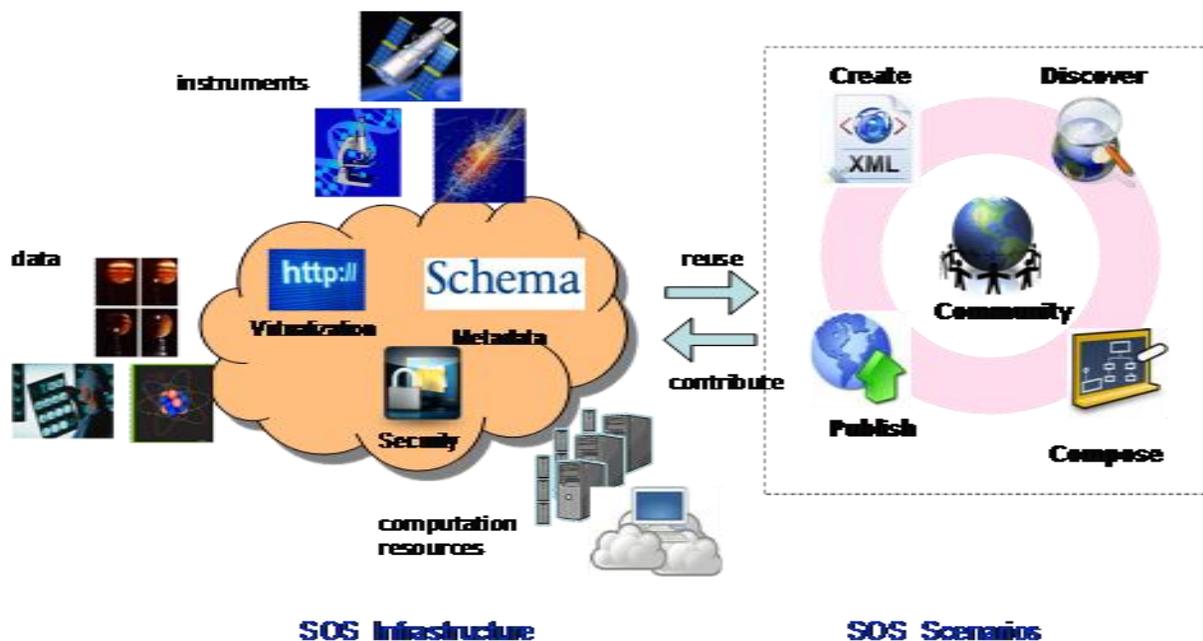


图 1: 面向服务的科学: 架构和典型应用场景

基于以上描述的架构, SOS 的典型应用场景包括创建、发现、发布服务和进行社区交互。

创建: 需要将数据、程序、仪器和计算资源的访问包装成 Web 服务, 可以使用开源工具包括 Apache Axis^[8]和 CXF^[9]。但是科学应用有自己的特殊需求, 如元数据和安全等, 所以一般要基于开源工具进行扩展和定制。这方面的工具有著名的网格工具包 Globus Toolkit^[10]以及 OMII-UK 项目提供的工具^[11]。

发现: 因为 SOS 往往以一个大系统的形式出现, 了解其有什么服务、如何从多个服务中选择合适的, 是一个难题。用户往往不知道很多服务的存在, 更不用说了解它们的功能和语义。因此, 一个 SOS 系统必须有一个共同的词汇来描述服务, 并维护一个服务的注册机构 (registry)。目前 W3C's 的资源描述框架

Resource Description Framework (RDF)^[12] 被广泛用于资源的描述。

组合: 科学家往往需要组合原子的服务来完成复杂的科学实验。 workflow 工具被广泛用于服务的组合^[13]。

发布: 一个可持续发展的 SOS 必须鼓励用户来贡献到系统中, 这可以看作是 Web 2.0 的原则在 SOS 中的体现。一个例子是微软研究院开发的 Trident^[14] 工具, 它提供一个 Microsoft Word 的插件, 使得科学家可以在发表的论文中嵌入一个科学 workflow。读者可以运行这个 workflow 来验证论文的结果, 也可以在自己的研究中重用它。

社区: 在创建、发现、组合和发布的过程中, 科学家可以通过社区的支持来交流和重用经验。面向科学家群体的社会网络目前也越来越得到重视。

SOS 的理念和技术已经被一些项目所采用, 包括英国的 myGrid^[15,16], 美国 NIH 支持的 caBIG/caGrid^[17], CVRG^[18]和 BIRN^[19]。下一节我们以癌症网格 caGrid 详细介绍。

3. 癌症网格 caGrid

癌症生物医学信息网格 (the Cancer Biomedical Informatics Grid, caBIG)^[20]是美国 NIH 下属的国家癌症研究所支持的一个项目。其目的是建立一个开放源代码、开放访问的、面向癌症研究的信息网络。开始于 2004 年, 2011 年 caBIG 拥有来源于超过 700 个机构的 2000 多名参加者。

它提供了 40 多个软件工具, 包括临床试验、生物图像处理、网格基础架构和元数据管理等。癌症网格 caGrid 是 caBIG 的面向服务的网格基础设施, 它基于 Globus Toolkit 网格中间件构造。caGrid 由核心基础架构, 基础架构上部署的服务, 以及 caGrid 的社区群体组成。

caGrid 核心基础架构使用 Globus Toolkit 构建, 并使用 Web Services Resource Framework (WSRF)^[10]作为资源描述、管理和访问的协议。caGrid 服务由其核心基础架构创建、部署和管理; 包括数据服务 (访问数据) 和分析服务 (访问计算资源) 两类。如图 2 所示, 在 2012 年 7

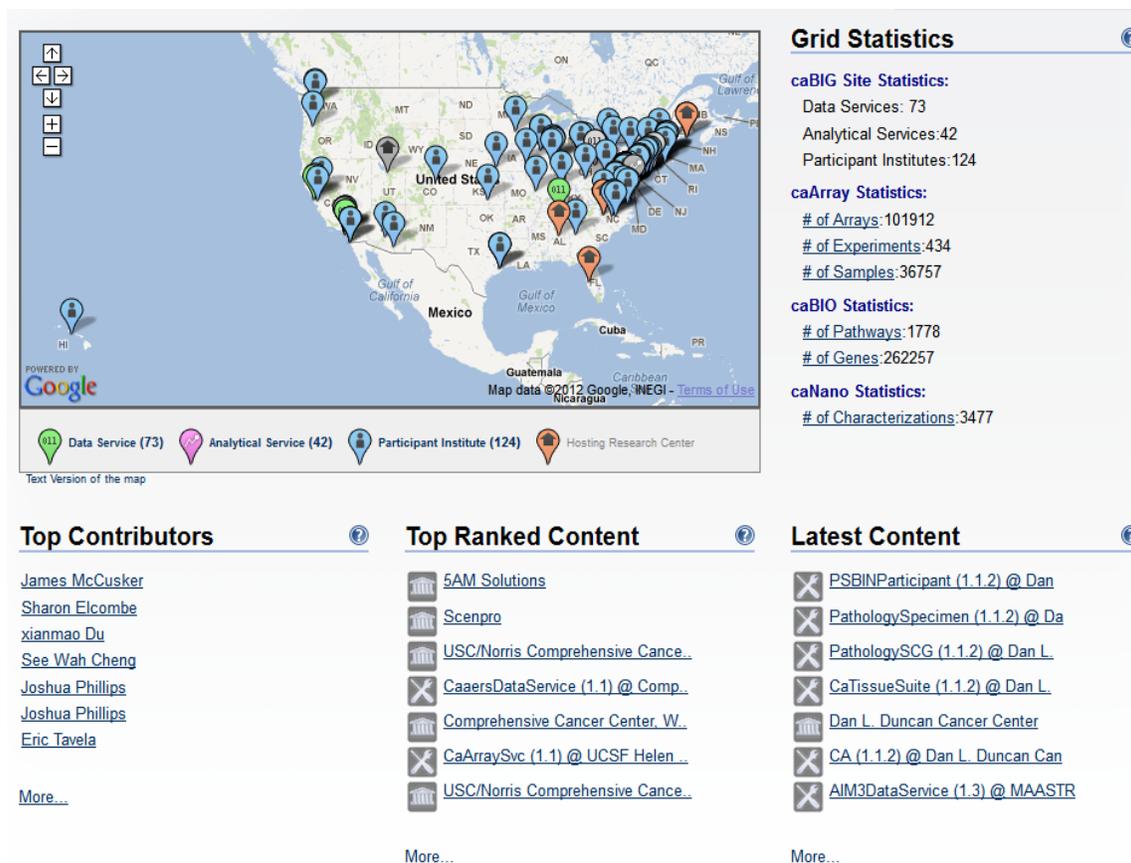


图 2: 癌症网格 caGrid 的资源分布 (图片来自 <http://cagrid-portal.nci.nih.gov/>, 采集自 July 15, 2012)

月 15 日, caGrid 上部署了 70 余个数据服务和 40 余个分析服务。caGrid 的社区群体包括 100 多个组织机构。

4 小结

在科学研究领域, 特别是生物医学和生物信息学, 面向服务的计算正在扮演数据和资源集成的关键角色。以美国 NIH 支持的项目为例, caBIG 和 CVRG 等项目都广泛使用了面向服务的基础架构和科学工作流等技术。本文介绍了面向服务的科学(service oriented science, SOS)的内涵, 并以癌症网格 caGrid 为例介绍了其技术细节。更多的内容可以参见作者的论文^[21-23]。

致谢

本文介绍的工作由多位同事合作完成, 包括美国芝加哥大学和 Argonne 国家实验室的 Ian Foster 教授, 英国 Manchester 大学的 Carole Goble 教授, 英国牛津大学的 Dave De Roure 教授等。

参考文献

[1]. Hey, T., S. Tansley, and K.i. Tolle, The fourth paradigm: data-intensive scientific discovery. 2009, Redmond, Washington: Microsoft Research.

[2]. Brumfiel, G., High-energy physics: Down the petabyte highway. *Nature*, 2011. 469(7330): p. 282-283.

[3]. PubMed. PubMed. April 1, 2012]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/>.

[4]. Galperin, M.Y. and G.R. Cochrane, The 2011 Nucleic Acids Research Database Issue and the online Molecular Biology

Database Collection. *Nucleic Acids Research*, 2011. 39(suppl 1): p. D1-D6.

[6]. Benson, D.A., et al., GenBank. *Nucleic Acids Research*, 2011. 39(suppl 1): p. D32-D37.

[7]. Bell, G., T. Hey, and A. Szalay, Beyond the Data Deluge. *Science*, 2009. 323(5919): p. 1297-1298.

[8]. Foster, I., Service-oriented science. *Science*, 2005. 308(5723): p. 814-817.

[9]. apache.org. Apache Axis. July 30, 2011]; Available from: <http://axis.apache.org/>.

[10]. apache.org. Apache CXF. July 30, 2011]; Available from: <http://cxf.apache.org/>.

[11]. Foster, I., Globus Toolkit Version 4: Software for Service-Oriented Systems. *Journal of Computer Science and Technology*, 2006. 21(4): p. 513-520.

[12]. OMII-UK. Open Middleware Infrastructure Institute. April-01-2012]; Available from: <http://www.omii.ac.uk/>.

[13]. W3C. Resource Description Framework (RDF). April 1, 2012]; Available from: <http://www.w3.org/RDF/>.

[14]. Taylor, I., et al., eds. Workflows for e-Science: Scientific Workflows for Grids. 2007, Springer.

[15]. Barga, R., et al., The trident scientific workflow workbench, in IEEE International Conference on eScience2008, IEEE: Indianapolis, IN, USA. p. 317-318.

[16]. Stevens, R.D., A.J. Robinson, and C.A. Goble, myGrid: personalised bioinformatics on the information grid. *Bioinformatics*, 2003. 19(suppl 1): p. i302-i304.

[17]. Goble, C., et al., Knowledge Discovery for Biology with Taverna: Producing and consuming semantics in the Web of Science, in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C.J.O. Baker and K.-H. Cheung, Editors. 2007, Springer US. p. 355-395.

[18]. Saltz, J., et al., e-Science, caGrid, and translational biomedical research. *IEEE Computer*, 2008. 41(11): p. 58-66.

[19]. cvrg.org. The CardioVascular Research Grid (CVRG). April 1, 2012]; Available from: <http://cvrgrid.org/>.

[20]. Helmer, K.G., et al., Enabling collaborative research using the Biomedical Informatics Research Network (BIRN). *Journal of the American Medical Informatics Association*, 2011. 18(4): p. 416-422.

[21]. Von Eschenbach, A.C. and K. Buetow, Cancer Informatics Vision: caBIG™. *Cancer informatics*, 2006(2): p. 22-24.

[22]. Tan, W., et al., CaGrid Workflow Toolkit: A taverna based workflow tool for cancer grid. *BMC Bioinformatics*, 2010. 11: p. 542.

[23]. Tan, W., J. Zhang, and I. Foster, Network Analysis of Scientific Workflows: a Gateway to Reuse. *IEEE Computer*, 2010. 43(9): p. 54-61.

[24]. Tan, W., et al., ServiceMap: Providing Map and GPS Assistance to Service Composition in Bioinformatics, in *IEEE International Conference on Services Computing (SCC)2011: Washington, DC*. p. 632-639.

作者简介



Wei Tan is currently a Research Staff Member at IBM T. J. Watson Research Center, USA. His research interest include workflow management, service oriented computing, Petri net, workload optimization, data centric computing, bioinformatics. He was a researcher at Computation Institute, University of Chicago and Argonne National Laboratory 2008 – 2010. At that time he was the technical lead of caGrid Workflow Toolkit, a service-based scientific workflow platform of cancer Biomedical Informatics Grid (caBIG) which is funded by National Cancer Institute. He published over 30 papers in leading journals and conferences, and PC member or reviewer for multiple journals and conferences. He is AE of *IEEE Transactions on Automation Science and Engineering*, and PC co-chair of the First *IEEE/ACM SN4CCGridS Workshop*. Wei Tan got his Ph.D in Control Science and Engineering from Tsinghua University in Beijing, China.

An overview on Pervasive Service Computing

Jiehan Zhou

The Department of Computer Science and Engineering, University of Oulu, Oulu, Finland

Jiehan.zhou@ee.oulu.fi

Abstract:

This paper overviews the research on the novel Web service-centric solution to pervasive computing, called Service-oriented Pervasive Computing (i.e. Pervasive Service Computing), which enables computer systems to deal with context in the user's environment, to dynamically discover and compose existing services, and to develop Internet-scale multimedia applications that support users' activities. The proposed improvements are achieved through enhancing three aspects of pervasive computing services, namely community coordinated multimedia, context-aware computing, and service-oriented computing.

Keywords

pervasive computing, service-oriented computing, community coordinated multimedia, context awareness, service composition, cloud computing

Introduction

Pervasive computing (i.e. ubiquitous computing) introduces a post-desktop computing environment that provides people with information and services when and where desired^[1]. Within this field, Pervasive Service Computing (PSC) is a Web service-centric solution to pervasive computing (i.e. Pervasive Service Computing) for

facilitating people's daily activities. The services provided include community coordinated multimedia, context awareness, and service composition. Community Coordinated Multimedia (CCM) enables maintaining virtual communities for the creation, aggregation, and consumption of community-created content. CCM offers enhancement of end user experience, with applications to enrich, extend, enhance, and share multimedia content. Context awareness enables applications to sense context, and adapt to changes in the pervasive environment. Service composition enables people to develop new applications through integrating existing loosely-coupled services via the Internet. In this paper, a service composition is a software application developed through integrating Web services.

The potential for Pervasive Service Computing has evolved beyond distributed computing and mobile computing platforms. It is fuelled by the proliferation of palm-sized computing devices, wireless networking standards such as IEEE 802.11 or Bluetooth, and micro/nano sensors embedded in artifacts. The kinds of pervasive computing envisioned by science fiction writers have been steadily moving into the real world. Well-known early developments and applications are the T-engine, by the Univ. of Tokyo^[2]; aware home, by Georgia Tech^[3]; infoPad by UC

Berkeley^[4]; Cups and smart-Its by Karlsruhe^[5]; and wearable computing by MIT^[6].

The following examples illustrate recent work on pervasive applications:

- MeasureIT, by the Open Wearable Computing Group^[7], is dedicated to measuring everything, everywhere, through the widespread dissemination of smart phones, in order to provide human-centered services that improve quality of life. Pervasive healthcare^[8] applies pervasive computing, proactive computing, and ambient intelligence technologies to enable healthcare and wellness management, and to make healthcare available everywhere, anytime.

- The program vSked^[9] focuses on using ubiquitous computing technologies to support parents in caring for premature infants.

- The ABC project^[10] utilizes pervasive computing to support the work of clinicians in the treatment and care of hospital patients.

- The Escort System^[11] is designed to protect people living with Alzheimer's disease. User-worn mesh-networked badges transmit indoor location information obtained in real time. A central server sends real-time short messaging service (SMS) alerts when a user might be at risk, giving caregivers information to address the situation before problems occur.

- Pervasive shopping, retail^[12], and advertising are also major areas for applications of pervasive computing^[13].

These applications include mobile shopping and checkout applications, price comparison tools, location-based services, and targeted mobile promotions, among other consumer-targeted technologies.

There are, of course, criticisms against this rise of pervasive computing applications. Many people raise concerns about the privacy implications of these devices, especially the cameras in public spaces that record much of people's everyday lives in the UK^[14].

Pervasive Service Computing^[15] incorporates service-oriented computing into a conventional pervasive computing paradigm, which takes advantage of options for registration, discoverability, composability, and open standards in a service orientation approach. This paper regards service (also called Web Service) as an autonomous, standards-based component, whose public interfaces are defined and described using a machine-processable format (specifically Web Services Description Language, or WSDL) that supports interoperable machine-to-machine interaction over a network using mainly Web-based standards¹. Service-orientation perceives function as distinct service units, which can be registered, discovered, and invoked over a network. Technically, a service consists of a contract, one or more interfaces, and an implementation. The service orientation principles for modern software system design are promoted through contemporary service-oriented architecture (SOA). This architecture

¹ <http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/>, accessed 01-08-2011.

introduces standardizations to service registration, semantic messaging platforms, and Web services technology^{[16][17]}.

A number of research efforts have focused on applying dynamic composition of services to support user activities in terms of service orchestration and choreography in mobile computing environments. A project named Ozone^[18] elaborated a task synthesis middleware for providing dynamic, situation-sensitive composition and reconfiguration of user tasks within an AmI (Ambient Intelligence) environment. The Aura project^[19,20] developed an architectural framework which explicitly represents user tasks as collections of services, and allows tasks to be configured in the context of a computing environment. Project Gaia^[21,22] introduced a mobile Gaia middleware for integrating resources of various devices. It provides functions such as forming and sharing resources among devices, and enables seamless service interactions.

Pervasive Service Computing incorporates multimedia and social networking to provide users with multimedia-intensive services and to strengthen the capability for coordination in pervasive applications. Digital convergence between audiovisual media, high-speed networks, and smart devices becomes a reality, and helps make media content more directly manageable by computers^[23]. Such digital convergence presents new opportunities for people to improve the ways they work. For example, hospitals will be able to provide multimedia-intensive medicating services at the patients' bedsides, even when the patients are being moved.

Social networking has already revolutionized the way users communicate and share information with one another in everyday life^[24]. But the software services people use to build online social networks for communities, or to share and explore interests and experiences, can be greatly enhanced. Most of these software services are primarily Web-based, such as video and voice chat, messaging, email, file sharing, blogging, discussion groups, and so on. All these can be integrated with context-sensitive, environment-aware systems.

2. Motivation and research problems

Existing challenges to development of distributed application and integration in pervasive computing include the lack of service discovery and composition mechanisms. Also, the available systems for information retrieval and sharing are insufficiently community-based. These shortcomings lead to the following overall research question which this research seeks to answer:

How can we automatically and efficiently develop and manage service compositions to adaptively support community-scale user activities?

This research aims to answer the above question through studying Pervasive Service Computing, community coordinated multimedia, context awareness, and service composition. This research suggests that the quality of human daily life (e.g. productivity and creativity) could be significantly improved through Pervasive Service Computing. The research suggests tackling the challenges in achieving the transition to

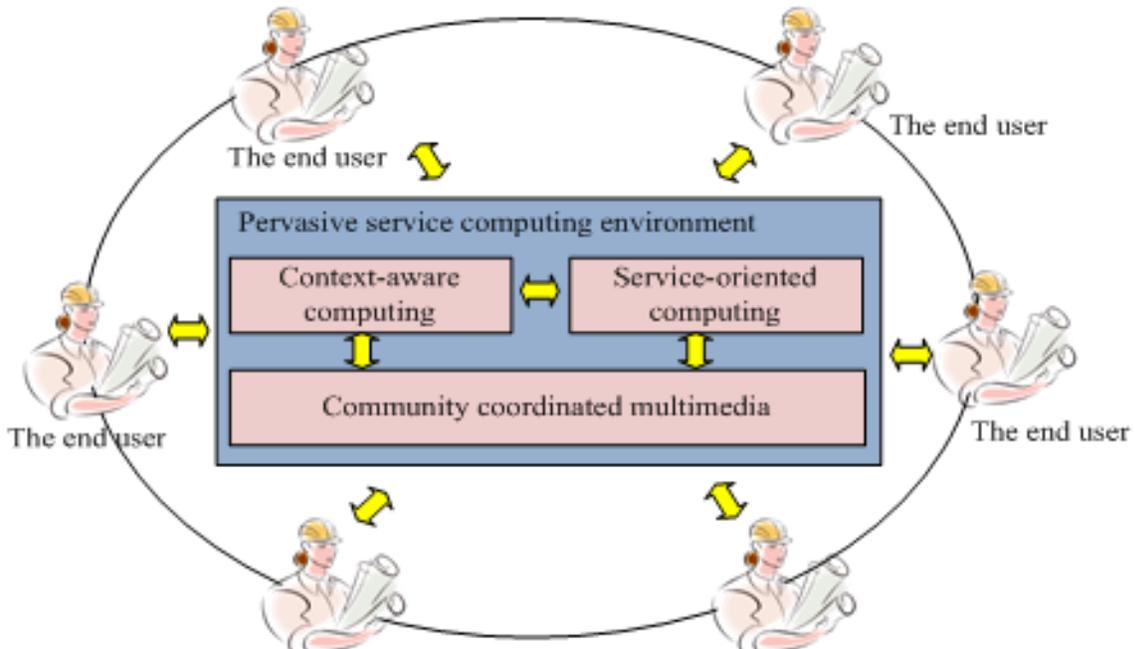


Fig. 1. Pervasive Service Computing Environment.

a service-intensive world by combining community coordinated multimedia, context awareness, and service composition. The concept of Pervasive Service Computing is illustrated in Fig.1.

The proposed improvements can be achieved through enhancing three aspects of pervasive computing services, namely community coordinated multimedia, context-aware computing, and service-oriented computing. Community coordinated multimedia is capable of maintaining virtual communities, and allowing them to create, aggregate, and use community-created multimedia content or services. Existing websites like YouTube offer some of this functionality. However they are proprietary and limited in functions (e.g. YouTube is only for video sharing). They lack multimedia processing intelligence, and lack service-oriented mechanisms (e.g. service discoverability and composability). Context-aware computing

enables applications to sense context and adapt to changes in pervasive environments. The capability of providing service composition is named composability. In other words, composability is the capability of a system to develop applications by combining existing services.

The overall research question of this research is divided into the following four sub-questions. The first sub-question serves as an umbrella for the other three, each of which focuses on one aspect of potential for Pervasive Service Computing.

Q1. What is Pervasive Service Computing, what is its paradigm, and why is it emerging as a significant aspect of pervasive computing?

Q2. What is community coordinated multimedia, what is Multimedia Application as a Web Service (MAWS), and how do we

apply such Web services into community coordinated multimedia applications?

Q3. What is context awareness in Pervasive Service Computing, what is context-aware pervasive service composition, and how can we combine these features in pervasive service systems?

Q4. How do we describe service composition in the context of Pervasive Service Computing, what is dynamic service composition in the Cloud, and how do we enable pervasive service composition in the Cloud?

3. Research contribution

The first contribution is to present a vision, a conceptualization, and a reference model of Pervasive Service Computing. Pervasive Service Computing (PSC) could act as a Web service-centric solution to facilitate modern daily activities, with an emphasis on service composition, building applications from services, and controlling the execution of these applications^[25] (suggest format modification for author to consider. If accepted, please modify the reference text in this session) . In^[26], the paper categorizes characteristics of Pervasive Service Computing and identifies challenges to the development of this field by examining four typical service-oriented computing scenarios under the umbrella name of “Pervasive Campus.” In^[27], the paper proposes a generic Reference Model of Pervasive Service Composition (PSC-RM) to guide Pervasive Service Computing architecture design and implementation.

The second contribution of this research is to explore the potential for Service-oriented Community Coordinated Multimedia (SCCM). This involves addressing research questions concerning the vision, conceptualization, modeling, and implementation of service-oriented, community coordinated multimedia. In^{[28][29]}, the paper explores how human experience is being extended and enhanced by collaboratively utilizing electronic and networked content and multimedia-intensive services. In^[29], the paper applies a service orientation approach with Web services technology to establish a service-oriented CCM architecture. This architecture enables us to tackle the requirements of scalability, discoverability, composability, layered abstraction, Quality of Service (QoS), and agility in distributed collaborative multimedia management. In^[30], the paper examines methodologies for metamodeling community coordinated multimedia, and introduces a community coordinated multimedia metamodel. This model was co-developed as a part of ITEA2 CAM4Home Metadata Framework. The community coordinated multimedia metamodel aims to provide a primitive constructor for instantiating, classifying, capturing, indexing, searching for information on program entities, and to provide supports for user decision-making concerning multimedia delivery options. In^{[31][32]}, the paper introduces a concept of multimedia applications as Web services, and then describes a multimedia content annotation end-user service.

The third contribution of the research is to expand the scope for context-aware

service composition. In^[33], the paper broadens the definitions of context and context-awareness in terms of pervasive service composition. The research then classifies context-aware service composition into adaptations at three levels: peer coordination, process service adaptation, and utility service adaptation. Next the paper specifies a design process for building context-aware pervasive service composition applications, then designs and develops the context-aware pervasive service composition architecture.

The fourth contribution of the research is to explore service composition modeling, description, and implementation in the Cloud. In^[34], the paper examines and formalizes service composition, with categories of service collaboration and coordination. In^[35], the paper explores ontology methodologies in Pervasive Service Computing, and generalizes the concept of ontology-driven pervasive service composition. This leads to the design of an ontology model for Pervasive Service Composition. In^[36], the paper explores dynamic service composition in the Cloud, which aims to provide users with a Cloud-based middleware for dynamic service composition, which can support on-demand service composition over the Cloud

4. Conclusions and future work

To explore Pervasive Service Computing, the paper focuses on addressing its three essential characteristics—community coordinated multimedia, context awareness, and service composition. These three capabilities empower Pervasive Service Computing to deliver on-the-air,

multimedia-rich services which support user activities—anytime, anyplace, on any device, and on any network.

Future work in the development of Pervasive Service Computing will focus on automatically, transparently, and intelligently managing integrated computing and communication resource services across networks, e.g. the Internet. This work will happen through developing and enhancing middleware technologies in the context of community coordinated multimedia, context awareness, and service composition.

References

- [1] M. WEISER, "The computer for the 21st century," *Scientific American*, vol. 265, pp. 94–104, 1991.
- [2]. J. Krikke, "T-Engine: Japan's ubiquitous computing architecture is ready for prime time," *Pervasive Computing, IEEE*, vol. 4, pp. 4-9, 2005.
- [3] C. D. Kidd, R. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, E. D. Mynatt, T. Starner and W. Newstetter, "The aware home: A living laboratory for ubiquitous computing research," in *CoBuild '99: Proceedings of the Second International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture*, 1999, pp. 191-198.
- [4] T. E. Truman, T. Pering, R. Doering and R. W. Brodersen, "The infopad multimedia terminal: a portable device for wireless information access," pp. 673-687, 2002.

[5] L. E. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl and H. Gellersen, "Smart-its friends: A technique for users to easily establish connections between smart artefacts," in *UbiComp '01: Proceedings of the 3rd International Conference on Ubiquitous Computing*, Atlanta, Georgia, USA, 2001, pp. 116-122.

[6] A. Smailagic, "Wearable computers: a new paradigm in computer systems and their applications," *Computers, IEEE Transactions on*, vol. 52; 52, pp. 977-978, 2003.

[7] M. Lawo, O. Herzog, M. Boronowsky and P. Knackfuss, "The Open Wearable Computing Group," *Pervasive Computing, IEEE*, vol. 10, pp. 78-81, 2011.

[8] A. K. Dey and D. Estrin, "Perspectives on Pervasive Health from Some of the Field's Leading Researchers," *Pervasive Computing, IEEE*, vol. 10, pp. 4-7, 2011.

[9] Hirano SH, Yeganyan MT, Marcu G, Nguyen DH, Boyd LA & Hayes GR, "vSked: Evaluation of a system to support classroom activities for children with autism. Proc International Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA:1633-1642." 2010.

[10] Jakob E. Bardram and Henrik B. Christensen, "Pervasive Computing Support for Hospitals: An overview of the Activity-Based Computing Project," *Pervasive Computing, IEEE*, vol. 6, pp. 44-51, 2007.

[11] D. M. Taub, S. B. Leeb, E. C. Lupton, R. T. Hinman, J. Zeisel and S. Blackler, "The Escort System: A Safety

Monitor for People Living with Alzheimer's Disease," *Pervasive Computing, IEEE*, vol. 10, pp. 68-77, 2011.

[12] C. Narayanaswami, A. Kruger and N. Marmasse, "Guest editors' introduction: Pervasive Retail," *Pervasive Computing, IEEE*, vol. 10, pp. 16-18, 2011.

[13] J. Krumm, "Ubiquitous Advertising: The Killer Application for the 21st Century," *Pervasive Computing, IEEE*, vol. 10, pp. 66-73, 2011.

[14] N. Davies, "What Will You Sacrifice?" *Pervasive Computing, IEEE*, vol. 10, pp. 2-3, 2011.

[15] Jiehan Zhou, " *Pervasive Service Computing: community coordinated multimedia, context awareness, and service composition*," PhD dissertation, University of Oulu, Faculty of Technology, Department of Computer Science and Engineering, 2011.

[16] Dirk Krafzig, Karl Banke and Dirk Slama, *Enterprise SOA: Service-Oriented Architecture Best Practices*. New Jersey: Prentice Hall, 2005.

[17] Thomas Erl, *Service-Oriented Architecture (SOA): Concepts, Technology, and Design*. Prentice Hall, 2005.

[18] N. Georgantas and V. Issarny, "User activity synthesis in ambient intelligence environments," in *EUSAI '04: Proceedings of the 2nd European Union Symposium on Ambient Intelligence*, Eindhoven, Netherlands, 2004, pp. 45-50.

[19] D. Garlan, D. P. Siewiorek, A. Smailagic and P. Steenkiste, "Project Aura:

Toward Distraction-Free Pervasive Computing," *IEEE Pervasive Computing*, vol. 1, pp. 22-31, 2002.

[20] Joao Pedro Sousa and David Garlan, "Aura: An architectural framework for user mobility in ubiquitous computing environments," *Proceedings of the 3rd Working IEEE/IFIP Conference on Software Architecture*, Montreal, August 2002, pp. 29-43.

[21] S. Chetan, "Mobile Gaia: a middleware for ad-hoc pervasive computing," *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pp. 223-228, 2005.

[22] M. Roman, C. Hess, R. Cerquiera, A. Ranganathan, R. H. Campbell and K. Nahrstedt, "A Middleware Infrastructure for Active Spaces," *IEEE Pervasive Computing*, vol. 1, pp. 74-83, 2002.

[23] ERCIM, "Multimedia Informatics," *ERCIM News*, 2005.

[24] Danah Boyd and Nicole Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, vol. 13, October 2007.

[25] J. Zhou, J. Riekkki and J. Sun, "Pervasive Service Computing toward Accommodating Service Coordination and Collaboration," *Frontier of Computer Science and Technology, 2009. FCST '09. Fourth International Conference on*, pp. 686-691, 2009.

[26] J. Zhou, Jukka Riekkki and Minyi Guo, "Pervasive Service Computing:

Visions and Challenges," *CIT 2010*, Bradford, UK.

[27] J. Zhou, Junzhao Sun, Mika Rautiainen, Oleg Davidyuk, Meirong Liu, Ekaterina Gilman, Xiang Su, Mika Ylianttila, Jukka Riekkki, "PSC-RM: Reference model for pervasive service composition," in *IEEE Proceedings of Pervasive Service Computing and Application 2009 in Conjunction with FCST 2009*, Shanghai, China, 2009.

[28] J. Zhou, M. Rautiainen and M. Ylianttila, "Community coordinated multimedia: Converging content-driven and service-driven models," in Hannover, Germany, Proc IEEE International Conference on Multimedia & Expo, Hannover, Germany, IEEE Press, pp. 365-368, 2008.

[29] J. Zhou, M. Rautiainen and M. Ylianttila, "SCCM: Service-oriented community coordinated multimedia architecture," in Turku, Finland, 2008, .

[30] J. Zhou, M. Rautiainen, M. Ylianttila, M. Foulonneau, P. Blandin, "Metamodeling for community coordinated multimedia and experience on metamodel-driven content annotation service prototype," *proc Congress on Services Part II, Beijing, China, IEEE Press*, pp. 88-95, 2008.

[31] J. Zhou, M. Rautiainen and M. Ylianttila, "P2P Service-Oriented Community Coordinated Multimedia: Modeling Multimedia Applications as Web Services and Experience," *Asia-Pacific*

Services Computing Conference, 2008. APSCC '08. IEEE, pp. 145-149, 2008.

[32] J. Zhou, M. Rautiainen, Zhonghong Ou and M. Ylianttila, "P2P SCCM: Service-oriented community coordinated multimedia over P2P and experience on multimedia annotation service development," in *Media in the Ubiquitous Era: Ambient, Social, and Gaming Media*, Artur Lugmayr, Heljä Franssila, Pertti Näreänen, Olli Sotamaa, and Jukka Vanhala (eds.), Ed. IGI Global, 2010, .

[33] J. Zhou, Ekaterina Gilman, Juha Palola, Jukka Riekk, M. Ylianttila and Junzhao Sun, "Context-Aware Pervasive Service Composition and its Implementation," *Journal of Personal and Ubiquitous Computing* 15 (3), ACM/Springer-Verlag, pp. 291–303.

[34] J. Zhou, Jukka Riekk and M. Ylianttila, "Modeling service composition and exploring its characteristics," *Proc. International Workshop on Web Service Composition and Adaptation*, LA, USA, IEEE Press, pp. 446–451, 2009.

[35] J. Zhou, E. Gilman, J. Riekk, M. Rautiainen and M. Ylianttila, "Ontology-driven pervasive service composition for everyday life," in *Pervasive Computing and*

Networking, Mieso Denko and Mohammad S Obaidat, Eds. Wiley, 2010.

[36] J. Zhou, Kumaripaba Athukorala, Ekaterina Gilman, Jukka Riekk & M. Ylianttila, " Cloud architecture for dynamic service composition," *International Journal of Grid and High Performance Computing*, vol. 4, pp. 17-31, 2012.

Biography



Jiehan Zhou is a senior research scientist in the Department of Computer Science and Engineering, University of Oulu, Oulu, Finland. He obtained his 2nd PhD in Network Information System from University of Oulu. He obtained his 1st PhD in Manufacturing and Automation from the Huazhong University of Science and Technology, Wuhan, China in 2001. He did postdoctoral and visiting research in many institutes and universities, including University of Toronto, Carleton University, VTT/Oulu Finland and INRIA/Sophia Antipolis France, Public Research Center Henri Tudor Luxembourg, Tsinghua University. His current research interests include Internet of Things, Cloud Computing, Pervasive Computing, Pervasive Service Computing, and Emotion-Oriented Computing.

哥伦比亚大学简介

沈琦

旅美科协总会副会长，大纽约分会副主席

Charles Shen <qs2005@columbia.edu>



位于纽约市的哥伦比亚大学源于创立于1754年的国王学院，是纽约州最古老的研究型大学和全美第五所历史最悠久的大学。哥伦比亚大学拥有16所本科和研究生学院以及四所附属学院，专业涵盖新闻、建筑、艺术、商业、法律、医学、科学、工程、教育、国际关系、宗教等诸多领域。哥大有22个图书馆，藏书超过1100万册并拥有大量数字化信息资源。哥大总共拥有近四万名学生和教职工，分别

来自于全球近150个国家。目前哥大还在阿曼、北京、伊斯坦布尔、巴黎、孟买、圣地亚哥、内罗毕设有七个全球中心。哥大是每年普利策奖的颁奖机构，也是美国

大学联盟14所创会成员之一，亦是全美第一所颁发M.D.学位的学校。全球共有80位诺贝尔奖获得者是哥大校友或在哥大工作过。哥大校友中还出了九位美国最高法院法官、25位奥斯卡奖获得者和29位国家元首，其中有包括奥巴马在内的三位美国总统。

哥伦比亚大学是世界上最重要的研究中心之一，也是一所有鲜明特色的优秀本科生和研究生培养基地。坐落在纽约城



中心，哥伦比亚大学有着独特的地理优势，其教育、研究活动依托、服务于这个大都会，也辐射到全世界。

如果你还想知道更多，请阅读
<http://www.columbia.edu/>

支援移动保密与隐私实施模型

Patrick C. K. Hung (洪澤權)

University of Ontario Institute of Technology, Canada

patrick.hung@uoit.ca

摘要

随着移动电话技术的发展以及与之相关的基础设施建设的完善,人们对随时随地获取网络服务的需求也在增加。这些被称作 m-service (Mobile Service) 的服务,可以把对个人电脑浏览器的支持通过网络扩展到了诸如移动电话 (Smart Phone)、个人数字助理、IP 电话系统等手持设备上。通常而言,一个 m-service 指一个独立的应用程序单元在手持设备,它通过无线技术与网络技术,在任何时间、地点,将电子商务功能或信息提供给服务的消费者。m-service 如 Web 服务 (Web Service) 的服务计算技术承担。Web 服务是基于一系列 XML (eXtensible Markup Language) 规范—如通用服务发布和发现规范 UDDI (Universal Description, Discovery, and Integration)^[1], Web 服务接口描述语言 WSDL (Web Services Description Language)^[2], 简单对象访问协议 SOAP (Simple Object Access Protocol)^[3,4]。近来, Web 服务得到了主流供应商 (诸如 IBM、微软、Sun、惠普、Bea) 的广泛支持。显然, m-service 计算的保密与隐私实施模型 (security and privacy enforcement model) 是服务计算中一个重要但同时充满挑战的研究领域。我们将从通讯和业务策略两个主要角度研究 m-service 计算环境下的移动保密与隐私实施模型。为了满足现实世界中的实际业务需要,在业务流程

集成与管理、服务协商与协议条件下, m-service 需要被高度定制来处理保密与隐私需求。

相关工作及最新研究进展

在 2003 年,“服务计算”作为一个新研究领域诞生了,它包含一套新的突破性的技术: Web 服务和面向服务的体系结构 SOA (Services Oriented Architecture)、业务流程整合与管理、效用计算/网格计算、自主计算、以及服务计算相关的商业和科研应用。为了处理这个新出现的、充满挑战的研究领域, IEEE 计算机学会 (Computer Society) 在 2003 年十一月成立了服务计算技术指导委员会 TSC-SC (Technical Steering Committee for Services Computing)。服务计算学科涵盖了弥合业务服务和 IT 服务差距的科学与技术。

在 Web 服务中,已经有一些工作将 XML 语言用于描述安全断言,但并非针对 m-service 的。这些基于 XML 的定义只是将访问限制在授权方上,而且保护了在松耦合执行环境下消息交换的完整性与可信性。特别地,现在有两种知名的基于 XML 的安全记号格式: 安全断言标记语言 SAML (Security Assertion Markup Language) 和可扩展权利标记语言 XrML (Extensible Rights Markup Language)。SAML 被用作 Web 服务的证

明与授权决定。^[13]Web 服务提供者提交 SAML 记号给安全服务器进行安全判断。SAML 是一种基于 XML 的框架,它通过对主体断言的方式交换安全证明。类似的, XrML 帮助服务的拥有者确定授权用户和单位的权利,明确授权用户与单位在何种条件下可以执行这些权利^[14]。接下来, WS-Security 描述了扩充 SOAP 消息的方式,通过消息完整性、消息可信性和单消息认证以提供保护的质量^[15]。基于 WS-Security, WS-Policy 提供了一种用于表达 Web 服务安全策略的语法^[16]。WS-Policy 包含了一系列安全策略断言用以支持在 WS-SecurityPolicy 中定义的 Web 服务安全规范^[17]。另外, WS-Authorization 定义了 Web 服务如何管理授权数据和策略^[18]。可扩展访问控制标记语言 XACML (eXtensible Access Control Markup Language) 定义了主体与资源间细粒度的授权与权利策略^[19]。总之,以上所列的语言或解决方案都没有解决 m-service 在协商与协议支持环境下的安全断言。另一方面,隐私技术在网络商业环境下的研究也已开展了一段时间^[20]。例如, W3C 的保密参数选择平台 P3P (Platform for Privacy Preferences Project) 工作组开发了 P3P 规范用于网站表述其隐私实践^[21]。P3P 用户代理使用户自动地被通知所访问站点的隐私实践,同时基于 Web 站点的隐私实践自动决策。P3P 框架并不是被设计用来处理服务计算隐私问题的。此外, WS-Privacy 在业界也已经提出了一段时间,它被设计用来规定 Web 服务的主体隐私参数和组织隐私安全声明^[18]。现在, WS-Privacy 规范还没有正式向公众发布。因而,企业内部或

企业间隐私控制的实际执行仍由企业隐私授权语言 EPAL (Enterprise Privacy Authorization Language) 技术进行规范化^[22]。因此,现在并没有一个全面的框架来处理 Web 服务甚至 m-service 的安全与隐私问题的完整框架。

提到移动计算的安全性,最基础的技术是 IEEE 802.1x 系列标准,这个标准通过纳入可扩展认证协议 (EAP) 连接 802.11 网络和多种认证服务, EAP 是一种认证对话,它是一种基于认证和会话密钥生成标准的加密解决方案。然而,关于应用程序的安全和多服务的隐密性几乎还没有研究。最近, IETF (互联网工程任务组) 商议了一个针对移动多宿主节点的隐秘威胁模型,但他们的研究焦点集中在对 MAC 和 IP 层隐私的威胁和可能的攻击。这些证据显示了将隐私加强技术与多服务安全机制整合方面的缺失。

从控制访问协议的观点出发, Xue 等人提出了一种用于支持在一个服务网格中的服务部署者和目标服务容器之间接口的解决方案^[24]。接下来, Buscemi 等人通过使用一种基于约束的模型,为基于 SOA 的电话服务提出了怎样指定、通讯和执行策略^[25]。Jerman Blazic 等人提出了采用信用管理,隐私协议和身份管理的,在内部实体通讯周期中,具有基础性和普遍性的中间件的隐私授权概念模型。然而,上述研究都未涉及移动服务中的访问控制机制。

谈到 XML 技术, ebXML 提供了一种在 CPP (Collaboration Protocol Profile) 环境下能够在两个部分之间就不同的技术问题实现自动化交流的过程,通过自动协

商达成交易伙伴间的合作协议协定 (Collaboration-Protocol Agreement, CPA)。然而, ebXML 主要关注的是 CPP 和 CPA 模板上, 而不是 m-service。

这些发展状况表明, 在可预见的将来, 移动服务计算研究的重要性将不断提高。移动服务相当于无线技术和面向服务架构 (SOA) 的一体化。移动服务的架构建立在一个不安全、不受监督和共享的环境下, 同时继承了无线技术和 SOA 架构在安全性上的弱点, 使得其安全性问题更加突出。例如, 很多时候, 移动服务处理的信息涉及商业机密、避免泄露机密信息相当重要。

移動服務模式

普适计算及其基础设施增加了在松耦合环境下, 不同业务流程间共享与协调使用 Web 服务的需求。一个业务流程包含一系列的代表业务任务和 Web 服务间交互的活动。工作流与 Web 服务的结合在研究领域与工业界已经变得十分流行。例如, Web 服务业务流程执行语言 BPEL4WS (Business Process Execution Language for Web Services) 最近已经在 W3C (World Wide Web Consortium) 被提议作为正式的业务流程与交互协议^[8]。如图 1 所示, m-service 可以视作无线技术与 Web 服务技术在服务计算环境下的的结合。

普适计算的发展对于 m-service 在不同商业环境下的协调和共享提出了更高的要求。这个研究计划的挑战比很多其他学科的挑战大得多。在建立移动服务框架模型时需满足以下五个基本属性:

1. 移动性: m-service 应该在有限的范围内, 由应用程序的业务逻辑决定。
2. 端对端: m-service 相互之间直接交互, 无需使用一个中央服务器。
3. 配置: 所有的应用程序之间的逻辑交互导致基于位置的用户之间的物理交互。
4. 合作: 并列的 m-service 之间易于合作。
5. 瞬时社区: m-service/用户可以在交互的任何时间加入和退出, 这样便产生了一种不断变化的映射。

安全与隐私的实施

m-service 架构构建于不安全的、无监视的共享环境上, 而这种环境易受到安全威胁事件的影响。同很多其他应用程序的情况一样, m-service 处理的数据可能是商业的敏感数据, 所以保护这些数据使其不受诸如暴露给未授权方之类的安全威胁是非常重要的。Web 服务安全领域的研究充满挑战, 包含了许多学科, 从验证/加密到访问控制管理/安全政策。在 B2B (Business-to-Business) 的 m-service 活动中, 隐私策略经常被明确且简洁地表述为要达

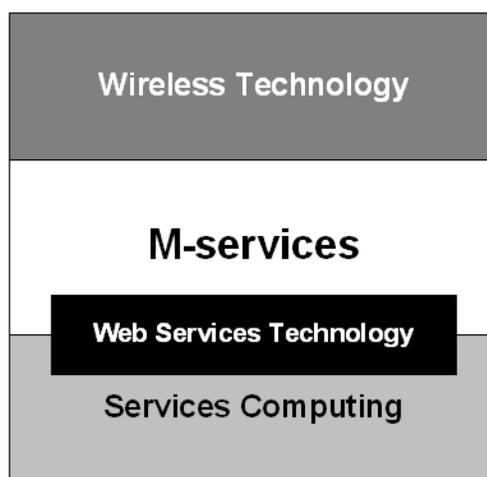


图 1. M-Services 的架构

成何种数据保护机制。因此，每一个信息保密系统都应该实施该机构的隐私策略^[9]。在这种环境下，信息保密机制也应该被包含在隐私增强技术中。所有这些证据都表明，基于处理电子商务安全与隐私方面的考虑，在 m-service 中将隐私的概念融入安全机制是非常重要的。

业务流程集成与管理

尽管 m-service 有着极大的吸引力，其复杂的技术问题和组织上的挑战仍未解决。Web 服务不能被直接扩展到移动环境上，就其本身而论，m-service 需要依照用户的喜好和其个人移动平台进行高度定制。因此，对这些 m-service 的管理形成了一个极具挑战性的问题。将 m-service 集成到 B2B 流程的时刻即将到来，安全与隐私依旧是 m-service 集成中一个重要的问题。

服务协商与协议

在一个 m-service 发现过程中，往往会有超过一个的服务能够拥有相同或非常相似的能力来满足请求者的需要。在发现流程过程中一个重要的问题就是对于 m-service 的提供者和请求者进行协商，从而发现一个对双方最理想的整合方案。在很多情况下，m-service 的提供者都需要与请求者在开始提供服务前协商如服务价格等服务成本问题，如服务响应时间、可用性、吞吐量和安全等服务质量问题^[10]。因此，我们需要一个更加复杂的、具有协商特性的电子商务模型以适应这个充满挑战的研究领域。我们计划通过加入一个具有不同领域特定词汇表的服务分级协议 SLA (Service Level Agreement) 模板

来支持不同类型的 m-service 协商^[11]。一个服务分级协议就是一份存在于 Web 服务请求者与提供者之间的、用于保证在限定等级上可量化的问题的正式合同^[12]。

基于前文讨论的移动服务框架,研究中的主题如下所述:

隐私策略的理论模型

隐私是有限度接近一个人的状态或者是条件，尤其是信息隐私和个人权利息息相关，个人有权决定如何、什么时候以及到什么程度向他人和机构来公布自己信息。隐私经常通过安全方法来得到保障^[23]。隐私策略主要说明什么隐私保护机制可以实现目标。相反，隐私策略通常是以自然语言描述和规范一个系统或者是机构如何保护隐私。一般来说，隐私策略描述一个机构的信息数据活动相关，如他们从个人收集什么信息，使用这些信息的目的，是否提供了信息入口，这些信息可能产生的后果，信息保留多久，以及发生纠纷时向谁报告。因此，隐私规则是执行自然语言隐私时所采用的系统级的规范。最终，每个信息系统像 m-service 一样都应该执行机构声明的隐私策略。

隐私入侵控制协商模型

信息安全和隐私保护中防范非法入侵是移动服务中的关键问题，这种情况下，移动服务的信息安全机制应该嵌入到隐私保护技术中。这一研究由于涉及到多学科，面临很大的挑战，从认证、加密到访问控制管理和隐私策略。在 m-service 中，访问控制主要是用于监视用户如果在成功认证之后，是否被容许进入特定资源或信息，

定；(3)将协商支持技术融入在移动服务框架上。

在研究项目中和移动服务的协作中的隐私访问控制是可伸缩的、可扩展，在变化和异构的移动环境下是可适应的。构建移动服务的隐私访问协商模型是一个很大的挑战，面临以下难点(1)协作中伸缩性和可验证性安全；(2)在协作中的安全和多个移动服务隐私策略协商的冲突调节。基于当前可信性协商的研究，移动服务之间的敏感交互发生在事先并不知道的情况下^[26]。因此每个移动服务能够选择自己的协商协议，策略以及相互独立的操作^[27]。为了构建移动服务的隐私协商模型，研究过程中由 IETF(Internet Engineering Task Force)定义策略执行抽象模型：

- 策略决策点(PDP)：做出决策的地方
- 策略实施点(PEP)：策略执行的地方
- 资源：像数据中一些有价值的东西，可以在访问授权之前的规则和服务策略标准首次应用。
- 策略管控点(PAP)：策略创建，修改存储的地方，策略是规则和服务的结合，

它定义了资源访问和使用的标准。

- 策略信息点(PIP)：作为属性值的源头，属性是一个主体，资源，动作或环境的特性，可引用为一个前置条件和目标。

一般而言，PEP 在移动服务请求方实现，而 PDP 实现是由服务提供方实现。应用和 PEP 可表述为一个主体，可以发出资源访问的请求。该研究项目将使用这一抽象模型作为基本的隐私访问控制协商模型。因此，该抽象模型不再考虑移动服务中的其它协商实体。图 3 中，在隐私访问控制策略管理架构中的策略协商模型，在 PEP 和 PDP 之间引入一个策略协商点(PNP)^[29]。该研究还包含隐私策略协商的不同阶段，并通过移动医疗应用一个例子来解释 PNP 如何在协商策略发挥作用^[30]。PNP 的研究将基于 WS-Negotiation 研究基础之上^[11]。WS-Negotiation 是一个独立 XML 语言，并且可以应用到不同类型的协议模板中。WS-Negotiation 包含三部分：协商消息，协商方之间消息交换的格式；协商协议，各协商方遵从的机制和规则；协商决策，是一个内部的基于效费比或其它策略的决策过程。通过扩展 WS-Negotiation 的范围来支持 m-service。此

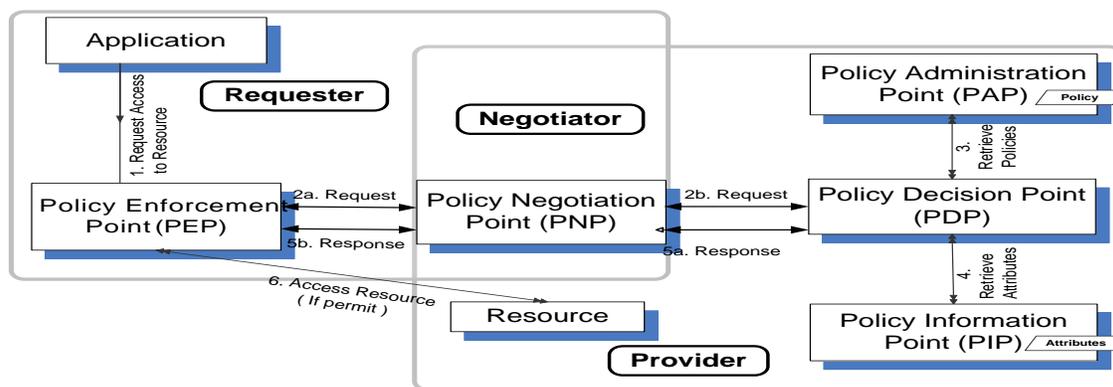


图 3. 隐私策略协商模型

外, 研究团队将检查隐私法律, 推导出隐私访问控制协商模型的需求^[31]。

参考文献

[1] Universal Description, Discovery and Integration (UDDI). “UDDI Version 3.0,” UDDI Spec Technical Committee Specification, 19 July 2002.

[2] World Wide Web Consortium (W3C). “Web Services Description Language (WSDL),” Version 1.2, W3C Working Draft, 9 July 2002.

[3] World Wide Web Consortium (W3C). “SOAP Version 1.2 Part 1: Messaging Framework,” W3C Proposed Recommendation, 07 May 2003.

[4] World Wide Web Consortium (W3C). “SOAP Version 1.2 Part 2: Adjuncts,” W3C Proposed Recommendation, 07 May 2003.

[5] IEEE Computer Society Technical Steering Committee for Services Computing (TSC-SC). Online: <http://tab.computer.org/tsc/>

[6] Joseph Reagle and Patrick C. K. Hung, “[P3P]: Beyond HTTP - P3P Task Force Report,” World Wide Web Consortium (W3C) document, July 2003.

[7] F. Leymann, D. Roller and M.-T. Schmidt, “Web services and business process management,” IBM Systems Journal, vol. 41, no. 2, pp. 198-211, 2002.

[8] IBM Corporation, “Business Process Execution Language for Web Services (BPEL4WS),” Version 1.0, 2002.

[9] S. Fischer-Hubner, “IT-Security and Privacy,” LNCS 1958, 2001.

[10] Patrick C. K. Hung and Haifei Li, “Web Services Discovery Based on the Trade-off between Quality and Cost of Service: A Token-based Approach,” ACM SIGecom Exchanges, vol. 4, no. 2, pp 20-26, September, 2003.

[11] Patrick C. K. Hung, Haifei Li and Jun-Jang Jeng, “WS-Negotiation: An Overview of Research Issues,” Proceedings of the IEEE Thirty-Seventh Hawaii International Conference on System Sciences (HICSS-37), Big Island, Hawaii, USA, January 5-8, 2004.

[12] A. Sahai, A. Durante and V. Machiraju, “Towards Automated SLA Management for Web Service,” HP Technical Report, 2002.

[13] OASIS, “SAML 1.0 Specification Set: Committee Specifications,” 2002.

[14] Contentguard, “eXtensible rights Markup Language (XrML),” Version 2.0, 2001.

[15] IBM, Microsoft and VeriSign, “Specification: Web Services Security (WS-Security),” Version 1.0 05, April 2002.

[16] IBM, BEA, Microsoft, SAP, Sonic Software and VeriSign, “Web Services Policy Framework (WS-Policy),” September 2004.

[17] Microsoft, VeriSign, IBM and RSA Security, “Specification: Web Services Security Policy (WS-SecurityPolicy),” Draft, 18 December 2002.

[18] IBM and Microsoft, “Security in a Web Services World: A Proposed Architecture and Roadmap,” White Paper, Version 1.0, 2002.

[19] OASIS, "OASIS eXtensible Access Control Markup Language (XACML)," OASIS Standard 1.0, 2002.

[20] V. Senicar, B. Jerman-Blazic and T. Klobucar, "Privacy-Enhancing Technologies – approaches and development," *Computer Standards & Interfaces*, vol. 25, pp. 147-158, 2003.

[21] World Wide Web Consortium (W3C), "The Platform for Privacy Preferences 1.0 (P3P1.0) Specification," W3C Recommendation, 16 April 2002.

[22] IBM, "Enterprise Privacy Authorization Language (EPAL)," IBM Research Report, 2003.

[23] Haddad, W., Nordmark, E., Dupont, F., Bagnulo, M., Park, S. S. D., Patil, B. and H. Tschofenig. Anonymous Identifiers (ALIEN): Privacy Threat Model for Mobile and Multi-Homed Nodes, the Internet Engineering Task Force (IETF), 2006.

[24] Xue, W., Huai, J. and Y. Liu. Access control policy negotiation for remote hot-deployed grid services. In the Proceedings of the First International Conference on e-Science and Grid Computing, Beijing, China, 2005.

[25] Buscemi, M. G., Ferrari, L., Moiso, C. and U. Montanari. Constraint-Based Policy Negotiation and Enforcement for Telco Services. In the Proceedings of the First Joint IEEE/IFIP Symposium on Theoretical Aspects of Software Engineering (TASE'07), China, Pages 463-472, 2007.

[26] Squicciarini, A., Bertino, E., Ferrari, E., Paci, F. and B. Thuraisingham. PP-trust-X: A system for privacy preserving trust negotiations. *ACM Transactions on*

Information and System Security (TISSEC), Volume 10, Issue 3, 2007.

[27] Yu, T., Winslett, M. and K. E. Seamons. (2003). Supporting structured credentials and sensitive policies through interoperable strategies for automated trust negotiation. *ACM Transactions on Information and System Security (TISSEC)*, Volume 6, Issue 1, Pages. 1-42.

[28] Yavatkar, R., Pendarakis, D. and R. Guerin. (2000). A framework for policy-based admission control. *IETF RFC 2753*, January, 2000.

[29] Cheng, V. S. Y., Hung, P. C. K. and D. K. W. Chiu. (2007). Enabling Web Services Policy Negotiation with Privacy preserved using XACML. In Proceedings of the IEEE Forty Hawaii International Conference on System Sciences (HICSS-40), Big Island, Hawaii, USA.

[30] Watson, M., Stanyon, W. and P. C. K. Hung. (2008). A Study of Ontario Privacy Act-Compliant Mobile Healthcare Access Control. To publish in the *International Journal of Health Information Systems and Informatics (IJHISI)*, in press.

[31] Yee, G. and L. Korba. (2004). Privacy policy compliance for Web services. In the Proceedings of the IEEE International Conference on Web Services (ICWS 2004), Orlando, USA, Pages. 158-165.

作者简介

Patrick C. K. Hung is an Associate Professor at the Faculty of Business and Information Technology in University of Ontario Institute of Technology (UOIT) in Canada and a Visiting Researcher at the University of Aizu in Japan in summer 2012. In addition, he is an Adjunct Professor at the State Key Laboratory of Software Engineering at Wuhan University in China. He has been an Adjunct Faculty Member at Department of Electrical and Computer Engineering in University of Waterloo in Canada, a Guest Professor at Institute of Computer Science in University of Innsbruck in Austria and Department of Information Engineering and Computer Science in University of Trento in Italy.

Hung has been working with Boeing Research and Technology at Seattle in the USA with a US patent on "Mobile Network Dynamic Workflow Exception Handling System." Before that, he was a Research Scientist with Commonwealth Scientific and Industrial Research Organization (CSIRO) at Canberra in Australia. He also has prior industrial experience in e-business projects in the North America and Hong Kong. He is a founding committee member of the IEEE International Conference of Web Services, IEEE International Conference on Services Computing, and IEEE Congress on Services. He is an associate editor of the IEEE Transactions on Services Computing, International Journal of Web Services Research and International Journal of Business Process and Integration Management.

Patrick has Ph.D. and Master of Philosophy Science in Computer Science from the Hong Kong University of Science and Technology in Hong Kong, Master of Applied Science in Management Sciences from the University of Waterloo in Canada, and Bachelor of Science in Computer Science at the University of New South Wales, Australia. He has also been a visiting Ph.D. student at the Graduate School of Informatics in Kyoto University in Japan and RSA Laboratories West at San Mateo, California in the USA.

云计算时代的信息技术安全和信息保证以及政府的作用

陈志雄 (Zhixiong Chen)

网络安全中心, 美国纽约玛希大学
(Cybersecurity Center, Mercy College, NY, USA)

摘要

云计算时代的信息技术安全和信息保证面临着更大的挑战和机遇。它们是企业和国家数字化安全发展的基石, 也是个人自我保护的核心。本文通过介绍一些重大的网络安全反例来说明云安全的重要性。在简单地介绍了云计算的核心内容, 技术和所要解决的问题后, 本文例举了云技术企业应用中的安全担忧。同时, 还介绍了信息技术安全和信息保证的最新进展及其将来的发展趋势。也阐述了公共政策和政府在推进全民信息技术安全和信息保证上有着不可替代的责任和作用。这里, 重点介绍了美国政府在信息技术安全和信息保证上制定的各项政策, 法规以及措施, 尤其是这几年推出的落实这些政策和法规的具体项目。通过本文的介绍, 希望信息技术安全和信息保证成为国家发展安全战略中的一个支柱, 落实到具体的项目上, 增强全民网络安全的意识, 培养迎接将来可知的和不可知的挑战的专业人才。

关键词

信息技术安全, 信息保证, 国家安全战略, 公共政策, 云计算

1. 信息技术安全和信息保障的必要性

信息技术安全 (IT Security) 和信息保证 (Information Assurance) 是网络安全 (Cybersecurity) 的核心。它们的重要性已为越来越多的大众和政治家们所理解。这首先得感谢媒体广为曝光的许多反面例子。远的不说, 最近就有 google、Yahoo、社交网 linkedin、团购网 Groupon、美国航空航天洛克希德马丁公司 (Lockheed Martin) 和日本任天堂游戏公司 (Nintendo) 的用户密码泄露; 索尼公司游戏网和娱乐网的全线崩溃, RSA 安全钥匙卡的私有钥泄密, DigiNotar 和 Comodo 的认证授权钥匙被攻陷; 和各式各样的病毒 (Malware) 以越来越复杂的面目出现, 比如, 著名的 stuxnet 蠕虫 (worm) 病毒。

其次, 在美国有越来越多的研究报告关注网络安全¹。在 2012 年七月由跨二党政策中心网路安全工作小组 (Bipartisan Policy Center task force)²出版的关于信息共享 (Public-Private Information Sharing) 的文章中提到, 就 2011 年十月到 2012 年二月间, 美国有案在录的就有五万次的网路攻击。其中, 86 次是攻击国家网络最关键的基础设施, 这包括电力网, 水系网, 公共交通网等。这些记录的案例只是实际

¹网络安全, 联邦计算机网络和信息技术研发资源库
<http://cybersecurity.nitrd.gov/>

²<http://bipartisanpolicy.org/>

攻击数目中很小的一部分，实际的攻击数目可能要比这个数目搞得多。当然，另一种可能就是夸大了可能的攻击数目。美国战略和国际研究中心 (Center for Strategic and International Studies) 也列举了自从 1996 年里来发生的各种网路攻击的典型案例。截止到 2012 年五月，网路攻击共有 97 起³。在其为美国总统布什 (George Bush) 撰写的一份报告中⁴提到网路安全涉及到国家安全，商业利益和个人隐私。警告美国必须时刻准备着最坏的网路战争。这份报告还建议了：

1. 协同组织和领导使得网络安全成为国际的首要重点；
2. 有权利要求对国家关键基础设施有更好的网络安全，同时也要开展与私有企业一起努力的方法；
3. 充分利用美国的能力通过新的外交政策来设立于计算和网络有关的规范和相应的破坏行为的结果；
4. 扩展军事和情报的能力来抵御外来的威胁；
5. 适应于数字技术的条款和规则来加强对隐私和自由的统管；
6. 增强对关键基础设施的证件验证；
7. 建立一个扩展的有相应的网络安全技能的劳动力；
8. 修改联邦的购买政策以促进更安全的产品和服务；
9. 修改政策和法律框架以指导国家的网络安全活动；
10. 聚焦网络安全的难题，发现这些难题的步骤和协调安排基金资助；

³ Significant Cyber Events,
<http://csis.org/publication/cyber-events-2006>

⁴ Cybersecurity Two years later, January, 2011

在这种大背景下，2009 年五月，奥巴马总统高调宣布，网路空间是攸关美国国家核心利益，是美国的重要财产。这就意味着，美国可以动用任何办法对发动网路攻击的国家和恐怖组织发动反击。

与此同时，政府也感觉到网络安全不能单靠私有企业来实现。这些私有企业拥有着近 80% 的至关国家命运的重要的基础设施 (Critical Infrastructure)。但是，私有企业和政府对网络安全的认识有着比较大的差异。政府是站在国家利益的角度，而私营企业考虑更多的是盈利，所采取的行为也大多为商业行为。这样，除了极少数要靠安全恐吓来发财的公司，私营企业一般比较理性，对民众的安全恐吓就比较少。最主要的是，私有企业对政府的规范有着天生的反感，认为政府的参与会扼杀网络的创新和发展，更不愿意与政府合作来共享他们用户的信息。这不仅对自己的业务和信誉有潜在的负面影响，而且会容易染上不必要的官司。这是因为政府很有可能会泄露他们的客户信息。在这种背景下，大规模的，需要政府和企业一起来推动和实施的网络安全项目就很难开展。

另一方面，政府也意识到虽然它有着许多企业所没有的资源，网络安全也不能靠政府自己解决。于是，政府发动媒体，教育全民关于网络安全的重要性。也利用自己手中的财力资源和立法手段提出各项公共政策，资助学术研究和提供奖学金给学网络安全的大学生和研究生，引导私营企业以更加合作的态度联合起来面对网络安全问题。

本文就作者这么多年研究和教育网络安全的心得和观察来介绍在云计算时代的信息技术安全和信息保证。在第二节，本文将具体介绍网络安全的一些典型的案例来了解网络安全的重要性，这些案例都是发生在近二年的时段里。在第三和第四节里，本文重点放在云计算里的信息技术安全和信息保证，就是云安全。它包括云数据保证和云基础设施的安全。提出了不仅要靠技术的进步来解决云安全，更要靠系统的 IT 安全审计去连接云计算中各个部件，以高屋建瓴的高度去处理云安全。在第五节，我们重点分析政府在促进网络安全的作用，介绍和例举美国政府在推进信息安全和保证的具体措施和结果。

希望通过本文的介绍能使大家对信息技术安全和信息保证有所认识，理解网络安全是国家安全中很重要的一部分。只有积极准备，才能应对各种状况和走在对手的前头。

2. 典型网络安全反例

2.1 私有钥和 PKI 认证机构的攻陷

我们知道，密码技术是现代电子商务的基石。在公钥密码 (PKI) 中，私有钥拥有者可以在网络社会里证明自己的身份。也可以利用私有钥来代表公司，政府或个人来签发数字证书和计算机程序。它们也可以用于许多认证身份的应用中，包括实时产生密码钥匙。如果私有钥被攻陷，则所有的所谓身份和密码无从谈起，更没有所谓网络安全。

一般讲，现在的私有钥和身分证书都是由大家认同的证书机构 (Certificate

Authority) 来验证和发送。它们是 PKI 系统的关键。也是电子商务的基石。这些证书机构通常是公认的相当安全的机构。但是，最近发生了的几个案例使人们对证书机构的安全信心发生了动摇。

2.1.1 荷兰 PKI 认证证书公司 DigiNotar

DigiNotar 是一个证书发放机构。也被荷兰政府认可和允许签发与政府有关服务的网络证书。2011 年七月到九月它被发现它发放了针对 google、Mozilla、Yahoo 及 Skype 等逾 500 个公司和机构的假证书。这些伪证书常常可以被用来进行中间人攻击 (man in the middle)。这是一个令人震惊的发现和很严重的安全事故。它的直接结果是荷兰政府使用的 DigiNotar 证书一下子变为不可信了。政府的网络行为就会被人怀疑。

九月后，微软和其他网络浏览器纷纷把 DigiNotar 的根证书放到不受信任的发行机构名单里面。

DigiNotar 没有公开它们是如何被骇客使用何种方法入侵成功的。在网上有自称是 Comodohacker 的骇客承认它在三月已经侵入成功。他的动机是为回教徒找回正义。但是，有人通过这些使用假证书的用户认为，伊朗政府可能是幕后的抄手。

与此相关联的是同一个骇客 Comodohacker 也承认它成功入侵另一个证书发放机构——Comodo Group。这是一个总部在美国的私有企业，也发放网络证书。Comodo 承认几十个主要是针对 google 的假造证书被发现。

2.1.2 RSA 安全身份

RSA，现在是 EMC 的一个安全部门，在美国被认为是一个最安全的机构。它是由创立 RSA 算法的研究人员成立，其后被并入 EMC。

在 2011 年三月 RSA 公开承认，它们的双因子认证令牌 (SecureID Token) 已泄露。双因子认证令牌被广泛地用来远程的安全进入公司或政府的内部网络。

RSA 没有说明是什么原因造成了泄露。但是在随后的根据 EMC 上传的病毒案例上分析，这是一起典型的网络钓鱼病毒。一份电子邮件送给了 EMC 一个员工，顺附送给了另外三个员工。电子邮件的标题是关于招聘结果并附上微软的表格文件。这个附件利用了 Adobe 快闪的漏洞植入了另一个文件，从而打开了一个后门。它可以给进攻者提供进一步的挖掘机会。

五月，国防部的承包商洛克希德·马丁 (Lockheed Martin) 就发现有人用偷来的 RSA 数据对其进行攻击。

2.2 密码被盗

我们知道，为了在网络上证明自己的身份，我们可以利用人们的所知道的，人们身上所拥有的，和人体的特征。而其中利用人们所知道的最常用的就是密码。几千年来，密码一直是人们用来进行进出控制。如果密码被人所知，则任何控制都是没用的。

2.2.1 LinkedIn

LinkedIn 是一个为大多数职场人士使用的职场社交网。它拥有大约一亿五千万个注册用户。2012 年六月，linkedin 承认，它的用户密码可能已泄露。要求用户修改密码。一个自称在苏联的黑客在网上公布了约六百五十万个密码。理论上讲，这些用户的个人信息也有可能同时被泄露。

目前为止，从公开的资料，还不知道是什么方法和手段造成了这一次的泄密。

2.2.2 Groupon

在 2011 年六月，美国著名的团购网 Groupon 在印度的一个自我经营的子公司在网上公开了一个数据库，其中有三十多万用户的电子邮件和没有加密的密码也被公开。虽然在人们发现后，这些数据被马上拿走，但是，由于象 google 这些搜索引擎还是会把在高速缓冲存储器里这些数据编入索引。他们的直接结果就是这些用户的所有帐号信息有被人侵入的危险。

有意思的是，人们不明白为什么这些密码没有用哪怕最简单的加码存入。这样，至少给人看到的是加过密的密码。

2.3 邮件账号被利用

2.3.1 Yahoo

最近，大家是否注意到，我们经常会收到自己朋友圈子里送来的莫名奇妙的邮件。这是由于他们的邮件账号被人用各种手段而窃取。作案者冒用这些人来送各种邮件，最常见的是连接到奇奇怪怪的网址；也有自述在国外被人打劫，请求金钱帮助的邮件。我最近就收到过一个邮件，要我

帮忙。他全家在英国度假时钱包被人抢了。人还好，要急着赶飞机，但是旅馆不能结账，需要急用钱。我还回了一个邮件，试试看这些邮件到底是怎么能够骗到钱的。还真收到回复，让我把钱汇到在全世界都有分支的西联分支。我还顺着问了我怎么知道你就是我的朋友那，你有什么让我相信的证据。那边还回邮件试图证明我们有共同认识的朋友。我估计是从我与我的朋友电子邮件里查到的。

2012年七月，Yahoo承认大约有40万用户的密码账号被偷。这也是因为有人在网上公布了这些信息。

2.4 数据采集与监视控制系统(SCADA)的感染

在美国，几乎所有的至关重要的基础工业设施网络都是用数据采集与监视控制系统(SCADA)来监视和控制的。比如，纽约的地铁系统，从上州到纽约城的供水系统，电源能源网，等等。

最初的SCADA系统是一个独立的与外界没有联系的系统。随着网络的发展，新的SCADA采集和控制元件也开始与网络有接口。已有的SCADA系统也在更新，开始与网络接口。这本是一个提高效益和全局控制的技术进步。但是，它也同时增加了被人进攻的潜在危险。由于SCADA系统涉及许多技术细节，一般人是很难掌握并发动攻击。但是，如果有强大的组织，财政支持，或国家行为，攻入SCADA系统还是有可能的。最典型的莫过于在伊朗核能SCADA系统的Stuxnet蠕虫。

2.4.1 Stuxnet 蠕虫

在2012年六月，纽约时报报道了，美国政府和以色列政府在2010年联合开展了代号为奥林匹克运动会行动⁵。他们通过视窗操作系统传播内核型蠕虫病毒。这些蠕虫专门针对西门子公司SCADA系统中可编程的逻辑控制器(PLC)内核。这是一个很复杂的感染过程。真正的具体细节还不清楚。现在有很多的分析。

据报，伊朗的几个有关铀浓缩系统遭到了感染。使得铀浓缩控制出现混乱。至少延迟他们的铀浓缩过程一年以上。

2.5 网络系统被攻陷

2.5.1 SONY

2011年四月，SONY(索尼)公司承认了它的网上游戏网被人攻陷。大约有七千七百万的用户信息包括信用卡信息可能被盗。五月，它又承认它的娱乐网也有可能被攻陷，它的二千四百万用户的信息也有可能被泄露。这是迄今为止最大的用户信息泄密案。它超出了2007年轰动一时的美国TJX连锁店的四千五百万信用卡被泄露案件。

由于整个网被攻陷，为了另建新网，索尼游戏网被迫停运了24天。

事件起因可能是由网上匿名组织，anonymous为了抗议索尼起诉公开索尼软件漏洞的黑客而发动的。不过，anonymous否认了这些指控。

⁵ 就是这篇文章使得美国参院和众院开始了是谁把这些重要的有关国家安全的情报捅了出去的听证会。

2.5.2 Google

Google 在 2011 年四月发布了他们发现许多 gmail 的用户，主要是一些名人，美国的高级官员和一些有争议的人物，被人用钓鱼的方式攻陷。黑客们除了下载相当数量的邮件，也一并下载许多文件。这件事是比较轰动的 google 文件被人下载案例。它的背后背景不简单。但是，目前许多有关的文章都是猜测，并没有具体的结果。

3. 云计算的安全和保证

云计算要解决的首要问题就是大数据。现在，我们每天收集和产生的数据都在 2.5 个艾未数(Exabyte or quintillion or 10^{18})。最近二年的数据量比以前所有的数据量总和都多出几倍⁶。我们把各种各样的数据利用各种方式收集起来并存储，不管现在有用或无用，因为看起来现在没用的或无能力处理的数据将来就不好说了。何况，数据就是金钱。

伴随着大数据是对大数据的解析处理。海量、实时、实用是对任何解析处理方式的基本要求。传统的数据储存、检索、管理已不能适应了。这时，Hadoop 并行处理技术和相应的 mapReduce 就适时成长。

伴随着大数据的解析处理就是管理技术。最主要的技术就是不断向前发展的虚拟化技术和管理。虚拟化技术和管理早已出现。最早可以追溯到 IBM 的大型主机。随后是操作系统与计算机硬件的虚拟化。

一台机器可以运行任何指导的操作系统，到最后计算资源的虚拟化。

同时，以服务为架构的中间件以及以松散型结合的应用软件也使得计算资源共享，存储共享和开发共享的云计算的抽象表述有了可靠的基石。

但是，云计算的安全和保证一直是一个让人担忧的大问题。许多公司望而却步。主要的担忧可以归类为二十多种⁷。比如，数据经过的途径和最后的地点；数据的生产者，发布者和拥有者；数据的保护和保证，以及实用的标准；数据在任何时候的孤立性；数据形式的标准，不管是结构的还是非结构；合规性(compliance)，自治性(governance)和风险性(risk)；危机处理和灾备；云计算退出机制等等。

只有这些问题在实用上真正的解决，才有真正意义上的商业(enterprise)云计算。

关键云基础设施(Critical Infrastructure for Cloud Computing)包括计算机网络、大型数据中心和能源。它们的安全和保证是云安全的前提。有许多标准，运作优化和法律跟进的工作需要解决。

4. IT 安全审计

越来越多的云计算专家，学者和第一线工作的专业人员意识到云计算的安全和保证是不能光靠技术的进步来取得。它们牵涉到许多知识领域。连接这些领域的节

⁶ <http://www-01.ibm.com/software/data/bigdata/>

⁷ Zhixiong Chen and John Yoon, IT Auditing to Assure a Secure Cloud Computing, In the proceedings of The 6th IEEE World Congress on Services, July 5-10, 2010, Miami, Florida, USA, pp253-259

点就是内部和外部的 IT 安全审计(Internal and External IT Security Auditing)⁸。

现代审计主要有二个支柱，一个还是传统的财务审计，另一个是最近几年发展起来的 IT 安全审计。IT 安全审计主要包括审计原理、指导框架、过程和各种辅助工具。比如，COSO (The Committee of Sponsoring Organizations of The National Commission of Fraudulent Financial Reporting)⁹原来主要是用在审计财务欺诈。现在它也被用在 IT 安全审计。CobIT 主要是 IT 自治管理的指导框架和内部控制的成熟模型¹⁰。它把任何 IT 活动定义为一个 IT 进程。对每个 IT 进程给出了一套要求，称为控制目标。通过这些控制可以提供一个比较可信的保证，商业目标就可以实现，不必要的事件就可以避免或及时发现。所以，CobIT 就给了 IT 安全审计员们一个指导纲领。其它的指导框架还有 ITIL¹¹。

IT 安全审计的过程可以简单地归类为，定义审计的范围，计划，在实地的考察和验证工作，撰写报告，记录所有的过程，以及审计以后的跟踪。

IT 安全审计是个很细致和复杂的工作，既要有对被审计的系统有所了解，要

有知识的积累，也要有与人交流的技能 and 写作的才能。需要把握很高的职业标准和道德底线。

5. 政府的作用

政府可以在推动云安全，或者一般的网络安全上起到应有的作用。他们可以通过立法来贯彻网络安全的基本要求，提出各种公共政策。也可以利用自己的财力资助各项研究，以及培养数字化和安全化的人才。

2009 年五月，美国总统奥巴马高调宣布，要发动全民运动去推动网络安全知识，扫除数字技术文盲。让网络安全从我们的象牙塔到课堂里，要打造 21 世纪的数字化武装的劳动大军。

在构建全民网络安全意识上，美国国防部在过去数年里在全美建立了十二个侧重于不同方向的卓越中心 (Center of Excellence)¹²。每个中心给与研究基金开展研究，教育和与其它院校和研究机构合作。比如，设在马里兰大学的恐怖主义研究和应对中心 (START)¹³是在 2005 年第一批成立的优秀中心。他们的目标就是要向有关国家安全的政策制定者和实际工作者提供以数据为依据的确实可靠的关于恐怖主义的人类原因和结果。他们建立了关于恐怖事件的数据库。纽约的 John Jay Criminal College 利用社会学原理也参与了数据库的建设。现在，每当任何一件恐怖事件，不管是国际还是国内的，他们都会在第一时间发布有关数据的采集和看法。

⁸ Zhixiong Chen, John Yoon and Yun Wang, Integrating IT Security Auditing to IA Curriculum, Proceedings from the 16th Colloquium (CISSE) (ISBN: 1-933510-95-1), Lake Buena Vista, Florida, June 11-13, 2012, pp149-156

⁹ <http://www.coso.org/>

¹⁰ ISACA (Information Systems Audit & Control Association, <http://www.isaca.org>), digital publication, COBIT4.1 (The Control Objectives for Information and related Technology. COBIT now is under the responsibility of ITGI (The IT Governance Institute, <http://www.itgi.org/>)

¹¹ <http://www.itil-officialsite.com/>

¹² www.dhs.gov/files/programs/editorial_0498.shtm

¹³ <http://www.start.umd.edu/start/>

国防部通过建立这些中心，把他们认为对美国安全至关重要的领域从政府推到了社会。利用全社会的人才来解决一些重大的课题。比如。解决陆地边境有效地抽查过边境的车辆和人员的算法就是在国防部的资助下完成的。

相对于国防部，美国国家安全局(NSA)的信息保证教育优秀中心(Centers of Academic Excellence in Information Assurance Education)则更具有广泛的代表性¹⁴。自从2004年开始，他已经授予了一百六十多个学校，教育和研究机构信息保证教育优秀中心。随后，他又开始了信息保证研究优秀中心和以社区学校为中心的信息保证教育优秀中心。2012年又开始了信息保证运作(Cyber Operation)优秀中心。首先，每个优秀中心的学科标准必须要通过国家安全系统委员会¹⁵制定的某项或多项标准。然后要通过对中心的团队，研究能力，学生素质的考查。

信息保证教育优秀中心并没有研究基金的支持。但是，由于它的广泛性，极大地推动了信息保证的教育。这是一个政府推动的很成功的例子。

虽然没有直接的基金支持，但是，只有信息保证教育优秀中心才可以有资格申请学生资助。它的要求是学生毕业后必须要为政府部门工作。时间的长短取决于学生要求资助的数目。这是一个很成功的项目。在工作不景气的状况下，许多学生还是愿意服务于政府。同时，中心也可以申请其他研究项目。

美国国家安全部还把每年的十月定为全民安全月¹⁶。迄今为止已办了七届了。通过这一全民月来提升全民的安全意识。理解和保护各种有形和无形的资产。

美国国家标准和技术(NIST)和国安局还启动了全国网络安全教育，出版了关于网络安全劳动力框架(The NICE Cybersecurity Workforce Framework)的咨询版。这份框架书是要解决关于网络安全工作没有公认的和统一的定义，就是同一个术语不同的人可以有不同的含义。这份框架书也想解决定义网络安全工作的性质和级别。这是一个很有意义的但只有类似国家标准和技术局才能推动的工作。这份框架书把网络安全工作分成七大类：

1. 安全(Securely Provision)：主要是关于安全IT系统的概念形成，设计和建立。也包括一些系统开发；

2. 运作和维护(Operate and Maintain)：主要是对IT系统的正常运行提供技术支持以保证这些系统高效率工作和安全；

3. 保护和防卫(Protect and Defend)：主要是对内部IT系统或网络进行身份识别，分析和缓解可能的威胁；

4. 调查(Investigate)：主要是对于IT系统，网络和数字证据有关的事件或罪犯进行调查；

5. 运作和采集(Operate and Collect)：主要是对很专业化的网络信息进行采集以期发展和得到任何有用的认知；

6. 分析(Analyze)：主要是对正来临的网络信息进行专业化的检阅和评估以期决定是否对认知的提高有帮助；

¹⁴ www.nsa.gov/ia/academic_outreach/nat_cae/index.shtml

¹⁵ <http://www.cnss.gov/>

¹⁶http://www.dhs.gov/files/programs/gc_1158611596104.shtm

7. 支持 (Support)：主要是提供网络支持以帮助别人完成他们的于网络安全有关的工作。

每一类又可以分为几大领域。比如，网络安全的合规性，有安全意识的软件开发，企业架构，技术演示，系统要求的规划，测试和评估和系统开发就归入了第一类。每一个领域定义了工作的条条和必须具备的技能，以及可能的工作头衔。通过这一框架性的工作就可以为政府或企业在网络安全工作上有了共同的语言。也为这些机构的技能分析提供了架构。

美国国防部和美国自然科学基金分别设立了特殊的奖学金以鼓励学生学习网络安全专用。学生们在毕业后必须要为政府服务。服务期限要看奖学金的数目和年限。许多学生或者年轻的退役兵就选择了这一很不错的途径，不仅解决了学费和生活费，也解决了就业的问题。有了在政府工作的经验就可以比较容易在私有企业找到工作。

其次，各种研究基金、参议院和众议院的基金也来支持网络安全的学习和研究。比如，美国自然科学基金的可信计算 (Trustworthy Computing) 就是比较有名的项目。

再次，资助全国性的网络安全竞赛。比如，Cyber watch¹⁷、Cyber Challenging¹⁸、Cyber Defense Competition¹⁹。与此同时，许多学校，机构和工业实验室被要求撰写各种新的比赛还在不断的出现。最近的是把安全和科学、技术、工程和数学 (STEM) 联系起来 (CyberStem)。

美国国防部还对美国公民开放一些实验室，学生们可以在那里进行学习和研究。比如，最早开放的 DC3²⁰。这个中心不仅提供实验设备和软件工具，还与学校合作，开发提供许多符合大学课程要求的课程。这些课程就可以转到各自的大学。不过，他要求学生具有美国国籍或至少是持有美国绿卡。

6. 结语

本文根据作者自己在网络安全的研究，教育和工作中把所观察到政府有作为的地方列举出来，希望能起到一个启示作用。

¹⁷ <http://www.cyberwatchcenter.org/>

¹⁸ <https://www.nbise.org/uscc>

¹⁹ <http://www.nationalccdc.org/>

²⁰ <http://www.dc3.mil/>

作者简介



Zhixiong Chen (陈志雄)

is a Professor and Director in Cybersecurity in the School of Liberal Arts at Mercy College, New York. He holds the president of Services Society, a non-profit professional organization promoting

research and technical collaboration on Services Computing based modern Services Science among academia and industrial professionals. He is also the Chair of Board Directors, Chinese Association of Science and Technology, USA, Greater New York Chapter. Dr. Chen's research interests are in information assurance and security within Services Computing including Cloud Computing. He has published more than 35 peer reviewed journal and conference papers, and served as program chairs, committee members in various IEEE conferences. He holds CISSP (Certified Information Systems Security Professionals), a senior member of IEEE and a member of ACM. He was the founding president of Services Society, a non-profit professional organization promoting research and technical collaboration on Services Science among academia and industrial profession.



中國旅美科技協會
Chinese Association for Science and Technology, USA (CAST-USA)
www.cast-usa.net