



2015 科研成果年报



中国科学院计算技术研究所



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

2015 科研成果年报

二〇一六年五月
中国科学院计算技术研究所科研处编

编委主任：孙凝晖

编 委：李锦涛 隋雪青 陈熙霖 程学旗

主 编：罗瑞丽

责任编辑：吴金凤

编 辑：孙 愿 李 丹 梁 丽 支 天

通 讯 员：董 慧 何玉晓 王春颖 黄 治 孙晓茜 李晓宇 田卫平 何文静 宋艳梅 任 菲

计算所简介

中国科学院计算技术研究所(简称计算所)创建于 1956 年,是中国第一个专门从事计算机科学技术综合性研究的学术机构。计算所研制成功了我国第一台通用数字电子计算机,并形成了我国高性能计算机的研发基地,我国首枚通用 CPU 芯片也诞生在这里。

在过去的五十多年里,计算所是我国计算机事业的摇篮。一直到 2030 年,计算所的定位是建成引领创新型战略高技术研究所,保障国家信息安全,引领产业技术方向,成为中国计算机产业人才与技术的源头。在计算技术学科的计算机系统、网络、智能技术三个主要研究领域,开展以体系结构与算法为特色的学术研究、技术创新、技术应用与技术转移,重点在龙芯处理器、曙光高性能计算机、天玑网络数据引擎上取得重大突破。截至 2015 年底,计算所获得国家、院、市、部级科技奖励 219 项,其中,国家级科技奖励 47 项(含非第一完成单位 10 项),院、市、部级科技奖励 172 项(含非第一完成单位 13 项)。

计算所坚持“基础性、战略性、前瞻性”的三性原则,坚持新时期办院方针,以“跻身国际前沿,关注国计民生,引领中国信息产业”为己任,落实“创新、求实”的理念,率先成为世界一流的科研学术机构。



目录

年度科研概况	1
科研成果与主要进展	5
重点 / 重大科研任务进展	5
海云计算系统关键技术与系统研究	5
神经网络处理器结构设计及应用研究	7
非精确计算的加速器结构与编程方法研究	9
医学影像大数据可计算性科学问题与关键技术研究	11
面向工业产品全生命周期的仿真公共服务平台	13
面向融合通信的多媒体内容审计平台	15
面向 LTE-Advanced 的终端软基带技术	17
面向 C-RAN 的低功耗通用处理器平台研发	20
面向服务的未来互联网体系结构与机制研究	22
网络大数据计算的基础理论及其应用研究	24
E 级超级计算机新型体系结构及关键技术路线研究	27
异构媒体数据的关联与挖掘研究	29
高精度低成本低功耗的海水电导率传感器研究与设计	31
基于信息流的多粒度云安全管控与追责技术	33
云环境下的图像视频群体协同表达与处理	35
在线社会关系网络的挖掘与分析	37
动态可配置的压缩感知成像系统	39
数据并行与线程并行合一的可伸缩处理器体系结构	41
高效通用数据中心体系结构研究	43
易编程的异构并行处理器结构	45
差错容忍计算器件基础理论与方法	46
面向气候和湍流模拟的百万量级异构众核可扩展并行算法与优化方法	48
大数据结构与关系的度量与简约计算	50
Web 信息检索与数据挖掘	52
多媒体内容分析	54
煤矿信息化关键技术及集成装备联合研发基地	56
新华社 713 实验室技术研究项目 - 大数据与智能信息处理	58
研究方向科研进展	60
获奖成果	72
视觉模式的局部建模及非线性特征获取理论与方法研究	72
社会化媒体数据的感知与处理关键技术及应用	74
远程信息无障碍交互关键技术及应用	76
结题验收的代表性成果	78
高通量计算系统的构建原理、支撑技术及云服务应用	78
3D 内容视觉获取技术及设备	80
超并行高效能计算机体系结构与设计方法研究	82
理论计算机科学	85
计算机体系结构	87
视觉模式分析与识别	88
大规模图像的内容分析与理解及其智能服务技术	89
多核编程框架研究项目	91
面向残疾人信息化的远程维权终端系统	93
基础研究项目群	95
龙芯 GS464E 处理器核的验证研究	97
科研基地进展	99
计算机体系结构国家重点实验室	99
移动计算与新型终端北京市重点实验室	102
中科院智能信息处理重点实验室	105
中科院网络数据科学与技术重点实验室	107
学术活动	109
知识产权	119
学位论文	140

年度科研概况

一、承担课题概况

2015年，计算所共有在研项目781项（包括新增项目214项）。其中，主持或参与中科院战略性先导科技专项课题A类1项B类3项（新增）、中科院STS项目3项（新增1项）、院重点部署项目5项（新增1项）；主持或参与工信部国家重大科技专项课题34项（新增2项）；主持科技部国家“973”项目3项、承担课题22项（新增2项），主持或参与科技部国家“863”课题44项（新增14项）、国家科技支撑课题17项（新增3项）；主持国家自然科学基金委创新群体项目1项、重大项目1项、重点类项目9项（新增5项）、国家杰出青年科学基金项目1项、优秀青年科学基金项目4项；承担横向项目200余项。

二、圆满完成所层面战略性工作

2015年，中科院对计算所十二五期间“一三五”任务完成情况的评价是：“定位准确合理，布局合理，高质量完成目标任务，取得了有重大影响的成果，保障措施与重大改革举措合理”。其中，重大突破“曙光高性能计算机”被评为优秀突破，重点培育“云计算服务器”被评为优秀

培育，科技布局与组织模式被评为亮点工作。

作为科教融合单元承办国科大计算机与控制学院，一年来，工作已经步入正轨。建立了院务会、学术委员会、教学委员会，成立了教研室，出台了《计算机与控制学院管理条例》。制定了本科生课程体系，建成了一批核心精品课程并修订完善了研究生课程体系，规范了教材。开设了2门公共必修课、3门专业基础课、4次科学讲座。

创新研究院申请获得批准立项，进入方案论证阶段。

按照院里的统一部署，开始制定计算所的“十三五”战略规划。

三、基础研究再夺硕果

国家基金委创新群体项目“超并行高效能计算机体系结构与设计方法研究”成功获得6年晋级9年的延续支持。科技部973项目“高通量计算系统的构建原理、支撑技术与云服务应用”顺利通过科技部验收。

基础前沿研究在细分领域走在了世界最前列。首个深度神经网络处理器——寒武纪，每秒能处理160亿神经元和2.56万亿突触，65nm验证片比Intel Xeon-E5能效比提升100倍，在



2015 年就形成了十余篇 CCF A 类顶级国际会议 / 期刊论文, 突破本领域最高期刊 TOCS (大陆首次), 入选了 Communication of the ACM 评出的研究焦点 (大陆首次), 并受到了 Intel、斯坦福大学, UCLA、苏黎世理工等欧美高水平机构的广泛引用, 开创的研究方向吸引了领域顶会 ISCA 5% 的投稿。

2015 年在基础研究方面取得的成果包括: (1) 面向 5G 的超级基站研究成果连续 10 个月在领域顶刊 IEEE Wireless Communications Letters 月下载量排名 TOP10; (2) 人体运动虚拟仿真成果发表在 ACM SIGGRAPH 2015, 同时全文被 ACM Transaction On Graphics 期刊收录 (中科院一区); (3) 高通量冷冻电镜数据处理成果发表在 Journal of Structural Biology 上, 相关软件 MarkerAuto 在生物物理所、清华大学、UCSD、NCMIR 等国内外 20 余家单位安装应用; (4) 面向 NVM 主存的数据管理研究发表在 VLDB; (5) 社交媒体大数据传播分析的成果发表在 Science、PNAS、SIGIR、WWW、AAAI 上, 近两年被包括 Nature、Science 等刊物引用 300 余次; (6) 在多个竞赛中取得优异成绩, 如获得阿里大规模图像搜索竞赛第 1 名; 获得 ICCV2015 举办的社会事件分类竞赛第 1 名和表观年龄估计竞赛第 2 名; 获得 FG2015 举办的视频人脸识别竞赛第 1 名; 获得 IJCAI 数据挖掘竞赛第 1 名; 获得 NIST 汉英机器翻译竞赛第 4 名; 获得 ImageNet 视觉识别竞赛第 4 名。

四、未来网络研究迈上新台阶

计算所作为主要单位共建的江苏未来网络创新研究院, 成为国家重大科技基础设施 CENI 项

目的法人建设单位, 将承担 SDN 主干网络总体设计、核心设备研制、联邦与试验管理系统研制等重要工作。面向服务未来网络体系结构在网络模型与路由机制取得进展, 在 CCF A 类会议与期刊发表 10 多篇学术论文, 相关关键技术 in 工业界逐步得到应用, 仅华为 2015 年就与本方向启动了 8 项合作研究项目, TCP 优化协议已经在华为 4K 高清项目获得应用, 吞吐量提升了 7 倍。以服务路由器为基础, 正与相关单位合作研制星载路由器。

五、移动通信技术取得系统性进展

面向天地一体化信息网络的重大需求, 形成从地面终端、基站到测试系统的地面系统整套核心装备, 并成功应用于某卫星通信系统。2015 年卫星终端芯片流片成功, 同时完成了专用测试系统的研制, 用于终端和芯片的认证测试。

六、海洋数据融合系统脱颖而出

面向海洋信息化, 研制的海洋数据融合系统汇聚全球近百万船只实时船位数据、海洋水文气象数据, 全球船只档案数据、近五年历史航迹及海洋锚地、水道、码头、气象等数据, 来自 20 多颗海洋观测卫星、16 个国际地面站及数千岸基站信息, 通过数据清洗、特征抽取、特征关联、异常发现及目标融合处理等关键技术, 形成按时间、区域、船型、国别、行业等定制实时分发能力, 可为海军、海警、渔政、海关等涉海重大应用提供数据引擎。

七、高通量 DPU 取得可喜进展

DPU-m 芯片面向数据中心的视频编解码应用, 片上集成高效的视频编解码单元, 以较低功

耗提供超高性能，在高密度视频处理场景下，跟 Intel 方案相比，在芯片级具有 20 倍以上的能效比，在系统级具有 50 倍以上的能效比。通过与国家网络安全中心合作，DPU-m 加速卡进入高通量视频监控平台；通过与歌华有线、视博云公司合作，推出了高通量有线电视平台。

八、创新人才不断涌现

陈熙霖研究员当选 IEEE Fellow，张勇东研究员入选国家基金委杰青，韩银和研究员、侯锐研究员、陈天石副研究员入选国家基金委优青，陈云霄研究员入选 2015 年度 MIT TR35（全球 35 位 35 岁以下杰出青年创新者），蒋树强研究员入选中组部万人青年拔尖人才计划，程学旗研究员和季统凯高工入选科技部创新人才推进计划，程学旗研究员和谢高岗研究员入选百千万人才工程国家级人选，山世光研究员获 CCF 青年科学家奖。

引进的孙贤和教授入选国家千人计划 B 类，李炼研究员获得中科院百人计划择优资助，此外，今年还引进了王元陶研究员，欧阳文涛、韩琥、郭琦三位副研究员和刘宇航助理研究员等海外人才。

九、科技奖励硕果累累

2015 年获得的科技奖励包括：（1）“视觉模式的局部建模及非线性特征获取理论与方法研究”获得国家自然科学奖二等奖；（2）“32 位星载容错控制计算机系统关键技术及应用”获得国家科技进步二等奖（第二完成单位）；（3）“基于大数据的互联网机器翻译核心技术及产业化”获得国家科技进步二等奖（第五完成单位）；（4）“普适计算关键技术及支撑平台”获得国家科技进步二等奖（第六完成单位）；（5）“社会

化媒体数据的感知与处理关键技术及应用”获得北京市科学技术奖（技术发明类）二等奖；（6）

“远程信息无障碍交互关键技术及应用”获得北京市科学技术奖（社会公益类）二等奖；（7）“互联网支付体系风险防控的行为分析技术及应用”获得上海市科技进步一等奖（第三完成单位）；（8）

“龙芯处理器抗辐照设计关键技术”获得 CCF 科技进步一等奖；（9）“大数据挖掘算法与云服务”获得吴文俊人工智能科学技术二等奖。此外，

“一种四路服务器主板”获得国家知识产权局中国专利奖（优秀奖）。

十、技术转移全面开花

计算所今年为中科晶上公司和中科天玑公司分别融资 4000 万和 1.5 亿。

龙芯中科公司销售首次过亿元，实现盈利；龙芯 3B 占据党政公文办公一期试点 85% 以上份额，销售芯片 1 万多颗；龙芯 1H 耐 175℃ 超高温，是石油勘探钻头的专用芯片，打破了国外的垄断。

与曙光公司合作研制了地球系统数值模拟系统原型——硅立方，在北京中关村软件园正式落成，原型系统的建设为国家大装置的建设提供了坚实的基础。基于 SimICT 平台的模拟产品在航天领域的应用进一步取得突破，为我国自主研制航天操作系统提供验证与设计平台。

中科睿芯公司荣获首届“德勤 - 中关村明日之星”奖。

中科蓝鲸公司面向全球提供大数据存储产品及解决方案，服务超过 80 多个国家的海外用户，在意大利国家电视台（RAI）和瑞士 SwissTV 部署容量约 500TB，服务超过 3 万个用户，在印度 ViaCom18 集团部署容量超过 1PB。

在鼓励创新创业方面，有 9 位员工进行创业



或参与创业团队，计算所向下属企业输送了 12 位员工。抓住新的《促进科技成果转化法》10 月 1 日颁布实施的机遇，以 180 万的总价成功出售 9 件专利。华为联合实验室 2015 年新达成项目 9 项，额度超过 2700 万。

十一、分所工作服务创新驱动战略

浙江分所以高性能永磁交流伺服电机为市场切入口，配套研发了电机的核心部件与自主伺服驱动，形成数字化工厂核心体系，年销售超过数万台。

济宁分所开发的 G+ 智慧家居战略平台是支持格兰仕全品系家电的云接入服务平台，依托“智城云”数据中心能实现百万级并发智能家居。

宁波分所创客空间激发了宁波地区的创新基因，打造了创新氛围。

太仓分所奥视佳数字影棚以绿幕棚为基础，解决了前期拍摄与后期制作的分离问题，先后支

持了央视《第 N 度空间》，湖南卫视《偶像来了》，TFboy MV 等节目的拍摄。

十二、国际合作交流日益活跃

计算所支持鼓励科研人员既要“走出去”，参加各类学术会议，又要“请进来”，邀请国内外知名学者到计算所来交流讨论，举办高水平国际学术会议等。本年度出访 196 人次，参加各类国际学术会议 109 人次、并作报告 111 人次；邀请国内外知名学者来所作学术报告 86 人次，举办 4 场国际学术会议；程学旗研究员应邀担任 WSDM'15 大会主席，陈云霁研究员应邀将担任 ASPLOS 2017 大会主席；继续引进国外专家和青年学者开展合作研究，2015 年引进了比利时的 Mathy Laurent 教授，被纳入中科院“国际人才项目访问学者计划”。

此外，计算所科研人员还积极承担或参与国家基金委和科技部的相关国际合作项目。

创新人才

创新文化

创新环境

科研成果与主要进展

重点 / 重大科研任务进展

海云计算系统关键技术与系统研究

“海云计算系统关键技术与系统研究”为中国科学院战略性先导科技专项 A 类项目课题（批准号：XDA06010400），起止时间是 2012 年 1 月至 2016 年 12 月。

以云计算模式为核心的现有信息服务系统在数据处理性能和能耗方面无法应对由终端规模急剧增长所带来的挑战。为适应“人-机-物”三元融合的趋势，以建立“感知中国”信息网络为背景，本任务提出海云计算系统的理念和关键技术，搭建海端和云端设备共同参与任务执行的计算系统，用以构筑一个海云协同计算和数据处理环境，从而为实现计算处理系统千倍效能比提升的目标探索一条有益途径。2015 年，本课题在计算领域取得了国际领先的技术突破，产生了重要的学术影响和产业影响。

首先，通过可重塑处理器设计思路，在国际上首次验证了“性能功耗比提升千倍”（1000 GOPS/W）的可行性，从部件层次破解了性能与

能耗挑战。在 2015 年可重塑处理器成功进行功能测试片流片，采用 65nm 工艺，功耗不超过 20 瓦。完成了基于 FPGA 的可重塑处理器样机，能运行 CNN/MLP/DNN 类神经网络算法，相比主流 CPU 提高了 100 倍性能功耗比。

其次，在 2015 年海云服务器持续进行关键技术优化和系统集成。通过消息式内存、一体化存储、数据中心网络、资源按需管理可编程体系结构及资源共享互连等技术原型，用以缓解数据中心资源利用率低，应对应用多样性，提高数据中心服务器内部各组件及整体性能。其中，采用消息式内存的程序执行速度达到标准 DDR3 总线的 4 倍。为降低远程资源的访问开销，海云服务器提出了支持细粒度与粗粒度全局高效共享访问的远程设备资源访问机制，结果表明在自研 FPGA 平台上应用级别跨芯片内存访问的通路延迟性能优于商业高端互连系统。

第三，海云存储完成高可靠、低冗余开销、

高访问性能的网络盘阵系统，空间冗余开销相比谷歌 GFS 的 3 倍降低了 1.66 倍。通过将冗余处理与用户 IO 访问解耦，以副本的方式处理用户写，在后台异步低开销转换为纠删码方式存储，避免了纯纠删码固有的低访问性能与冷数据冗余转化高额的性能开销，有效解决了存储系统的高可靠、高性能及存储资源利用率三大目标间的矛盾问题。该系统已达到产品级可应用的成熟形态，在电信

环境中已有应用，并获得了包括 Intel 的国内外投资人的投资。

上述部分科研成果已发表于 2015 年的 ISCA、ASPLOS、CGO 等多个顶级国际会议，申请了多项国内外专利，其中代表性工作参与了 2015 年 8 月的院先导中期检查成果展示，并获得了由国内知名高校和相关行业的院士和专家所组成的评审专家组高度认可。



图 1 2015 年 8 月院先导中期检查成果展示现场



图 2 可重塑处理器功能验证片样片

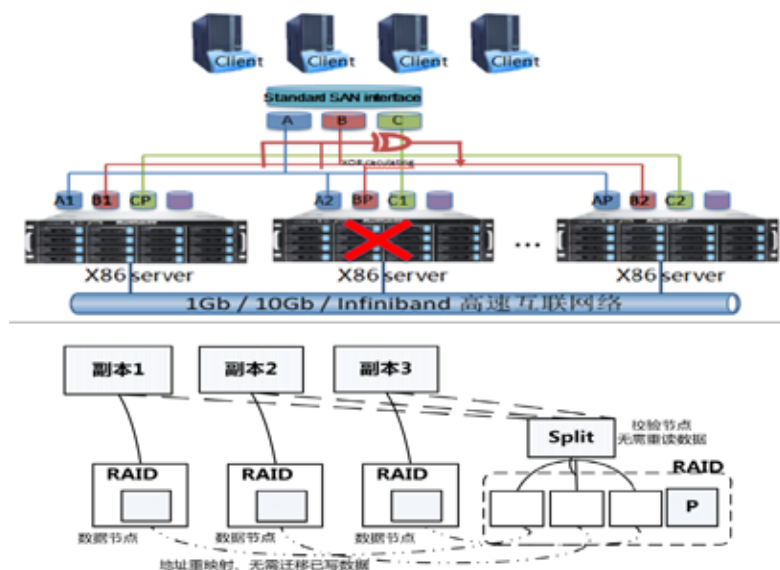


图 3 海云存储高可靠网络盘阵系统

重点 / 重大科研任务进展

神经网络处理器结构设计及应用研究

“神经网络处理器结构设计及应用研究”是中国科学院战略性先导科技专项 B 类项目子课题（批准号：XDB02040009），起止时间是 2015 年 1 月至 2016 年 12 月。2015 年度，在神经网络处理器体系结构研究、设计实现及测试等方面进行了探索研究，取得了一系列成果，完成了预期目标。

下面介绍本项目各课题本年度取得的突出进展。

在本项目中的神经网络处理器的编程关键技术研究，本项目主要研究了如何让程序员克服编程墙，方便地让多个处理器核高效协同工作。深度神经网络规模大、结构复杂，不可能直接映射到处理器上进行运算，而需要分块分时映射到不同处理器核进行处理。不当的分块方案可能影响整个处理器的实际性能，而普通程序员直接对神经网络进行分块很难获得好的效率。因此，我们提出了自动化分块调度技术以解决上述问题，通过对神经网络算法的自动化分析，提取出数据重用特性，以此为基础进行分块。

在 2015 年度，本项目在核心科学问题的研究方面完成了预期目标。共发表论文 14 篇，

其中，SCI 收录 11 篇，包括 IEEE Transactions on Computers、IEEE Micro、ACM Transactions on Architecture and Code、IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems、ACM Transactions on Design Automation of Electronic、IEEE Transactions on Circuits and Systems、IEEE Transactions on Evolutionary Computation、IEEE Transactions on Parallel and Distributed、等国际期刊和 ISCA、The International Symposium on Code Generation and Optimization、Design Automation & Test in Europe 等国际会议。申请并受理发明专利 15 项。所取得的阶段性研究成果对本项目明年的研究奠定了坚实基础。

产生的学术影响：

2015 年，在神经网络处理器结构设计这一领域的不同方面进行了深入研究。其中发表的 14 篇文章中，包括了 CCF 的 A 类国际学术期刊 TOCS 一篇，TOC 两篇，TCAD 一篇，CCF 的 B 类国际学术期刊 TACO 一篇，TODAES 一篇，发



表的国际会议文章包括了 CCF 的 A 类国际会议 ISCA 一篇, ASPLOS 一篇, MICRO 一篇, CCF 的 B 类国际会议 DATE 一篇。这些发表在国际顶级期刊和会议上的众多文章, 具有很大影响力。因此陈云霁研究员入选 2015 年度 MIT TR35, 并受邀担任了 CCF 的 A 类国际会议国际微体系结构会议 (MICRO) 程序委员会成员和 IEEE 国际计算机体系结构 (ISCA) 技术程序委员会成员。

其中在神经网络处理器的体系结构研究中的成果产生的论文 PuDianNao: A Polyvalent Machine Learning Accelerator 发表在了 ASPLOS 2015 (CCF 评级 A 类的国际会议) 上, 在已有的面向单一场景和单一算法的神经网络处理器的基础上, 选择差异化的算法进行加

速, 并根据了算法的计算模式和数据局部性进行了体系结构加速设计, 设计出的加速器架构可以适用于通用神经网络处理应用中。另两篇论文 ShiDianNao: shifting vision processing closer to the sensor 和 Neuromorphic Accelerators: A Comparison Between Neuroscience and Machine-Learning Approaches 发表在了 ISCA 2015 (CCF 评级 A 类的国际会议) 和 MICRO 2015 (CCF 评级 A 类的国际会议) 上, 其针对于特定相关神经网络应用领域进行了加速, 取得了突出成果。以上在神经网络处理器领域的研究取得了突出性的成果, 在热门的人工智能领域研究的体系结构设计领域, 得到了国内外同行的认可。

坚持“基础性、战略性、前瞻性”
的三性原则

重点 / 重大科研任务进展

非精确计算的加速器结构与编程方法研究

“非精确计算的加速器结构与编程方法研究”是中国科学院国际合作局对外合作重点项目（批准号：171111KYSB20130002），起止时间是2014年1月至2016年12月。2015年度，在非精确计算的高能效加速器的结构设计原理与方法，以及与之相适应的编程方法和系统等方面进行了探索研究，取得了一系列成果，完成了预期目标。

下面介绍本项目各课题本年度取得的突出进展。

在2015年度，在面向非精确计算的加速器结构与编程方法研究领域，对于误差传播理论模型，误差控制方法和误差反推技术进行了深入研究。通过误差的局部化，确保在非精确电路达到一定规模的前提下依然能将误差控制在给定范围内。实现了通过对电路的行为的研究，从而了解误差的传播规律，并能根据加速器的最终误差输出，结合加速器结构，反推出各个子模块边界处的误差累计情况。

在2015年度，本项目在核心科学问题的研究方面完成了预期目标。共发表论文14篇，其中，SCI收录11篇，包括IEEE Transactions

on Computers、IEEE Micro、ACM Transactions on Architecture and Code、IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems、ACM Transactions on Design Automation of Electronic、IEEE Transactions on Circuits and Systems、IEEE Transactions on Evolutionary Computation、IEEE Transactions on Parallel and Distributed、等国际期刊和ISCA、The International Symposium on Code Generation and Optimization、Design Automation & Test in Europe等国际会议。申请并受理发明专利15项。所取得的阶段性研究成果对本项目明年的研究奠定了坚实基础。

产生的学术影响：

2015年，在非精确计算的加速器结构与编程方法这一领域的不同方面进行了深入研究。其中发表的14篇文章中，包括了CCF的A类国际学术期刊TOCS一篇，TOC两篇，TCAD一篇，CCF的B类国际学术期刊TACO一篇，TODAES一篇，发表的国际会议文章包括了CCF的A类



国际会议 ISCA 一篇，ASPLOS 一篇，MICRO 一篇，CCF 的 B 类国际会议 DATE 一篇。这些发表在国际顶级期刊和会议上的众多文章，具有很大影响力。因此陈云霁研究员入选 2015 年度 MIT TR35，并受邀担任了 CCF 的 A 类国际会议国际微体系结构会议 (MICRO) 程序委员会成员和 IEEE 国际计算机体系结构 (ISCA) 技术程序委员会成员。

在“非精确计算的加速器结构与编程方法研究”的课题下，陈云霁研究员和陈天石副研究员带领的课题组完成了非精确计算神经网络

多核处理器 (Neural Network Multicore Processor, NNMP) 的体系结构设计，提出适用于神经网络多核处理器的编程关键技术，为其实用化铺路。目前以上论文及专利研究的成果已经进行了流片，并通过了基本功能测试。并搭建了非精确加速器的 RTL 设计及验证体系机构模型，构建非精确加速器软件抽象层。以上成果在处理器结构设计的创新，有望在未来使得每个手机这样的移动终端具备谷歌大脑级别的认知处理能力。

跻身国际前沿

关注国计民生

引领中国信息产业

重点 / 重大科研任务进展

医学影像大数据可计算性科学问题与关键技术研究

本课题是院重点部署项目“医学影像信息大数据若干关键科学问题研究”的子课题（批准号：KGZD-EW-T03-2），起止时间是2014年5月至2017年4月。旨在通过对医学影像信息大数据的可计算性关键科学问题的深入研究，突破制约影像信息大数据在临床诊治应用、医疗资源共享及医疗新技术发展方面的底层高效计算框架、并行存储关键技术和快速访问技术瓶颈，为医学影像大数据技术的发展提供高效存储与计算框架的支撑。

2015年度主要研究内容如下：

1. 键值存储系统 ICTBase 性能和功能优化

ICTBase 系统设计的初衷是满足大规模图像、医疗相关联多维信息的高效存取，基于分布式 K-V 模型和类 SQL 分布式关系存储模型，ICTBase 拥有良好的性能、扩展性和灵活性，在内层存储架构上，ICTBase 选用 LSM 树作为存储引擎。为了进一步满足 ICTBase 对大规模影像数据存储的场景需求，一方面，本课题针对 ICTBase 既有 LSM 存储引擎在查询性能上的不足，研究了一种新的可动态适应的合并算法，该合并算法通过减少每次查询涉及的文件数量，进而提高查询效率，另

一方面，针对 ICTBase 系统索引单一依靠主键来支持定位目标数据，系统原有的查询方式不支持全文查询，本课题研究并开发了 ICTBase 系统的全文索引，由此支持存储医疗相关数据的内置全文检索。

2. 通用大规模并行计算平台架构体系

大规模医学影像数据处理应用对并行计算框架提出了很高的要求，不仅需要解决批量数据的处理问题，还需要解决实时数据的大规模计算问题。本年度课题组研究了一套支持离线计算、实时计算分布式内存计算等多种分布式计算模式的大数据计算框架，并在此基础上形成了天玑大数据分析引擎系统（Big Data Analysis, BDA）。BDA 系统将大数据分析计算所需要的 ETL、模型训练、评估和在线预测整合成流水线，并提供图形化配置和运行监控界面，实现大数据分析任务一站式配置和运行。同时，系统提供离线数据批量分析、流数据分析、在线数据实时分析等多种应用模式，帮助用户快速实现大数据处理流程，降低普通用户开发大数据分析应用的门槛，提高开发效率。

3. 大规模医学图像并行计算应用研究

课题在上述大数据并行处理技术的基础上开展了基于内容的医学影像检索应用研究。

基于内容的医学影像检索 (Content-Based Medical Image Retrieval, CBMIR) 利用图像处理技术从医学影像中提取灰度分布、纹理、形

状和拓扑等底层特征, 构建能有效表征影像特征的数据向量, 进而通过特征匹配、相似度计算等找出“相似”的医学影像。该方法可以实现影像之间的量化分析与比较, 具有较强的客观性, 能够快速有效地进行大量数据的分析与搜索, 可以广泛应用于 PACES、HIS、RIS 等医院信息网络系统。

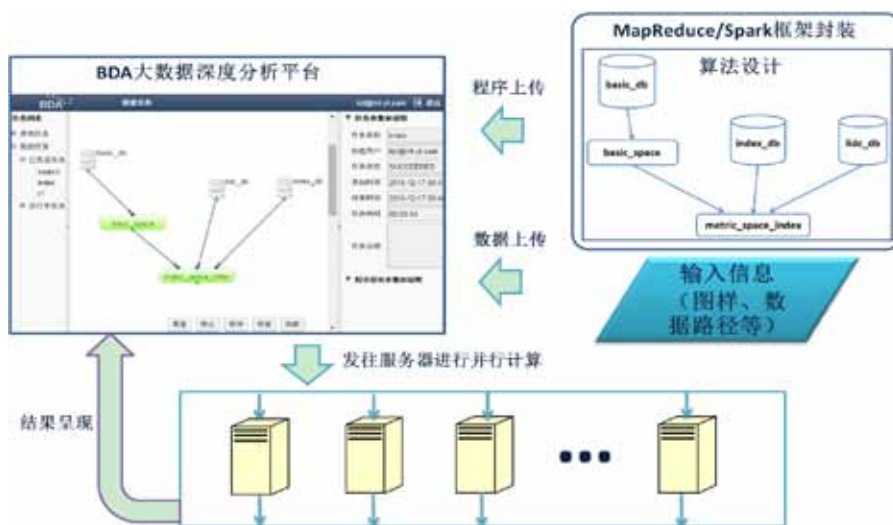


图 1 CBMIR 系统的分布式架构设计



图 2 医学影像样本索引建立分布式算法执行



图 3 医学影像图像搜索分布式算法执行

重点 / 重大科研任务进展

面向工业产品全生命周期的仿真公共服务平台

“面向工业产品全生命周期的仿真公共服务平台”是中国科学院科技服务网络计划的专项课题（批准号：KFJ-EW-STS-045），起止时间为2014年5月至2015年12月，主要任务是通过搭建工业产品全生命周期仿真公共服务平台，构建能覆盖产品全生命周期的仿真公共服务能力体系。

2015年该项目的主要进展和成果如下：

1. 面向工业产品全生命周期的资源池建设

在软硬件资源方面：项目组已购置1台工业级3D打印机和1台工业级3D扫描仪；

在数据资源建设方面：项目组搭建了规范库，涵盖主要机械制造业国家标准和行业标准，共计900多条；项目组为工业典型产品的结构、流体以及热等多物理场的仿真构建了案例库，共计257个；建设零部件3D模型库，涵盖1102个工业常规零部件；

在智力资源建设方面：完成与业内6家研究院、10所高校的智力资源共享与整合，平台现有的专家库中约358位专家。

2. 面向工业产品的全生命周期仿真云服务平台

基于计算机辅助工程（CAE）仿真与结构优化技术、快速成型以及虚拟展示技术，项目组构建了一个面向工业产品全生命周期仿真的云服务体系。主要内容如下：

1) 面向工业产品设计的智慧仿真云模块

该模块提供产品几何设计、结构仿真与优化服务以及个性化定制设计、仿真以及报告书的平台化接口，设计师对产品的外观、结构、功能提出设计方案，用户在该模块按提示进行操作，下载计算报告书，获取对该设计方案的合理性评价的相关信息以及对产品原始设计方案的优化方法。

2) 面向工业产品生产的智慧工业云服务网络系统

智慧工业云服务网络系统（图1）嵌入到企业接单、设计、生产、销售的全产业链，用户通过云服务终端，能随时随地按需获取制造资源 and 能力，最终实现对产品全生命周期涉及的人及物进行智慧地感知、互联、协同、分析。

3) 面向工业产品交易的“奥视界”三维虚

拟展示系统

“奥视景”三维虚拟展示系统（图 2）由云服务、“奥视景”软件、智能转盘、专业摄影灯箱组成。可实现快速拍摄、制作产品的全方位展示动画，同时，通过“奥视景”云服务快速对接现有移动互联网营销平台，实现即时制作、快速分享。

3. 构建了面向工业产品全生命周期的仿真公共服务体系

为推动平台在本地区进一步落地，联合苏州市优质工业企业资源，成立了苏州市工业设计与 3D 打印产业联盟，依托成立的联盟，项目组为一百多家企业客户，针对产品从研发、设计、生产、销售等相关的产业链，提供了技术服务。

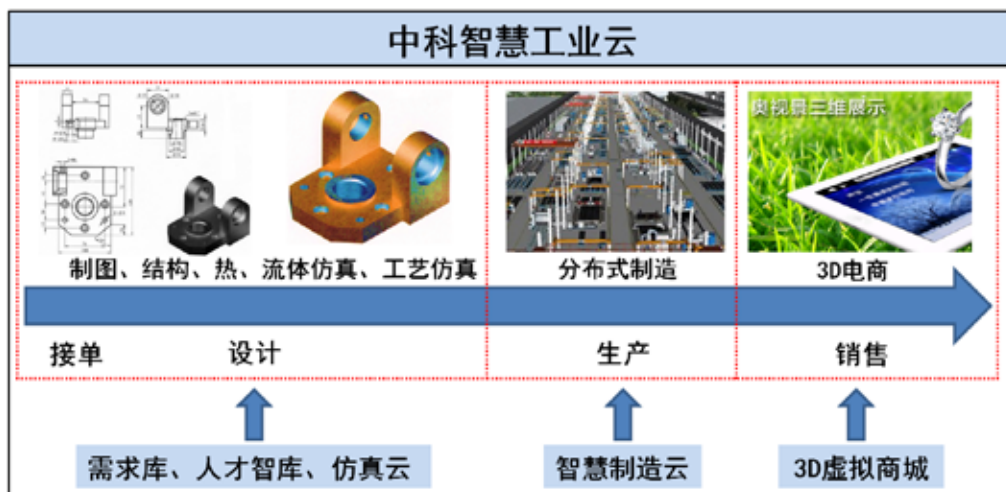


图 1 智慧工业云嵌入全产业链



图 2 “奥视景”三维虚拟展示系统

重点 / 重大科研任务进展

面向融合通信的多媒体内容审计平台

“面向融合通信的多媒体内容审计平台”是中国科学院科技服务网络计划的专项课题，起止时间为 2015 年 1 月至 2016 年 12 月。中央全面深化改革领导小组第四次会议上通过的《关于推动传统媒体和新兴媒体融合发展的指导意见》明确指出：以微博、微信为代表的网络媒体已经成为舆论斗争的主战场，直接关系到我国意识形态安全和政权安全。所以加强对网络媒体的监管已经迫在眉睫。

本课题针对网络媒体发现难、处置难的特点，有针对性的提出一个面向融合通信的多媒体内容审计平台，如图 1 所示，对融合通信中的多媒体数据实现自动审计，提高融合通信中多媒体数据内容监管效率，降低劳动力成本，提高融合通信的服务质量，确保融合通信业务的健康发展。

此平台已完成的两项核心技术：

1. 符合融合通信规范的数据接入技术

融合通信业务数据格式来自中国移动的相应的一个或多个业务系统，其中可能包含不同类型的数据传输协议、压缩格式，在接入内容审计平台前需要进行统一的协议剥离，格式转换。转换

后的数据格式规范由中国科学院大学与中国科学院计算技术研究所共同协商制定，并形成统一的数据接口、传输规范文档。最后由双方共同开发符合规范文档的数据交互系统模块。

2. 高性能视频解码技术

高效的视频解码算法是提高系统吞吐能力的先决条件。随着视频解码标准的发展，解码算法的复杂度持续提升，目前传统的单核处理器已经很难满足先进的视频解码技术对计算能力的要求。在这种情况下，我们历经多年研究，成功研制出基于众核 (Many-core) 处理器的并行视频解码专用设备，实现了软硬件协同的高效多媒体数据解码，如图 2。同时，为应对计算能力受限问题，提出了复杂度可分级多媒体数据解码技术；为应对网络数据丢包、错包问题，提出了基于多级语法校验的容错解码技术。这些技术能够有效保证审计平台的数据解码的效率和准确率。

目前本课题已经部署在线网应用，a. 在中国移动（洛阳）信息安全运营中心系统所处理流量的 20%，b. 进行多媒体内容审计试运营；c. 单节点处理能力达到 100MBps，d. 并发任务数 24；

e. 多媒体格式覆盖常见音视频的封装及压缩格式；f. 基于视觉特征的审计准确率达到 90%，特征库规模可达 10 万小时；g. 基于视觉特征的审计准确率提高到 99%，视觉语义概念检测准确率

达到 90%。除此之外，本课题在视频编解码、相似图像视频检测、视觉语义检测等核心技术方面具有自主知识产权和国际先进水平。

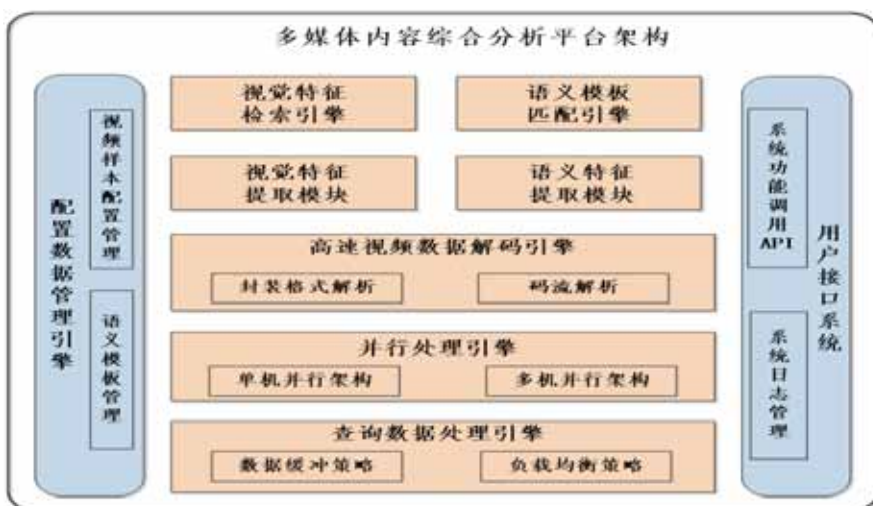


图 1 面向融合通信的多媒体内容综合分析平台架构图

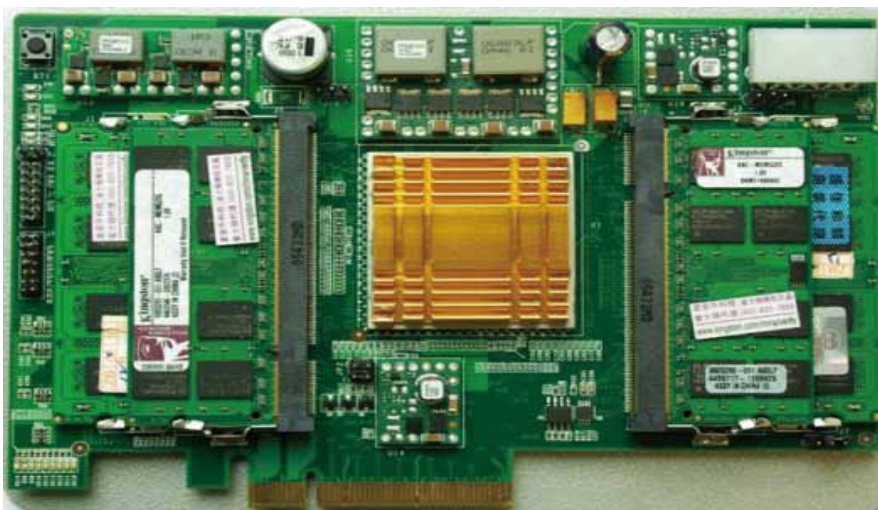


图 2 基于众核处理器的并行音视频解码卡

重点 / 重大科研任务进展

面向 LTE-Advanced 的终端软基带技术

“面向 LTE-Advanced 的终端软基带技术”是计算所牵头的国家科技重大专项“新一代宽带无线移动通信网”课题（批准号：2013ZX03003014），课题执行时间为 2013 年 6 月至 2015 年 12 月。

本课题以矢量处理器内核的研究和相关工具链开发为基础，对 2G/3G/4G/LTE-Advanced 的基带处理算法的并行化和矢量化、矢量和标量处理的优化等展开研究，设计并实现多模终端软基带处理器芯片架构，并通过原型验证平台和工程样片研制对课题研究成果进行验证和测试。

目前，本课题已完成全部研究内容，进入课题验收阶段。2015 年度，本课题取得的主要研究成果如下：

1. 高性能的自主矢量处理器内核

课题组根据 2G/3G/4G/LTE-Advanced 基带算法运算特点，设计了基带专用矢量处理器核及工具链。该自主矢量处理器核在指令集上支持 200 余条标量和矢量指令，其中近 150 条是针对基带处理优化的矢量指令集；在微结构上采用“SIMD+VLIW”混合架构，支持一条矢量运算指令完成 512bit 数据的并行运算，以及 6 条标量和向量指令的并行执行。在 SIMC 55nm 工艺下，单个矢量处理器核主频在 400MHz 时处理能力高达 100GOPS，完全满足 LTE-Advanced 的基带处理能力要求。矢量处理器核的具体技术指标如下表所示。

设计架构	微结构	△ SIMD/VLIW 混合架构 △ 6 VLIW 数据通路, 512-bit SIMD △ 9 级流水线 △ 512-bit 矢量运算; 32 路 16-bit 复数定点 MAC 运算 △ 硬件循环控制、多路 Load/store 结构, 增强 I/O △ AMBA AXI 接口 △ 增强 DMA
	指令集	△ 200 条指令 △ 近 150 条基带优化的矢量指令 △ 50 余条对程序流程优化的控制指令
性能指标	面积	△ 130 万逻辑门 △ 336K/208K Byte 紧耦合 RAM △ 5.37 mm ² @ 55nm CMOS 工艺
	功耗	△ 300mw @ 430MHz, 55nm CMOS 工艺
	时序	△ 430MHz @ 55nm CMOS 工艺

2. 多核软基带处理器芯片

课题组在矢量处理器内核及工具链研发的基础上, 通过集成 7 个并行处理的矢量处理器核和硬件加速单元, 设计并实现了面向 LTE-Advanced 的多模软基带芯片。该芯片的硬件架构如图 1 所示, 并在 SIMC 55nm 工艺下流片成功,

工程样片如图 2 所示。单颗多模软基带处理器芯片在 430MHz 主频下提供高达 1100GOPS 的定点运算能力 (含硬件加速单元的运算能力), 满足 2G/3G/4G/LTE-A 对基带芯片运算能力的要求, 达到了业界先进水平。

课题研发的终端软基带芯片在如下三个方面

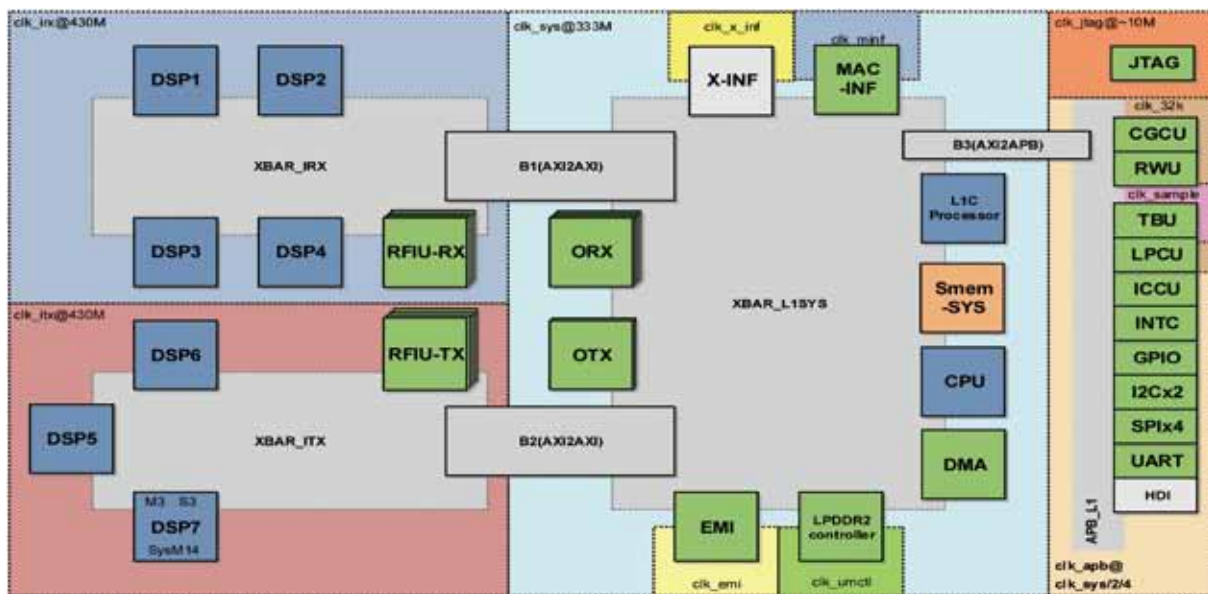


图 1 LTE-A 基带芯片架构



图 2 动芯系列终端软基带芯片的工程样片

取得了技术突破：

1) 性能：课题组从任务级、指令级和数据级对基带处理算法进行了矢量化和并行化设计，

以及标量和矢量运算的优化分配与调度，在 LTE-Advanced 的四载波聚合模式实现了下行接收 600Mbps、上行发送 300Mbps 的峰值数据速率，是国内首颗支持如此高速率的自主化的软基带芯片。

2) 功耗：课题组综合运用电路级、结构级和系统级的低功耗设计技术对芯片功耗进行优化，在峰值速率下芯片平均功耗小于 3W，达到了课题指标和实用化的要求。

3) 灵活性：软基带芯片可通过动态加载 2G/3G/4G/LTE-Advanced 等协议标准的基带软件来支持多模可重构计算，满足终端低成本和平滑设计的需要。

创 新

求 实

团 结

高 效

重点 / 重大科研任务进展

面向 C-RAN 的低功耗通用处理器平台研发

“面向 C-RAN 的低功耗通用处理器平台研发”是计算所牵头承担的“新一代宽带无线移动通信网”国家科技重大专项课题（批准号：2014ZX03003004），起止时间为 2014 年 1 月至 2016 年 12 月。

本课题目标为面向新型集中式 RAN 架构，研究并实现基于通用处理器平台的 C-RAN 系统基站池关键技术，包括：软基带 DSP 的物理层硬件加速器研制、基于集中式基站池的实时虚拟化技术、载波的动态迁移及扩容，突破目前在通用处理器平台上实现 C-RAN 基站池面临的功耗高等技术瓶颈，提高系统单位功耗载波处理能力，降低系统功耗，从而满足未来异构融合、绿色、智能蜂窝通信网络建设和运营的需求。

2015 年度的主要工作进展如下：

1. 超级基站原型研制

本课题提出的超级基站设备面向未来移动通信产业发展需求，融合了超级计算技术与无线通信技术，属于移动通信网络架构创新。课题组已经完成了超级基站工程样机（如图 1 所示）的研发，本年度在支持资源虚拟化、系统动态可重构等方

面进行了研发，并从协议和管控两个层面完善了异常处理和故障处理机制，提升了系统稳定性和可靠性。目前，该工程样机已经通过了第三方测试，在功能、接口、异常、压力、时延等各项测试中均超出研制指标要求。

同时，基于已完成的超级基站工程样机，本年度完成了面向未来移动通信系统的教学实训原型平台；采用软件无线电和软硬件解耦的思想，提供面向移动通信系统的开放、集成的开发环境；支持 2G/3G/4G/5G/ 宽带无线接入 / 卫星等各种无线通信标准的实现、实验网络的快速部署与相关技术验证；支持射频、基带、协议、核心网等不同层次的灵活可配置、可替换、可扩展。



图 1 超级基站工程样机系统

2. 超级基站资源智能管控

基于超级基站架构，针对潮汐效应带来的负载差异性和负载动态变化，以及资源在虚拟映射时的资源利用率等问题，进行了如下研究：首先，提出一种和网络负载变化相适应的处理资源分配方法，研究了负载差异性条件下的集中式最优资源映射机制；第二，提出了多协议处理实体在同一个处理器上执行时的协议组织和运行机制，通过动态地对下行数据处理进行提前，在不降低处理资源利用率的前提下，保证协议处理的实时性；

第三，研究了小区休眠技术在超级基站中的实现方法，如图 2 所示，超级基站可以通过集中式管控单元（OMU）对高虚拟化程度的处理资源和 RRH 进行动态映射和开关，同时，对比传统分布式蜂窝网络架构下的 traffic-aware 休眠机制，提出了 centralized-best 休眠机制，不仅能够通过优化算法得到全局最优部署方案，而且通过对业务的预测和休眠模型的建立，降低了休眠检测与决策的系统开销，减少了频繁切换带来的问题，提高了整体系统的稳定性。

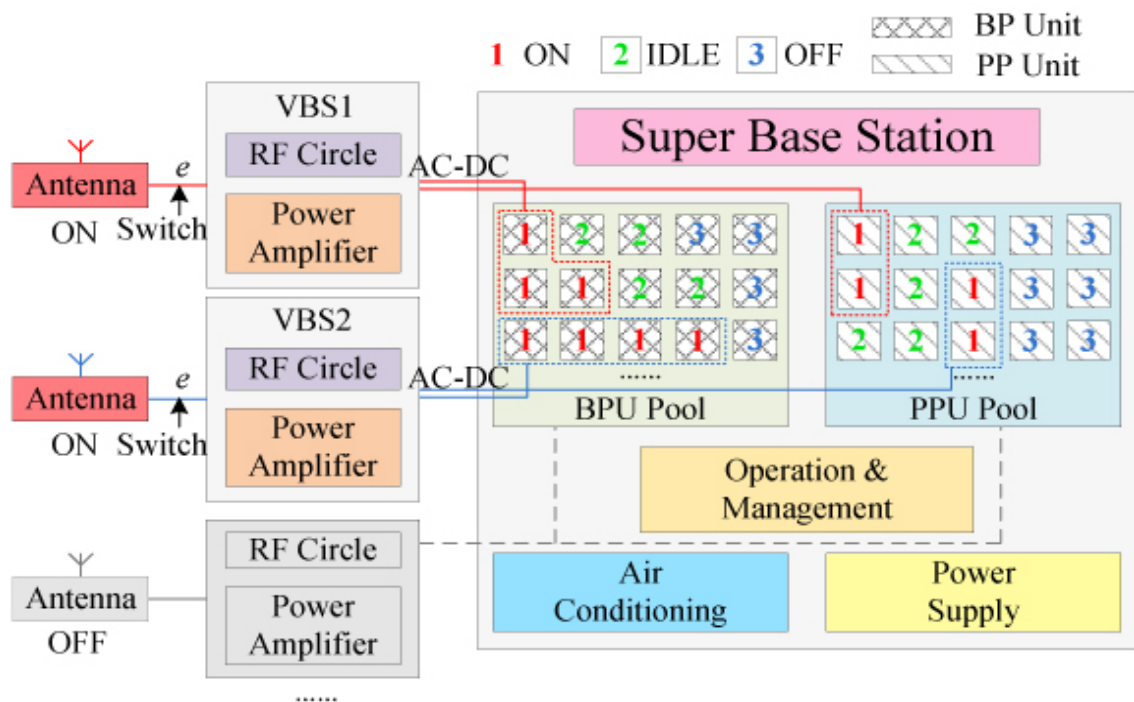


图 2 超级基站休眠机制示意图

重点 / 重大科研任务进展

面向服务的未来互联网体系结构与机制研究

“面向服务的未来互联网体系结构与机制研究”是计算所牵头的国家 973 计划项目（批准号：2012CB315800），起止时间为 2012 年 1 月至 2016 年 8 月。本项目针对互联网面临的可扩展性、动态性、安全可控性等问题，围绕着传输与服务的动态复杂耦合、海量差异化服务的透明映射以及可扩展服务路由与高效传输等三个关键科学问题，研究未来互联网体系结构，在网络设计与运行机理层面取得重大突破，形成一系列核心理论和机理方法。

截止 2015 年 12 月，本项目在体系结构模型、关键技术研发、研究平台建设等三个方面取得的主要进展如下：

1. 未来互联网体系结构的理论模型方面

通过对互联网最为典型的内容分发应用的深入分析、测量与建模，推导出能够反映内容流行度分布和变化趋势的理论模型，在此基础上创新性提出了面向服务的未来互联网体系结构，其核心思想是将网络中对数据的处理抽象为服务，通过灵活的服务命名、寻址和实例化等实现用户对服务的高效获取，推动互联网从以地址为

中心逐步演化到以服务为中心，从而更好地支持互联网应用模式的创新，有效应对现行互联网面临的挑战。

2. 未来互联网体系结构的核心机制方面

以服务为中心的互联网体系结构需要解决服务命名与寻址、最优化服务部署、高效服务缓存等核心机制。针对服务命名与寻址问题，提出了 SAIL 高效名字查找算法，将整个路由查找按照处理过程和名字前缀长度进行分类，可有效减少核心路由片上内存使用，极大提升名字查找效率；针对最优化服务部署问题，提出了基于全局视图的服务部署策略，通过节点间的有效协同进行服务部署，实现全网最优的服务访问效率；针对服务实例的缓存效率问题，提出了基于网络编码的缓存机制，通过中间网络节点对服务请求的随机转发和对数据块的随机编码，实现最优的隐式协同缓存效率。

3. 未来互联网体系结构的验证与应用方面

试验验证是网络技术研究的最基本手段。在研究平台建设方面，针对未来互联网的性能评估

和验证需求，项目建设和支持软件定义与虚拟化的未来互联网试验床，满足在真实网络环境中对项目提出的体系结构及核心机制进行实验验证的需求。试验床部署了网络功能虚拟化系统，能够对用户提供定制化网络服务，目前已经支持网络安全检测和防御、面向服务缓存的文件传输等多项技术试验，为项目成果的综合集成与应用提供了重要的基础平台。

基于上述研究，本年度共发表国际论文 84 篇、国内论文 36 篇（其中 SCI 收录 45 篇、EI 收录 59 篇）；申请国家发明专利 21 项，获专利授权 5 项。在本项目资助下，共累计培养博士研究生 37 名、硕士研究生 105 名。后续将重点在流量的智能调度与传输、恶意流量的检测与识别、新型服务的商业模式等方面开展研究，并将项目成果在未来网络试验床上进行部署与验证。

一流人才

一流管理

一流成果

一流效益

重点 / 重大科研任务进展

网络大数据计算的基础理论及其应用研究

“网络大数据计算的基础理论及其应用研究”是计算所牵头的国家 973 计划项目（批准号：2014CB340400）。本项目从网络大数据计算的面临的数据复杂性、计算复杂性以及系统复杂性三个挑战性问题出发，提出“网络大数据复杂性的内在机理”、“网络大数据可计算性及新型计算范式”、“网络大数据处理系统的效能评价与优化”三个关键科学问题，并设置了 6 个课题。今年完成项目的中期评估和后三年调整方案。

2015 年度，在国际、国内期刊会议上共发表论文 117 篇，其中 SCI 收录论文 27 篇，EI 收录论文 35 篇，在国际著名期刊如 TKDE、PNAS 等上发表论文 12 篇，顶级国际会议论文 38 篇。国际 / 国内会议特邀报告 3 篇，一般学术报告 19 次，申请专利 18 项（其中 4 项已授权），软著 1 项，培养博士后 1 名，博士生 35 名，硕士生 41 名。课题二负责人郑纬民教授获得 ISO/IEC 19763 系列标准特别贡献奖。

项目组承办了国内大数据相关会议 2 次，包括 2015 年全国信息检索学术会议（CCIR 2014）、大数据技术与学术大会（BDTC

2015），其中大数据技术与学术大会参会人数超过了 1500 人。举办的第三届中国大数据技术创新大赛有 800 多支队伍参赛，涵盖了国内所有的 985 高校和大部分 211 高校，大会还发布了《中国大数据技术与产业发展白皮书（2015）》，吸引了学术界、产业界的共同关注以及众多媒体的报道；组织了国际研讨会（workshop）2 次；组织了其他技术研讨与交流会 5 次，邀请了包括裴健、邢波、翟成祥等国际知名学者访问和交流 6 人次。

1. 大数据基础理论方面

针对网络大数据复杂性规律发现、核数据导向的网络大数据约简、深度学习和基于核数据的简约计算这四个研究中都有所收获。在网络大数据复杂规律发现方面，重点研究了社交网络中的信息传播存在什么样的内在规律，发现了其中用户间的每次交互会带来新的激励，使消息传播呈现出级联效应；在核数据导向的网络大数据约简方面，研究了核数据的挖掘方法和如何对冗余特征进行约简，提出了双重网络中核数据的约简方

法以及基于 filter 特征的特征约简方法；在深度学习方面，主要研究了文本大数据的表达学习以及匹配，提出了大规模文本数据的词表示方法和匹配特征学习方法等；在基于核数据的简约计算方面，主要研究了面向机器学习任务的近似计算和优化算法，提出了面向不平衡数据，非独立同分布数据和组合优化 NP 难问题的近似优化算法。

2. 大数据计算系统方面

针对流式大数据计算系统，提出了一种流式大数据计算中的资源调度方法；提出了采用源节点、目标节点两级混洗划分的大图数据快速计算方法，提高了图数据访问的局部性；提出了使用 GPU 设备的迭代计算并行加速方法，在保持计算精度不变的情况下，与采用 CPU 的计算模式相比，计算效率提高 6-10 倍。

3. 大数据计算的基准测试方法方面

建立了有代表性的大数据基准测试程序集，提出一种应用负载的特征表达及状态约简方法，将 77 个大数据应用负载约简为 17 个代表性负载，比大数据测试集 CloudSuite 等覆盖面更广，且更精简。

4. 大数据感知与获取方面

建立了一个实时智能收集主题相关的网络数据的感知和获取框架的互联网传感器，并研究了实用的网络大数据信息抽取算法；提出了基于深度学习与稀疏学习的图文表示学习方法；基于图文统一表示，还提出了解决图像语义分割、视频目标检测等难题的新方法；针对结构非结构化数据的融合表示，提出了基于超球相关性保持的网络图像重排序方法、基于跨尺度滤波器共享的张

量深度学习方法、基于稀疏和连通约束的增量视频目标感知算法。

5. 大数据内容建模与语义理解方面

研究特征结构模型上的语义操作问题，通过扩展本体代数的逻辑运算、集合运算，定义特征语义获取、特征语义更新和特征语义查询算子，利用迁移学习技术，提出了一种基于维基百科词条的 Web 服务聚类方法，引入标签辅助的双作者主题模型，解决服务描述数据和维基百科辅助数据的不一致性问题；开展基于网络大数据的属性学习理论研究，提出了一种扩展的中文分词算法，抽取多种主题知识及其结构化信息；开展面向知识图谱的领域知识迭代式获取研究，通过文本图像的联合语义标注，实现对网络大数据的深层内容理解。

6. 网络大数据关联关系方面

针对网络大数据关联关系分析和挖掘，以推荐系统和社交网络为研究载体，展开了多个层面的研究，给出了关系推导、隐性关联发现、信息有效融合的方法。针对网络大数据的模式发现和演化，主要对不同性质网络（如同构和异构网络）的静态特点和动态变化规律进行了研究，并对网络演化中的稳定性进行了分析；同时还发现信息传播的中观分析可以取得很好的效果。

7. 平台建设和大数据引擎开发方面

研发了 MPP 海量大数据仓 BDE-DSQL，支持对多源异构数据的融合查询，可在同一条 SQL 语句中整合 HDFS、RDBMS、Hive、Hbase、LocalFS 等数据源，同时支持诸如 join、groupby 等关联操作，实现基于 SQL 模

型的数据整合。可支持千亿级数据记录，在项目 PoC 测试中，基于实际 200 亿条数据条件查询时延 $<0.6s$ 。完成“流式处理引擎中动态自适应滑动窗口技术”研发，有效兼顾吞吐率和处理时延，在保证数据不丢的情况下达到最快的响应速度，从而实现高吞吐率流式处理引擎。目前 HTS1.0 单节点吞吐率达到 4.6Gbps；流式分析结果 1s 内可见。实现了面向海量多通道网络信息的分布式、可扩展的高性能获取系统架构，解决了 JavaScript、AJAX 等 Deep Web 动态网页的采集与解析还原问题，支持新闻、论坛、博客、微博、社交网络等多通道网络信息中特定网络犯罪行为数据的主动采集和网页结构化特征抽取。

项目的学术交流会三届数据科学与大数据分析 Workshop (DSBDA-2015) 在国际顶级会议 ICDM-2015 上顺利召开，共收到论文 35 篇，录用 11 篇。所有被录用文章的作者都到场做了报告，

每个报告包括提问的 30 分钟都很充实。参加本届 Workshop 的学者学生共有 50 多人，大家在交流讨论中畅所欲言，开阔了学术视野，拓宽了研究思路。

项目中期总结会在中科院计算所召开，项目首席科学家华云生教授主持了此次会议。科技部基础司重大项目处的副处长李非莅临会议并致辞。参加会议的还有中科院前沿局技术科学处处长孔明辉，项目特邀专家李国杰院士、梅宏院士，973 项目责任专家戴国忠研究员、唐常杰教授、彭群生教授，项目专家李学龙研究员、李忠诚研究员、刘志勇研究员、过敏意教授、周傲英教授，以及各课题负责人和学术骨干共计 50 余人。会议期间，各个课题负责人向专家组汇报了课题进展和遇到的问题与挑战，与会专家对各个课题的成果进行了讨论，明确了项目的研究方向，并对后三年项目实施方案等提出了建设性意见。



项目中期总结全体人员合影

重点 / 重大科研任务进展

E 级超级计算机新型体系结构及关键技术路线研究

“E 级超级计算机新型体系结构及关键技术路线研究”是国家 863 计划课题（批准号：2015AA01A301），由计算所牵头，联合江南计算技术研究所、国防科技大学、北京航空航天大学、清华大学等单位共同承担，起止时间为 2015 年 1 月至 2016 年 12 月。2015 年度重点开展了 E 级应用、E 级硬件系统、E 级软件系统等的研究工作，并顺利通过了 863 专家组组织的课题中期检查。本年度课题完成的主要创新成果如下图所示，具体包括：

1.E 级应用分析方面

主要包括 E 级应用特征分析和 benchmark 构建。我们结合国际 E 计算社区共同关注的问题和中国高性能计算的实际应用战略需求，选取了具有 E 级计算需求的应用，包括大气环流模式模拟、激光聚变模拟、湍流模拟、材料损伤性能模拟、生物信息数据分析等。本年度完成了针对上述应用的计算、存储、通信等方面特征及可扩展性分析，并基于典型应用，构建了支持容错的 E 级 benchmark。





2.E 级硬件系统研究方面

主要从处理器结构、互连网络、整机基础架构三个方面开展了研究。处理器研究的核心是能效比约束，本课题提出了高性能 GPDSP、数据流 SPU、异构通用众核等多种不同技术路线分别开展研究，在 14nm 工艺下，处理器能效比有望达到 30-60GFLOPS/W。互连网络研究主要提出了两种不同的技术路线，分别是高维可扩展互连网络和光电混合互连网络，基于两种不同架构分别提出了有效支持 10 万个节点规模的高速互连方案。整机基础架构方面重点针对散热技术开展了研究，提出了包括肋片型强化换热液冷冷板和相变冷板等新型散热技术，可以有效满足 E 级环境下系统散热体积功耗密度达到 100KW/M³ 以上的要求。

3.E 级软件系统研究方面

本年度主要从操作系统和运行时系统、编程

环境、系统管控与容错、软件调试和性能分析这四个部分开展研究工作。操作系统研究方面，完成了面向海量 IO 并发的用户级多层存储文件系统和高效的 E 级用户级存储中间件技术研发，可以支持高效的海量 IO 并发请求，相关技术已在 O(1000+) 规模机群系统上得到实实验证。在编程模型方面，提出了面向 E 级计算的 HPC 中间件技术，主要包括面向 E 级的软件编程及调优架构，可以有效屏蔽 E 级计算系统硬件的多样化、确保 HPC 用户不丢失。在系统管控与容错方面，提出了动态按需的系统管控、主动式动态故障检测、以及基于 PCA 的快速故障诊断技术，有效应对硬件系统多层异构、故障难以定位的问题。在软件调试和性能分析研究方面，提出了以快照序列为基础进行逐步细化和错误定位技术进行软件确定性调试支持，提出了基于低效行为模式的在线过滤方法以减少性能数据规模。

科研为国分忧

创新与民造福

重点 / 重大科研任务进展

异构媒体数据的关联与挖掘研究

“异构媒体数据的关联与挖掘研究”是计算所牵头承担的国家 863 课题（批准号：2014AA015202），起止时间是 2014 年 1 月至 2016 年 12 月。本课题的总体目标紧密围绕基于媒体大数据的信息消费服务的需求，开展异构媒体数据的关联与挖掘技术研究，力争有效突破异构媒体数据的关联与挖掘技术，为基于媒体大数据的信息消费服务提供关键技术支撑。

2015 年度本课题按照预定计划重点开展了异构媒体数据的多模态关联分析和层次化语义关联分析研究，以及构建相应的验证系统。主要研究成果如下：

1. 多模态关联分析方面 提出了基于局部感染免疫过程的抗噪并行聚类方法、基于稀疏表示的大规模检索方法，以及一种规整化的多模态关联图构建方法；在此基础上，着重研究了异构媒体的视觉、文本和音频模态及社交媒体上下文信息的关联挖掘，提出了相应的异构媒体关联挖掘方法和基于统一表达的跨模态哈希检索方法。

2. 层次化关联分析方面 提出了基于层次化语义的复杂多层特征学习方法、多源融合和跨模态

层次化语义关联挖掘方法，以及基于深度学习的视频复杂事件检测方法。

3. 系统构建方面 研究开发了基于稀疏表示的大规模检索系统、基于深度稀疏集成学习的语义概念检测系统、基于复杂语义表示的跨模态相关学习系统 COLAR 和 TINA。研发的多项相关系统已经在工信部、国家广电总局、西南电子电信技术研究所、新华社、以及国内多家公司等获得大规模应用，应用效果显著。

在 2015 年度，本课题发表（含录用）学术论文 33 篇，包括国际顶级期刊（CCF A 类期刊）IEEE TIP 2 篇，CCF B 类期刊 IEEE TMM 2 篇和 ACM TOMCCAP 1 篇；国际顶级会议（CCF A 类会议）CVPR 2 篇、ICCV 1 篇、VLDB 1 篇。一篇论文获 IEEE 多媒体信号处理国际会议 MMSP 2015 Top 10% 论文奖。通过课题研究，培养博士研究生 4 名、硕士研究生 4 名；锻炼和培养了中青年学术骨干，增强了研究队伍的整体实力。

在本课题的研发过程中，课题组与所属 863 项目“基于媒体大数据的信息消费服务关键技术及应用示范”的其它课题承担单位如西北大学、



北京航空航天大学、央视国际和浙大网新、中科院高能所等单位，多次召开技术交流会，共同商讨各课题承担单位之间如何协作进行系统对接，以构建异构媒体关联挖掘实验验证系统。

综上所述，本课题遵循既定研究思路，已

圆满完成了 2015 年度计划的主要研究任务。课题于 2015 年 10 月顺利通过了项目中期检查，专家组的考核结论为“课题执行情况良好”，为进一步深入开展课题研究工作打下了良好的基础。



863 项目五个课题承担单位之间的中期合作交流会议（西安）

重点 / 重大科研任务进展

高精度低成本低功耗的海水电导率传感器研究与设计

“高精度低成本低功耗的海水电导率传感器研究与设计”是计算所独立承担的国家 863 计划课题（批准号：2014AA093402），起止时间为 2014 年 1 月至 2016 年 12 月。2015 年度课题重点开展了的传感器样机的稳定性测试、标定校准、以及实地部署测试等研究工作，并顺利通过项目组的中期考评。主要进展有：

1. 稳定性测试

在本年度的工作中，我们重点解决了传感器的测量漂移问题。在测试过程中，我们发现传感器样机存在银基底氯化、铂电极厚度不足、以及在低盐度条件下电极表面发生电化学反应等问题。这些问题影响了传感器的测量稳定性，造成较大漂移。为此，我们通过数月时间，改进了打磨与清洁工艺，解决了银基底氯化问题；并通过大量实验测试，最后发现将电镀时间设定在 1 个小时，将测量频率设定在 7.8kHz，将测量电流设定在 125uA 时，能够有效避免电极损耗和电化学反应的问题。

经过上述改进后，我们将两个样机在水箱中

进行了 3 天的一致性测试，测试结果如图 1 所示。由图可见两个传感器测量的盐度值具有很好的一致性，整体差别在 $\pm 0.005\text{psu}$ 之内，并且没有明显的漂移趋势，能够满足课题 $\pm 0.02\text{psu}$ 测量准确度的要求。

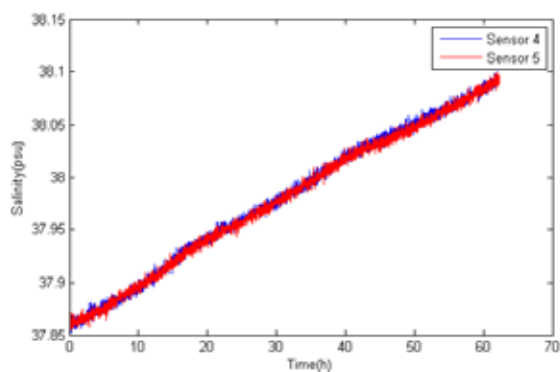


图 1 两个传感器的一致性测试数据

2. 标定校准

在取得上述稳定性测试结果之后，我们在国家海洋技术中心的恒温水槽中对样机进行了标定校准和水下持续测试。测试照片与结果如图 2 所示。恒温水槽通过改变内置海水的温度来改变海

水的电导率，并通过一个 SeaBird SBE49 CTD 传感器提供温盐对比数据。我们的样机在 9 月 29 日、10 月 9 日、和 10 月 15 日共进行了三次校准测试。我们将 9 月 29 日的数据用于传感器的初始校准，并用其余两次的测试数据用于测定传感器的重复性误差。如图 2 所示，在采用 2 阶多项式对电导率传感器进行拟合校准时，样机的盐度校准误差在 $\pm 0.003\text{psu}$ 之内，而重复性测试误差在 $\pm 0.01\text{psu}$ 之内，能够达到课题中 $\pm 0.02\text{psu}$ 的考核指标，并且测量结果没有明显的漂移趋势。

3. 传感器实地部署测试

为测试传感器在实际海洋环境中的运行情况，我们和中科院电子所合作，向其提供了 5 套 CTD 传感器样机，参与其在南海的实地部署测试。如图 3 所示，样机在 200 米深的海床上部署 5 天后，其外观保持完好，未受水压的破坏，并且传感器表面未受到海洋生物污染。样机采集到的温度、盐度、深度等数据与历史经验值接近，其中，在温度快速变化超过 10°C 时，盐度测量数据仍然

保持相对稳定，而深度数据则体现了潮汐造成的深度变化，这表明本课题研制的传感器能够在实际海洋环境中正常工作。

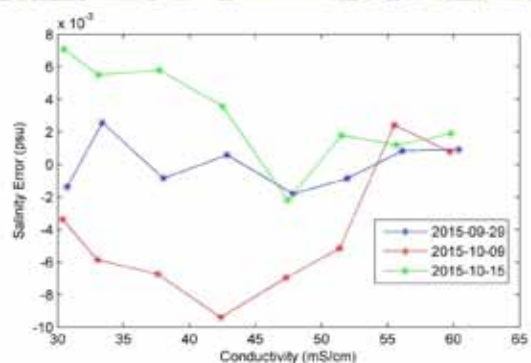


图 2 传感器样机校准测试照片（上）与测试数据（下）

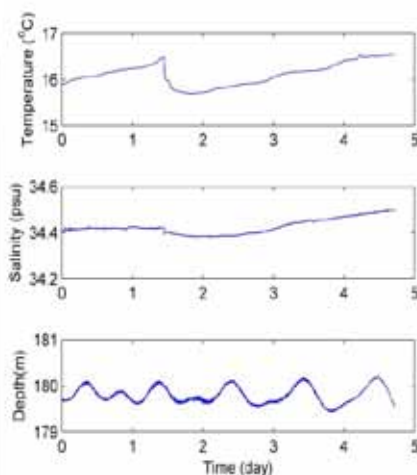


图 3 参与实地部署测试后的传感器样机照片（左）与测试数据（右）

重点 / 重大科研任务进展

基于信息流的多粒度云安全管控与追责技术

“基于信息流的多粒度云安全管控与追责技术”是计算所牵头承担的国家 863 计划信息技术领域课题（批准号：2015AA016005），起止时间为 2015 年 4 月至 2018 年 4 月。该项目的总体研究目标是研究基于信息流的云安全追责、管控技术；通过程序级、系统级、网络级的三层管控体系，实现以数据安全为中心的多粒度、全程式信息流管控和追责。突破数据安全标记快速生成与检索、虚拟机内细粒度信息流透明跟踪、虚拟机间的数据操作行为研判、基于标记的事件检测引擎、基于海量管控日志的追责证据链生成等一批关键技术，获得一批具有自主知识产权的成果，为我国云计算平台的可控安全提供有力的技术支撑。

2015 年度已开展的工作包括：

1. 云数据安全标记规范草案。该草案将数据安全标记引入云平台，定义了云数据安全标记格式，规范了安全标记的生成和管理，并结合 PaaS 平台特点建立了云数据安全标记框架，提出了一种云数据安全标记系统的实现方法，该方

法有效降低了标记操作对虚拟机执行效率的影响。目前该草案正在通标协云安全子组进行讨论。

2. 细粒度的标签能力传播模型研究。本年度提出了一种多层次多粒度的标签能力传播模型，该模型克服了传统分布式信息流模型中无法实现细粒度授权，以及无法对能力进行撤销的缺陷。

3. 虚拟机自省技术研究。虚拟机自省是在底层对虚拟机的状态、事件等进行观测，并从获取的原始二进制数据中抽取、还原出操作系统级相应的状态信息以进行监控的过程。由于虚拟机状态、事件在宿主机中是以二进制形式呈现的，无法反映操作系统级相应的语义信息，称此语义之间的差异为语义鸿沟，跨越此鸿沟是虚拟机自省的核心关键技术。本年度对跨越语义鸿沟的原理及实现方法进行了分析，将虚拟机自省技术分为“守株待兔”法、“顺藤摸瓜”法、以及“按图索骥”三种方法。

4. 网络级标记过滤系统框架设计。该框架由 NISG、vFW、NTAM、SCM、CTM 等五个功能模块构成，其中涉及负载均衡、vGuard



虚拟化防护、NEL 网络事件描述语言等关键技术。负载均衡技术提升了网络级管控子系统的可靠性和性能。vGuard 虚拟化防护技术简化了网络级管控子系统的部署，同时增强了网络级管控子系统在虚拟化环境中的防护能力。NEL 将规则内部组织方式大幅优化，将运行时所必需的运算降低到最低限度，大大提高了引擎的执行效率。

5. 基于云平台信息流的追责技术研究。信息流管控日志分析与追责主要针对云平台信息流管控过程中出现的信息泄漏、进程阻断，网络截流等管控行为进行发现和还原，并依据统一的规范形成追责证据，包括日志采集、日志聚类、冗余过滤、噪声过滤、关联分析、证据链生成等几

个步骤。本年度完成了日志信息的采集、预处理、存储与关联分析关键技术的研究，包括基于时空过滤日志压缩技术、基于语义关联压缩日志技术、基于特征属性的日志流关联分析技术等。

截止目前，该项目已发表论文 9 篇，其中 SCI 收录 1 篇，EI 收录 6 篇，申请专利 5 项，申请行业标准 1 项。

本项目各参与单位在 2015 年 4 月召开了项目启动会，在 2015 年开展了云数据标记方法与技术、分布式信息流控制模型、虚拟机自省技术、网络级标记过滤系统框架、基于云平台信息流的追责等关键技术的研究。在下一步研究中，将进一步加强项目中各单位之间的交流与合作，建立原型系统，开展云计算安全示范应用。

一流人才

一流管理

一流成果

一流效益

重点 / 重大科研任务进展

云环境下的图像视频群体协同表达与处理

国家自然科学基金重大项目“云环境下的图像视频群体协同表达与处理”（批准号：61390510）执行时间是2014年1月至2018年12月。该项目由五个课题构成，承担单位分别为中国科学院计算技术研究所、北京工业大学、同济大学、中国科学技术大学、北京大学。本项目旨在研究云环境下的图像视频群体协同表达与处理，发展以群体化、感知化为基础的信息理论，建立云环境下图像视频群体表达、编码、传输、评价方法，形成具有自主知识产权的AVS云媒体标准，促进我国信息产业由制造优势向技术优势、数据优势的转变。

在过去一年中，各课题均按照预定计划进展顺利，主要成果如下：

1. 在图像视频的群体数据协同结构化表达与处理方面，提出了一种新型黎曼流形度量学习框架，有效保持了黎曼流形的几何结构；提出了采用混合高斯模型进行稳定的集合建模，将集合间的匹配问题形式化为高斯分布黎曼流形上的判别学习；针对不同领域间的知识迁移问题，提出了

可双向迁移的自编码网络，实现领域间样本的相互转换与知识迁移；提出一种跨欧氏-黎曼空间的异质哈希学习框架，支持图像和视频在统一的二值编码空间中进行精简表示和高效比对；进一步，提出一种嵌入语义属性的多功能二值编码联合学习方法，有效支持多层次、多粒度的图像检索和属性预测任务。

2. 在群体化图像视频编码方面，提出了一种通过比较预测候选块和当前图像块的哈希特征实现快速运动估计的方法，显著减少了运动估计时间，通过使用保距的哈希函数，改善了量化噪声和采集噪声对图像块的影响；提出了一种混合的图像超分辨率重建方法来同时兼顾数据和模型驱动的机制；将二维稀疏表示模型与非局部相似性相结合，给出了二维非局部稀疏表示模型，并在图像去噪、单帧图像超分辨率重建等方面验证了该模型的有效性；提出了利用拉普拉斯尺度混合模型对多维稀疏信号进行建模。

3. 在群体化图像视频传输方面，围绕视觉失真可容忍的群体化图像视频传输开展研究。针

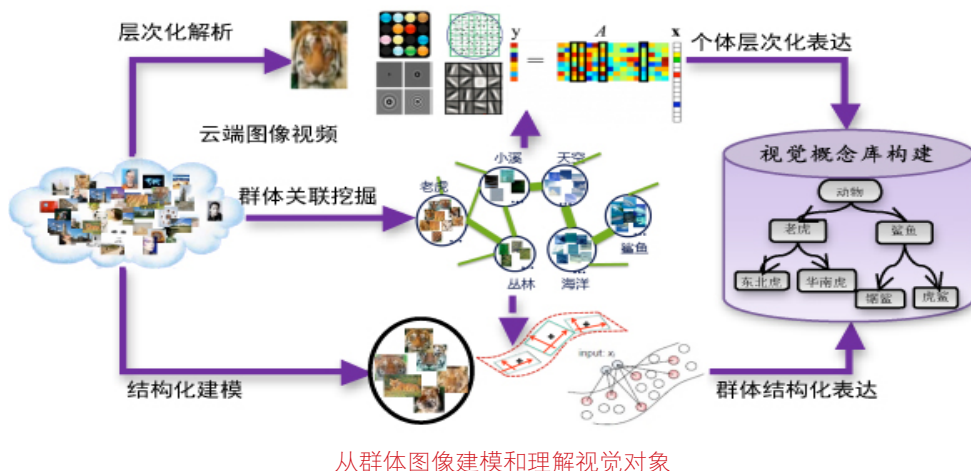
对多用户下的伪模拟传输场景，提出了不同优化分配目标下的资源分配。提出的无线视频层播框架 LayerCast；提出了一种联合使用分层的陪集编码和群体信息下的图像降噪进行群体信息辅助下的图像传输方法。在发送功率固定的情况下，可以有效的将低频大系数节省下来的能量分配给高频系数，因此更加有效的保护高频系数；提出一种基于群体相关的无线视频传输技术 (KMV-Cast)，将海量视频中的相关数据融合成了先验信息，从而最大化接收端恢复视频质量。此外，完成了自研的 2x2MIMO 软件无线电平台调试，解决了敏捷无线收发芯片 AD9361 的配置及校准等关键技术问题。

4. 在图像视频质量评价方面，旨在发展与主观感知一致的图像质量评价理论方法，为图像视频编码、图像视频检索等应用提供评价准则和优化目标。提出了一种无参考的锐利度评估准则用于盲图像模糊评价方法，并扩展到基于“双目竞争”的立体图像中；提出了一种基于稀疏分解的全参考的图像质量评价方法，以此获得测试图像的主观质量；提出一种有参考的重定向图像的质量评价算法，考虑了图像局部相似性、图像内容重要信息丢失以及结构失真三个方面，分别对应

图像重定向算法的三个主要目标。提出一种将图像的语义元素融入图像质量评价的方法，融入了目标检测得分、非极大值抑制、局部特征信息等，并且可整合到现有方法里，显著提高其性能。

5. 在面向群体数据的 AVS 云媒体标准及验证平台方面，主导制定的国际标准 MPEG CDVS 正式颁布，CDVS 国际标准简化了视觉特征抽取算法，不同大小的紧凑视觉特征可以互操作，带宽适应性强。面向大规模视觉词典的高性能检索问题，提出了兼容汉明距离度量的特征量化方法 HCQ；提出了亲和保持的特征量化方法 APQ，应对高维浮点特征向量投影空间的非自适应划分问题；在基于背景模型的监控视频高效编码方面，提出了背景图像质量可伸缩编码方法；探索了基于背景库的监控视频编码算法；提出了基于互补显著先验学习的对象分割算法；基于背景模型的监控视频编码技术已经被新一代视频编解码标准 AVS2 采纳，目前标准已经送审。

上述研究，2015 年度共发表国际刊物论文 26 篇，国际会议论文 40 篇，申请专利 10 项，相关前期研究成果获得国家自然科学奖二等奖两项，电子学会技术发明一等奖一项。



重点 / 重大科研任务进展

在线社会关系网络的挖掘与分析

本项目为国家基金重点项目（批准号：61232010），起止时间是2013年1月至2017年12月。本年度在网络信息传播、网络社区发现、排序学习、表达学习、推荐算法、情感分类、高维视觉信息处理等方面展开了研究并形成了一系列有国际学术影响的研究成果。

1. 网络信息传播方面

针对传统信息传播点对点模型极易遇到过表达和过拟合的问题，提出了一种基于用户隐影响力和易感性的信息传播模型，解决了传统建模方式的问题，并克服了传统模型依赖社交网络的限制。针对传统的消息流行度模型无法处理社交网络中消息级联效应的问题，提出了基于自激励霍克斯过程的流行度动力学模型，刻画了消息传播过程中每一次转发所带来的激励作用，更好地预测了消息的流行度。针对传统图聚类网络社区发现方法找到的种子节点集合中存在冗余或边缘节点的问题，提出了一种新的模型，利用多样性图排序的方法选择排序 top-k 的节点作为种子节点集合，克服了传统算法在随机种子节点选择方面的问题。研究成果

发表在 WWW 2015、AAAI 2015 等网络信息检索和人工智能领域的国际会议上。

2. 排序与表达学习方面

针对传统信息检索中多样化排序学习算法无法同时满足在线约束和离线优化的问题，提出了一种新的排序学习算法，在排序过程中遵循最大边际相关度约束，同时将多样化排序问题转化为文档列表上的二值分类问题，并可以采用任意的多样化排序评价指标作为目标函数进行优化。针对信息检索中传统单词表达学习模型仅考虑单词在上下文中的共现关系或单词具有相似上下文的替代关系的问题，提出了一种新的表达学习模型，以并列或层次的方式同时建模单词间的两种关系，并在单词相似度、语义、句法等任务中的效果显著超越了传统模型。研究成果发表在 SIGIR 2015、ACL 2015 等信息检索和自然语言处理领域的国际会议上。

3. 高维视觉信息处理方面

针对传统高维视觉数据分析从数据的统计



假设出发，忽略了数据固有结构关系的问题，提出了基于数据结构分布、采用流形排序挖掘数据的显著程度的方法，提高了社交网络数据中高维信息表达的准确性。针对传统高维视觉数据分析中邻域结构信息难以有效表征个体与群体的关系的问题，构建了数据间的顶点 - 边缘权重图模型，增强了异常检测中目标和背景像素的辨别性。针对传统多模型数据处理方法仅考虑不同模型之间的差异分布，较少考虑模型内部的分布结构的问题，依据物理学中粒子间作用力模型构建了结构描述子，增强了个体与群体的差异信息，准确描述了场景全局信息。研究成果发表在 IEEE Transactions on Cybernetics、IEEE Transactions on Neural Networks and Learning Systems、IEEE Transactions on Geoscience and Remote Sensing 等 SCI 检索国际期刊上。

4. 学术活动

1) 项目负责人程学旗研究员组织了中英、中澳、中韩大数据论坛，推动各方之间的学术和技术交流与合作，建立了良好的沟通通道和合作机制。

2) 项目负责人程学旗研究员组织了第三届数据科学和大数据分析学术研讨会 (DSBDA)。

3) 项目负责人程学旗研究员共同组织了第二十一届全国信息检索学会会议 (CCIR 2015)。

4) 项目负责人程学旗研究员应邀在国内多家科研单位做了关于在线社会关系网络挖掘与分析的学术报告，扩大了本项目在国内的影响力。

5) 项目负责人程学旗研究员组织和参加了一系列大数据分析相关的学术会议和高峰论坛，介绍了本项目的研究成果。

6) 项目骨干成员沈华伟副研究员作为青年科学家共同组织了中德前沿技术探索圆桌会议 (ERTC 2015)。

创 新

求 实

团 结

高 效

重点 / 重大科研任务进展

动态可配置的压缩感知成像系统

本项目是国家自然科学基金科学仪器基础研究专项（批准号：61327013），起止时间是2014年1月至2017年12月，前期得到中科院装备研制项目的支持。本项目旨在面向压缩感知成像研究的实际需求，提供任意值测量矩阵的配置，用以验证不同的测量矩阵在光学成像中的实际效果；实现图像采集、恢复，呈现一体化，让研究人员高效快捷地验证自己的恢复算法和成像效果。2015年研究进展如下：

1. 初步实现了动态可配置的压缩感知成像装置，采集方式包括和单像素成像，阵面成像；成像方式包括无透镜成像和有透镜成像；采集对象包括图像和视频；测量矩阵实现包括液晶与DMD；采集频谱包括红外与可见光。并且实现了采集恢复流程一体化。

2. 对压缩感知图像重建算法方面进行了深入的研究，提出了全局测量局部稀疏的压缩感知图像恢复算法，基于非局部相似度模型的压缩感知图像恢复算法，基于秩极小化理论的压缩感知图像恢复算法等压缩感知图像恢复算法，提高压缩感知图像重建质量10%~20%。

3. 对压缩感知重建算法进行了深入研究，提出了给定的标准AGCM原子的OMP算法，AGCM原子的自适应无字典OMP算法，低秩矩阵拟合算法和正交秩1矩阵追踪法等压缩感知重建算法，已经广泛应用于地震波信号的恢复。

4. 在压缩感知图像重建硬件加速方面，实现了GPU并行的压缩感知图像恢复算法，与CPU相比，恢复速度提高了100倍以上。基于Xilinx V7 2000T，实现了一套完整的硬件化压缩感知数据采集重建架构，实现从数据采集到重建全过程的硬件化，数据同步和控制精度达到纳秒级，对外数据带宽达到10Gbps，实现高速采集和重建结果高速回传。

目前项目组在压缩感知成像方面发表论文11篇，其中在SCI期刊上发表论文7篇，在国际会议上发表论文4篇。培养研究生5名，申请5项国家发明专利。提出新的问题和研究思路成功申请了国家自然科学基金面上项目（61471343）。在项目实施过程中，项目组与国内外研究机构紧密合作，与美国德州大学奥斯汀分校，加拿大阿尔伯

特大学，德国哥根廷大学等国外著名大学研究者合作发表文章多篇。项目组主要成员参与组织SEG压缩感知-勘探地球物理应用新技术研讨

会，并担任技术委员会联合主席，在3天的会议中，国内外多名研究者交流了压缩感知的最新研究成果。



图1 压缩感知成像设备内部结构



图2 场景



图3 采样过程（掩膜+场景）

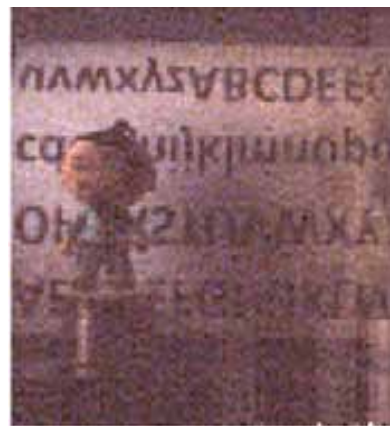


图4 单像素重建图像

重点 / 重大科研任务进展

数据并行与线程并行合一的可伸缩处理器体系结构

本项目为国家基金重点项目（批准号：61332009），起止时间是2014年1月至2018年12月。本项目旨在发展一种新的通用微处理器体系结构，它不同于当前常见的基于动态超标量技术和静态SIMD部件的通用处理器结构，而是用同一套控制和运算部件来同时支持线程并行和数据并行，因而具有结构简单灵活、片上资源利用率高、性能价格比高、性能功耗比高等特点。除了单独应用外，它还可以作为众核处理器的基本构件，用于实现高性能云数据中心服务器，也可以作为嵌入到SoC芯片中的处理器核心，用于实现各类云终端装置。基本思路是：通过深度流水提高工作频率、通过多线程和长向量容忍延迟、通过静态结构和资源复用降低功耗和成本。在该体系结构中，处理器具有三种执行方式，即纯多线程并行、纯数据并行、多线程并行和数据并行轮流切换。项目已于2016年1月通过基金委组织的中期检查，在2015年度完成的工作主要体现在以下三点：

1. ETDPIC 指令系统研究

研发出了一种能兼具线程并行和数据并行的

ETDPIC 指令系统，它采用了数据流思想，线程并行体现在执行过程中同时支持多个线程进行切换，线程流水进入处理单元；而每个线程则同时支持SIMD，同时对多个数据进行处理，体现数据并行思想。

ETDPIC 实现高效的线程并行和数据并行的同时给访存网络带来较大的压力，本研究采用了SIMD空间拆分的方式减少访存请求的数量，减少访存网络压力。具体实现是将单线程对应的多个SIMD的地址空间拆分，并将一次访存产生的多个地址合并到同一Bank上，如此可以让PE产生的SIMD访存消息的多个地址对应片上同一个缓存，提高访存效率。

2. 处理器体系结构模拟器研究

在研究中，模拟器将起到不可或缺的作用，因为对于数据并行和线程并行的设计和开销的权衡需要在模拟器进行复杂的结构设计以及做大量的实验，通过实验数据来确定本课题研究的可伸缩处理器体系结构的最终设计方案和数据支撑。

本研究开发了基于并行离散事件模拟的模拟器，其核心思想是将模拟的目标系统和模拟技术



本身分离，前者我们抽象成一个个的组件，后者则称为模拟框架，组件包含了目标模拟系统所含的模块，框架则包含了模拟本身相关的功能，如配置功能、并行加速功能、模块通信功能等。该模拟平台的构建为本项目相关结构研究提供了可靠保证。

3. 基准测试程序研究

基准测试程序是评价项目设计的新型处理器体系结构与已有的常规方案在性能、功耗、成本等方面的优劣的主要依靠标准。为了后续

微体系结构设计的研究和评估，本研究开发了一套 Micro-Benchmark。Benchmark 的提取主要从两个方面开展：一是分析 HPC 领域的典型 Workload，包括大气模拟、流体动力学、石油勘探、核物理、生物信息、基因、材料、金融保险、密码分析等重要领域的核心算法；二是分析 Big Data 领域的典型 Workload，包括大数据分析、数据挖掘、机器学习、网页搜索、社交网络、数据库、电子商务、大数据多媒体、可视化等重要领域的核心算法。

跻身国际前沿

关注国计民生

引领中国信息产业

重点 / 重大科研任务进展

高效通用数据中心体系结构研究

本项目为国家自然科学基金重点国际（地区）合作研究项目（批准号：61420106013），起止时间是 2015 年 1 月至 2019 年 12 月。本项目的研究目标是设计并实现一种新型高效通用数据中心体系结构，在保障应用服务质量的基础上将数据中心计算资源利用率提高到 60% 以上。

1. 2015 年度在以下四方面取得了进展：

（1）感知多应用的硬件资源管理机制方面，解决了两个主要的挑战：一是为 Cache 写回请求的标签机制；二是如何为 I/O 请求和中断响应请求打上标签。（2）硬件资源数据平面与控制平面设计方面，为多种组件设计了一个基本的控制平面结构，包括三个由标签索引的控制表，分别是：用于存放资源分配策略的参数表（parameter table），用于记录资源使用信息的统计表（statistics table），以及用于存放性能触发条件的触发表（trigger table）。（3）低复杂度低开销软件栈设计与实现方面，利用服务器中已有的 IPMI/BMC 模块实现集中式的平台资源管理器，实现了无虚拟化层的全硬件支持虚拟化功能，同时对性能隔离及其编程接口进行支持。

（4）数据中心作业 / 资源调度技术方面，设计与实现了基于 OpenStack 的支持 QoS 的硬件资源管理框架，通过对资源的精细化管理，实现硬件感知业务。

2. 研究初步成果与影响力

- （1）实现了两套原型系统以验证该体系结构的效果，分别是基于 gem5 的体系结构模拟器和 FPGA 原型系统（如图 1 所示）。研究初步研究成果发表于体系结构顶级会议 ASPLOS 2015；
- （2）因为国际学术界对 PARD 研究工作的肯定，课题执行负责人包云岗研究员收到邀请参加由 UC Berkeley 的 David Patterson 教授、EPFL 的 EcoCloud 中心主任 Babak Falsafi 教授等组织在德国举行的为期一周的关于数据中心架构（Rack-scale Computing）的论坛 Dagstuhl Seminar，与会者来自 Stanford、MIT、Berkeley、IBM、Microsoft、HP 等国际专家（如图 2 所示）；
- （3）在工业界，PARD 项目得到华为的大力支持，计算所与华为联合申请了包括多个高价值专利的专利群，作为华为全球合作五个代表性

成果之一入选华为 2015 年报；

3. 国际合作交流情况

在本课题的资助下，我们邀请了 13 位国外专家到国内进行学术交流，包括美国普林斯顿大学 Kai Li 教授、美国加州大学圣地哥分校 Yuanyuan Zhou 教授、美国德州大学奥斯丁分校 Mattan Erez 教授、美国杜克大学 Benjamin Lee 教授、美国高通公司 Dileep Bhandarkar

博士等世界知名学者专家。同时，项目组成员分别于 2015 年 3 月 9-10 日访问美国莱斯大学，于 2015 年 11 月 11-14 日访问芝加哥大学，并作学术报告：A New Perspective on Supporting QoS in Computer Architecture: The Computer as a Network。此外，本课题进一步扩大与普林斯顿大学的合作范围，与电子系 Ruby Lee 教授团队也开展了合作面向云计算安全的系统架构研究。

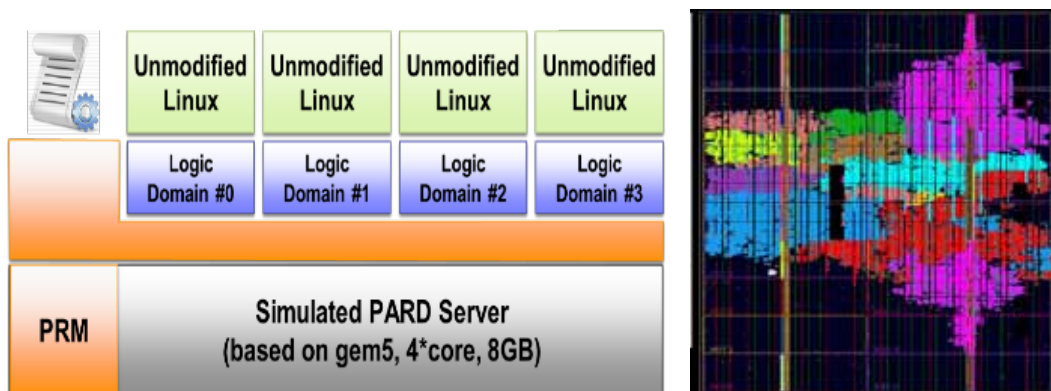


图 1 基于软件模拟器与 FPGA 的 PARD 原型系统



图 2 参加“Rack-Scale Computing”
Dagstuhl Seminar 论坛

重点 / 重大科研任务进展

易编程的异构并行处理器结构

“易编程的异构并行处理器结构”是国家自然科学基金重点项目（批准号：61432016），起止时间是2015年1月至2019年12月。2015年度，在异构并行理论模型、异构处理器总体架构、加速核结构、编程模型和运行时系统等五个方面进行了探索研究，取得了一系列成果，完成了年度预期目标。

下面介绍本项目本年度取得的突出进展。

在2015年度，在易编程的异构并行处理器结构研究领域，本项目使用了C/C++模拟器对易编程的异构并行处理器的结构参数和能效设计空间进行探索，该模拟器拟包含多个处理器核，并完成了RTL代码平台的搭建和验证平台的搭建，从而更加全面的评估其体系结构设计参数与能效设计空间的相互联系。本年度所

取得的阶段性研究成果对本项目明年的研究奠定了坚实基础。

2015年，在异构并行编程模型这一领域的不同方面进行了深入研究。其中发表的14篇文章中，包括了CCF的A类国际学术期刊TOCS一篇，TOC两篇，TCAD一篇，CCF的B类国际学术期刊TACO一篇，TODAES一篇，发表的国际会议文章包括了CCF的A类国际会议ISCA一篇，ASPLOS一篇，CCF的B类国际会议DATE一篇。这些发表在国际顶级期刊和会议上的众多文章，具有很大影响力。因此陈云霁研究员入选2015年度MIT TR35，并受邀担任了CCF的A类国际会议国际微体系结构会议(MICRO)程序委员会成员和IEEE国际计算机体系结构(ISCA)技术程序委员会成员。

重点 / 重大科研任务进展

差错容忍计算器件基础理论与方法

“差错容忍计算器件基础理论与方法”是国家自然科学基金重点项目（批准号：61432017），起止时间是2015年1月至2019年12月，计算所是该项目的承担单位，香港中文大学和同济大学是该项目的合作单位。

该项目2015年启动，主要在运行时结果质量可控的体系结构、运行时结果质量可控的电路设计、高能效存储器的近数据计算等方面取得了

重要进展，本年度发表和录用SCI期刊论文8篇（含IEEE/ACM Trans. 论文6篇）、EI国际会议论文15篇（含DAC 3篇，DATE 3篇，ICCAD 1篇）、EI核心期刊论文6篇。

在运行时结果质量可控的体系结构方面，主要针对片外存储器的访问进行了研究。很多大数据处理的应用需要频繁存取片外存储器内存放的数据，而很多这样的应用执行时并不要求

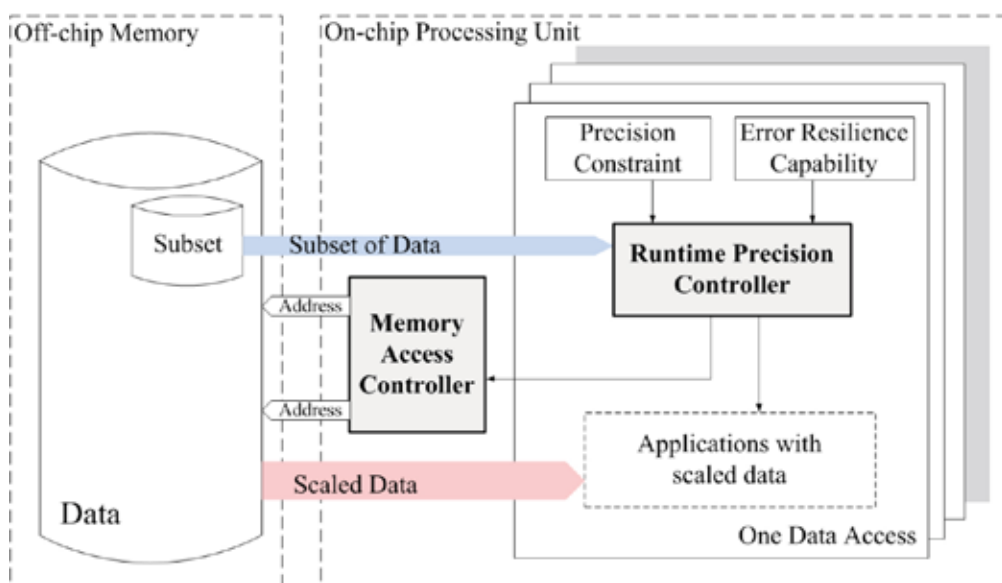


图1 ApproxMA: 近似存储访问

完整的数据精度，通过精度缩放可以在得到可控质量结果的同时极大减少能耗。针对这个问题，提出了针对片外存储器存取的精度缩放方法 ApproxMA，可调整片外存储器的访问精度从而取得更为可观的能耗节约，在保证计算质量的同时取得尽可能多的能耗节约，并成功地把 ApproxMA 应用到聚类算法中，结果证实了这种近似存储器访问的有效性。

在运行时结果质量可控的电路设计方面，主要针对人工神经网络类应用进行了研究。神经网络类应用以计算量大可容错能力高而著称。通过替换电路中一部分加法器或乘法器为可近似的版本，可以取得可观的能耗节约；然而以往的研究并没有考虑频繁数据读取对能耗和计算质量造成的影响，并且不能对计算结果的质量有所保证。针对这个问题，提出了一个针对神经网络应用的

系统性的电路设计方法 ApproxANN，综合考虑了对计算和数据读取的近似，由于数据读取所占能量消耗比重更大，可以比传统只考虑计算的方法取得更显著的能耗节约；进而提出了一个进行系统性设计的优化过程，在一定质量阈值下取得最大的能耗节约。

在高效存储器的近数据计算方面，主要研究将计算与数据存储资源紧密地集成到单芯片系统当中以解决存储墙问题，利用新型相变存储器设计出了一种低开销的近数据存储内加速器 ProPRAM，兼容现有主存架构实现近数据计算，有效地加速了存储密集型应用，从算法层与编译层重新优化一系列面向大数据处理的非计算密集型应用及算法，使其映射到所设计的近数据加速器之上，并且无需处理器参与任务执行，从而极大地提高加速能效比。

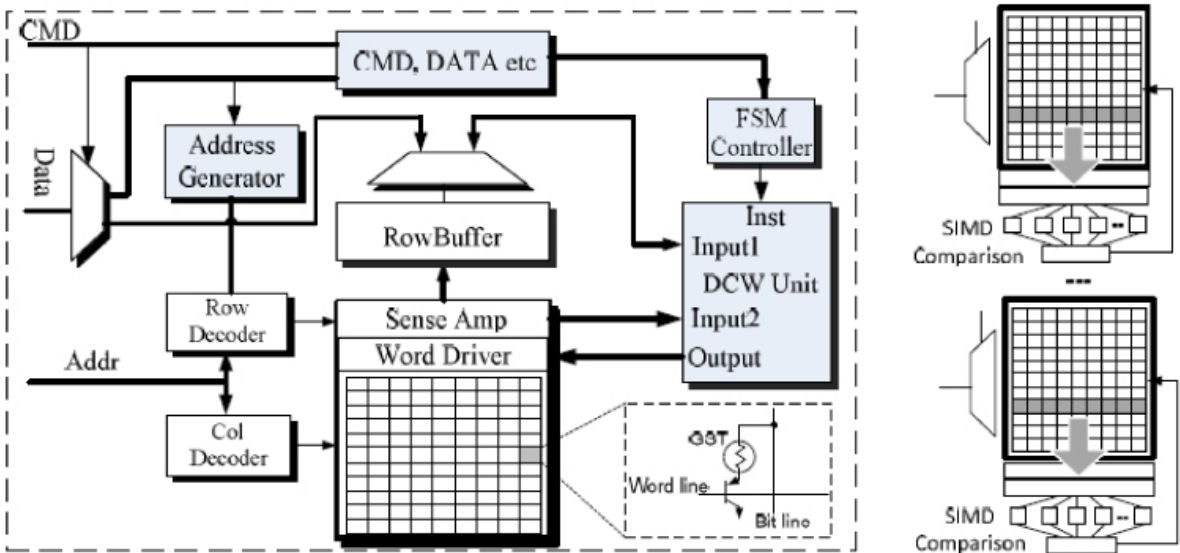


图 2 近数据存储内加速器 ProPRAM

重点 / 重大科研任务进展

面向气候和湍流模拟的百万量级异构众核可扩展并行算法与优化方法

“面向气候和湍流模拟的百万量级异构众核可扩展并行算法与优化方法”是计算所牵头承担的国家自然科学基金重点项目（批准号：61432018），起止时间为2015年1月到2019年12月。

本课题选取全球气候模拟和直接法湍流计算为研究对象，针对其中的结构与非结构网格、FFT、稠密矩阵计算和稀疏线性代数等典型计算和通信模式，开展网络拓扑和通信避免可扩展并行算法、并行编程框架和基于模板的自适应优化技术，从大规模并行应用软件研制角度出发，解决并行软件面临的性能难以连续线性可扩展和可移植的挑战。

2015年度主要进展如下：

1. 大规模异构众核并行的建模理论与可扩展性方法研究

进一步系统深入剖析了多层次不连续非线性可扩展现象；对两个科学计算核心关键算法的一类软错误给出了新的理论结果以及容错算法。在

深入调研国内外相关研究现状和进展的基础上，进一步对本项目拟研究的高性能计算的多个层次，即物理模型、并行算法设计和并行软件性能优化中的不连续非线性可扩展现象进行深入的扩展、细化和梳理。首先，将这一概念扩展到硬件层面，指出计算机硬件系统的发展也存在不连续非线性现象，在纵向层次扩展这一现象；其次，对于每种层次的不连续和非线性进行细化，在每个层次中横向列举更多实例；最后，对各种具体实例进行整体的分类梳理，完善形成最终的多层次不连续非线性可扩展现象综述。

2. 几个重点应用领域的可扩展并行算法设计研究

进一步在国产平台上对选取的两类应用进行大规模测试，深入分析了大规模应用的特征和瓶颈，为后续优化设计工作奠定了基础。在全球气候模拟应用方面。首先， $0.5^\circ \times 0.5^\circ$ 分辨率的 IAP AGCM4.1 在天河1上完成了10年（1979-1990）的 AMIP 积分试验，初步结果表明模式

的计算结果是合理可靠的。其次，初步分析了 $1.4^\circ \times 1.4^\circ$ 及 $0.5^\circ \times 0.5^\circ$ 两种分辨率的情况下不同过程占用的计算时间的比例，为后续计算性能的提升提供了依据。在 1.4° 的分辨率下，随着 CPU 核数的增加，物理过程计算时间所占的比重迅速下降，从串行的 57% 下降到 512 核的 17.2%，表明模式的物理过程有较好的并行可扩展性。最后，在曙光新一代高性能计算机硅立方上完成了 CESM 大气分量模式 CAM5 的万核测试，为今后开展 IAP AGCM4.1 的万核测试积累的经验。对单独的动力框架而言，无论对 -O2 还是 -O0 都具有较好的并行可扩展性，-O2 在 8192 核时仍具有较高的并行效率；-O0 在 1 万核仍保持高的并行效率。加入物理过程后，在 -O0 的情况下计算速度变化不大，且至 8192 核一直保持较高的加速比。由于计算资源的限制，有三组试验尚未开展万核以上的测试，下年度将继续进行该方面的工作。

在大规模 3 维 FFT 及直接法湍流模拟应用方面，调研国内外与本课题研究内容相关的现状和趋势。在国产异构众核并行系统平台上移植湍流模拟软件由 MPI 实现移植为基于 Parray 的 Parallel C 语言实现，测试规模至 16384^3 的软件性能及可扩展性。研究和建立异构众核并行计算系统的拓扑结构和层次存储感知的性能模型，以 GPU 集群为例分析其各存储部件间数据传输性能和各计算单元计算性能，为后续并行算法设计提供理论依据。进一步，对核心 FFT 问题，根据异构集群多层次异构存储的特点，决定了在其

上进行计算时，数据需要进行更多维度的划分。我们利用介绍多维数组维度转置的表示方法和课题组实现的 Parray 语言，可以对异构集群复杂数据维度变换的数据操作进行清晰表示。同时介绍基于数组维度类型程序设计方法和 Parray 语言实现的天河 1A 系统上的大规模 3 维 FFT，该算法代码实现简洁，同时得到了良好的性能和可扩展性。相关工作发表在 2015 年软件学报上。

3. 面向重点应用领域的异构众核并行编程框架研究

通过实例剖析和验证 Parray 语言特性并进一步提出了一种分片的语言机制，设计一种众核结构上基于制导的共享内存相关优化框架，设计了一种针对结构网格应用的并行编程语言及其检查工具。课题组提出的 Parray 使用数组类型对数据的物理存储和逻辑结构进行分离，使用统一的线程数组类型表示各种进程（线程）的创建以及它们之间的控制流转。使用 Parray 实现的高性能 GPU 矩阵乘法，在天河 1A 单节点上的测试性能和 CUBLAS4.0 相当。进一步分析表明，通常的编程机制，会被集成到多种编程语言中。然而，现有的各种通用编程语言机制没有反映异构体系结构及其应用的体系结构特征。我们进一步提出称为分片的语言机制，用来衔接不同层次的存储结构的编程：外层的对应于更大的、更慢的和分布的存储器以及内层的对应于更小的、更快的局部的存储器机制。相关论文发表在 2015 年 PPopP 和 CCGrid 国际会议上。

重点 / 重大科研任务进展

大数据结构与关系的度量与简约计算

“大数据结构与关系的度量与简约计算”是国家自然科学基金重点项目（批准号：61433014），起止时间为2015年1月到2019年12月。

本年度研究工作按照项目任务书的计划进行，围绕着网络链路预测、网络小世界模型、动态演化网络的相关算法等问题展开研究，并形成了一系列有国际学术影响的研究成果。代表性研究成果概述如下：

1. 网络链路预测

链路预测是网络科学中一个基本问题，为了衡量网络可被预测的难易程度，我们提出假设：网络越具有规律性越容易被预测。如果随机从网络中抽取出一小部分链路，网络的特征向量空间受到的影响很小，就说明网络是具有规律性的。借鉴量子力学中一阶微扰的方法，提出了结构一致性来度量这个差异，它可以直接用来刻画网络的“可被预测的程度”。大量的模拟网络和真实网络实验都支持了我们的结论：结构一致性越强的网络越容易被准确预测丢失的链路。我们利用结构一致性，提出了一种新的名为“结构微扰法”的链路预测方法。这个方法在预测丢失的链路，

以及甄别网络中添加的噪音边两方面都明显超过了当前主流的方法，包括知名的层次结构法和随机分块法。

2. 动态演化网络算法研究

排序作为最基本而经典的算法问题，在大数据时代依然是众多关键应用的基石。我们研究了访问受限的动态数据模型下的排序和查找问题。我们借鉴 Anagnostopoulos 等人提出的动态数据的模型，采用 Kendall tau 距离作为衡量算法性能的指标。我们研究了 top-k-selection 问题：在每个时刻 t ，找出 top k 的元素并将其排序。之前 Anagnostopoulos 等人的工作只研究了 $k=1$ 和 $k=n$ 。我们的主要贡献是确定了该问题的“相变点” k^* ，即当 $k=o(k^*)$ 时，该问题可以以 $1-o(1)$ 的概率无差错地解决。同时我们证明了当 k 超过 k^* 时，对于任何算法，所求得顺序与真实顺序的 Kendall tau 距离都至少是 k^2/n 的，而且我们的算法表明这个界是紧的。我们还研究了 top-k-set 问题：在这一问题中我们只需要确认 top- k 的元素，而不需要确定它们的顺序。我们证明了对任意的 k ，top-k-set 问题都可以以 $1-o(1)$ 的概率无差错的求解。

3. 基于博弈论的小世界模型

02 年 Kleinberg 提出了适于通行的小世界网络模型，当模型中的长边幂率分布系数等于 r 基准网络的维度时小世界网络才是可通行的。实证研究印证了现实网络的 r 确实接近于网络的有效维度。我们从博弈论角度出发将网络中的节点看作玩家，长边幂率分布系数 r 是其策略。引入了新的效用函数，使得每个节点的效用是她随机长

边的平均距离与随机长边有反向边的平均概率的乘积，后者反应节点倾向于连边的互惠性以使联系更加稳定。我们从理论上证明了该博弈仅有两个纳什均衡，适于通行的小世界网络是唯一的稳定均衡。模拟实验验证了即使节点不了解其他节点的连接偏好也同样会收敛到适于通行的小世界网络。我们还用人人网和美国 Live Journal 两个实际网络进行了验证。

一流人才

一流管理

一流成果

一流效益



重点 / 重大科研任务进展

Web 信息检索与数据挖掘

本项目为国家杰出青年科学基金项目（批准号：61425016），起止时间是2015年1月至2019年12月。本年度在信息传播预测、网络社区发现、推荐算法、文本表达学习、排序学习等方面展开了研究并形成了一系列有国际学术影响的研究成果。

1. 信息传播预测方面

针对社交媒体中信息传播呈现出的多中心和多峰现象给信息传播预测带来的挑战，提出了基于自激励霍克斯（Hawkes）过程的信息传播预测模型。该模型自适应寻找网络中触发传播的中心节点，进而对每个中心节点触发的传播过程进行建模，有效刻画了传播过程中每一次转发所带来的激励作用。该模型一方面克服了传统的单源传播模型无法建模信息多次传播带来的问题，另一方面通过寻找少数几个中心节点作为传播源大大降低了传统自激励过程的高复杂度问题。系列研究成果发表在 AAAI 2015、WWW 2015 等本领域著名国际学术会议上。

2. 推荐算法方面

如何对用户下一次的购物数据进行预测是市场分析里一个重要的问题。针对传统方法面临在同时建模个体购买时间序列和群体购买历史记录方面面临的问题，提出了层次化表达模型（Hierarchical Representation Model）来完成用户的购物推荐，将用户的表达和商品的表达建模在同一个连续的空间中，商品的表达代表用户购物的顺序行为，用户的表达代表用户的整体兴趣。在预测用户下一步的购买商品上，提出的模型显著超过了现有最好的方法。系列研究成果发表在 SIGIR 2015、WSDM 2016 等本领域重要学术会议上。

3. 文本语义匹配方面

文本匹配是感知数据的一个基础工作，对文本匹配的深度优化，会增强数据的主动获取能力。现有深度模型将文本分别表达成两个等长的向量然后直接计算文本间的相似度，面临两个问题：（1）难以表达细粒度的匹配信号，导致没法充分捕获细粒度的匹配信号；（2）两

段文本在最后才产生交互，计算相似度的方式过于简单，没法进行度量学习。针对上述问题，提出了 MatchPyramid 模型从最基本的匹配信号，然后多粒度、层次化得提取两段文本的匹配信号。MatchPyramid 模型的性能能高于传统模型和现有的深度学习模型。系列研究成果发表在 ACL 2015、AAAI 2016 等重要学术会议上。

4. 社区问答方面

社区问答是以网络为载体，公共的知识平台，它的价值在于重建人与信息的关系。如何评估社区问答中的答案的质量成了主动获取数据的一个关键问题。现有模型的性能受限于其基于的建设，即问题所在的表示空间和答案的表示空间不对应。为此，我们提出了 SPAN 模型，将候选答案和问题的匹配与支持答案的匹配设计成了一个直接可以端到端训练的模型。每个问题和答案我们都用较为常用的卷积句子模型进行建模，在得到句子的表达之后通过点积表达句子之间的相似度。SPAN 显著超过了传统

的问题和候选答案直接匹配的模型。研究成果发表在重要学术会议 AAAI 2016 上。

5. 其它方面

1) 项目负责人程学旗牵头的“社会化媒体数据的感知与处理关键技术及应用”成果，获得北京市科学技术二等奖；

2) 项目负责人程学旗受湛庐文化的委托，翻译了大数据人文方面 (Culturomics) 的著作《Uncharted》，译作由浙江人民出版社出版，中文名称为《可视化未来》；

3) 项目负责人程学旗作为 CCF 大数据专委会秘书长，共同组织了大数据学术大会、大数据技术大会和第三届 CCF 全国青年大数据创新大赛，在学界和业界形成了很好的影响。

4) 项目负责人程学旗组织了中英、中澳、中韩大数据论坛，推动各方之间的学术和技术交流与合作，建立了良好的沟通通道和合作机制。

5) 项目负责人程学旗组织了第三届数据科学和大数据分析学术研讨会 (DSBDA)。

集中力量办大事

重点 / 重大科研任务进展

多媒体内容分析

“多媒体内容分析”是国家基金优秀青年基金项目（批准号：61322212），起止时间是2014年1月至2016年12月。2015年度，我们在场景图像分类以及面向移动和智能终端的图像识别等方面开展了多项深入研究，分析了视觉场景中的物体和场景信息的差异性，建立了神经网络融合方法，并在多个公共数据集上取得了最优性能；研究了RGB-D手持物体识别技术，结合深度神经网络和传统特征表示的特点，建立了多特征融合方法，进一步提升了识别性，能并将此功能融合到机器博士综合演示平台中；进一步研究了基于地理上下文的图像分类问题，并将此技术集成到ISIA在线图像搜索系统中；不断更新和完善机器博士综合展示平台，并在2015年进行多次演示。下面具体介绍本项目在2015年度的主要进展。

1. 场景识别

1) 基于卷积神经网络的场景识别：物体，尺度以及数据偏差

卷积神经网络的尺度不变性对物体分类有重要的作用，特别是应用于场景识别时，由于场景图像由大量的不同尺度的物体组成，这种属性就尤为重要。同时，由于数据偏差的原因，在不同数据集

上训练得到的模型对物体、尺度等有不同的敏感度。我们提出了融合多尺度、多网络的模型，在MIT67和SUN397上分别达到了86.04%与70.17%的性能。该工作被国际会议CVPR2016录用。

2) 基于概念共生建模的大规模场景识别

场景图像往往是由多种物体或区域组成，具有高度抽象性。不同场景可能会存在相似的物体或区域，这种现象被称为场景共生。对场景共生建模能有效的减弱场景共生产生的歧义性，进一步提高识别准确率。我们基于局部描述并融合多种视觉特征建立了场景共生模型，并提出了多种语义描述。该工作被国际期刊Pattern Recognition 录用。

3) 在语义流形上利用多尺度卷积及神经网络的深度上下文关系建模

语义流形是一种描述图像概念分布的概率空间。基于卷积神经网络，学习具有更强描述能力的语义描述，同时基于马尔可夫随机场提出了一种融合多尺度，多网络，多相邻区域上下文关系的模型。该方法在场景识别标准数据集MIT67和SUN397分别达到了86.9%和69.4%的识别率。

4) 大规模视觉识别挑战赛 (ILSVRC2015)

通过融合多网络多尺度卷积神经网络特征，在ILSVRC2015有额外数据的大规模场景分类任

务中，取得了第二名的成绩。

2. RGB-D 手持物体识别

1) 基于异质特征融合的 RGB-D 手持物体识别

我们利用 Kinect 提供的人体骨骼信息和深度信息进行物体分割，综合深度信息和 RGB 信息构建点云提取多种手工特征，同时对 RGB 信息和深度信息提取卷积神经网络特征，使用多核学习与直接串联相结合的方式特征融合并分类。实验结果表明，我们的方法可以有效消除背景噪声，并且效果远好于单特征，在 Seen 和 Unseen 两种测试条件下分别达到了 88.59% 和 75.31% 的准确率。我们还设计不同深度的神经网络和不同的学习策略，使用两层全连接网络进行特征融合分类，该方法有效提高了在 Unseen 条件下的识别准确率。这些工作发表于 Journal of Computer Science and Technology(JCST)2015 上，以及被国际期刊 Multimedia Tools and Applications 录用。

2) 基于 RGB-D 的手持物体识别系统

我们基于手持物体识别技术开发了演示系统 Hand-Object Sense，该工作以技术展示的形式发表在 ACM Multimedia2015 上，系统界面如下图所示。该技术还集成到机器博士系统中，其在人机交互过程中辅助视觉问答模块，同时可以实现与

人的属性和用户问题的信息关联，增强交互体验。

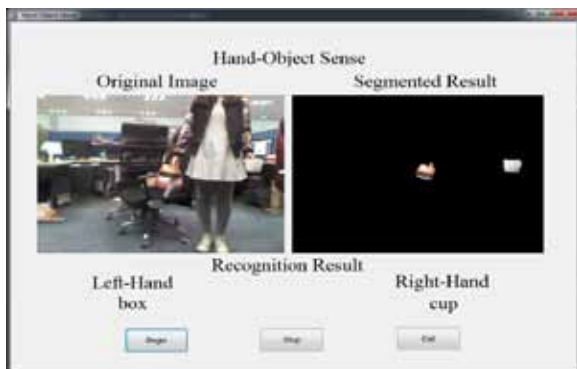


图 1 Hand-Object Sense 系统界面图

3. 机器博士系统

机器博士系统的研发目标是集成相对成熟的智能信息（图像、语音、文字等）处理技术，对领域问答、机器翻译、物体识别、人脸识别、三维交互、智能协调等技术进行了有针对性的研究与开发，并在此基础上进行了系统集成，将各功能模块融合到一个系统中。在 2015 年，我们完善了原有系统，特别改进了视觉识别功能，定制了大白机器人，改进了语料，增加了新闻时事以及计算所在读学生资料，开发实现了初步的综合调度版本，同时还开发实现了移动版机器博士。此外还进行了十余次机器博士的技术展示，包括计算所开放日、接待大中小学学生等。



图 2 机器博士功能演示

重点 / 重大科研任务进展

煤矿信息化关键技术 及集成装备联合研发基地

2012年，为了增强在煤矿信息化关键技术及集成装备联合研制实力，常州联力自动化科技有限公司与计算所协商一致共同成立了“常州联力-中科院计算所基于云平台的煤矿信息化关键技术及集成装备联合研发基地”，在煤矿信息化领域展开长期而深入的合作。

基地成立以来，团队瞄准煤矿信息化过程中安全生产、经营管理和决策数字化等需求，引申出了针对目前网络基础设施架构以及相关业务功能子系统的设计和改造，最终形成了融合传感器数据采集、新一代移动通信无线传输、云计算智能信息处理等技术的数字矿山系统解决方案。

数字矿山系统总体设计包括接入传输层、信息处理层和业务层三部分：

1、接入传输层：基于融合 TD-SCDMA 和 WiFi 技术的多模 Femtocell 基站及 3G 专用网关系统构建小型化、易部署、低成本、可运营的数字矿山通信系统。

2、信息处理层：面向“集团化”构建私有移动云平台，实现集团各个矿井大数据的集中管理、分析、共享、处理等；实现数字矿山集团数据汇聚、挖掘分析等智能信息处理，为科学决策

及应急联动提供支撑。

3、业务层：面向煤矿业务的系统软件，包括在云端的分布式服务子系统，分别对煤矿业务的各类终端和中控软件进行支撑，并为煤矿的安全生产提供了有利、直观的保障。

数字矿山系统解决方案在面向应急救援的脱网直通技术、用户分级管控及服务技术、井下人员逻辑定位技术、移动智能终端业务推送技术、面向“集团化”部署的移动云计算服务平台五个方面亮点突出。

1、面向应急救援的脱网直通技术：面向“矿难”等特殊场景设计应急通信方案，在井上、井下通信链路断裂即脱网后，井下仍可进行组网通信，并可通过自组网技术快速建立井上井下“生命通信通道”，提升救援效率。

2、用户分级管控及服务技术：实现用户的分组以及分级权限管理，根据权限进行业务选通，实现生产作业的精细化管理。

3、井下人员逻辑定位技术：基于 Femtocell 基站小区位置信息及无线信号信息实现井下人员的定位，扩展实现基于位置信息的应急救援、业务感知与推送、考勤等作业管理等功能。

4、移动智能终端业务推送技术：将数字矿山业务系统与移动智能终端相结合，实现移动业务推送，突破接入管理点的地域限制，满足“数字矿山”随时随地的移动办公要求。

5、面向“集团化”部署的移动云计算服务平台：面向“集团化”构建私有移动云平台，实现集团各个矿井大数据的集中管理、分析、共享、处理等，实现数字矿山集团数据汇聚、挖掘分析等智能信息处理，为科学决策及应急联动提

供支撑。

在联合研发基地的支持下，团队面向“数字矿山”应用需求，提出了脱网直通、用户分级管控、人员定位、移动业务推送、集团化移动云计算平台等创新解决方案，目前已经成功与神华集团、中煤集团、晋煤集团、徐矿集团等开展合作，并应用到神东“锦界煤矿”、“塔然高勒煤矿”、徐矿集团“秦源煤矿”等重点煤矿中，取得了良好经济效益和社会效益。



图 1 数字矿山系统总体设计



图 2 矿用调度系统

重点 / 重大科研任务进展

新华社 713 实验室技术研究项目 — 大数据与智能信息处理

本项目的目标是通过大数据与智能信息处理技术的研究和示范应用，突破新华社新闻信息集成服务平台的关键技术，从而提升整体服务水平，促进新华社生产传播模式转型升级，为实现媒体融合发展奠定重要基础。项目起止时间是 2015 年 1 月至 2016 年 12 月。

2015 年本项目已经完成研发工作，部分关键技术研发成果如下：

1. 分布式列式数据管理系统

传统的大数据处理方式一个方面是关注于数据规模的庞大，这是数据存储本身的内容，对于计算融合性不够，因此不能够进行高效处理。这个方面就需要进行体系结构的改进，将存储与计算进行融合，从而可以建立高效的大数据处理框架。分布式列式数据管理系统（又称“天玑 Base”）是适合大数据场景的海量分布式数据库系统。在架构上兼容了 Hadoop 体系的高可扩展性，可支持上百个节点，PB 级存储规模的线性扩展。性能上其针对大数据应用场景进行深度定制和调优，尤其对于高吞吐率入库和实时检索场景。

功能上除了支持标准的 K-V 访问接口之外，还可兼容 SQL 标准及 JDBC 接口，可以很好的与既有数据分析业务对接，并且支持索引性能优化以及对后缀和全文索引的支持。

2. 新闻个性化推荐系统

如何利用大数据来对用户进行个性化新闻推荐技术对促进网络新闻媒体的发展有重大意义。通过对大量用户历史浏览行为的分析，利用人工智能技术来发现用户的兴趣，为用户选择并推荐合适的新闻，从而帮助用户更快更好的获取感兴趣的内容，为用户提供更优质的服务。新闻个性化推荐系统的训练模型采用了业界领先的 SVDFeature 算法，此算法在 KDD 2012 年推荐国际竞赛获得冠军，是业界领先的推荐算法。本系统在实现过程中针对新华社实际的数据和系统情况进行了适应性改造，进一步提升了算法的精度。在算法的实现过程中采用基于分布式计算框架 Spark 进行实现，成功将 SVDFeature 训练算法扩展到分布式集群，实现了分布式的训练，并采用业界成熟的数据分布和模型平均方案，成功

解决了大数据环境下 SVDFeature 模型训练的速度和规模问题，并通过了压力测试，表明算法的运行时间与训练数据模型成线性增长，与运行节

点数目成线性下架，充分说明了技术方案先进、合理。此技术方案也可以成为其它机器学习分布式实现的参考。

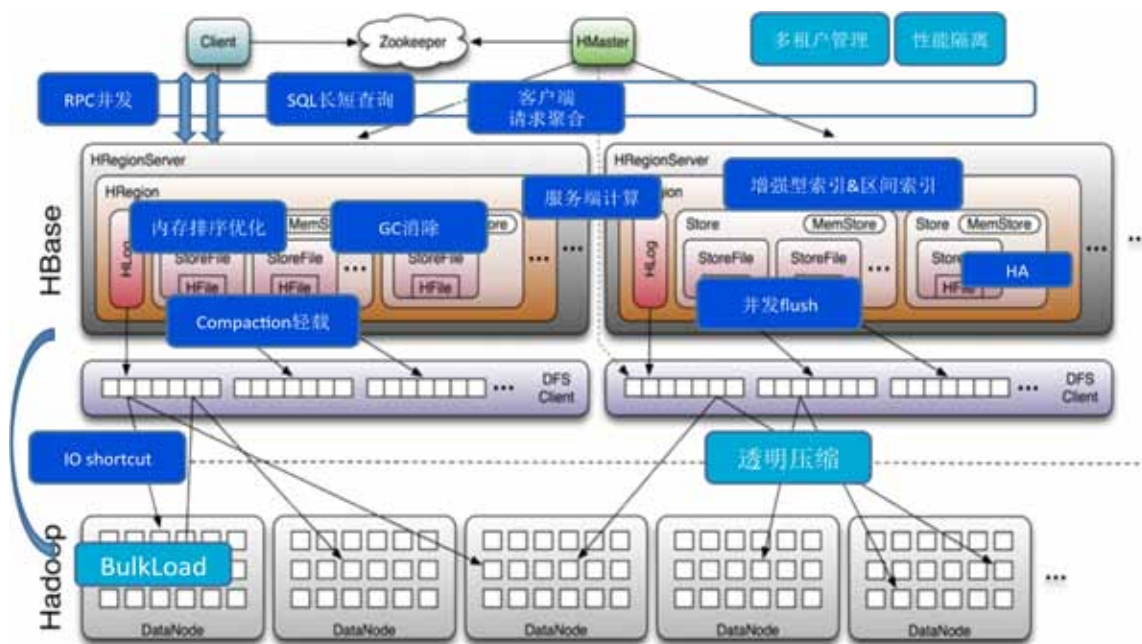


图 1 分布式列式数据管理系统关键技术

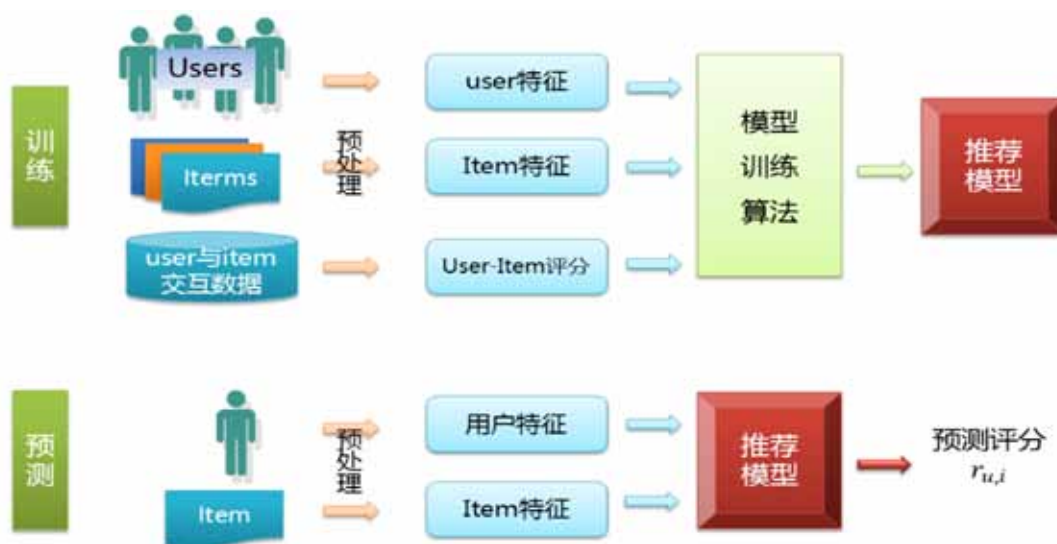


图 2 新闻个性化推荐系统算法流程



研究方向科研进展

研究方向科研进展

编译与编程

研究围绕两个关键问题展开：一是如何降低多核处理器的编程和优化难度，二是如何有效改善复杂软件系统的安全性和可靠性。在高通量编程方面，本研究团队从算法、编程、运行时环境三个方面协同解决高通量应用的可编程性和效率问题。在算法层次，我们研究了在算法层面发觉并消除冗余计算的方法，可显著改善调度中心类算法的效率（4.6 倍的性能提升）；在编程层次，我们提供 Hadoop+ 来屏蔽底层异构芯片对用户所造成的编程困难，为用户呈现统一的编程界面；在运行时环境层次，我们通过构建同构和异构的资源竞争模型，来为高通量应用分配合理的资源。在内存资源的有效管理方面，提出了：(1) “垂直”内存管理机制，协同管理 Cache 和 DRAM Bank；(2) OS 内核态的访存行为监控机制，建模对各层次内存资源的需求；(3) “非对称”的多级内存资源分配方式，依据程序的访存特点，动态协调 Cache 和 DRAM Bank 的资源配比。上述方法在真实机器上测试表明，其性能优于 Linux 内存管理系统，整体性能平均提升 15%。在面向软件健壮性方法，本团队提出了一个新的方法有效结合路径敏感的数据流分析和经典的指针分析

来精确分析程序中复杂的指针关系，大大提高了已有静态分析检错方法的能力；实现了一个初步的静态分析检错工具原型，可以有效地检测出已有工具和方法无法检测出的一些 use-after-free 类型错误；实现了一个自动分析 HIVE 日志提取执行流的工具，目前正在研究通用方法来自动分析不同系统中的日志信息。在面向安全的编译方面，本团队研究的内构安全编译平台将安全能力设计到应用程序中，从而减少 / 消除软件中潜在的安全缺陷，并阻止攻击者对软件缺陷的利用，本年度本团队创新地提出了二进制应用程序持续随机化思想，并在平台中实现；在维护 DEP 机制的情况下，持续地对二进制代码进行随机化，使得攻击者即使同时获得了内存泄露和控制流劫持相关漏洞，也无法进行攻击；并在代码混淆和 ROP 等技术方面取得了研究成果。上述研究得到了国家 973、863、自然科学基金项目的支持。

VLSI 测试验证、可靠设计与容错体系结构

围绕 VLSI 测试验证、可靠性设计、容错体系结构等开展深入研究，取得的突出进展如下：

1) 在 VLSI 测试验证方面，提出了基于通路约束

求解的半形式化激励生成方法，显著提升了验证的质量和效率；提出了一种基于芯片保持时间测试结果的细粒度刷新 DRAM 重组技术，有效降低了 DRAM 刷新功耗；2) 在可靠性设计方面，提出一种混合弱 PUF 和强 PUF 的物理不可克隆函数设计，使随机性接近理想值，而稳定性达到 97%，有效提高了电路的抗建模攻击能力；3) 在 VLSI 设计与容错方面，提出了实现存算一体化的计算型内存体系架构，针对大数据图计算应用可将访存性能提高 2 倍；提出了函数近似计算方法和芯核分级策略，大幅降低计算能耗和提高了处理器的可靠性。相关研究成果发表在 DAC、DATE 等国际知名会议，以及 IEEE Trans. on VLSI Systems 等国际知名期刊上。

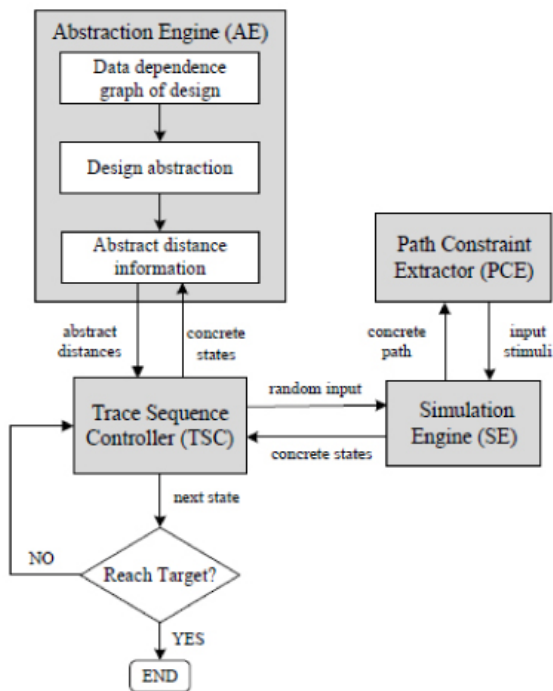


图 1 基于通路约束求解的半形式化激励生成方法

类脑计算机

让计算机能以类似人脑神经网络的方式高效

地进行学习等认知活动，是计算机科学的终极目标之一。但通常的计算机在模拟包含百亿个甚至更多的神经元及突触的深度神经网络进行学习时，处理速度和效率非常低下。未来计算团队在类脑计算机方向取得重要的进展。该团队基于在 2014 年已提出的寒武纪 1 号神经网络处理器架构，在 2015 年完成了寒武纪神经网络处理器的芯片设计、前后端和硬件平台验证，并已交由流片。此外，该团队提出了支持目前几乎所有主流的机器学习算法的首个通用机器学习处理器（其功耗为主流 GPU 的 1/200 而性能是主流 GPU 的 1.14 倍，发表在本领域最好的国际会议之一 ASPLOS）和基于机器学习的摄像头智能识别 IP（其能效是主流 GPU 的 4688 倍，发表在本领域顶级会议之一 ISCA2015）。

数据管理研究

数据管理研究组围绕大数据对系统平台的三大主要挑战（数据量大 Volume，数据的产生处理速度快 Velocity，数据的种类繁多 Variety）展开深入研究。针对 Volume 和 Velocity，主要工作包括硬件优化和增量计算两方面。2015 年，面向新兴的非易失主存优化广为使用的 B+Tree 索引结构，论文发表在 VLDB' 15，并申请获得了一项国家自然科学基金项目，继续探索面向 NVM 主存的数据管理。在增量计算方面，发表了 TKDE' 15 论文，针对大数据迭代计算进行增量计算优化，实现了一个 KIS（Key-value Incremental System）系统，其性能显著超越华为项目规定的优秀标准。针对 Variety，开展面向多种数据类型的大数据管理系统的研究，进一步优化处理半结构化树状数据的数据分析系统 Steed，实现了更加丰富的运算和多线程的执行，

性能比 MongoDB 快 2-100 倍。此外，分布式多线程同步图计算平台 GraphLite，在国科大春季课程“大数据管理系统”中，有 70 人在作业中使用，检验了系统的稳定性。

并行软件研究

1. 在并行计算模型方面，重新从一种全局观点解释时间重用和空间重用间统计变换，提出等价的另外两种公式，并提出了序列局部性概念。

2. 众核平台核心算法优化方面，对 Sharpness 开展了一系列优化方法研究，GPU 加速比最高达到 69.3 倍，论文发表在 ICPP 上。在众核架构上优化了快速卷积运算，提出了多通道卷积神经网络的 2D 卷积操作的有效解决方法，增加了数据的复用率，在两个 GPU 平台上性能提升 33% 和 28%，论文发表在 HPC 上。对动力学蒙特卡罗 KMC 并行算法进行通信聚合与通信调度优化方法显著提升了性能，相关工作发表在 HPCMS 上。

3. 在算法自适应框架调优方面。提出一种新的多核集群上稀疏矩阵向量乘法的自动调优框架，减少内存开销及节点间通信量，最高获得 1.5 倍性能提升。相关论文发表在《中国科学》上。

4. 在并行程序设计框架研究方面。提出了一种面向图像和无线通信领域、基于并行优化模式的高性能 OpenCL 编程框架 ParaCV，并行优化框架将高层语言程序源源变换为针对特定硬件平台的高性能 OpenCL 程序。

操作系统研究

研究焦点是构建高并发、高可扩展、高性能、高包容性、高灵活性大数据计算平台 RainbowH，其设计目标：提供多 PB 级源数据（结

构化、半结构化、非结构化）存储能力；支持多种分布式文件系统存储灵活扩展；支持千万用户级并发访问服务；业务运行时系统全部功能模块支持自动恢复；支持用户权限管理和在线日志定制，为此围绕 4 个挑战问题展开研究：平台加速方法、平台数据统一布局方法、平台高并发支持方法和平台可靠性方法，其性能在 SQL 操作方面较 SparkSQL 提升 100 倍，研发了多级缓存动态融合及其数据放置策略等十余种关键技术，申请了 6 项专利和发表了 2 篇论文。RainbowH 也在不断发展和完善如高效自动部署等方面，也不断推动 RainbowH 的商业应用。培养了若干名硕士生和博士生。

人脸识别

本年度课题组围绕真实复杂场景下的人脸识别与群体化图像视频分析这两个方向开展研究，针对图像视频结构与内在关联特性相关的表示模型这一核心问题，在大规模人脸识别多领域适应学习、跨图像视频信号与特征协同的层次化表示、图像视频关联数据的结构化表示等方面均取得了一些重要进展。针对视频人脸识别的建模和分类问题，提出新型的黎曼流形度量学习框架，在公开的视频人脸识别和物体分类数据库上进行了验证；面向大规模视频人脸检索中的图像和视频之间的跨模态比对问题，提出跨欧式黎曼空间的异质哈希学习方法，有效地将图像和视频在统一的二值编码空间中进行精简表示和高效比对。在此基础上，提出基于层次化特征提取表示的深度监督哈希学习方法和面向多任务检索的多功能二值编码学习方法，以解决大规模图像检索和多任务检索的问题。

课题组完成了人脸检测、特征定位与识别技

术的全面深度学习化，从而使得系统性能大幅提升，并因此成功争取到了中移在线、华为、亮风台等产业公司的物体检测与识别项目。

基于上述研究成果，本年度课题组发表论文 27 篇，其中主流国际期刊论文 9 篇（包括 IEEE Trans. PAMI 论文 1 篇，IEEE Trans. Image Processing 论文 2 篇），国际会议论文 18 篇（包括 CCF A 类会议 CVPR 论文 4 篇，ICCV 论文 5 篇，ICML 论文 1 篇）。

课题组还参加了 IEEE Face and Gesture Recognition 大会组织的视频人脸识别竞赛冠军，分别获得 ICCV2015 ChaLearn: Looking at People 两个竞赛的冠军和亚军。博士生刘昕、刘梦怡和李岩还组队参加阿里图像搜索大赛，获得了第一名的好成绩。

此外，为了课题研究开展需要，在大规模、层次化图像数据储备方面开展了大量工作，初步分级、分类构建了 1000 类具有语义类别标注的高清图像数据库，并初步完成了相应的视觉语义属性标注，用于后续层次化、结构化关联与表示方法模型的研究。

自然语言处理

自然语言研究方向整体稳中求进，在持续发展统计语言处理与机器翻译的同时，积极开展深度学习研究，在神经语言处理与机器翻译上也取得了显著进展，具体表现在以下四个方面：在联合成分句法与依存句法分析这一更复杂的语言分析问题上，提出跨语言联合成分依存句法推导方法，显著提升了成分句法分析精度和依存句法分析精度。针对文本匹配问题，提出基于文本子结构和深度学习技术的文本匹配模型，有效地求解任意长度文本的语义相似度匹配问题，在公开的

对话数据集上，我们的模型取得了当前最好的结果，显著超过其它方法。面向词序列的预测问题，提出全新的基于卷积神经网络的语言模型以处理任意长度的句子，更好的抓住语言的语义信息。对于统计机器翻译的解码，提出卷积神经玩过架构来动态地编码源语言句子的相关信息，有效预测目标语言生成序列的源语言部分，并将其与源语言整个句子一起编码为连续空间表示。

图像视频建模

以机器学习的模型和方法为基础，围绕计算机视觉领域的图像视频建模问题展开研究工作，特别关注图像视频的描述、度量、预测和增强等问题，主要应用场景包括图像视频检索和智能视频监控。在过去的一年中，该研究方向上的主要研究点和研究成果简述如下：在机器学习方法方面，提出了属性学习的新模型和新方法，并成功的应用于解决图像集合描述等问题；在基于深度学习的图像视频表示方面，提出了图像中层特征表示的框架模型；在智能视频监控相关的问题上，致力于研究行人 / 车辆检测、跟踪和再识别等关键问题，实现了基于字典学习的跟踪方法、基于深度模型的目标检测和行人再识别方法，在公开数据集上的测试结果表明我们的方法已经取得了国际领先的性能。

智能人机交互

本年度课题组仍坚持围绕手势交互、手语识别方向展开研究工作。在连续手语识别以及非特定人手语识别方面，取得了一定的突破。课题组提出基于层级格拉斯曼协方差矩阵（HGCM）手语描述，结合多时域置信传播（MTBP）策略，进行连续手语句子的识别。此外，针对连续手语识

别中一直存在的运动插入 (ME) 问题, 课题组提出虚拟 ME 生成方法, 利用 LSTM 进行手语句子的识别和分割。该方法与已有方法有显著的不同, 无需对 ME 进行建模, 在识别时也不需要 ME 进行判别。在 450 句真实句子集合上的实验结果表明, 本方法的召回率和精度均比传统的 HMM 方法有显著的提高。对于非特定人手语识别问题, 课题组提出迭代的参考驱动的度量学习方法 (iRDML), 每一个词汇用一个参考特征向量来进行描述, 进而迭代优化参考向量及对应的距离度量矩阵。该方法在 1000 词的非特定人手语识别任务上, 可得到 76.3% 的识别精度, 相比其他主流方法有显著的性能改进, 并且该方法所需的时间代价极低, 适合实时的手语识别应用。

除了在手语识别核心算法方面的研究, 课题组还着眼于对手势交互方向的布局。尝试利用层级自回归网络进行 3D 手部重建的研究, 并取得初步的成效。后续希望结合现有算法开展精细的手势交互任务的研究。

数据挖掘与机器学习

本年度课题组继续坚持原始算法的创新, 引入深度学习用于学习语义特征, 提出基于自动编码器的迁移学习算法, 实现了深度学习与迁移学习的融合。在复杂事件关联方面, 针对互联网上复杂事件关联, 且事件在不断地更新和涌现, 设计了一种紧凑、高效的数据结构来存储事件场景, 提出一种新的在线频繁场景挖掘算法, 实现了高效、在线、增量序列关联分析挖掘。在贝叶斯大间隔学习方面, 结合主成分分析 (PCA) 和大间隔学习, 提出了一种基于后验正则化的贝叶斯方法, 实现了基于数据分布的新型主成分分析方法。在深度学习方面, 提出了深度极限学习机模

型。该模型采用极限学习机作为基本的层构建方法, 以随机平移与核化作为构建的基本元素。在模型的每一层, 将数据通过极限学习机获得的预测结果进行随机投影, 然后应用特定的核函数从而生成最终的特征。上述算法研究相关成果分别发表在 IJCAI 2015、ICDE 2015、AAAI 2015 以及 Neurocomputing 上。

在大数据挖掘算法应用方面, 实现了高效并行的关联交易发现算法用于上海证券。此外, 庄福振指导学生参加了 IJCAI 2015 的数据挖掘竞赛, 提出一套完整的方案对可能重复购买的用户进行预测, 从来自 28 个国家 753 个参赛队中胜出获得 IJCAI 数据挖掘竞赛第一名。

网络数据

本年度有三项较为突出的成果。一是数据复杂性规律发现方面, 提出了社会媒体的自适应情感分类方法, 该模型在协同监督的框架下, 利用话题相关情感词的倾向性动态修正和估计、用户层及提及关系网络的倾向相关性等, 对社会媒体大数据在不同的话题情感空间进行自适应的、非文本与文本特征的协同度量和建模, 从而有效地进行情感分类。论文发表于 IEEE TKDE。二是在大规模文本数据的单词表达方面, 提出了综合纵横关系和构词方法的单词表达方法, 在单词相似度和单词类推等任务上均取得了很好的效果, 相关论文分表发表于 ACL 和 AAAI。三是在大规模文本数据的匹配学习方面, 提出了基于 CNN 的层次化模型和 LSTM 的多角度匹配模型, 在问答, 信息检索和对话等应用中均取得了良好的效果, 系列相关工作发表于 AAAI。

● 社交媒体自适应情感分类

社交媒体数据的情感分类由于话题的敏感

性,使得分类的数据特征具有复杂性,很难利用单一的方法进行有效度量。主要表现在一个话题上提取的情感特征及训练的模型,在其他的模型上往往表现得很差。尤其在对社会媒体的情感分析上面表现得更加明显,这是由于社会媒体话题非常多样,对所有的话题数据训练一个普适的分类器是不可能的。同时,与在线产品评论不同的是,类似微博的社会媒体服务往往没有一个打分机制来收集情感标注,并且极稀疏的社会媒体文本内容也使得一般分类器的训练效果不好。因此,针对社会媒体大数据在情感空间上所表现的话题敏感、文本内容极稀疏的复杂特征,本工作提出了一种话题自适应的情感分类模型(TASC)。该模型在协同监督的框架下,利用话题相关情感词的倾向性动态修正和估计、用户层及提及关系网络的倾向相关性等,对社会媒体大数据在不同的话题情感空间进行自适应的、非文本与文本特征的协同度量和建模,从而有效地进行情感分类。相比其他监督的和半监督的分类模型,在6个话题的公开数据集上实验,TASC取得了较好的改进效果。同时,考虑到社会媒体话题的动态演变特性,我们在原有模型基础上设计了随时间线动态自适应度量的框架(TASC-t),该方法也同样取得了较突出表现。

● 大规模文本数据的单词表达学习

单词表示一直是人工智能、自然语言处理、信息检索等领域的一个基本核心问题。分布式表达将单词表示为低维连续实数向量。受Bengio 2003年在神经网络语言模型中使用单词分布式表达所取得成功所启发与鼓舞,单词的分布式表达正成为自然语言处理领域最火热的一个研究点。同时,它也被广泛使用在自然语言处理的各方面,如语言模型,机器翻译,情感分类等。数十年

来,各种不同的单词表示方法层出不穷,如矩阵分解(LSI, GloVe),概率话题模型(LDA),神经网络(NNLM, SENNA, Word2Vec)。这些单词表达学习算法大体都是基于一个同样的假设——分布式假设。其假设一个单词的语义由其周围的上下文决定。由于单词之间存在着横向(syntagmatic)和纵向(paradigmatic)两种关系,如图2所示。其中,横向关系主要关注的是词与词之间在上下文中的共现关系;而纵向关系则关注的是词与词之间具有相似上下文的关系。现有模型都只单独考虑了一种关系。如隐式语义索引(latent semantic indexing, LSI)利用了横向关系;而Word2Vec这类方法利用了纵向关系。

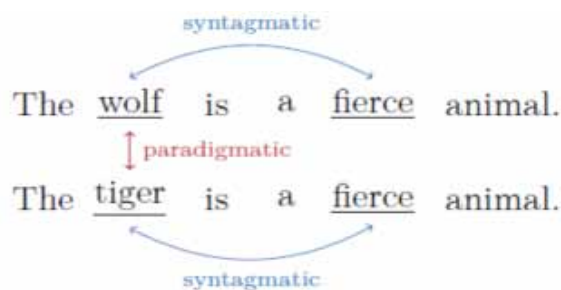


图2 paradigmatic 与 syntagmatic 关系示例

在ACL2015的工作提出了两种新的单词表达学习模型,有别于现有模型只建模单词间的横向关系或纵向关系,以并列(PDC模型)或层次(HDC模型)的方式同时建模这两种关系,以得到更好的单词表达。PDC模型和HDC模型对应地扩展了Word2Vec中CBOW和HDC模型,在其基础上,利用文档表达来预测文档中出现的单词,以捕捉单词间的横向关系。这两个模型在单词类比任务和单词相似度任务中都取得了state-of-the-art的结果。

然而,基于分布式假设的模型往往难以处理

那些出现次数很少的单词。这些单词的上下文稀缺，不足以学出良好的表达。极端情况下，如果一个单词没有任何上下文，这时分布语义假设无能无力。另一方面，分布语义假设，也难以捕捉单词见细粒度的语义关联。如，基于分布语义假设的模型，难以回答“buys 与 sells 哪个单词与 buy 更相似”这样的问题。使用单词的构词信息 (morphological information) 可以很容易的回答这类问题。如，没有任何上下文，依然可以仅通过“unbreakable”的词素：“un”、“break”和“able”得到它的语义。也能很容易的回答“slow: slower as fast: ___”这样的问题。但是仅仅利用词素信息无法捕捉两个不共享词素单词之间的关联。

因此，为了学习更好的单词表达，需要同时利用单词的内部词素与外部上下文信息。与以往工作使用复杂的组合方式利用词素信息构造神经网络语言模型不同，在 AAAI 2016 的工作使用简单通用的预测模型，同时建模单词的内部词素信息与外部上下文信息。此工作分别扩展了 Word2Vec 的 CBOW 与 SG 模型，得到对应的 BEING 与 SEING 模型，模型框架图见图 3。该方法在单词类比和相似度等任务上取得了更好的结果。

● 大规模文本数据的匹配学习

文本语义匹配是诸如信息检索和问答系统等核心任务之一。比如在信息检索中，文本语义匹配就是匹配一个查询和一篇文档。现在已有的文本匹配的模型主要分成两大类，传统模型 (BM25) 和新兴的深度学习模型。深度学习模型有自动学习和提取特征的优势，在很多任务上已经超过传统模型。在目前对深度模型的研究当中，往往会讲文本分别表达成两个等长的向量，然后直接计算文本间的相似度 (余弦距离，欧式距离)。这类模型会遇到如下问题：(1) 难以表达细粒度的匹配信号，导致没法充分捕获细粒度的匹配信号；(2) 两段文本在最后才产生交互，计算相似度的方式过于简单，没法进行度量学习。为了解决这些问题，我们提出了 MatchPyramid 模型从最基本的匹配信号，然后多粒度、层次化得提取两段文本的匹配信号。实验表明，MatchPyramid 模型的性能高于传统模型，提高了 6%，相对现有的深度学习模型 (ARC-I) 提高了 4%。MatchPyramid 借鉴了在图像识别领域最成功的卷积神经网络 (CNN) 模型的建模思想，希望能和 CNN 模型一样从最基本的元素开始，进行层次化建模，下一层是上一层的进一步抽象，

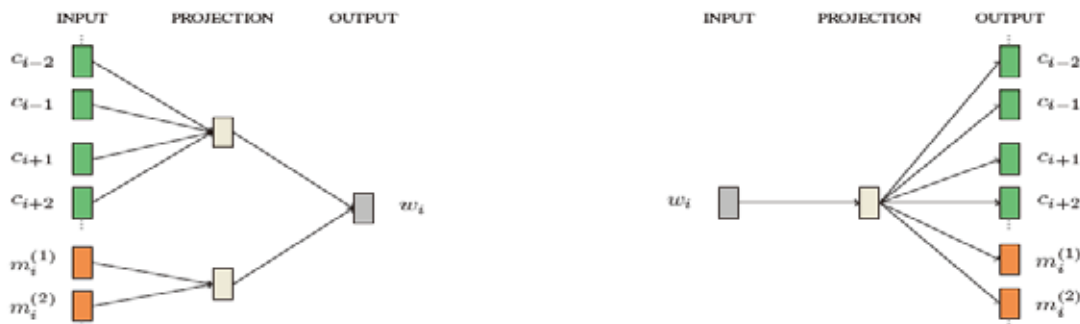


图 3 BEING 与 SEING 模型

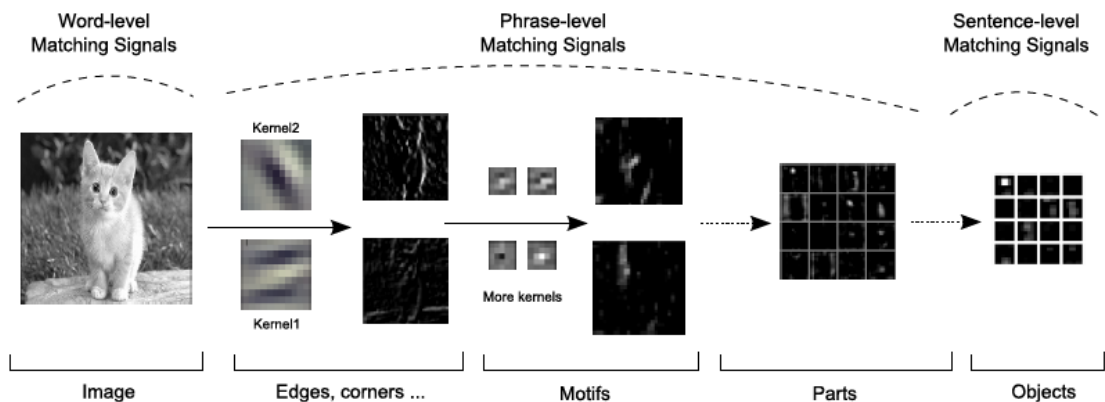


图 4 文本匹配类比图像识别的各个部分

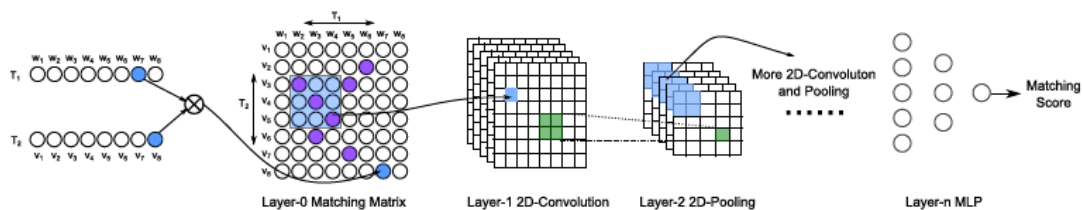


图 5 MatchPyramid 模型示意图

进而从词、短语、句子级别来表达两段文本的匹配。

MatchPyramid 模型从最基础的词的匹配信号开始，完整保留了两段文本的匹配信息，而后面的层次化卷积操作能够提取复杂的匹配特征，可以更好的完成文本匹配任务。

在文本匹配中，局部化信息和丰富的上下文信息对决定最终的匹配度非常重要。然而，现有的深度学习模型并不能同时兼顾这两个重要的因素。我们提出位置化句子表达（positional sentence representation）的概念以及一个新的深度模型（MV-LSTM）来解决这个问题。一个位置化句子表达就是聚焦于一个具体单词(位置)上的整个句子的表达。在 MV-LSTM 中，我们先基于双向 LSTM 模型学习文本语义表达。由于双向 LSTM 模型的特点，其学习到的在每个位置上

的中间表达就是一个位置化句子表达。然后我们使用 neural tensor network 来建模两个句子不同的位置化句子表达间的交互，并综合得到最终匹配得分。我们主要基于一个社区问答数据集进行实验，具体任务是匹配一个问题和其可能的答案。实验结果表明，我们的模型能够找到句子重要的匹配位置进行匹配，并相比现有的深度学习模型取得了更好的效果。

跨媒体计算

2015 年度积极推进科研仪器研制项目，坚固在国家互联网音视频监管任务中的地位，在学术研究、科研仪器研制、关键技术应用三个方面取得了显著的进展。学术研究方面，在图像恢复、排序学习、深度学习、目标检测技术的研究上取

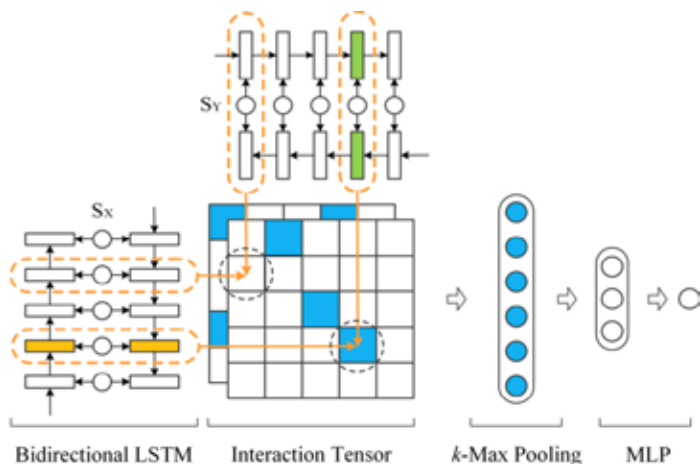


图6 MV-LSTM示意图

得了突破：提出了全局测量局部稀疏的图像恢复算法，提高图像重建质量 10%-20%；提出了一种在线排序学习方法，显著提高了排序模型的训练效率；提出了基于 Sparse CNN 的稀疏集成学习方法、基于密集区域融合的目标定位方法，提高了目标检测的精度，并在全球 70 多支队伍参加的国际著名 ImageNet 竞赛 ILSVRC 2015 中取得优异成绩：目标分类定位任务第四（国内第二）；目标检测任务排第五（国内第一）。发表多篇高水平论文：包括国际顶级期刊 IEEE TIP 1 篇（CCF A 类期刊），IEEE TMM 1 篇；国际顶级会议 IJCAI、CVPR、ACL（CCF A 类会议）论文 3 篇。科研仪器研制方面，初步实现了动态可配置的压缩感知成像装置，设计了一套完整的硬件化采集重建架构，实现高速采集和重建结果高速回传。关键技术应用方面，相关核心技术在互联网音视频分析和大规模社交媒体数据分析中得到了良好应用。研发的大规模近似图像检索识别系统，已经被西南电子电信技术研究所引进集成于国家某重大专项任务中，在全国多地部署应用，应用效果显著。

虚拟现实

2015 年度围绕 VR 数据获取、建模仿真两大核心问题开展研究，并注重研究成果在核电应急、电子商务等领域的应用。在数据获取方面，研究基于多源感知信息融合的场景深度图获取技术，采用主动式深度传感器和被动式立体视觉相互融合的方式，获取大规模、复杂场景下的高精度深度图。深度图估计方法在国际公认的 Middlebury 深度图估计评价数据集上的测试成绩排名第一。

在自然现象建模仿真方面，提出启发式层析图像火焰重建算法，通过能量约束优化模型使得重建图像的能量尽量逼近拍摄图像的能量，实现不同燃料、容器、外力和粒子的复杂火焰建模。在群体运动仿真方面，研究并提出情绪驱动的社会集体行为建模方法、适用于大规模交通仿真的路网表示等方法，并完成在核电领域多个场址的示范应用。

人体运动仿真

2015 年度在自然人体运动风格建模方面取得研究突破，研究论文发表在计算机图形学与交互式技术国际会议 SIGGRAPH 2015，被 ACM Transaction On Graphics 期刊（图形学领域唯一 SCI 一区刊物）收录。自然人体运动包含各种结构异质的具体动作，以及多样化的风格，因此，该研究工作主要存在两个方面的技术挑战：不同风格运动之间的时空关系异常复杂难以表达；人体运动中的动作和风格的识别和理解极其困难。我们提出一种混合自回归模型实时在线建模不同风格运动之间存在的复杂时空关系，并且自动转换未标记动作和风格的人体运动。该算法适用于

不同来源的输入数据。目前,已有来自美国、英国、法国、韩国等国家的科研团队向我们索要论文、代码和数据以开展进一步的研究工作。

生物信息学

生物信息学方向,2015年度继续坚持在计算蛋白质组学及其相关方向开展工作。蛋白质交联鉴定算法 pLink 相关成果于 2015 年再次突破 Nature Methods (五年影响因子 27.2), pFind 和 pLink 软件累计注册下载用户分别达 815 和 545,遍布世界六大洲的 34 个国家和地区,在国内外同行的 113 项研究成果中得以应用。在冷冻电镜图像三维重构方面,提出一种全新的冷冻电镜图像对位算法,将原图像特征识别问题转化为采样与分类问题,有效地解决了图像对位的精度和处理速度问题。相应的软件 MarkerAuto 已成为冷冻电镜图像对位处理领域的首选软件,已在中科院生物物理所、美国加州大学圣地亚哥分校 UCSD、美国 Direct Electron 得到实际应用。与美国国立卫生研究院 (NIH) 国立心、肺、血液病研究所合作研究,在国际著名学术期刊《Cell Metabolism》发表文章“A Liver-Enriched Long Non-Coding RNA, lncLSTR, Regulates

Systemic Lipid Metabolism in Mice”,在国际上首次绘制了长非编码 RNA (lncLSTR) 在脂类代谢过程中的精细调控机制,为该领域的进一步研究开启了新的思路。

计算机基础算法

2015 年度主要围绕着动态网络相关算法研究、博弈论相关算法研究、通信复杂度、判定树复杂度等方面展开,提出了动态数据模型下排序与查找算法,并确定了可以无差错求解排序的最高阈值,解决了带外部影响的蛋糕分配问题,提出了新的规约技术证明了单向通信模型下图联通性、二部性等问题最优的通信复杂度上下界,证明了对任意异或型判定树在以牺牲一个平均敏感度为代价,可以转化为传统的判定树。15 年共发表会议论文 5 篇,SCI 期刊论文 1 篇,包括 ISAAC, TAMC, COCOON, RANDOM-APPROX, ICDMW 等会议,Theoretical computer science 等期刊,另有 1 篇论文被 AAAI'16 接收。

微处理器

在微处理器结构设计方面,研制了新一代龙芯 GS464E 处理核,并在此基础上研制成功

Cell Metab. 2015 Mar 3;21(3):455-67. doi: 10.1016/j.cmet.2015.02.004.

A liver-enriched long non-coding RNA, lncLSTR, regulates systemic lipid metabolism in mice.

Li P¹, Ruan X¹, Yang L¹, Kiesewetter K¹, Zhao Y², Luo H², Chen Y³, Gucek M³, Zhu J⁴, Cao H⁵.

Author information

Abstract

Long non-coding RNAs (lncRNAs) constitute a significant portion of mammalian genome, yet the physiological importance of lncRNAs is largely unknown. Here, we identify a liver-enriched lncRNA in mouse that we term liver-specific triglyceride regulator (lncLSTR). Mice with a liver-specific depletion of lncLSTR exhibit a marked reduction in plasma triglyceride levels. We show that lncLSTR depletion enhances apoC2 expression, leading to robust lipoprotein lipase activation and increased plasma triglyceride clearance. We further demonstrate that the regulation of apoC2 expression occurs through an FXR-mediated pathway. lncLSTR forms a molecular complex with TDP-43 to regulate expression of Cyp8b1, a key enzyme in the bile acid synthesis pathway, and engenders an in vivo bile pool that induces apoC2 expression through FXR. Finally, we demonstrate that lncLSTR depletion can reduce triglyceride levels in a hyperlipidemia mouse model. Taken together, these data support a model in which lncLSTR regulates a TDP-43/FXR/apoC2-dependent pathway to maintain systemic lipid homeostasis.

Copyright © 2015 Elsevier Inc. All rights reserved.

PMID: 25738460 [PubMed - indexed for MEDLINE] PMID: PMC4350020 Free PMC Article

图8 生物信息非编码与复杂网络组发表在 Cell Metabolism 的文章

四核龙芯 3A2000 芯片。GS464E 处理器核基于 LoongsonISA 指令集，采用 4 发射乱序执行结构，集成三级 Cache(一级指令 / 数据 Cache 各 64KB、二级 Cache 为 256KB、三级 Cache 为 4MB)，具有高性能的特点。GS464E 处理器核是在 GS464 核基础上改进的，具体包括：优化了访存性能和分支预测准确率；实现了 MIPS DSP 指令集和虚拟机支持；增大了处理器中各项队列的项数，并增大了 Cache 容量和 TLB 容量；支持 3 级 Cache 结构，每一级都采用 LRU 替换策略，可以支持多核缓存一致性协议；支持多层次的硬件预取机制。首款集成 4 个 GS464E 处理器

核的采用 SMIC 40 纳米工艺的龙芯 3A2000 已流片成功，在主频 800MHz 时实测 SPEC CPU2006 定点分数为 5.89 分，在同主频下性能达到国际上 AMD 主流处理器相当的水平。这表明 GS464E 处理器核架构设计已经达到世界先进水平。目前正在研制基于 28 纳米工艺的更高主频性能的龙芯 3A3000 芯片。在应用推广方面，目前已经开发了基于龙芯 3A2000 的桌面计算机及服务器，以及多款应用于军工领域的计算机板卡等，在国家工信部的党政军示范项目中得到应用。基于 GS464E 处理器核的多核芯片完全能够满足国产信息化建设的需要，具有很好的应用前景。

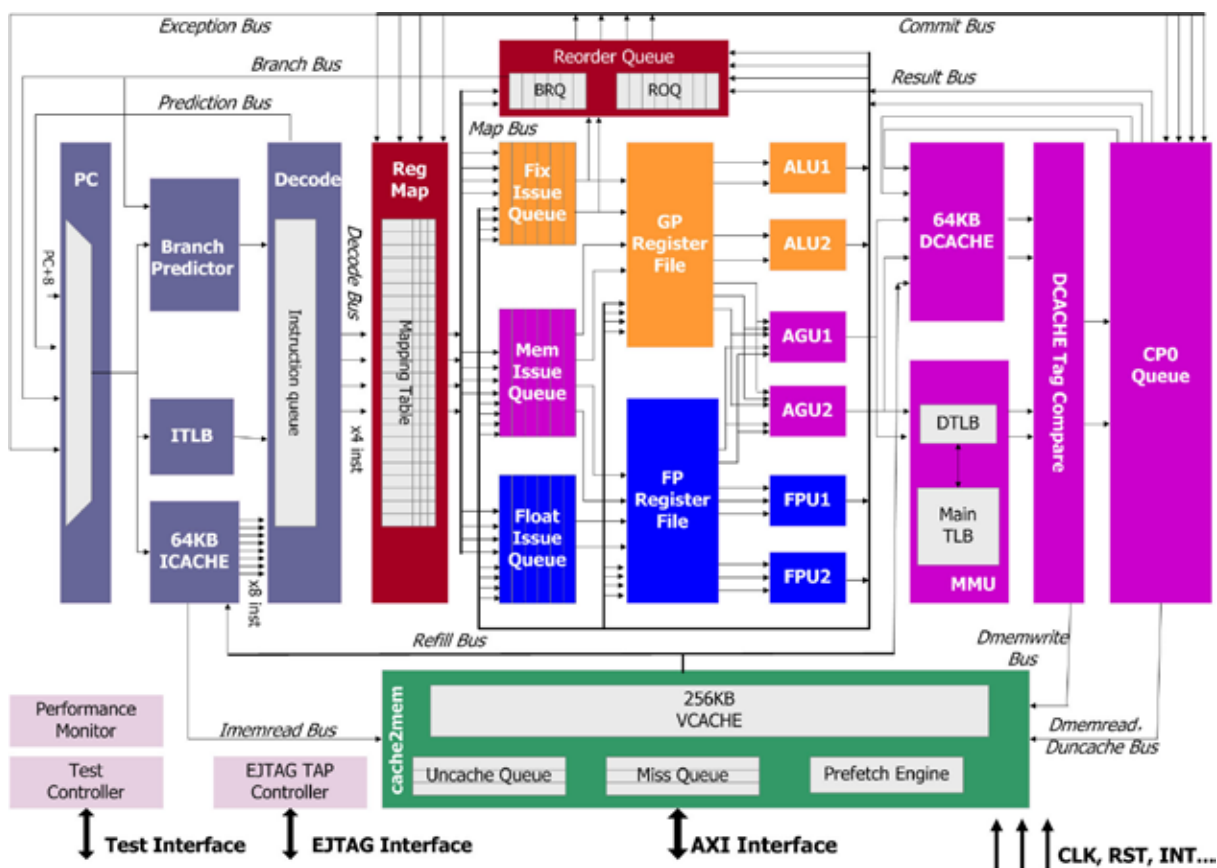


图 8 GS464E 微结构图

数据存储

本年度数据存储技术在支持万亿文件扩展的元数据服务集群、面向海量小文件低延迟访问的授权型客户端缓存优化技术方面具有阶段性的进展。元数据集群在完成高效分布、定位、分布式操作一致性保证的基础上，完成了无阻塞的元数据负载高效迁移、低开销的元数据服务器失效接替技术。相比于 Ceph 系统，负载迁移对应用的阻塞时延从数秒级降低到毫秒级，元数据服务器

失效接替时延从十秒级下降到秒级。下一步将搭建支持万亿文件的分布式文件系统原型，对上述技术进行整体验证。针对海量小文件高并发低延迟的访问需求，完成了授权型客户端缓存优化架构，基于目录授权的缓存一致性保护，可在客户端内自主完成文件创建、资源分配、异步批量（元）数据回刷、激进大粒度（元）数据预取等，有效降低了海量小文件访问中同步元数据访问延迟，以大粒度访问方式提升了存储效率。

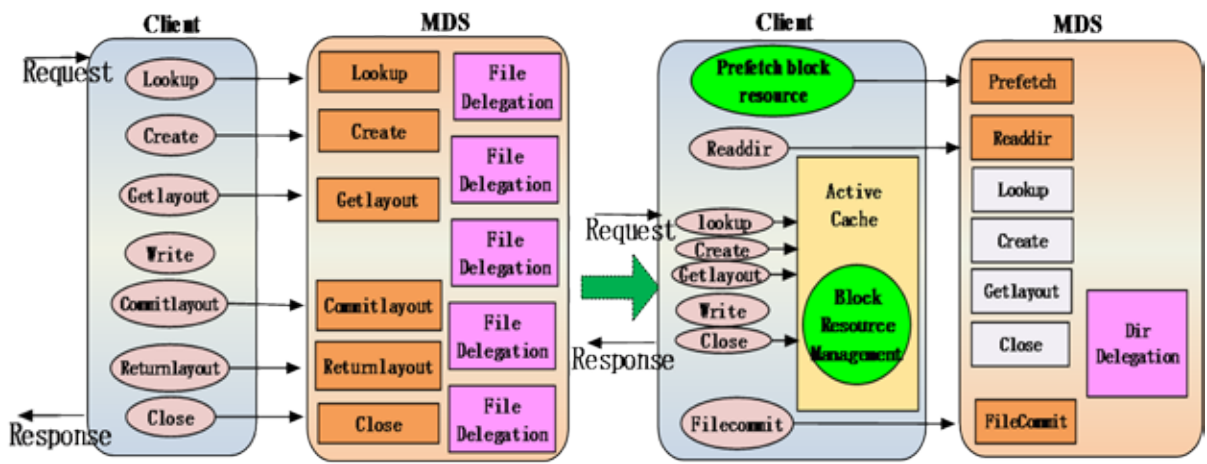


图 9 客户端预授权元数据服务示意图

获奖成果

视觉模式的局部建模及非线性特征获取理论与方法研究

图像、视频等视觉信息的有效建模是模式识别、计算机图像处理和计算机视觉的核心问题。成像过程中成像条件、成像对象的变化会导致物体表现的显著变化。视觉模式的建模就是要获取其中反映本质特性的描述,从而为解决后续识别、检索等应用中的瓶颈问题提供基础性的支撑。项目组在国家自然科学基金等的连续支持下,围绕这一基础性问题进行了长达十余年的研究,取得的主要研究成果如下:

1. 提出了人类视觉特性启发的局部模式建模方法。借鉴人类视觉的选择性响应特性和感受器的相对性特性,提出了仿初级感受野特性和反映视感知相对性的局部特征描述方法,在包括人脸识别、人体识别、纹理分析、目标检测与跟踪等方面得到了广泛的应用。

2. 提出了视觉内在非线性信息的提取与度量方法。通过对数据分布流形的学习,获取视觉信息的非线性结构,从而获得数据分布的本征维度。在此基础上,建立了流形间的对齐和距离度量方法。上述方法在图像集合、视频识别以及超分辨率等方

面获得了广泛的应用。

3. 提出了逆成像过程分析的建模方法。通过分析成像过程,对光照、姿态等影响成像结果的过程建模,进而通过逆过程分析,获得稳定的特征表示,并在光照补偿、跨姿态识别等方面得到了广泛的应用。

上述成果从仿生物视觉系统的角度,借鉴人类生物视觉系统的“局部视觉编码”特性,丰富了局部编码范式,从而推进了“视觉表示”这一基础科学问题的研究。同时从流形几何的角

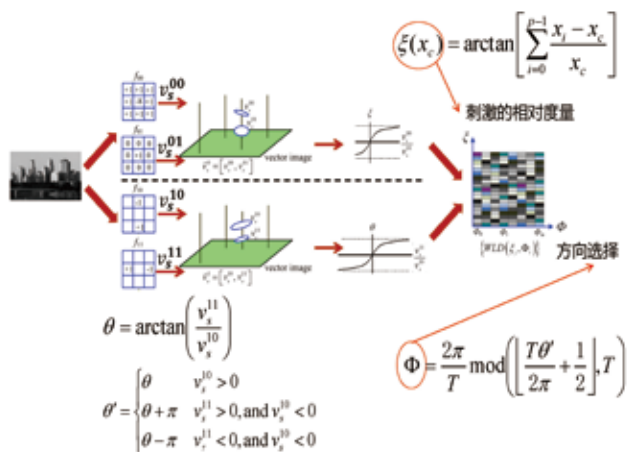


图1 韦伯局部描述子建模方法概要

度，提出了高维视觉数据内在非线性本征维度提取方法，以及相应的流形距离度量方法，逐渐引领形成了一个新的“流形表示与度量”分支研究方向。

所提出的方法被国际同行广泛用于多种识别和检索任务中，因而获得广泛的引用和良好的评价。

研究成果所发表的 20 篇主要论文中，包括 IEEE Transactions on PAMI, IEEE Transactions on Image Processing 等本领域的主流期刊和 ICCV, CVPR 等主流会议。据

Google Scholar 统计，截至 2014 年 12 月，这 20 篇论文被引 3498 次。此外，研究成果获得发明专利授权 7 项。

项目成果得到了广泛的产业应用。银晨科技将项目研究的面部比对核心技术应用于公安部出入境管理局，有效遏制了多重身份泛滥的问题“百度寻人”公益平台采用了项目组免费提供的人脸识别技术，取得了良好的社会效益。

本项目由来自中国科学院计算技术研究所、北京大学和哈尔滨工业大学的研究者共同完成，荣获 2015 年度国家自然科学二等奖。



图 2 代表性工作的部分引用与评价



图 3 项目组部分成员参加颁奖大会

获奖成果

社会化媒体数据的感知与处理

关键技术及应用

当前，蓬勃发展的网络空间社会化媒体已成为信息发布、文化传播等的重要平台，微博、社交网站等社会化媒体产生着大量的实时数据，包含着丰富的知识和线索，对大规模社会化媒体数据的实时感知、内容监测和信息研判直接



关系到网络空间秩序和社会的稳定。同时，如何提高社会化媒体数据的个性化服务质量与优化传输效率是推动我国新媒体产业可持续发展的重要方向。本项目紧紧围绕维护社会稳定和推动经济可持续发展的重大需求，以社会化媒体数据的分析和检索服务为应用目标，研究适用于大规模社会化媒体数据处理的新型计算模型、算法等关键技术。在大规模社会化媒体数据的感知与获取、存储与融合，分析与挖掘、监测与预警等方面取得了重要原创性突破。为维护我国网络空间安全和社会稳定，重大事件线索发现和有害信息的监测与预警提供核心技术和系统服务。

项目在国家自然科学基金、北京市科学基金

和多个合作项目的支持下,针对大规模网络社会化媒体数据的感知与监测分析所面临的挑战性问题开展研究,成果覆盖了社会化媒体数据的感知与获取、知识融合与检索、人物与事件分析、内容监测与预警等关键环节。系统在多通道实时

感知与主动获取、自适应知识融合与社会化检索、跨领域群体与事件的研判与分析、跨媒体融合的内容监测与分析等方面取得了重要突破。研制完成的线索发现与情报分析系统和不良内容监测与实时预警等应用系统及其关键技术在网络空间的信息内容安全监管与舆论引导、线索情报分析与挖掘和违法犯罪内容监测与预警等重大战略性任务中得到了规模化应用。主要系统架构及发明点如下图所示。

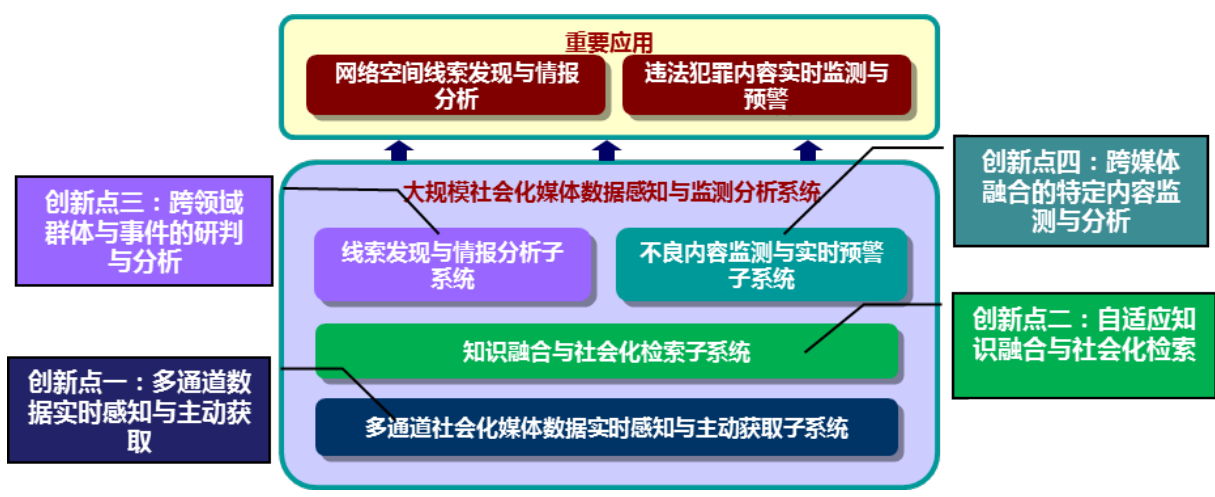
项目主要创新点有：发明了社交媒体数据的多通道实时感知与主动获取技术；发明了基于概率图模型、融合多样性与相关性排序优化的自适应知识融合与检索方法；发明了跨领域线索与

目标的研判与分析方法；发明了跨媒体融合的内容监测与分析技术。

本成果获授权发明专利 22 项，软件著作权 23 项，发表论文 172 篇，3 项核心技术在多个国际权威评测中排名第一。相关技术成果在上海瀛联、山东广电等重要企业系统中得到广泛使用，并且，在国家互联网信息办公室、公安部等 20 余

家重要业务部门，在重要特殊时期的网络信息监测和重大突发事件等重要时期安全保障中发挥了关键作用。

本项目由中国科学院计算技术研究所、公安部第三研究所共同完成，荣获 2015 年度北京市科学技术奖二等奖。



大规模社会化媒体数据感知与监测分析系统架构及主要创新点示意图

从 优 秀 到 卓 越

获奖成果

远程信息无障碍交互关键技术及应用

远程信息无障碍交互技术及应用项目主要是面向《国务院关于印发“十二五”国家战略性新兴产业发展规划的通知》中关于增强民生领域信息服务能力的战略需求，通过研发远程信息无障碍交互关键技术及系统，以实现残疾人就近就便得到信访、维权、就业、教育、康复、社会保障、情感交流等各项无障碍服务，同时改善远程务工人员与亲人之间的交流现状，促进跨地区交流和合作的效率。

本项目受到国家科技支撑计划，国家自然科学基金项目，中国科学院计算技术研究所创新项目和北京市经信委项目的支撑，主要突破的关键技术包含以下三项：

1) 沉浸式交互技术

针对视频双方由于空间障碍无法进行眼神交流的问题，给出了一种普适化的

基于虚拟视角的视线矫正方法，实现视线平视效果；针对容易产生误码的异构网络，提出健



壮性高、压缩效率高、易分级的基于视频信号分布特征的高效传输编码方法，大幅度节省带宽，削弱马赛克效应。

2) 无障碍交互技术

针对盲人与信息化设备之间存在的人机交互障碍问题

以及聋哑人手语翻译系统生成动画真实感不足的问题，研发了面向无障碍人群的交互技术，包括符合盲人日常书写习惯的盲文点字输入技术和智能盲汉转换技术以及高逼真感的多通道手语合成技术，用于帮助盲人方便地操控信息化设备，促进聋哑人与普通人的正常交流。

3) 远端协同交互技术

针对远程视频交互时用户位于不同视频空间造成的情感隔阂问题，研发了在线视频前景分割技术和沉浸式视频合成技术，为远程用户在虚拟空间提供统一呈现和协同交互。针对用户对传统键盘、鼠标交互的排斥问题，研发了自然人机交互技术，包括基于多源信息融合的用户自适应手

势识别技术以及鲁棒语音识别技术，为用户提供手势和语音等多种自然交互手段。

本项目已获得授权国家专利 21 项，受理专利 5 项，发表论文 102 篇，软件著作权 4 项。本项目开发的残疾人信息无障碍系统部署应用到北京市 16 个区县的 460 个残疾人温馨家园，惠及北京市 46 万残疾人，从很大程度上解决了残疾人出行、康复、交流以及参与社会活动的困难。另外，本项目开发的爱心小屋系统成功应用于北京和河南

等多个省市的远程务工群体和留守儿童的亲情互动，对于改善民生，构建和谐社会具有重要意义。本项目还将继续深入探索新型感知计算体系架构，构建大规模远程信息无障碍交互平台，让尖端科技贴近朴实民生。

本项目由中国科学院计算技术研究所，北京市残疾人联合会共同完成，获得 2015 年度北京市科学技术奖二等奖。

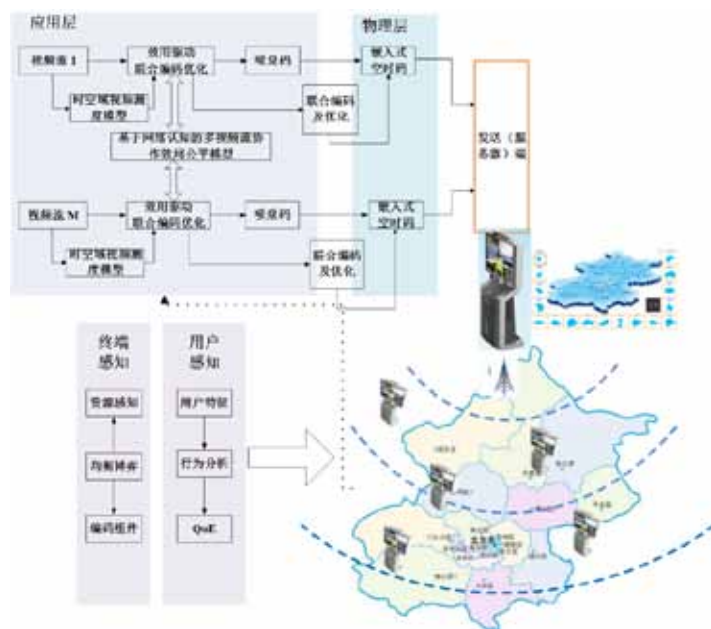


图 1 基于视频信号分布特征的高效传输编码方法



图 2 远端协同交互技术应用实例



结题验收的代表性成果

高通量计算体系的构建原理、支撑技术及云服务应用

“高通量计算系统的构建原理、支撑技术及云服务应用”是计算所牵头承担的国家 973 项目（批准号：2011CB302500），于 2015 年 10 月通过了课题验收，11 月顺利通过了科技部的项目验收。

本项目在核心科学问题的研究方面超额完成了预期目标：

共发表论文 325 篇，其中，SCI 收录 92 篇，包括 IEEE TVLSI, IEEE TACO, IEEE TCAD, COMMUN ACM, IEEE Micro, IEEE TC, IEEE TPDS, IEEE TMC 等顶级国际期刊，和 ASPLOS、

PLDI、ISCA、HPCA、FSE、DASFFA、CGO、PACT、DAC、PPoPP、ICS、SC、Hot Chips、ICCAD 等顶级国际会议；

申请并受理发明专利 265 项，并有多项专利申请并受理为国际专利，与华为开展了高通量服务器和数据中心技术的战略合作；

出版专著“Distributed and Cloud Computing”，获美国图书馆协会最佳学术著作奖；

本项目成果还获得省部级科学技术一等奖、二等奖各一项。

内容	预期目标	实际完成情况	成果体现
面向高通量计算的可扩展、高效能并行微结构	提出大规模线程并行的体系结构，构造千线程芯片的原型，展示其可扩展、高通量的特性。在千线程芯片原型的后硅调试、失效预测和检测、面向高通量应用的测试等方面提出创新方法。	已完成。 完成了高通量数据处理器 DPU 的结构设计及验证测试工作，并成功流片原型芯片。	发表论文 85 篇，其中 SCI 检索论文 14 篇，EI 检索论文 76 篇，本课题的核心成果共申请专利 41 项，其中已授权专利 6 项，国际专利 16 项。

内容	预期目标	实际完成情况	成果体现
面向 EB 级数据并发处理的新系统结构设计及评价方法	支持海量并发请求和数据并发流动的高效能、可扩展的并行系统结构，提出面向高通量计算系统应用的计算机系统执行模型，开发支持高并发、海量数据流动、低功耗、应用感知的数据通道和存储系统。	已完成。 提出了高通量处理器片上资源的分区分配和管理机制；提出有效支持片上逻辑处理器快速调整的任务调度框架，提出了逻辑核资源分配算法；提出了面向公平性的进程调度算法；提出了片上末级 Cache 的分配和管理机制。	共申请国家发明专利 116 项，其中申请 PCT 专利 28 项、进入实审美国专利 6 项；发表（或录用）了 58 篇论文，其中 SCI 期刊收录 13 篇，EI 收录 50 篇。
大规模高通量计算系统的可靠性设计	提出电路级和指令级的故障检测和容错方法，提出轻量级硬件检查点和回卷方法，提出基于高效片上互连结构的线程迁移方法；提出软硬协同的高效健康状态检查方法，提出数据备份存储方法，提出基于程序语义的软件故障调试和容错方法	已完成。 突破了传统基于“冗余”的可靠性设计的框架，在达到既定的服务质量水平下，为“高计算通量”为特征的新型大规模系统提供了低开销的可靠性保障机制。	发表学术论文总计 52 篇，包括国际顶级会议 ISCA, HPCA, DAC, 顶级期刊 IEEE Transactions on Computers, IEEE Transactions on VLSI 等；专利 22 项。
面向亿级并发负载的编程模型与支撑环境	面向亿级并发负载的编程模型与支撑环境：提出易编程的计算并行模式和数据并行模式的编程抽象及其表示方法；提出通过编程语言和运行时系统协同映射高层编程抽象到底层执行模型，充分挖掘和利用体系结构并行性以及局部性的方法。	已完成。 完成了高通量计算系统软件支撑相关的研究内容，提出了支持亿级并发负载的编程环境 Hadoop+。Hadoop+ 在国际上首次实现了在异构系统中以同构的视图编程，同时率先提出了一系列系统规模和负载数量不敏感的动态优化技术。	发表学术论文 22 篇，包括 CCF A 类国际会议和期刊 ISCA、ASPLOS、TPDS 等，申请发明专利 21 项。
高通量计算系统的云计算服务环境	高通量计算系统的云计算服务环境：实现能力服务流动机制、资源虚拟化及其提供机制和面向领域的能力服务提供机制与原理。	已完成。 构建了虚拟资源和物理资源混合管理调度的云服务模型；建立了云服务模型的基准测试环境；提出了软件透明的云平台中混合内存一致性保证机制；提出了云平台中任务顺序不变量。	发表学术论文 56 篇；授权发明专利 3 项、受理发明专利 20 项。
高效可扩展移动服务关键技术及应用	面向移动互联网，针对云计算服务资源的高效管理和调度问题、云计算环境中服务提供关键技术以及基于云的开放移动互联网业务平台的关键技术进行研究。	提出了云计算中面向服务的多维资源管理；虚拟网络资源的映射算法；云计算大规模任务调度关键技术；移动多媒体服务质量保证的关键技术等。	发表学术论文 51 篇；申请发明专利 42 项，获得授权 9 项；已经向国际标准组织提交提案 10 篇。

结题验收的代表性成果

3D 内容视觉获取技术及设备

“3D 内容视觉获取技术及设备”是由中国科学院计算技术研究所牵头承担的国家 863 计划课题（批准号：2012AA011501），起止时间为 2012 年 1 月至 2014 年 12 月，于 2015 年 4 月顺利通过科技部验收。

围绕项目任务书的研究规划，本项目就基于红外散斑的成像系统及其三维几何重建、基于多透镜阵列的成像系统及其三维几何重建、基于物理的启发式三维流体重建展开研究工作，完成了原型系统的搭建和相关技术验证。

项目的主要研究内容及具体成果：

1、基于红外散斑的成像系统及其三维几何重建

设计出一套红外散斑成像系统，该系统由散斑投影模块和散斑成像模块组成。放置在散斑投影模块中的激光器发出红外光束照射到散光板后，产生伪随机分布的散射光斑，并被投射到被测目标物上；利用散斑成像模块，将散斑图成像到 CCD 图像传感器上，采取零均值归一化互相关函数作为三维重建算法的匹配代价函数，利用 GPU 加速技术实现了 1024*768 分辨率下的实时场景三维重建，平均相对深度误差为 5%。



成像系统的内部结构和外观

2、基于多透镜阵列的成像系统及其三维几何重建

设计出一套多透镜阵列成像系统，该系统利用二次成像技术，先通过成像透镜将目标成像到微透镜阵列的前端面，再经过微透镜阵列成像，在微透镜阵列后端面形成一个像的阵列，微透镜阵列直接和图像传感器紧贴在一起，在图像传感器上形成多个视角的图像；采取在对积图像上的先验似然法，计算轮廓边缘及内部区域的深度，然后利用指数距离度量函数求得轮廓边缘的深度，为内部区域提供先验，通过先验加似然，将深度从轮廓边缘扩散到整个区域，从而恢复出整个区域的深度值，平均相对深度误差为物距的 6%。

3、基于物理的启发式三维流体重建

提出了一种基于能量约束的火焰三维重建模型，并结合热力学逆辐射过程重建火焰的三维温度场。首先，利用 CCD 相机同步采集火焰数据，实现不同角度的火焰数据获取；然后，基于拍摄的火焰图像，通过定义能量优化模型和基于光线投射的自适应温度调整算法，迭代求解，实现火焰温度场的重建，火焰场在 R、G、B 三个通道上的平均重建误差为 13.1。

项目组在任务执行期间，申请国家发明专利 11 项、国家实用新型专利 3 项，提交国家标准 1 项并报送国家标准化管理委员会立项，全面完成了项目预定目标。

跨越发展

持续发展

结题验收的代表性成果

超并行高效能计算机体系结构与设计方法研究

“超并行高效能计算机体系结构与设计方法研究”是国家自然科学基金委员会创新研究群体科学基金项目，负责人是李国杰院士。本创新群体已经顺利通过六进九考核，获得第三期资助（经费 600 万）。第二期起止时间是 2013 年 1 月至 2015 年 12 月。

本项目在第二期的主要研究工作取得了一系列成果，并得到了国际同行的肯定。截止 2015 年为止，共发表学术论文 142 篇，其中 EI 索引 90 篇、SCI 索引 51 篇，期刊论文 74 篇、会议论文 68 篇；发表专著 1 部；申请发明专利 48 项；授权发明专利 17 项。我们的工作主要发表在 IEEE TVLSI、IEEE TOC、IEEE TPDS、IEEE TCAD、IEEE TEVC、ACM TACO、ACM TOCS、JCST 等知名期刊上，以及 ISCA、PLDI、ASPLOS、MICRO、ICS、SC、CGO、PACT、DAC、DATE 等领域顶级国际会议中。2010-2014 年在体系结构四大顶级会议 ISCA、HPCA、MICRO、ASPLOS 发表文章 15 篇，位于全球第 17 名，国内第 1。群体的研究工作还分别摘得 ASPLOS' 14 以及 MICRO' 14 的最佳论文奖，前者是亚洲科研机构首次获此奖，后者则是 MICRO 创办 47 年来首次由美国之外的国家获

得此奖。

本群体完成了多项引领国际的学术成果。在项目的第一个执行期中，我们率先在国际上提出了高通量计算机的概念，并深入开展了面向高通量计算的软硬件系统关键技术研究，目前，在高通量计算的应用特征分析、评价体系，高通量计算的基准测试程序集，高通量计算芯片及内存系统、存储系统等核心使能技术方面形成了系统性的研究成果，并在 ASPLOS、ISCA、HPCA 等顶级会议上发表了一系列有影响力的研究论文。在项目的第二个执行期中，我们积极开展了智能计算体系结构的研究工作，群体中陈云霁研究员所带领的研究团队在国际上开创性地提出了深度学习处理器这一新的领域，其研究成果有效解决了深度学习计算速度和能耗的困难，并通过体系结构创新解决了如何用有限硬件资源处理近乎无限规模深度人工神经网络的科学难题。研究成果相继获得 ASPLOS 和 MICRO 最佳论文奖，此外，国际计算机学会（ACM）通讯也将研究成果中的寒武纪 1 号评选为 2014 年的计算机领域研究焦点，这是大陆研究工作首次入选。

本创新群体通过在多种技术、方法和设计上的创新，为多个国家重大课题提供了有力的技术

支持。在国家科技重大专项的核高基项目中，本群体所提出的众核处理器微体系结构、片上显示存储结构管理方案、低功耗片上网络结构设计等都得到了认可与应用，通过这些技术应用于万亿次级 CPU 的设计方案，使之后续转化为适用于未来应用的国产芯片，为我国高性能计算机研发机构提供核心运算部件和关键技术，以实现“产学研用”的良好结合。在国家 863 项目中，本群体解决了曙光高效能计算机中系统规模、计算密度、系统效能的三大难题，实现了高效能计算机从十万亿次提升到百万亿次、千万亿次跨越，曙光高性能计算机国内市场份额连续 6 年排名第一，实现了我国高性能计算机产业的历史性跨越。

通过项目的执行，创新群体自身建设也取得了长足的进步。在项目执行期内，群体中 2 人（胡伟武、陈云霁）入选了中组部万人计划、1 人（张立新）入选了中组部千人计划、1 人（李晓维）入选国家百千万人才工程、1 人（陈云霁）入选中组部青年拔尖人才、5 人（孙晓明、陈云霁、韩银和、侯锐、陈天石）获得国家优秀青年科学基金、2 人（孙凝晖、陈世敏）入选中科院百人计划、1 人（李晓维）入选“科技北京”百名领军人才培养工程、1 人（范东睿）入选北京市科技新星计划、1 人（陈云霁）入选 MIT2015 年 TR35 杰出青年创新人物、1 人（范东睿）获得中科院卓越青年科学家称号、1 人（鄢贵海）获得了中国计算机学会优秀博士论文奖。从海外引进研究员 3 人（陈世敏、李炼、王元陶）。

群体还获得多项国家和省部级奖励，其中：

“曙光高效能计算机系统关键技术及应用”获得 2013 年度国家科技进步奖二等奖；

“星载微处理器系统验证 - 测试 - 恢复技术及应用”获得 2012 年度国家技术发明奖二等奖；

“32 位星载容错控制计算机系统关键技术及应用”（与北京控制工程研究所合作完成）获得 2014 年度北京市科学技术奖一等奖，2015 年度国家科技进步奖二等奖；

“高性能众核结构设计及验证技术”获得 2013 年度北京市科学技术奖二等奖；

“跨指令集虚拟机的性能优化技术”获得 2012 年度北京市科学技术奖二等奖。

经过近 6 年的项目执行，本群体已经成为一支科研目标明确、创新能力强、团结协作的研究群体，也是国际计算机体系结构领域研究能力最强的团队之一。在项目执行过程中，本群体所依托的计算机体系结构国家重点实验室也顺利通过验收，正式运行。

国际学术合作交流是团队研究力量建设的重要手段。通过交流，了解国际学术前沿、促进群体的研究工作，同时扩大自己的影响。项目组成员胡瑜研究员应邀在 ITC2013 上做了题为“Capturing Post-Silicon Variations by Layout-Aware Path-Delay Testing”的大会报告，范东睿研究员在美国特拉华大学做了关于 Godson-T 的特邀报告。项目组还聘任瑞士洛桑理工大学的 Paolo Ienne 教授和法国 INRIA 的 Olivier 教授为中科院计算所特聘研究员，进行多核处理器片上存储和众核结构设计的关键问题进行研究，并合作完成多篇学术论文。同时，还邀请美国加州大学生芭芭拉分校的 Fred Chong 教授、哈佛大学的 David Brooks 教授、加州大学圣迭戈分校的 Dean Tullsen 教授以及周源源教授、伊利诺伊理工学院的孙贤和教授、宾州州立大学的 John Sampson 教授，瑞士 EFPL 的 Babak Falsafi 教授等国际知名学者多次来访计算所，很好地促进了我们一些研究工作的开展。



本群体有多位研究人员在学术组织担任学术职务，如李国杰院士、孙凝晖研究员、李晓维研究员、唐志敏研究员、张云泉研究员、李华伟研究员等在中国计算机学会及高性能、容错、信息存储、体系结构等专业委员会担任理事长、主任、常务理事、秘书长等职务。李晓维研究员、张立新研究员、李华伟研究员、韩银和研究员、陈云霁研究员、侯锐研究员、武成岗正高工、张磊副研究员、包云岗研究员等分别担任 HPCA、CGO、APPT、ATS、WRLT、APSys 等知名国际会议的指导委员会主席、大会主席、程序委员会主席等职务。李国杰院士、孙凝晖研究员、徐志

伟研究员、李晓维研究员等分别担任 JCST、《计算机学报》、《计算机研究与发展》等期刊的主编和副主编等职务。

综上所述，通过本创新群体项目的执行，本群体在高效能计算机的超并行体系结构和设计方法，多核 / 众核处理器的新型体系结构及低功耗设计，多核处理器的测试、验证与高可靠性设计方法，高性能并行编译技术等方面取得了一批重要科研成果，产生了重要国际影响，群体成员的学术水平与国际学术影响有明显提升，圆满完成了项目的预期目标。

跻身国际前沿

关注国计民生

引领中国信息产业

结题验收的代表性成果

理论计算机科学

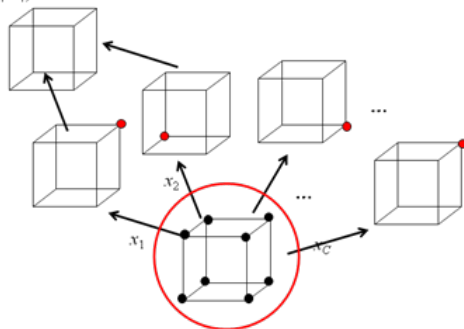
本项目是国家基金委优秀青年基金项目（批准号：61222202），起止时间是2013年1月至2015年12月。

本项目针对判定树复杂性、流式算法的空间复杂度等方面展开研究，在判定树复杂度中敏感度-块敏感度复杂性猜想、图性质函数的判定树复杂性、流式计算模型下求解线性方程组的空间复杂度、社会网络优化的近似算法等方面的研究工作中取得了进展，在ICALP, KDD, AAAI, IJCAI 和 Theoretical Computer Science 等知名会议和期刊上发表论文19篇，相关工作分别得到了美国科学院院士 Michael Jordan 等国际同行的关注。下面针对研究工作中的几个亮点成果展开具体介绍。

1. 敏感度猜想的研究

Sensitivity 复杂度最早由图灵奖得主 Cook 和 Dwork 等提出，是判定树复杂度研究中的一个重要度量。Nisan 和 Szegedy 在 92 年提出关于 Sensitivity vs. Block Sensitivity 的猜想，这一猜想断言：对于任意布尔函数 $f: \{0,1\}^n \rightarrow \{0,1\}$ ，其 sensitivity 复杂度和 block sensitivity 复杂度之间是多项式量级相关的。

Isoperimetric Inequality: $E(A, \bar{A}) \geq |A| (n - \log |A|)$



在研究工作中我们引入关于超立方体的等周不等式、关于布尔函数的傅立叶变换等数学工具，通过考察一个单值超立方体的向外两层邻居结构（如图所示），证明了如下定理：

定理：对于任何的布尔函数 f ，其块敏感度和敏感度复杂度满足 $bs(f) \leq C(f) \leq s(f) 2s(f) - 1$ 。

之前关于用敏感度复杂度来表达的块敏感度复杂度的最好上界是 Kenyon 和 Kutin 在 04 年证明的 $bs(f) \leq O(\sqrt{n} e^{s(f)})$ ，我们的结果将上界的指数从 $e^{s(f)}$ 改进到了 $2^{s(f)}$ 。这一工作发表在 ICALP 会议上，得到了国际同行的积极评价。

2. 图性质的判定树复杂度研究

图性质函数是判定树理论中一个重要的研究对象，Aanderaa – Karp – Rosenberg 猜想断言：任何单调的图性质都是诡的，即其判定树复杂度为 $\binom{n}{2}$ 。Rivest 和 Vuillemin 证明了其判定树复杂度至少是 $\Omega(n^2)$ 的。我们在研究中首次将这一结论推广到 3 均匀超图性质，证明了如下弱“诡函数猜想”：

定理：对于任何有 n 个顶点的单调非平凡 3-均匀超图性质函数，其确定性的判定树复杂度都是 $\Omega(n^3)$ 的。

在证明技术上我们将 Rivest-Vuillemin 所提出的基于组合数论的证明技术与 Kahn-Saks-Sturtevant 所提出的代数拓扑技术相结合，证明中还使用了 Vinogradov 所证明的关于奇数的哥德巴赫猜想，即任意充分大的奇数都可以表示成 3 个素数之和。此工作已经发表在 TAMC

2013 会议上。

3. 求解线性方程组问题的通信复杂度

通信复杂度最早由姚期智教授提出，是理论计算机科学中最重要的证明下界的工具之一，也是研究流式计算空间复杂度的重要方法。我们在研究工作中讨论了在被动通信模型下求解线性方程组等线性代数问题的通信复杂度，提出了一般性的将多人通信问题归约到两方通信问题的方法，证明了 s 人的多方随机通信复杂度至少是 s 倍的两方通信复杂度。证明的关键是利用了线性代数问题的对称性，构造了对称的困难实例的分布；另外我们还给出了多个新的两方通信复杂度下界，结合这两方面，我们证明了求解线性方程组等问题紧的通信复杂度下界。我们的下界还适用于判定矩阵的秩是 a 还是 b 等问题。这一工作已经发表在 DISC 2014 会议上。

自主创新

科技强国

结题验收的代表性成果

计算机体系结构

“计算机体系结构”是国家基金委优秀青年基金项目(批准号: 61222204), 起止时间是2013年1月至2015年12月, 于2015年底顺利结束。过去三年中, 项目组按照预期的研究计划, 通过软件(图计算分布式计算框架)和硬件(图计算领域专用高性能处理器)的协同设计, 挖掘图计算中数据的请求局部性, 从而达到了提升图计算的目标。在实施的过程中, 围绕图计算专用高性能处理器体系结构设计这一方向, 从四个紧密相关的主题开展研究工作, 分别取得了如下研究进展:

1. **在图计算处理器的请求高速缓存:** 在处理器中加入一个新的独立于数据高速缓存、指令高速缓存的请求高速缓存。请求高速缓存能紧致地存储大量请求, 并根据请求之间数据是否可以重用来对请求进行调度, 从而降低访存开销。

2. **可塑图计算加速器核:** 引入访存延迟不敏感的弹性电路以在可重构的灵活性和性能功耗面积等因素之间进行权衡, 从而设计出了适合不同应用的可塑图计算加速器核。

3. **图处理器编程范式:** 倡导程序员明确分工、相互协作的编程范式。在此范式下, 程序员被划分为不同的类型。普通程序员注重如何快速构建图计算应用, 而专家程序员则专心实现图计算的

高性能实现。两者职责更加明确, 约定更加规范, 相应的编译框架和运行时也将在此约定下为程序提供自动的正确性、性能和可移植性的保证。

4. **请求局部化的分布式图计算框架:** 面向请求局部性的分布式图计算框架可以动态调整子图的分割方式, 尽可能把拥有请求局部性的数据聚拢在一个子图中, 充分保障数据的有效重用。

基于上述工作, 本课题共发表论文14篇, 其中, SCI收录11篇, 包括IEEE Transactions on Computers、IEEE Micro、ACM Transactions on Architecture and Code、IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems、ACM Transactions on Design Automation of Electronic、IEEE Transactions on Circuits and Systems、IEEE Transactions on Evolutionary Computation、IEEE Transactions on Parallel and Distributed、等国际期刊和ISCA、The International Symposium on Code Generation and Optimization、Design Automation & Test in Europe等国际会议。申请并受理发明专利15项。

项目培养了相关方向博士生5名(全部毕业), 硕士生4人(全部毕业)。

结题验收的代表性成果

视觉模式分析与识别

本项目为国家基金委优秀青年基金项目（批准号：61222211），起止时间是 2013 年 1 月至 2015 年 12 月，于 2015 年底顺利结束。过去三年中，项目组按照预期的研究计划，围绕视觉模式分析与识别这一方向，从四个紧密相关的主题开展研究工作，分别取得了如下研究进展：

1. 在深度层级视觉模型方面，重点研究了小数据条件下深度层级模型的学习问题，提出了由粗到细的多阶段方法、渐进自编码方法等，并应用于特征定位、表情识别、超分辨率和人脸识别等任务；

2. 在视频建模方面，系统深入的研究了视频的流形建模、对齐、匹配和测度学习等问题，提出了通用流形模型、无监督通用流形对齐、异质空间流形距离测度学习、黎曼流形降维和测度学习等新方法；

3. 在跨域和多视模式识别方面，以寻找公共判别子空间为思想，提出了多视判别分析、自适应判别分析、源域目标化判别分析以及双向迁移自编码方法等新的算法或技术方案；

4. 在非可控条件人脸识别方面，提出了多

局部区域描述子协同测度学习方法、结合 3D 几何和学习的可变形场模型、耦合对齐和识别的视频人脸识别方法等，还建设并发布了一个 2000 人规模的视频人脸数据库。

基于上述工作，课题组在领域重要国际期刊上发表论文 12 篇（其中 CCF A 类国际刊物论文 5 篇），重要国际会议论文 29 篇（其中 CCF A 类国际会议论文 17 篇）。所发表的论文中，单篇最高被 Google Scholar 引用 93 次。已申请发明专利 5 项。

此外，项目负责人获得了 2015 年度国家自然科学基金（第 2 完成人）。带领团队获得了 ACM EmotionW2014 的冠军，获得 ICCV2015 ChaLearn: Looking at People 两个竞赛的冠军（Culture Event Recognition 竞赛）和亚军（Apparent Age Estimation 竞赛）。

项目培养了相关方向博士 10 名（毕业 5 名，在读 5 人），硕士 12 人（毕业 6 人，在读 6 人）。其中，1 人次获计算机学会优博，1 人次获中科院优博。

结题验收的代表性成果

大规模图像的内容分析与理解 及其智能服务技术

本项目是中科院计算所-联想联合实验室青年科学家计划项目，执行周期为2012年12月至2015年11月。本项目面向海量多媒体网络环境下的图像数据庞杂、数据内在关联多样的挑战，以建立基于视觉内容的图像内容分析与匹配方法、基于语义关联的内在层次增强方法、和网络多模态语义增强方法为科学目标，重点研究视觉特征提取与学习、多模态数据聚类、图像标注优化、以及高效的图像索引与搜索技术等理论与方法。在此基础上建立海量图像智能搜索与标注的原型系统。项目组在大规模图像检索、基于类别和实例的物体识别、基于RGB-D的图像识别、场景分类等方面开展了多项研究，取得了一系列研究成果。

在图像检索方面，提出了一种旋转不变性的局部相似图像检索算法。该方法主要包含两个部分：1) 方向和位置关系的不变性；2) 图模型的一致性。所提方法解决了传统局部匹配算法存在的非匹配区域噪声过多，匹配区域特征过少以及匹配区域旋转、变换等方面的问题，通过在多个大规模图像数据集上的测试均表现

出了很好的检索性能。以该方法为基础的论文发表在IEEE Transactions on Multimedia2013和IEEE Multimedia2013等国际期刊上。在此基础上，研究提出了实例级物体检索与识别数据集，数据集主页：<http://vipl.ict.ac.cn/isia/instre/home.html>，数据库包含2万张超过200类的实例图像，并对图像中实例物体的位置和物体类别进行标注，此工作发表在期刊ACM ToMM2015上。项目组还提出了基于相似检索的多目标物体识别系统：ObjectSense（系统截图如下），该系统获得国际会议ACM ICMR2013 Best Demo Award。

在图像分类方面，提出了一种利用马尔可夫随机场对局部空间和多视觉特征上下文关系建模的方法，以增强图像语义描述，并将该语义描述用于场景图像分类，所提方法增强了语义流形中图像的语义描述，并在大规模的场景公开数据集上得到了较好的分类正确率，此工作发表在CVPR2015、Pattern Recognition2016等期刊和会议上。在传统的基于图像视觉内容的基础上引入了图像的上下文信息来进行菜品图像的识

别，提出了基于地理约束的视觉分类模型，以该方法为基础的论文发表在 IEEE Transactions on Multimedia 2015 上。此外，项目组还研究了基于 RGB-D 的场景分类和手持物体识别技术，有效地融合多种特征，达到了比较好的识别性能。

以上研究的成果分别应用在 ISIA 大规模视觉检索系统（界面如下）和机器博士系统中。在项目执行期间，共在 IEEE Transactions on Multimedia、CVPR 等国际期刊和会议上发表

学术论文 30 余篇。在本项目支持下，项目负责人获得国家自然科学基金委优青项目支持、获 2013 年度中国科学院青年科学家国际合作奖、入选 2014 年度中组部万人计划青年拔尖人才计划；项目负责人所在团队还获得国际会议 ACM ICMR2013 Best Demo Award、ImageCLEF2013 Robot Vision 竞赛的冠军、以及国际会议 Gamenets2014 Best Paper Award 等学术奖励。

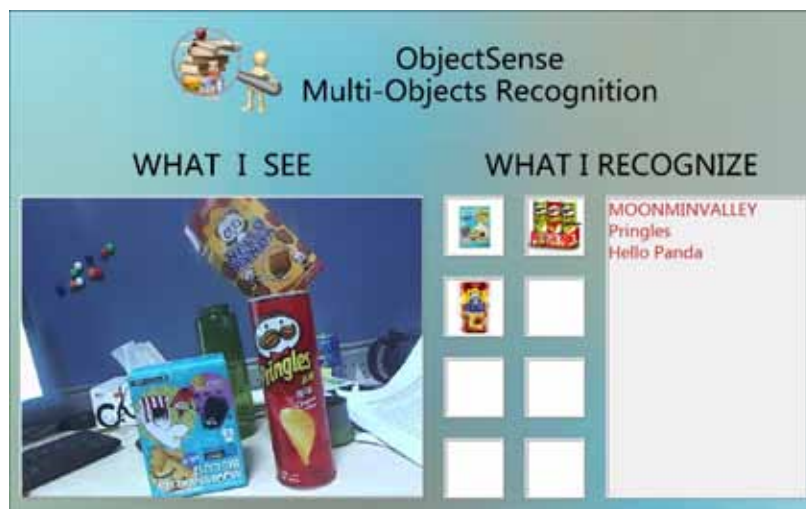


图 1 ObjectSense 系统界面



图 2 ISIA 图像搜索系统界面

结题验收的代表性成果

多核编程框架研究项目

多核和众核已经成为处理器的主要结构，越来越复杂的硬件并行层次，给软件带来了巨大的编程、优化挑战。针对上述问题计算所与华为公司中央软件研究院在 2014 年 8 月签署了多核编程模型项目，目标是以华为公司的典型应用场景出发，为其提供新型的多核编程框架和技术方案，以降低并行编程和优化的难度，在多核编程领域开展创新性研究并取得技术突破。此项目是计算所与华为公司的第二个 A 类合作项目，项目总金额为 937 万。2015 年 12 月本项目验收通过，圆满完成项目合同要求，共输出 10 项技术专利，其中两项被华为公司评价为高价值专利，本项目成为首个直接落地到华为产品线的 A 类合作项目，赢得了华为公司高度评价，同时获得了华为电信软件事业部的关注并与其洽谈合作事宜。本项目共分为三个课题，课题一是面向华为的“管”类应用的分布式并行编程模型，课题二是面向异构计算环境的自适应编程模型，课题三是软硬件协同设计。

分布式计算是近些年最重要的软件技术之一，是互联网应用、大数据、云计算应用的基

础，其高速发展受益于分布式编程模型和框架的发展和运用，业界的分布式编程框架能够极大的改善分布式开发的生产效率，对系统的容错性、可移植和易于部署、提供完备高效的异步通信机制等方面提供了良好的支持。然而当前的分布式编程模型并非尽善尽美，并且不能直接适配华为的软件产品。为此，我们针对华为应用场景的特点，在吸纳业界先进的分布式编程模型的优点的基础上，提出了一种新型的基于消息的编程框架 MOLA (Message Oriented Language Architecture)。MOLA 是基于 Actor 模型的微服务分布式编程框架，利用编译和运行时系统来屏蔽分布式并发和并行编程的细节。MOLA 根据华为程序员的传统的基于消息流式的编程习惯设计了隐式 actor 模型，相比业界方案具备最小的学习成本，通过模块化抽象，让用户只关心业务逻辑，并屏蔽 actor 创建、销毁、迁移、任务分发、负载均衡调度、锁及同步的分布式并行编程细节，使得普通程序员可以轻松的开发分布式应用，从而提升产能。同时 MOLA 支持 actor 的自动多实例并行，自动异步同步任务调度，部署位置自

动优化，SG 高级并行模式等编译框架，以及高性能的 IDL 消息处理系统等先进技术和特性。使得 MOLA 获得了超越业界标杆产能 2 倍以上，性能好于 20% 以上的突出成绩（MOLA 的产能比微软的 Orleans 系统好 2.1 倍，性能好 6 倍；比 CAF 产能好 2.5 倍，性能好 3 倍）！目前分布式并行编程模型 MOLA 及其 IDL 系统 Pigeon 已经直接落地到华为的存储产品线中。

由包含传统处理器（X86 和 ARM）和加速部件（GPGPU 等）的异构计算结构，已广泛应用于移动智能终端（智能手机、Pad）、桌面计算机、服务器和数据中心。异构平台上的加速部件虽然能够满足华为对产品性能的要求，但是在生产效率、性能优化和移植方面提出了巨大的挑战。为了解决这些问题，业界提出了很多基于领域编程模型的解决方案。尽管这些方案能够提高编程效率、降低开发难度，但不能满足华为在图像处理和无线基站等生产场景的需求。为此，我们针对华为应用场景的特点和华为公司程序员的工作习惯，设计了一种基于 C 语言的领域编程环境 ParaC。其基于对图像处理领域算法特点的总结和抽象，提供了简洁的语言扩展和领域相关的数据类型以降低编程复杂度，并将算法描述和性能优化分开，用户只需要负责描述算法，性能优化则由编译器利用专家优化经验自动完成。ParaC

已经完成了本期项目要求的设计开发，在华为指定的场景下，开发效率是 OpenCL 的 5 倍以上，性能处于 OpenCL 手工优化版本程序的 90% 到 500% 之间。华为方对子课题二自适应编程模型的成果非常满意，并已达成了与华为公司的第二期合作意向。

软硬件协同设计是未来重要的技术，其是华为 ICT 战略 SoftCOM 软件要突破的关键山顶技术，也是华为终端和 IT 部门指导硬件设计提高性能功耗比的重要业务需求。为此，计算所在面向系统级软件定义硬件，从全系统角度探索软硬件协同设计的方法和工具等方面开展了探索性研究，以实现软硬件垂直整合的最佳目标。经过本期合作项目，目前已经完成了基于高层次计算模式和特征提取的软硬件协同设计框架，以及针对华为应用场景抽象了 map-reduce 计算模式，并设计了软硬件协同的编程接口和加速部件规范。并提交了包括《系统级软硬件系统综合分析方法研究报告》、《硬件演进与定制路线选择建议》、《通用加速器调研报告与定制建议》、《计算密集型 benchmark 通用加速分析方法》、《通用加速器核 SPEC》、《通用加速器和功耗与面积评估方法》、《通用加速器配套工具链及效果验证方法说明书》和《计算密集型 benchmark 通用加速评测报告》等详细的技术分析和方案设计报告。

结题验收的代表性成果

面向残疾人信息化的远程维权终端系统

该项目是计算所与北京市残疾人信息联合会合作建立残疾人的远程信息化服务系统的一个重要项目，课题执行时间为 2014 年 7 月至 2015 年 12 月。

我国有 8502 万残疾人。他们大部分生活在交通不便的农村，由于身体状况、经济条件和受教育程度的限制，残疾人群在维护自身合法权益时，普遍面临着比普通人更多的困难，亟需更实用的面向残疾人的远程信息化服务系统，以期让残疾人不出社区、村，就可得到信访维权等各项无障碍服务。



图 1 远程维权示例

立足现实，针对需求，普适计算研究中心研发了残疾人远程维权系统，可以方便快捷地为残疾人等弱势群体提供法律咨询，心理疏导，信访咨询，维权交流以及领导视频接访等服务工作（如图 1 所示），从而保障残疾人等弱势群体特别是行动不便的弱势群体的合法权益。具体地，系统可为残疾人以及残疾人服务机构量身定制“交谈”、“文本”、“广播”和“管理”四大功能。其中“交谈”功能包括一对一的“面对面”交谈功能；一对二的“三点协作”交谈功能，实现市、区、街道\村三级机构协作交流；一对三的“广播答疑”交谈功能。“文本”功能包括纸质文本远程提交功能和电子文本在线讲解功能，方便残疾人递交信访材料，同时帮助残疾人服务机构向残疾人提供方便快捷的信访、法律咨询服务。“广播”功能主要是实现一对多视频广播通知功能，可以随时召集和举行会议，缩短残疾人管理服务机构与基层服务机构之间的距离，节省大量的差旅费用，降低成本，同时减少在旅途中的时间，提高工作效率。“管理”功能包括监控管理功能、可视化

分层管理功能、排队管理功能、服务器管理功能、日志和统计功能以及历史音视频数据存储和检索功能，帮助残疾人管理机构对残疾人服务机构的工作情况进行监督、检查，统计和查询，使残疾人服务工作制度化、规范化，确保残疾人享受到优质优惠服务。

目前，该系统已在北京市 16 个郊区县区的残

残疾人温馨家园进行了推广应用（如图 2 所示），惠及全北京市 46 万残疾人。通过为残疾人提供不出社区、村情况下的无障碍信访维权服务，同时为残疾人服务机构实现高效的视频广播和监控管理功能，有效改善了残疾人维权上访难的实际社会问题，积极推动了残疾人信息化维权事业的发展。



图 2 残疾人远程无障碍交流系统终端部署示意图

结题验收的代表性成果

基础研究项目群

2012 年经计算所所务会讨论决定，部署所创新课题基础研究项目群，包含 5 个课题，课题名称分别是“高性能功耗比的通用处理器结构”、“分布式计算中的几个问题研究”、“新一代蛋白质鉴定引擎 pFind 3.0 和蛋白质交联鉴定引擎 pLink 2.0 研究”、“蛋白质结构模体识别及结构预测算法研究”、“安全博弈论关键理论研究”。2015 年 8 月科研处组织相关专家对这些课题进行了验收。各课题取得的成果如下：

高性能功耗比的通用处理器结构：本课题的目标是对一类新的通用微处理器体系结构及其设计方法开展预研。取得的主要研究成果包括：1. 支持长指令字执行的超标量体系结构。本课题提出并设计了采用超标量和长指令字合一的可配置的数据通路的处理器流水线，该处理器在超标量结构的基础上，增加了长指令字执行的功能。2. 基于最大时间差的指令流水线设计。本课题提出通过简单的控制结构加上深度流水，来提高主频进而提高性能。3. 支持数据并行的低功耗多线程处理器。本课题提出了数据并行与线程并行合一的

新型处理器结构，通过深度流水提高工作频率、通过多线程和长向量容忍延迟、通过访存优化提高片上存储器效率、通过静态结构和资源复用降低功耗和成本。课题执行期间，共发表高质量论文 13 篇，申请专利 9 项，培养博士生 7 名。在本课题研究的基础上，已成功申请到一项由计算所独立承担的国家自然科学基金重点项目。

分布式计算中的几个问题研究：围绕着敏感度复杂性下界、分布式流模型下的通信复杂度、社会网络中影响力最大化这三个主要科学问题展开了深入的研究，达到预期目标。取得的主要研究成果包括：10 年来首次改进了敏感度复杂性猜想的上下界；推广 Karp-Yao 猜想到超图性质，证明了单调超图性质的最优判定树复杂度下界；提出了带空间限制的新通信复杂性模型，发展了借助傅立叶变换的对偶下界证明技术；将渗流理论引入到多方安全计算的研究中，给出了多项式时间的多方安全通信协议；提出了带概率保证的社会网络影响力传播最大化问题和新算法，解决了 Tenenholtz 等提出的关于并行秘书问题在线

算法的未解问题。代表性成果发表在理论计算机科学领域重要的国际会议和期刊上，其中会议论文 18 篇，SCI 期刊论文 4 篇。依托本课题所做的大量前期工作，成功申请到国家自然科学基金重点项目一项。

新一代蛋白质鉴定引擎 pFind 3.0 和蛋白质交联鉴定引擎 pLink 2.0 研究：本课题完成了新一代常规蛋白质鉴定搜索引擎 pFind 3.0 和交联蛋白质鉴定搜索引擎 pLink 2.0，灵敏度、准确度、速度均达到本领域当前最佳水平。pLink 相关合作成果于 2012 年和 2015 年连续发表在 Nature Methods（五年影响因子 27.2），前者也是中国蛋白质组学团队在该期刊上的首次突破。pFind 和 pLink 软件累计注册下载用户分别达 792 和 534，遍布世界六大洲的 34 个国家和地区，在国内外同行的 109 项研究成果中得以应用，其中多项成果发表于 Nature、Science 等著名期刊。此外，课题组倡议并主办了第二届和第三届中国计算蛋白质组学研讨会，促进了国内外计算蛋白质组学领域的交流与合作。

蛋白质结构模体识别及结构预测算法研究：本课题提出了蛋白质折叠过程的“精英绑架”观点，

即总体看残基的结构信号比较弱，但是存在少量残基具有较强的结构信号；弱信号区域是允许折叠发生的必要条件。在此观点指引下，发现了螺旋区域的 A1-A4/A5 强信号模体。其次，提出了优化能量函数设计的“扩大吸引盆”方法，通过逆向采样（reverse sampling）技术，并迭代允许线性规划不断优化能量项的权重；最后，提出了测量势函数表面“崎岖”程度的度量，采用逆向采样，能够准确地测量出势函数表面崎岖程度。

安全博弈论关键理论研究：过去几年，基于斯塔克伯格博弈的安全博弈论在安全领域的资源分配及调度方面的理论逐渐建立并且在现实环境中得到成功应用，但现有安全博弈理论存在若干重要局限。本课题研发了新一代安全博弈论应用的关键技术和算法（如求解大规模博弈、动态收益博弈、带资源外部性博弈、有限次观测博弈），同时探索了安全博弈在更多重大社会问题中的应用。共在人工智能顶级会议 AAAI, IJCAI, 和 AAMAS 上发表论文 8 篇，其中计算所为第一单位的文章 6 篇。项目成果的创新性和重要性得到了国内外领域专家的认可，一些关键技术有望在某些领域应用。

从 优 秀 到 卓 越

结题验收的代表性成果

龙芯 GS464E 处理器核的验证研究

“龙芯 GS464E 处理器核的验证研究”是计算所 2013 年部署的所创新课题，起止时间是 2013 年 9 月至 2015 年 8 月。2015 年 8 月通过了计算所科研处组织的验收。

本课题提出并建立了龙芯 GS464E 处理器核的功能验证、性能验证、功耗验证的流程及环境，完成了 GS464E 处理器核的验证工作，确保了龙芯新一代 GS464E 处理器核架构的研制成功。主要包括：

在功能验证方面，建立完整的 GS464E 功能验证流程和验证环境。龙芯功能验证流程，首先是仿真验证，在覆盖率的指导下，完成处理器核模块及处理器核的随机指令验证；对一些关键模块进行形式化的验证，如浮点乘加模块；在硬件加速平台上和 FPGA 上进行系统级验证，能够运行较大型的程序进行验证；基于样机的硅后验证环境可以运行大量程序进行验证。并建立了与龙芯功能验证流程配套的功能验证环境，包括：RTL 级仿真验证环境、形式化验证环境、指令级随机验证环境、基于 Zebu-

Server 仿真加速器的硬件仿真加速验证环境、FPGA 系统验证环境、基于样机的硅后验证环境等。龙芯验证流程全面，整个验证流程涵盖了从仿真验证到形式化验证、从软件仿真验证到硬件加速 /FPGA 验证、从模块级验证到系统级验证的全方位验证。

在性能验证方面，建立了完整的 GS464E 性能验证流程和验证环境。龙芯性能验证流程包含三个主要实施阶段，依照开展的时间先后顺序，依次为：RTL 级仿真性能评估阶段、系统级性能评估阶段和硅后性能评估阶段。其中前两阶段均在芯片流片前开展，可根据评估结果对 RTL 设计予以及时调整。硅后性能验证阶段在实际样机上运行大量的测试程序和实际应用，并通过与 Intel、AMD 处理器运行结果进行横向对比，可以定位到 GS464E 处理器核结构设计不足，针对未来架构改进提出具体措施。另外，建立了相应性能验证环境建设，主要包括：RTL 仿真性能验证平台、仿真加速性能验证平台、FPGA 性能验证平台、基于样机的硅后性能验



证平台等。龙芯性能验证流程在 GS464E 的研制中发挥了重要作用。

在功耗验证方面，建立了 GS464E 的功耗验证流程和环境。GS464E 处理器核的功耗验证在设计初期采用了基于 RTL 代码的功耗分析流程，后期采用基于后仿真的功耗分析流程。

龙芯 GS464E 处理器核是在原有 GS464 架构基础上改进的，支持 64 位 LoongsonISA 指

令集，采用 4 发射乱序执行结构，集成片上三级 Cache、支持多层次硬件预取等。采用上述验证技术后，采用 SMIC 40 纳米工艺设计的集成 4 个 GS464E 处理器核的龙芯 3A2000 一次性流片成功，该芯片在主频 800MHz 时实测 SPEC CPU2006 定点分数为 5.89 分，在同主频下性能达到国际上 AMD 主流处理器相当的水平。

正 气

大 气

骨 气

科研基地进展

计算机体系结构国家重点实验室

获奖及科研项目方面：国重集成电路团队的“32位星载容错控制计算机系统关键技术及应用”获得了2015年度国家科技进步二等奖（第二完成单位），国重操作系统团队的“网络支付系统风险防控关键技术及其应用”获得了2015年度上海市科技进步一等奖（第三完成单位）。

2015年国家自然科学基金资助项目公布，实验室获批资助14项，其中重点项目2项，重点国际（地区）合作与交流项目1项，优秀青年科学基金项目3项，面上项目2项，青年科学基金项目6项。

973项目“高通量计算系统的构建原理、支撑技术及云服务应用”，在高通量计算系统的微结构设计、系统结构设计、可靠设计、云计算服务与应用等方面进行了探索研究，完成了预期目标，已于2015年11月顺利通过了科技部的验收。创新研究群体科学基金“超并行高效能计算机体系结构与设计方法研究”成功获得6年晋级9年的延续支持，进入第三个执行期（2016年到2018年）。

实验室2015年共承担各类研究课题98项，

其中973计划2项（课题4项）；863计划2项（子课题4项）；国家自然科学基金项目44项（其中创新群体1项，优秀青年基金1项，面上基金11项，重点基金5项，青年基金19项）；科学院项目14项；横向项目26项。纵向经费总计11203.9万元，实到经费3641.15万元；横向经费总计5841.77万元，实到经费2201.62万元。本年度新增课题30项，申请获批基金15项。

学术进展方面：首个深度神经网络处理器——寒武纪，在2015年就形成了十余篇CCF A类顶级国际会议/期刊论文，突破本领域最高期刊TOCS（大陆首次），入选了《Communication of the ACM》评出的研究焦点（大陆首次），65nm验证片比Intel Xeon-E5能效比提升100倍，并受到了Intel、斯坦福大学、UCLA、苏黎世理工等欧美高水平机构的广泛引用，开创的研究方向吸引了领域顶会ISCA 5%的投稿。

实验室持续发表了ISCA-2015、MICRO-2015、ASPLOS-2015、ICSE-2015、VLDB-2015 DAC-2015等顶级会议论文9篇，在TOCS、TC、TPDS、

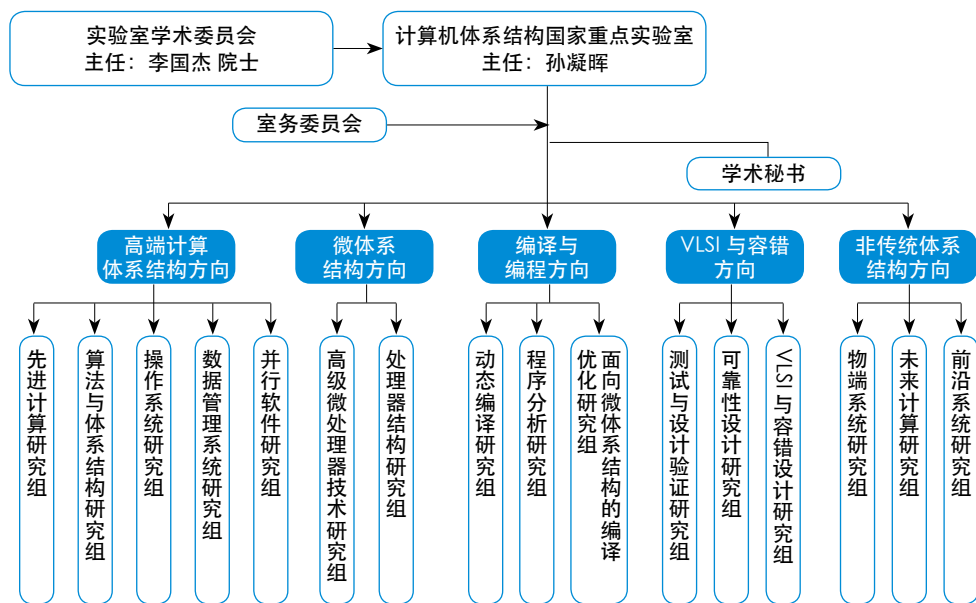


图1 计算机体系结构国家重点实验室组织结构图

TKDE、TODS、TCAD、TVLSI等顶级期刊上发表了18篇论文。2015年度在本领域四大顶会 (ISCA、HPCA、MICRO、ASPLOS) 发文量排名全球第11位，近五年比前五年前进了15名 (2011-2015年度共发表14篇，全球排名第17名；2006-2010年度共发表5篇，全球排名第32名)。

本年度实验室研究人员发表论文100篇，其中EI收录56篇，SCI收录27篇。授权发明专利14项，申请专利34项。公众开放2次，大中小学及社会人士参观访问共计1000余人。

国内外学术交流方面：本年度内，国内外学者来实验室做学术交流报告共14次，参与国际会议38人次 (其中参加CCF A类会议10人次)，在国际学术会议上做报告20人次，合作研究类出访6人次。

2015年度实验室通过组织专家进行评审，择优资助十项开放课题的申请。国家重点实验室自成立以来 (2011年至今)，合计择优资助了

40家单位、50项开放课题。受资助的单位中，主体是高校，占比93%，其余是院所和企业。受资助的高校中985院校占40%，211院校占到三分之二。

人才引进和培养方面：2015年，在人才引进方面，李炼获得中科院百人计划择优资助，引进了王元陶研究员，郭崎副研等海外人才。在人才培养方面，韩银和、陈天石和侯锐三人入选优青，陈云霄入选2015年度MIT TR35杰出青年创新人物。包云岗、侯锐副研究员晋升为研究员，武成岗副研究员晋升为正研级高工。房双德荣获院长优秀奖，路航荣获中科院朱李月华博士奖。

科研人员在国际舞台日益活跃，陈云霄担任ASPLOS 2017联合主席，在体系结构顶会ASPLOS'15、ISCA'16、MICRO'16担任PC，陈世敏在数据管理主要会议SIGMOD'15、SIGMOD'16、ICDE'16、VLDB'16、ICDCS'16、PVLDB'17担任PC，李华伟担任国际学术期刊

IEEE TVLSI 编委。

国重多位老师在中国计算机学会 (CCF) 中担任要职, 五位老师当选为 CCF 第十一届理事会理事, 分别是主任孙凝晖研究员、常务副主任李晓维研究员、张云泉研究员、李华伟研究员、韩银和研究员, 其中孙凝晖研究员当选为第十一届 CCF 副理事长, 张云泉研究员当选为 CCF 第十一届理事会常务理事。

自研课题和研究平台建设方面: 2015 年 6 月, 国重组织了 2013-2014 年度自主研究课题结题验收会。18 项自主研究课题 (包括主任基金 2 项、重点研究课题 7 项、自由探索课题 7 项和人才引进课题 2 项, 执行期两年, 批准金额共计 465 万元) 均进展良好, 并取得了一系列优秀的成果, 18 个

课题全部通过验收。

此外, 四个特色自研平台进展良好。由鄢贵海副研究员负责的“高效能可重构大数据处理加速引擎研究平台”、由陈云霁研究员负责的“机器学习加速器的结构和编程研究平台”、由陈世敏研究员负责的“面向测量信息的日志记录分析平台”、由张云泉研究员负责的“众核处理器并行编程方法研究平台”, 均按照计划进展良好。

国重建设方面: 实验室主任: 孙凝晖研究员 (兼); 副主任: 李晓维研究员 (常务)、冯晓兵研究员; 学术委员会主任: 李国杰院士。实验室包括十六个研究组 (如图 1), 部署在五大研究方向——高端计算体系结构、微体系结构、编译和编程、VLSI 与容错计算、非传统体系结构。



图 2 国重学术委员会第五次会议合影



科研基地进展

移动计算与新型终端北京市重点实验室

移动计算与新型终端北京市重点实验室依托于中国科学院计算技术研究所组建，2011年5月由北京市科委正式认定。2015年，实验室承担省部级以上课题及横向课题共计53项，在IEEE Communications Magazine、IEEE Transactions on Multimedia等期刊会议上发表论文47篇，新申请发明专利32项，新授权15项，获得软件著作权6项。本年度实验室获得国家科技进步二等奖、北京市科技发明二等奖、中国科学院北京分院科技成果转化特等奖等奖项。此外，实验室通过了北京市科委组织的北京市重点实验室三年绩效考评，考评成绩优秀（绩效考评获得优秀比率为13.9%）。本年度实验室继续以移动计算与新型终端发展方向的引领性关键基础理论与技术实现方法为核心开展研究，在新型终端、智能应用和移动计算三个方向都凝结出丰硕的成果。

1、新型终端方向

1.1 LTE 小型基站基带芯片

完成了国内首颗自主4G小型基站基带芯片

的研制，该芯片拥有完全自主知识产权，并按照中国移动相关测试规范要求进行了内场测试，并将在工信部电信研究院组织的内外场测试基础上开展公网以及行业专网的试用推广。基于DX001-0.5芯片研发成果，开展面向LTE-Hi小型化基站的SoC芯片研发。该芯片符合3G、LTE和LTE-Advanced等通信规范，支持室内密集组网。芯片架构上采用8个自主矢量DSP核和2个高性能CPU核，可完成所有的基站数字域处理，大幅降低了小型基站的体积和成本，性能上将超过博通、Intel、高通等的同类型芯片。

1.2 新型终端开发平台

成功研制出低成本、沉浸式、易操作、高保真的基于人机物三元融合端计算技术的远程信息无障碍交流平台。平台以听力残疾人的基本需求为导向，从听力残疾人参与社会活动的实际需求，以及残疾人管理机构的需求出发研制的远程沉浸式交互定制系统。平台在沉浸式交互、无障碍交互和机—物协同等方面进行了系列理论创新和技术突破。2015年，该平台荣获北京市科学技术奖。

1.3 盲人人机交互技术

基于新型智能汉盲转换算法，实验室人机交互研究组实现了一个汉盲转换云服务平台，基于 HTTP 接口对外提供汉盲转换服务。已与中国盲文出版社达成一致，后续将以该平台为基础开展示范应用，向盲用智能终端提供实时汉盲转换服务，并为中国盲文出版社盲文编辑排版平台提供服务支撑。基于 RGBD 传感器的视觉导盲避障技术，在现有的基于单帧 RGBD 图像处理的基础上，针对 RGBD 视频序列，结合时空信息，提出了基于历史数据分析的障碍物检测和识别方法，已将该方法融入智能导盲避障原型系统中，研究成果已在 2015 年的普适计算顶级会议 UbiComp 上发表，并在第 11 届中国信息无障碍论坛上做了报告。

2、智能应用

2.1 虚拟 3D 展示

基于多源感知信息融合的场景深度图获取技术，实验室采用主动式深度传感器和被动式立体视觉相互融合的方式，获取大规模、复杂场景下的高精度深度图。该方法在国际公认的 Middlebury 深度图估计评价数据集上的测试成绩排名第一，并在国际期刊 Sensors 上发表了论文 Reliable Fusion of Stereo Matching and Depth Sensor for High Quality Dense Depth Maps。基于深度图的虚拟视点绘制技术，实验室利用已获取的深度图信息，采用基于 3D object 和深度图边缘自适应校正的算法，绘制高质量虚拟视点画面，有效减少空洞、“鬼影”和叠影等问题的出现。利用上述两种技术形成了一套完整的“奥视景”三维自由视点展示系统，该系统可实现对物品水平 360° 或者水平 360° × 俯视 90° 的三维立体展示，并能够支持 PC、Web、IOS 等多个平台。

2.2 智能感知

根据饮食中的咀嚼行为与头部咀嚼肌肉具有密切的关联运动关系，实验室通过加速度信号来识别用户的咀嚼行为、提取肌肉的运动周期来计算用户的咀嚼次数。相关成果发表在国际重要会议 UbiComp' 15 (Poster) 上。实验室研发了动态情境自适应的增量行为学习模型，在动态蓝牙数据的基础上，利用模糊 ISODATA 聚类算法挖掘新的行为类别，提取新类行为的特征数据；同时结合 OSELM 学习算法构建增量学习模型，实现模型对新行为类别的自适应增量学习。相关成果已在国际重要会议 UbiComp' 15 (Poster) 上发表。实验室利用移动电话作为个人数据管理平台，开发了一套开放普适可穿戴数据融合平台 WearableHUB。WearableHUB 可以通过蓝牙与异构可穿戴设备进行数据交换。根据用户的动态情境信息，自动从后台服务器下载对应的 APP 到终端，为用户提供智能化的服务。相关成果已在国际重要会议 PerCom' 15 (Demo) 上发表。

3、移动计算

3.1 超级基站原型研制

实验室提出的超级基站设备面向未来移动通信产业发展需求，融合了超级计算技术与无线通信技术，属于移动通信网络架构创新。实验室已经完成了超级基站工程样机的研发，同时完成了面向未来移动通信系统的教学实训原型平台，采用软件无线电和软硬件解耦的思想，提供面向移动通信系统的开放、集成的开发环境，支持 2G/3G/4G/5G/ 宽带无线接入 / 卫星等各种无线通信标准的实现、实验网络的快速部署与相关技术验证，支持射频、基带、协议、核心网等不同层次的灵活可配置、可替换、可扩展。

3.2 超级基站资源智能管控

基于超级基站架构，针对潮汐效应带来的负载差异性和负载动态变化，以及资源在虚拟映射时的资源利用率等问题，研究了负载差异性条件下的集中式最优资源映射机制、提出了多协议处理实体在同一个处理器上执行时的协议组织和运行机制、研究了小区休眠技术在超级基站中的实现方法。超级基站可以通过集中式管控单元（OMU）对高虚拟化程度的处理资源和 RRH 进行动态映射和开关，同时，对比传统分布式蜂窝网络架构下的 traffic-aware 休眠机制，提出了 centralized-best 休眠机制，不仅能够通过优化算法得到全局最优部署方案，更通过对业务的预测和休眠模型的建立，降低了休眠检测与决策的系统开销，减少了频繁切换带来的问题，提高了

整体系统的稳定性。

3.3 基于多拟阵类优化的无线视频传输技术

实验室开发了基于拟阵理论的最优化无线传输方案，以及基于边际用户的定价策略。通过边际用户理论，可以联合视频内容和终端特性，对每个内容进行定价。同时提出了拟阵优化传输方法，具有优秀的几何和图论性质，尤其在解决资源分配和调度等问题具有独特优势，可以很大降低计算复杂度。该方法可以广泛地应用到各种无线视频传输系统中，尤其在线点播等系统，可以保证用户端保证一定质量的前提下，WSP 获得最大收益。相关成果发表在移动计算领域顶级国际期刊 IEEE Trans. Mobile Computing 2015，以及多媒体领域顶级国际期刊 IEEE Trans. Multimedia 2015 上。



学术委员会会议现场

中科院智能信息处理重点实验室

中国科学院智能信息处理重点实验室（以下简称“实验室”）依托于中国科学院计算技术研究所，成立于1987年。实验室在学术委员会的学术指导和计算技术研究所的大力支持下，经过多年建设，已逐渐发展成为我国智能信息处理领域的重要基础研究基地。

2015年度，智能信息处理重点实验室坚持以信息的智能处理为主线，依托计算技术研究所 在芯片、系统、网络等方面的优势，从信息的获取、表示、理解与转换等几个方面进行布局，构成了一个互为依托的有机整体。实验室的主要研究方向，包括智能科学、大规模知识处理、自然语言处理、知识网格、视觉信息处理与学习、多媒体技术、智能人机接口、生物信息学等。

在科研布局和方向上，实验室始终围绕着三个基本的目标进行——满足国家重大需求、服务国民经济建设和建设和谐社会，同时充分利用计算所在高性能计算和网络方面的优势。

在基础研究方面，本年度实验室继续保持了良好的发展势头，完成了多项创新性研究成果，相关成果累计发表论文90余篇，其中IEEE Trans.

on IP/ CSVT/MM 等国内外刊物论文30余篇，CVPR、ICCV、ICML等一流国际会议论文30余篇。新申请专利十余项、出版专著1本。

在“智能计算”领域，继续坚持机器翻译、智能科学、知识网格和大数据挖掘等学科方向。自然语言处理研究组集中力量积极开展深度学习研究，提出基于文本子结构和深度学习技术的文本匹配模型，有效求解任意长度文本的语义相似度匹配问题。在词序列的预测上，课题组设计了专门针对词序列生成的精细的门结构，更好的控制词序预测的局部和全局信息，在语言生成和机器翻译的n-best重排序上都远远超过了基线系统。相关工作发表于领域顶级刊物Computational Linguistics。在大规模数据挖掘和机器学习方面，

提出了深度极限学习机模型，该模型采用极限学习机作为基本的层构建方法，以随机平移与核化作为构建的基本元素。在模型的每一层，我们将数据通过极限学习机获得的预测结果进行随机投影，然后应用特定的核函数从而生成最终的特征。相关工作发表在Neurocomputing，并



成为该期刊 2015 年度下载次数最多的三篇论文之一。在大数据挖掘算法应用方面, 实现了高效并行的关联交易发现算法用于上海证券, 并获得 IJCAI 数据挖掘竞赛第一名。

在“视觉计算”领域, 坚持智能交互、视觉模式分析、多媒体等方向。人脸识别研究组继续围绕真实复杂场景下的人脸检测、人脸识别、基于视频的人脸识别三个方向开展研究工作, 提出了多领域适应学习、异质哈希学习、深度哈希学习、多功能二值编码学习等方法, 发表了 10 篇 CCF A 类论文 (包括国际刊物 TIP 一篇, 国际会议 ICCV3 篇、CVPR 论文 4 篇、ICML1 篇)。此外, 课题组获得 ICCV2015 ChaLearn: Looking at People 两个竞赛的冠军和亚军。在视频建模方面, 提出了属性学习、图像中层特征表示的模型框架。在智能视频监控问题上, 实现了基于字典学习的跟踪方法、基于深度模型的目标检测和行人再识别。智能人机交互研究组围绕 3D 手语识别及其应用展开研究工作。在连续手语识别和非特定人手语识别方面, 进行算法的创新, 分别提出虚拟运动增音预测及迭代的参考驱动度量学习

方法, 在对应问题上均取得优于现有方法的结果。

在科研项目承担与争取方面, 实验室本年度获批自然科学基金重点项目 1 项 (负责人为蒋树强研究院)、广东省重大科技项目 1 项 (参与单位负责人为何清研究员)、面上 / 青年基金项目多项。此外, 还持续得到了百度、联想、华为、高通等产业公司的支持。2015 年底可用经费 3000 余万元, 2015 年收入 2000 余万元。

在科研奖励方面, 实验室本年度作为项目主要完成单位获得国家自然科学二等奖 1 项 (第一完成人为陈熙霖研究员)、作为项目参与单位获得国家科技进步奖二等奖 1 项 (完成人刘群研究员)。此外何清研究员获得中国人工智能学会吴文俊人工智能科学技术创新奖二等奖、山世光研究员获得中国计算机学会优秀青年科学家奖。

在实验室管理方面, 实验室召开了学术委员会、学术规划会、室务会和学术年会, 批准了 7 项重点实验室开放课题。在公共经费使用方面, 形成了透明的制度, 通过为导师缴纳学生培养费、支持博士生学术交流、按学术论文进行经费拨付等方式, 尽可能为研究人员提供稳定的支持。

自主创新
科技强国

科研基地进展

中科院网络数据科学与技术 重点实验室

2015 年度，中科院网络数据科学与技术重点实验室继续坚持已有定位：推动网络数据学科发展，突破 ZB 级网络数据感知、传输、存储、管理与分析体系架构，研究网络数据界的溯源、定位、预测与控制方法，支撑安全大数据、情报大数据、金融大数据、商业与媒体消费大数据等系列应用。在基础研究、大项目争取、学术交流和推动大数据研究与发展等方面都取得了进展。获得北京市科学技术技术发明类二等奖。

具体包括：在基础研究方面，围绕着网络数据研究，发表论文近 50 篇，译著 1 部，。其中，发表在 PNAS 等 A 类期刊上的论文有 5 篇，发表在 WWW、SIGIR 等 A 类会议上的论文有 篇。同时我们组织了大数据分析的国际 workshop，撰写了一系列与大数据相关的前瞻性论文，完成了大数据学科的布局与推进。

在重要项目争取方面，获得国家自然科学杰出青年基金项目“web 信息检索与数据挖掘”。国家自然科学基金项目 5 项，国家 863 计划子课

题 3 项。

在重大项目进展方面，完成了国家 973 计划项目“网络大数据计算的基础理论及其应用研究”完成中期评估。项目首席科学家华云生教授主持了此次会议。科技部基础司重大项目处的副处长李非莅临会议并致辞。参加会议的还有中科院前沿局技术科学处处长孔明辉，项目特邀专家李国杰院士、梅宏院士，973 项目责任专家戴国忠研究员、唐常杰教授、彭群生教授，项目专家李学龙研究员、李忠诚研究员、刘志勇研究员、过敏意教授、周傲英教授，以及各课题负责人和学术骨干共计 50 余人。会议期间，各个课题负责人向专家组汇报了课题进展和遇到的问题与挑战，与会专家对各个课题的成果进行了讨论，明确了项目的研究方向，并对实施方案等提出了建设性意见。

在推动大数据学科与产业发展方面，牵头组织了国内规模最大、最具影响的大数据领域技术盛会——2015 中国大数据技术与学术大会 (Big



Data Technology Conference 2015, BDTC 2015)。本次大会历时3天，共包含大数据大赛，大数据生态系统、大数据基础设施、大数据政策法规和标准化、数据库技术与实践、大数据与人工智能、大数据与推荐系统、大数据安全技术论坛，金融业与大数据、工业制造业与大数据、交通旅游与大数据、互联网大数据等行业应用论坛。

为迎接大数据、云计算与“互联网+”等新兴产业浪潮，促进大数据专业人才的培养，推进“大众创业、万众创新”，举办了“中国好创意”CCF全国青年大数据创新大赛。此次大赛通过打造一个集数据共享、管理、建模与分析于一体的在线大数据云平台，汇聚企事业单位和政府机构的大数据资源与大数据分析处理的实际应用需求。据悉，比赛将分为“大数据创新程序大赛”和“大数据创意应用大赛”两部分，分别由企业和政府

提供数据和题目，参赛者们可以通过“众包”模式，针对问题设计高效算法、提出创意方案或研发应用原型系统。

在学术交流方面，举办了第三届数据科学与大数据分析workshop，本届Workshop内容涵盖了：大数据的采集、表示、索引、存储管理；大数据处理与预处理；大数据挖掘与理解的模型、算法和方法；大数据知识发现与语义挖掘；大数据可视化分析与组织；网络大数据内容挖掘；基于网络大数据的社会计算；大数据挖掘的工业和科学应用。本届Workshop共收到论文35篇，录用11篇。所有被录用文章的作者都到场做了报告，每个报告包括提问的30分钟都很充实。参加本届Workshop的学者学生共有50多人，大家在交流讨论中畅所欲言，开阔了学术视野，拓宽了研究思路。

正 气

大 气

骨 气

学术活动

计算所非常重视对外合作交流，鼓励科研人员既要“走出去”，参加各类学术会议，又要“请进来”，每年邀请国内外知名学者到计算所来交流讨论。2015 年度参加各类国际学术会议并作报告 111 人次；邀请国内外知名学者来所作学术报告 86 人次，举办 4 场国际学术会议，提升了计算所在相关学科领域的学术影响力。

□ 学术报告

序号	报告人	报告题目	报告人单位	报告日期
1	Shi-Yan Hu	Smart Home Cybersecurity: Threat and Defense in a Cyberphysical System	Michigan Technological University	2015-1-7
2	房佑寒	Bayesian Sampling Using Stochastic Gradient Thermostats	Purdue University	2015-1-8
3	杨超	新型基准测试程序 HPCG 在天河 2 上的算法设计与性能优化	中科院软件所	2015-1-8
4	张永乐	An Effort to Prevent Catastrophic Failures and Detect Performance Problems in Distributed Systems	University of Toronto	2015-1-16
5	唐光明	Bit-Slice Arithmetic Logic Unit and Multiplier for a 32-bit Rapid Single-Flux-Quantum Microprocessor	Kyoto University	2015-1-20
6	Shao-Shan Liu	Fast Big Data Analytics with Spark on Tachyon	University of California, Berkeley	2015-1-28
7	王元陶	The Little Engine that Could	英特尔中国研究院	2015-2-3
8	曹政	大规模互连网络研究中的热点问题	中科院计算所	2015-2-5
9	程明明	Efficient Image Scene Analysis and Applications	南开大学	2015-3-9

序号	报告人	报告题目	报告人单位	报告日期
10	胡伟武	君子先器识而后文章——兼谈如何做博士论文	中科院计算所	2015-3-10
11	韩伟力	移动终端感知安全与管控	复旦大学	2015-3-13
12	翟成祥	Statistical Methods for Integration and Analysis of Opinionated Text Data	University of Illinois at Urbana-Champaign	2015-3-31
13	Chang-Wen Chen	Assessing Photo Quality with Geo-Context and Crowdsourced Photos	University at Buffalo, the State University of New York	2015-4-1
14	樊莉亚	OS4M: Achieving Global Load Balance of MapReduce Workload by Scheduling at the Operation Level	IBM 中国研究院	2015-4-2
15	Wei Chen	Combinatorial Learning for Combinatorial Optimizations --- A Trilogy	微软亚洲研究院	2015-4-8
16	ErykDutkiewicz	Research Challenges of Dynamic Spectrum Sharing for 5G Networks	Macquarie University	2015-4-13
17	方更法	The Next Generation Mobile Networks and LTE over Unlicensed Bands	Macquarie University	2015-4-13
18	Xiao-Song Ma	Framework for Automatic Extraction of Parallel Benchmarks from HPC Applications	Qatar Computing Research Institute	2015-4-13
19	魏朝晖	An Introduction to PSD-rank and its Applications	Nanyang Technological University	2015-4-16
20	陈天石	体系结构与机器学习的交叉领域研究	中科院计算所	2015-4-17
21	郑伟诗	视频监控下的交互行为分析研究	中山大学	2015-4-20
22	Pin-Yan Lu	Approximate Counting via Correlation Decay	微软亚洲研究院	2015-4-23
23	Wei Chen	开源，开放，面向未来的 IBM HPC 之路 - 关于 IBM 未来几年 HPC 技术发展的思路和构想	IBM	2015-4-27
24	Periklis A. Papakonstantinou	Big Data as a Powerful Source of Randomness	清华大学	2015-5-7
25	邢国良	基于智能手机的非打扰移动健康监测系统: 以生物钟监测为例	Michigan Technological University	2015-5-8

序号	报告人	报告题目	报告人单位	报告日期
26	Xiao-Yi Lu	Accelerating Big Data Processing through HPC Technologies and Techniques	The Ohio State University	2015-5-10
27	孙贤和	Parallelism for High Performance Data Processing: A Rethinking	Illinois Institute of Technology	2015-5-15
28	Nei Kato	D2D: Research Trend and Future Perspective	Tohoku University	2015-5-19
29	Qin MA	Development and Application of Bioinformatic Tools Driven by Genomic and Transcriptomic Data	Georgia State University	2015-5-21
30	Hai-Bin Ling	High-order Tensor Analysis for Groupwise Correspondence and Its Applications in Computer Vision	Temple University	2015-5-21
31	Michel Dubois	Synchronization, Coherence and Event Ordering in Multiprocessors	University of Southern California	2015-5-22
32	Jiong-Dong Wang	Big Media Data: Search and Management	微软亚洲研究院	2015-5-27
33	DileepBhandarkar	A Journey Through History From Mainframes to Smartphones	美国高通公司	2015-5-28
34	Qi Tian	Discriminative Visual Representations for Image Search and Classification	University of Texas at San Antoni	2015-6-2
35	Jie Wu	Algorithmic Crowdsourcing: Current Status and Future Perspective	Temple University	2015-6-17
36	Gang Zhou	Reducing Smartphone Application Delay & Energy Consumption through Storage I/O Optimization	College of William and Mary	2015-6-18
37	Benjamin C. Lee	Economic Mechanisms for Managing Risk in Datacenters	Duke University	2015-6-19
38	Xin-Yu Zhang	Millimeter-wave Wireless Networking and Sensing: A Unified Perspective	University of Wisconsin-Madison	2015-6-30
39	MattanErez	Adapt or Die: One Architect's Perspective on Memory Technology	University of Texas at Austin	2015-6-30
40	Xuan Zhang	Designing Micro-Robotic "Brain": a System-on-Chip (SoC) Approach	Washington University in St. Louis	2015-7-2
41	Sun-Yuan Kung	Visualization of Big Data	Princeton University	2015-7-8



序号	报告人	报告题目	报告人单位	报告日期
42	Zhu Han	Resource Allocation for Full-Duplex Communication and Networks	University of Houston	2015-7-9
43	李秋明	美丽的数据——效率、发现与说服	财新传媒	2015-7-17
44	Rose Hu	Towards Spectrum Efficient, Energy Efficient and QoE Aware 5G Wireless Systems	Utah State University	2015-7-23
45	Chen Ding	Theory and Practice of Multicore Cache Optimization	University of Rochester	2015-7-23
46	Terrence Mak	Dynamic Programming Networks for on-Chip Communications	University of Southampton	2015-7-23
47	罗杰波	Big data, Better life	University of Rochester	2015-8-13
48	Yu-Chen Zhang	Distributed Estimation of Generalized Matrix Rank: Efficient Algorithms and Lower Bounds	University of California	2015-8-13
49	Qu Gang	Challenges and Opportunities for the Design of Secure, Trusted and Energy Efficient Systems	University of Maryland	2015-8-18
50	Ran Duan	Approximate and Exact Algorithms for Maximum Weight Matching	清华大学	2015-8-27
51	Tao Qin	Cloud Computing Pricing: Algorithms and Mechanisms	微软亚洲研究院	2015-8-31
52	陈勇	高性能计算和科学大数据问题的一些探讨	Texas Tech University	2015-9-2
53	John Leidel	Goblin-Core-64 RISC-V Extension for Data Intensive Computing	Texas Tech University	2015-9-2
54	宁振波	诊断分类——号脉中国制造	中航工业 301 研究所	2015-9-9
55	李凯	新一代计算系统的思考	Princeton University	2015-9-10
56	张晓东	Impact of Big Data on Access Performance	The Ohio State University	2015-9-10
57	史元春	人机交互技术最新进展	清华大学	2015-9-15
58	陈益强	可穿戴计算与 E-Health	中科院计算所	2015-9-15
59	Run-Yao Duan	On Zero-error Communication via Quantum Channels in the Presence of Noiseless Feedback	University of Technology Sydney	2015-9-15
60	李德毅	脑认知的形式化	北京邮电大学	2015-9-17

序号	报告人	报告题目	报告人单位	报告日期
61	李建	Learning Arbitrary Statistical Mixtures of Discrete Distributions	清华大学	2015-9-24
62	刘青山	Face Detection and Alignment	南京信息工程大学	2015-9-24
63	陶大程	Multi-view Learning	University of Technology, Sydney	2015-9-24
64	徐东	SVM Based Approaches for Visual Domain Adaptation	Nanyang Technological University	2015-9-28
65	刘威	Learning to Hash by Discrete Optimization	滴滴快滴研究院	2015-9-28
66	Jim K. Omura	Data Storage and Management Applications of Information Technology in United States Healthcare	DataFascia Corporation	2015-10-12
67	Hai-Wu HE	Toward BigData Network Security for CSTNET	中科院网络中心	2015-10-14
68	赵地	基于 GPU 计算的并行深度学习	中科院网络中心	2015-10-14
69	段海新	谁动了我的饼干 —— Web Cookie 的完整性和现实威胁	清华大学	2015-10-16
70	梁伟	1. 选择云服务, 何去何从; 2. 云测试之 SPECvirt	中国电信研究院	2015-10-26
71	施巍松	Toward Energy-Efficient Computing	Wayne State University	2015-10-30
72	孙贤和	Utilizing Concurrency in Data Access: Thinking from a data-centric view	Illinois Institute of Technology	2015-11-2
73	DimitrisGizopoulos	Microarchitecture-Level Reliability Assessment for CPUs and GPUs	University of Athens	2015-11-4
74	John Kim	Security Vulnerability of Processor-Interconnect in NUMA Architecture	Korea Advanced Institute of Science and Technology	2015-11-10
75	魏朝晖	Assessing the Dimension of an Unknown Quantum System in a Device-Independent Manner	Centre for Quantum Technologies in Singapore	2015-11-12
76	Christos Faloutsos	Mining Large Graphs: Patterns, Anomalies, and Fraud Detection	Carnegie Mellon University	2015-11-19



序号	报告人	报告题目	报告人单位	报告日期
77	Jie Wu	Dual-Centric Data Center Network Architectures	Temple University	2015-11-24
78	何炳胜	When HPC Meets Big Data: Emerging HPC Technologies for High-Performance Data Management Systems	Nanyang Technological University	2015-12-1
79	赵伟	从物网到物联网及更远 -From Net of Things to Internet of Things and Beyond	澳门大学	2015-12-8
80	钟林	Base stations are interesting computer systems	Rice University	2015-12-11
81	凌少平	Redefining Cancer Genomics with Population and Evolutionary Genetics	中科院北京基因组所	2015-12-11
82	周源源	The Gap between Research and Practice	University of California, San Diego	2015-12-16
83	Juan Carlos San Miguel	Self-reconfigurable Cameras for Multi-sensor Target Tracking in Video	Universidad Autónoma de Madrid (西班牙马德里自治学府)	2015-12-17
84	诸葛海	Dimensions on Summarisation	中科院计算所	2015-12-24
85	高飞	量子密码简介	北京邮电大学	2015-12-24
86	张姗姗	How Far are We from Solving Pedestrian Detection?	MPI für Informatik (德国马克斯 - 普朗克计算机科学研究所以)	2015-12-24

人人是才

人人幸福

□ 重要学术会议

大数据 "Big Data in China" 中澳论坛

8月10日,网络数据实验室承办了在Big Data Summit 2015上组织“Big Data in China”论坛。本次论坛有幸邀请到李国杰、高文、梅宏、怀进鹏、鄂维南等院士担任顾问。论坛还邀请了CCF大数据专家委员会委员清华大学朱文武教授、北京大学袁晓如教授、美国罗格斯-新泽西州立大学熊辉教授等国内外大数据领域的专家学者作特色报告。同时,来自华为的张宝峰委员、阿里巴巴的闵万里博士、蚂蚁金服的陈继东委员等国内IT与互联网龙头企业的代表以Showcase的形式向参会人员介绍了他们在大数据领域的规划与措施,以及开展的具体研究工作。



论坛于8月10日下午在悉尼希尔顿酒店举行,由论坛共同主席朱文武教授与靳小龙副研究员主持。清华大学朱文武教授重点介绍了网络-物理-人类(CPH)大数据计算的主要概念、挑战和科学问题、方法和途径,并使用网络社会数据的例子来说明如何进行社交媒体大数据计算。美国罗格斯-新泽西州立大学熊辉教授分

享了他们针对动态的商业环境中顺序模式应用分析所提出的新方法。北京大学袁晓如教授的报告介绍了不同类型城市数据的可视化案例,包括出租车的全球定位系统数据、车辆识别数据、地铁IC卡数据与社会媒体数据,并展示如何集成不同的数据集来进行高级可视化分析。清华大学唐杰副教授以AMiner.org为例介绍了几个重要的异构图挖掘算法,包括节点主题建模、路径相似性度量、信息子图挖掘、发现关联规则等。靳小龙副研究员在论坛上介绍了中科院网络数据科学与技术重点实验室在处理大数据的数据复杂性、计算复杂性和系统复杂性方面取得的成果。在论坛的Showcase部分,华为的张宝峰委员分享了电信行业中解决实际问题的经验和见解。阿里巴巴的闵万里博士讨论了电子商务中的几个应用案例,并强调数据挖掘技术在这些应用案例中的重要作用。蚂蚁金服的陈继东博士介绍了基于大数据的支付宝网络安全和欺诈风险管理。

尽管论坛与KDD 2015的研讨会和短期课程同时进行,但仍然吸引了一百多位听众参会并与讲者们积极互动,还就大数据如何促进中国“互联网+”国家战略,面对大数据往往掌握在企业界手中的现实,企业界与学术界如何合作等议题进行了热烈讨论。与会的国外专家分享了他们的丰富经验,也给出了宝贵建议。

《通信与计算机科学技术融合》第二届全球华人学术研讨会

数据的爆炸式增长要求数据传输、存储和处理能力分别提升1000倍以上,为了实现绿色发展,信息系统的性能功耗也要提升1000倍以上,只有高效融合的3C技术才能应对这样的挑战。

在大数据和“互联网+”应用的牵引下，计算、通信和控制技术必将进入3C融合的时代。什么理论可以指导3C融合？3C融合的突破口在哪里？

为了促进通信和计算机领域科研人员的交流合作，同时为推进国际合作的进程，由中国科学院计算技术研究所主办，中科院计算所无线通信技术研究中心承办的《通信与计算机科学技术融合》第二届全球华人学术研讨年会于2015年10月15-16日在北京香山饭店成功举办。此次会议的举办受到中国工程院、北京市科学技术委员会和北京市自然科学基金委员会等领导部门的大力支持。本次研讨会以3C融合为主题，旨在开放式地探讨3C融合和虚拟云服务下的未来移动通信网络体系架构，以及支撑这种体系架构的3C融合高通量计算和通信的关键技术与系统装备等。会议中，国内外通信与计算机领域的知名专家学者就3C融合进行了深入的技术交流，探讨了3C融合对互联网+、大数据等应用和未来5G、空天地一体化等系统发展的深远作用，对3C融合提出了指导性意见。



本次会期两天，会议主要包括大数据时代的未来网络体系架构、3C融合的空天地一体化网

络、面向移动通信的3C融合关键技术三大部分，与会专家开放式地探讨了在3C融合时代，未来的地面与空间移动通信网络如何融合计算机网络、高通量计算等前沿技术，寻求3C融合的发展新路。

2015年10月15日上午会议正式开始，会议由中国科学院计算技术研究所无线中心副主任周一青研究员主持。大会主席李国杰院士和北京市自然科学基金委办公室江南副主任分别致辞，欢迎各位领导专家莅临，并希望通过此次会议的交流讨论，碰撞通信思维与计算机思维，探索移动通信的科学前沿技术，为首都发展树立品牌与标杆，建设全球科技创新中心。



第三届中国计算机学会大数据学术会议

2015年10月20日至21日，中国计算机学会（CCF）主办的第三届大数据学术会议在合肥隆重召开。会议由CCF大数据专家委员会（大专委）、中国科学技术大学及安徽大学联合承办，安徽大数据应用协同创新中心协办，中国科学院网络数据科学与技术重点实验室参与组织。大专委主任、中国工程院李国杰院士、香港中文大学副校长华云生教授担任大会名誉主席，中国科学技术大学陈国良院士、上海交通大学副校长梅宏院士担任大会共同主席，中国科学技术大学陈恩

红教授、清华大学王建民教授、国防科学技术大学肖侬教授、兰州大学胡斌教授担任大会程序委员会共同主席。大会开幕式于 20 日上午举行，由中国科学技术大学计算机学院副院长陈恩红教授主持。开幕式上，中国科学技术大学周先意副校长代表承办单位致辞。安徽大学王群京副校长在致辞中对各位参会嘉宾和代表表示诚挚谢意与欢迎。



CCF 大数据专家委员会秘书长、中国科学院网络数据科学与技术重点实验室主任程学旗在代表 CCF 和大专委致辞时表示，大专委自成立以来，在推动大数据学科方向的建设与发展，构建面向大数据产学研用的学术交流、技术合作与数据共享平台，以及为政府相关部门提供战略性意见与建议方面做了大量工作，每年一度“大数据白皮书”的发布等活动，在学术与产业界，以及政府决策部门产生了重要的影响。

第三届数据科学与大数据分析 Workshop (DSBDA-2015)

2015 年 11 月 14 日，由中国科学院网络数据科学与技术重点实验室在数据挖掘领域的国际顶级会议 ICDM-2015 上组织的“第三届数据科学与大数据分析 Workshop (DSBDA-2015)”

在美国大西洋城召开。

本届 Workshop 内容涵盖了：大数据的采集、表示、索引、存储管理；大数据处理与预处理；大数据挖掘与理解的模型、算法和方法；大数据知识发现与语义挖掘；大数据可视化分析与组织；网络大数据内容挖掘；基于网络大数据的社会计算；大数据挖掘的工业和科学应用。

本届 Workshop 共收到论文 35 篇，录用 11 篇。所有被录用文章的作者都到场做了报告，每个报告包括提问的 30 分钟都很充实。参加本届 Workshop 的学者学生共有 50 多人，大家在交流讨论中畅所欲言，开阔了学术视野，拓宽了研究思路。

第二届中韩大数据论坛在韩国大邱召开

第二届中韩大数据论坛于 2015 年 10 月 30

日在韩国大邱召开。本届论坛由中国计算机学会大数据专家委员会、韩中科学技术合作中心、韩国 WATEF(World Association for Triple Helix & Future Strategy Studies)、韩国数据科学学会(KADS)、韩国生命工学研究院 (KRIBB)、韩国信息化振兴院 (NIA) 共同主办。

来自中国科学院计算机网络信息中心沈烁委

员、中国科学院网络数据科学与技术重点实验室靳小龙委员、中国科学院计算技术研究所生物信息研究组孙世伟副研究员、贵阳大数据交易所执行总裁王叁寿等以及韩国蔚山科学技术院、韩国安全性评价研究所、韩国生命工学研究院、国立群山大学、韩国首尔大学、韩国三星集团的专家学者和企业高管参会并作特色主题报告。



从优秀到卓越

知识产权

2015 年，计算所申请新专利 145 件，获得新授权专利 79 件，软件登记 49 项，获得国家知识产权局“中国专利优秀奖”1 项，在专利布局和组合申请方面加强规划和深度挖掘，改变单点申请的局面，在智能加速芯片方向形成 13 项专利组合申请。在标准制定方面，全年主导制定 3 项通信行业标准、主导 2 项信息编码国家标准、参与 3 项定位技术国家标准。

2015 年继续在知识产权转移转化方面狠下功夫，苦练内功，深入推进专利价值分析和分级分类管理工作，通过专利评价、评级，科学合理判断专利使用价值，找出适合市场需求主体的专利，为促进专利转移转化进行有益尝试。今年计算所牵头承担了中科院科技促进发展局的 STS 项目—中科院专利价值分析试点工作，联合长春应化所、大连化物所、微电子所等试点所，探索专利价值分析在知识产权管理和技术转移中的示范作用；同时抓住新的科技促进发展法颁布的契机，加大专利转移转化的力度，结合专利价值分析结果，成功转让 9 件专利，获得 180 万的收益。

此外，今年还参与了中科院科技促进发展局的 STS 项目—中科院知识产权管理规范贯标试点工作，通过该工作使研究所的知识产权工作更加规范和高效。

□ 授权发明专利列表

序号	授权号	专 利 名 称	发 明 / 设 计 人
1	ZL201010233950.0	基于信息关联的网络安全态势感知系统及其方法	席荣荣、金舒原、吴进、董昭
2	ZL201210172844.5	一种用于多根共享系统的 I/O 重映射方法和装置	刘小丽、曹政、张佩珩、安学军、孙凝晖、王展、苏勇、刘飞龙
3	ZL201210208518.5	基于设备能力的多终端数据同步方法和系统	马久跃、刘立坤、严得辰、李旭
4	ZL201210270595.3	一种三维人体运动的实时在线获取方法及其系统	夏时洪、费心宇、苏乐



序号	授权号	专利名称	发明 / 设计人
5	ZL201110005686.X	基于 DS 推理的无线传感器网络多目标跟踪数据关联方法	孙荣丽、王睿、张磊、崔莉
6	ZL201210447321.7	一种动态断点的自动生成方法及系统	李丰、霍玮、陈聪明、李龙、袁璐洁、冯晓兵
7	ZL201210068457.7	一种迭代式概念属性名称自动获取方法和系统	曹存根、汪平仄
8	ZL201210160516.3	基于 TPM 的本地验证式启动方法	高云伟、薛栋梁、孙毓忠
9	ZL201210113442.8	一种基于指纹特征的通用用户注册认证方法及系统	史红周、何遥、费德林
10	ZL201310123570.5	基于二维宏观流模型描述交通流行为的方法及模拟系统	毛天露、王华、王兆其
11	ZL201210376184.2	一种无反馈的优化码率的分布式视频编解码方法及其系统	纪雯、陈益强
12	ZL201110134284.X	一种安全加密协处理器及无线传感器网络节点芯片	王义、赵泽、崔莉
13	ZL201210401238.6	一种面向增量式翻译的结构化语言模型构建方法及系统	于恒、米海涛、刘群
14	ZL201210320575.2	一种面向中文搜索引擎混杂语言的查询纠错方法及系统	程舒杨、熊锦华、公帅、颢悦、张成、程学旗
15	ZL201110156631.9	移动通信终端及其实现方法	黄亮、石晶林
16	ZL201110328212.9	多汇聚节点无线传感器网络组网方法及其系统	张招亮、张乐、黄庭培、李栋、崔莉
17	ZL201210208874.7	基于物理脉冲耦合的无线传感器网络时间同步方法	洪学海、徐勇军、安竹林、林伟
18	ZL201210376220.5	一种无反馈的分布式视频编解码方法及其系统	纪雯、陈益强
19	ZL201210160589.2	基于 TPM/VTPM 的可信程序列表生成方法及系统	高云伟、薛栋梁、孙毓忠
20	ZL201210530581.0	一种基于端口偏射的云路由数据处理方法及系统	李彦君、张国清
21	ZL201210160339.9	一种机群文件系统客户端存储资源延迟分配方法和系统	张军伟、刘超、齐颖、郑彩平、刘振军

序号	授权号	专利名称	发明 / 设计人
22	ZL201210322527.7	一种 LTE 基站侧系统信息管理和发送的方法和装置	石晶林、李毅潇、钱蔓藜、韦伟、杨喜宁、胡金龙
23	ZL201210299144.2	一种基于多源实测数据的火焰建模方法及其系统	朱登明、沈亮、魏毅、王兆其
24	ZL201310143346.2	一种面向微博客平台的社交朋友圈的挖掘方法及系统	程学旗、李静远、贺志明、伍大勇、王元卓
25	ZL201210211713.3	一种基于业务满意度的无线资源分配方法与系统	王煜炜、唐培栋、张杨杨、刘敏
26	ZL201210171920.0	一种用于多根共享系统的直接 I/O 虚拟化方法和装置	曹政、刘小丽、张佩珩、安学军、孙凝晖、王展、苏勇、刘飞龙
27	ZL201210162995.2	一种时空轨迹相似度计算方法及系统	叶剑、朱珍民、张筱旋、王冠男、姚昱旻、杜静
28	ZL201210579563.1	一种基于电流分解的非侵入式负载监测方法及系统	刘晶杰、徐志伟、聂磊
29	ZL201210563218.9	基于网络中间点的网页响应时间被动测量方法及系统	徐翔、张广兴、谢高岗
30	ZL201210570089.6	一种分布式文件系统客户端缓存中文件创建方法及其系统	张军伟、齐颖、刘振军
31	ZL201210401317.7	用于对倒排索引进行压缩的文档序号重排方法及其系统	史亮、王斌、卫冰洁、李锐、张帅、张冠元
32	ZL201210438946.7	一种基于 TPM 的 Linux 内核初始化中的数据保护方法及系统	薛栋梁、高云伟、杨鹏斐、展旭升、孙毓忠
33	ZL201210536318.2	一种基于迟滞调度的网络流量冲突避免方法及系统	李彦君、张国清、郭庆萍
34	ZL201210495728.7	一种基于授权方式的元数据的处理方法和系统	张欢、冯硕、马一力、刘培峰
35	ZL201210519595.2	一种用于视频会议的眼神交互方法及系统	尹苓琳、陈益强、纪雯
36	ZL201310220559.0	一种通过加速度传感器计算步行速度的移动终端及方法	吴琳、徐勇军、安竹林、黄晁
37	ZL201310230933.5	一种个性化论文推荐方法及其系统	程学旗、郭嘉丰、薛欢、廖华明、曹雷



序号	授权号	专利名称	发明 / 设计人
38	ZL201310153333.3	一种基于加密模糊关键字的机会网络路由方法及系统	王飞、吴琳、徐勇军、安竹林、黄晁
39	ZL201010224004.X	一种业务流程间交互冲突的半自动修正方法及其系统	公帅、熊锦华、刘志勇
40	ZL201310016206.9	一种跨虚拟化数据中心的虚拟机在线迁移方法	杨鹏斐、杨坤、宋莹、赵霞、孙毓忠
41	ZL201210260600.2	一种按键行为的特征数据的处理装置及处理方法	陈振宇、王双全、陈益强、唐焱
42	ZL201310311420.7	一种单 SOC 芯片及单 SOC 芯片多工作模式的复用方法	吴少校、苏孟豪
43	ZL201210210044.8	虚拟机间安全通信方法	高云伟、邬小龙、孙毓忠
44	ZL201210347942.8	一种人机交互戒指	张博宁、钱跃良、陈益强、王向东
45	ZL201210555210.8	一种多维度用电行为的感知方法及系统	刘晶杰、徐志伟、聂磊
46	ZL201310038675.0	一种基于 PCIe 数据交换的通信方法及系统	孙凝晖、曹政、刘小丽、安学军、张佩珩、
47	ZL200910083522.1	文本情感倾向性分析方法	吴琼、谭松波、程学旗
48	ZL201110315576.3	网络安全协同联动系统及方法	云晓春、张永铮、孙建亮、臧天宁
49	ZL201210011178.7	微博客数据采集方法及系统	程学旗、房伟伟、李静远、陈根宝、邢国亮、张凯、金波、方滨兴
50	ZL201210033069.5	用于检测 LTE 系统基站发射天线数量的方法及装置	苏泳涛、黄守俊、何莹、姚彦斌、石晶林
51	ZL201210051468.4	一种离散事件网络模拟环境的时钟同步方法	林思明、王景飞、程学旗、张冬、周洲仪、李金明、王元卓
52	ZL201210222862.X	一种获取无线台管理连接标识的方法	田霖、杨育波、孙毅、周继华
53	ZL201210228113.8	用于更新动态场景的 Voronoi 图的方法及设备	方巍、毛天露、蒋浩、李杨、王兆其
54	ZL201110293416.3	面向高速铁路的 LTE 通信系统中的切换方法及系统	李娟、田霖、周一青、石晶林

序号	授权号	专利名称	发明 / 设计人
55	ZL201310081984.6	基于开放知识库的短文本语义概念自动化扩展方法及系统	程学旗、刘盛华、肖永磊、王元卓、刘悦
56	ZL201210118200.8	GPS 软件接收机中的跟踪环路信号处理方法及装置	姚相振、崔绍龙、方金云
57	ZL201310146037.0	查询多意图识别方法和系统	程学旗、熊锦华、程舒杨、廖华明、王元卓、公帅
58	ZL201210017962.9	获取对象级访存行为的方法及装置	陈荔城、包云岗、崔泽汉、黄永兵、陈明宇
59	ZL201110194206.9	一种视频质量客观评价方法	黄庆明、许倩倩、苏荔、秦磊、蒋树强
60	ZL201310136896.1	面向网络流式数据的事件实时过滤方法和系统	程学旗、刘盛华、邱文一、王元卓、刘悦、莫溢、黄展坤
61	ZL201310149482.2	基于错误模式挖掘的中文搜索引擎查询纠错方法及系统	熊锦华、程舒杨、程学旗、公帅、颢悦、张成、廖华明
62	ZL201110046896.3	一种语音浏览方法及浏览器	邓铸辉、陈启华、王向东、钱跃良、林守勋
63	ZL201310121491.0	一种频率资源分配方法	周一青、刘航、田霖、陈海华、石晶林
64	ZL201110240878.9	一种手机蠕虫检测方法	卢维清、崔翔、翟立东、廖鹏
65	ZL201310049317.X	面向领域的暗网资源采集方法和系统	熊锦华、林海伦、程学旗、张永超、廖华明
66	ZL201210144144.5	CoMP 上行链路中时间提前调整参考节点选取的方法及装置	周一青、陈海华、石晶林、史岗
67	ZL201310082990.3	一种短文本数据的事件演化分析方法	程学旗、刘盛华、李福鑫、王元卓、刘悦
68	ZL201310134933.5	用于搜索引擎的色情用户查询识别方法及设备	程学旗、熊锦华、公帅、张成、廖华明、王元卓
69	ZL201210587473.7	协作式缓存集群中面向视频点播服务的协作式缓存方法及系统	石刘、张军伟、刘振军、韩晓明、许鲁
70	ZL201210592755.6	两阶段协作多播方法及系统	周一青、刘航、田霖、陈海华、石晶林
71	ZL201310257668.X	一种用于 OFDM 系统的多普勒变化率估计方法与系统	周一青、侯占伟、田霖、王妮娜、苏泳涛、石晶林



序号	授权号	专 利 名 称	发 明 / 设 计 人
72	ZL201310005883.0	一种块级快照系统及基于该系统的用户读写方法	许鲁、郭高峰、赵猛
73	ZL201310165997.1	大规模群体表演动画合成方法及设备	吕蕾、毛天露、王兆其
74	ZL201210244438.5	一种识别动态特征应用流量的方法	郑红霞、张广兴、杨建华
75	ZL201310393446.0	一种服务器风扇功耗测量方法	江志雄、陆春阳、姜志颖、侯锐
76	ZL201310199292.1	一种虚拟机功耗测量方法	江志雄、陆春阳、马崇亚、姜志颖
77	ZL200810227105.5	一种基于关键词的文本情感分类器的训练方法和分类方法	谭松波、程学旗
78	ZL201210088888.X	一种用于宽带无线通信数字基带处理器的原型验证平台	石晶林、高明晋、韩娟、史岗、马英矫、李辉、王秋菊、任为、陈洋
79	ZL201210585790.5	一种块设备异步数据迁移方法和系统	许鲁、刘昌

正 气

大 气

骨 气

□ 软件登记列表

序号	登记号	软件名称
1	2015SR017235	Redmine 统计报表系统
2	2015SR017267	互联网公开数据采集质量巡查系统
3	2015SR017043	融合用户反馈的交互式语音识别系统
4	2015SR017048	天倪云博士云端学习机软件
5	2015SR017066	天倪云计算平台软件
6	2015SR017230	天玑微博监测与演化分析软件
7	2015SR017119	天玑大规模图数据存储与查询引擎系统
8	2015SR020220	大规模图像检索软件系统
9	2015SR020775	ICTBase 大规模分布式数据在线分析软件系统
10	2015SR070947	微魔镜微博预测预警系统
11	2015SR071532	颅面部解剖形态 3D 透视显示系统
12	2015SR071374	地市级的突发事件态势推演软件系统
13	2015SR071535	城市人口全生命周期数据挖掘系统
14	2015SR071211	残疾人远程维权服务平台软件
15	2015SR071060	千核万线程模拟器系统
16	2015SR120648	天玑网络数据引擎配置管理分系统
17	2015SR120464	天玑网络数据引擎系统
18	2015SR121995	新闻列表页发现软件
19	2015SR122099	基于包装器的通用信息抽取软件
20	2015SR122092	分布式 JavaScript/Ajax 网页解析软件
21	2015SR122090	多记录页面自动信息抽取软件
22	2015SR121989	博客采集软件
23	2015SR121978	模版自动生成软件
24	2015SR121906	单记录页面自动信息抽取软件
25	2015SR121970	新闻采集软件



序号	登记号	软件名称
26	2015SR121900	基于模板的通用信息抽取软件
27	2015SR121722	天玑网络数据引擎多通道采集分系统
28	2015SR121718	论坛采集软件
29	2015SR121669	天玑大规模数据检索系统
30	2015SR121691	ECFS 机群文件系统
31	2015SR121413	安全车联网 V2X 系统仿真平台 - 同步 DSRC 通信仿真子系统
32	2015SR120775	安全车联网 V2X 系统仿真平台 -LTE 通信仿真子系统
33	2015SR120768	安全车联网 V2X 系统仿真平台 - 异步 DSRC 通信仿真子系统
34	2015SR121355	集体人类行为虚拟现实仿真软件系统
35	2015SR121339	蛋白质质谱分析的交联鉴定软件
36	2015SR121225	单粒子效应防护效果分析软件
37	2015SR128726	基于电视内容实时监测的节目智能导视
38	2015SR127959	基于电视内容实时监测的导视服务平台系统
39	2015SR195929	知识产权管理系统
40	2015SR195975	专利价值评估系统
41	2015SR195762	安卓移动平台常见功耗设置系统
42	2015SR195714	安卓移动平台处理器和全球定位系统功耗测量软件
43	2015SR195920	安卓移动平台功耗分析软件
44	2015SR196784	安卓移动平台功耗分析系统环境参数检测系统
45	2015SR195891	安卓移动平台功耗相关组件测试系统
46	2015SR195765	迁移学习系统
47	2015SR195916	天玑互联网社交网络热词挖掘和词义扩展系统
48	2015SR195923	基于 LDA 的子话题生成及其关键词词组生成系统
49	2015SR229936	高随机性能 R6 虚拟化软件

□ 标准列表

序号	标准号	中文名称	标准范围
1	YD/T 2814-2015	IPv6 路由协议 适用于低功耗有损网络的 IPv6 多播路由协议	行业标准
2	YD/T 2815-2015	IPv6 技术要求 代理移动 IPv6 路由优化	行业标准
3	2013-2466T-YD	IPv6 技术要求 基于代理移动 IPv6 的组播	行业标准
4	GB/T 20090.11-2015	信息技术 先进音视频编码 第 11 部分：同步文本	国家标准
5	GB/T 20090.12-2015	信息技术 先进音视频编码 第 12 部分：综合场景	国家标准
6	GB/T 30996.1-2014	信息技术 实时定位系统 第 1 部分：应用程序接口	国家标准
7	GB/T 29261.5-2014	信息技术 自动识别和数据采集技术 词汇 第 5 部分：定位系统	国家标准
8	GB/T 31101-2014	信息技术 自动识别和数据采集技术 词汇 实时定位系统性能测试方法	国家标准



□ 代表性论文列表

会议论文：

CCF A 类：

1. Jiuyue Ma, Xiufeng Sui, Ninghui Sun, Yupeng Li, Zhihao Yu, Bowen Huang, TianiXu, Zhicheng Yao, Yun Chen, Haibin Wang, Lixing Zhang, Yungang Bao, Supporting Differentiated Services in Computers via Programmable Architecture for Resourcing-on-Demand (PARD) , ASPLOS, 2015.
2. Daofu Liu, Tianshi Chen, Shaoli Liu, Jinhong Zhou, Shengyuan Zhou, Olivier Temam, XiaobingFeng, Xuehai Zhou, and Yunji Chen, “PuDianNao: A Polyvalent Machine Learning Accelerator” , ASPLOS, 2015.
3. Zidong Du, Daniel D Ben-Dayyan Rubin, Yunji Chen, Liqiang He, Tianshi Chen, Lei Zhang, Chengyong Wu, and Olivier Temam, “Neuromorphic Accelerators: A Comparison Between Neuroscience and Machine-Learning Approaches” , MICRO, 2015.
4. Jianbo Dong, et al. "Venice: Exploring server architectures for effective resource sharing." IEEE International Symposium on High Performance Computer Architecture 2016.
5. Li, Zhenyu, et al. "User Behavior Characterization of a Large-scale Mobile Live Streaming System." Proceedings of the 24th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 2015.
6. Xiang Yuan, Chenggang Wu*, Zhenjiang Wang, Jianjun Li, Pen-Chung Yew, Jeff Huang, XiaobingFeng, YanyanLan, Yunji Chen and Yong Guan: Reproducing Concurrency Bugs Using Local Clocks. ICSE, 2015.
7. Zidong Du, Robert Fasthuber, Tianshi Chen, Paolo lenne, Ling Li, Tao Luo, XiaobingFeng, Yunji Chen, and Olivier Temam, “ShiDianNao: Shifting Vision Processing Closer to the Sensor” ,ISCA, 2015.
8. Shimin Chen, Qin Jin, "Persistent B+ Trees in Non-Volatile Main Memory". VLDB’ 15.

9. PengBao, Hua-Wei Shen, Xiaolong Jin, Xue-Qi Cheng. Modeling and predicting popularity dynamics of Microblogs using self-excited Hawkes processes, WWW 2015.
10. Zheng Lin, Weiping Wang, Xiaolong Jin, Dan Meng. A Word Vector and Matrix Factorization Based Method for Opinion Lexicon Extraction, WWW 2015.
11. Long Xia, Jun Xu, YanyanLan, JiafengGuo and Xueqi Cheng. Learning Maximal Marginal Relevance Model via Directly Optimizing Diversity Evaluation Measures, SIGIR 2015.
12. Pengfei Wang, JiafengGuo, YanyanLan, Jun Xu and Xueqi Cheng. Learning Hierarchical Representation Model for Next Basket Recommendation, SIGIR 2015.
13. Ji Wan, Pengcheng Wu, Steven C. H. Hoi, Peilin Zhao, Xingyu Gao, Dayong Wang, Yongdong Zhang, Jintao Li, "Online Learning to Rank for Content-Based Image Retrieval", IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, Pages: 2284-2290.
14. Wu Liu, Tao Mei, Yongdong Zhang, Cherry Che, Jiebo Luo, "Multi-task deep visual-semantic embedding for video thumbnail selection", CVPR 2015, Boston, Massachusetts, United States, June 8-10, 2015, Pages: 3707-3715.
15. Lei Zhang, Yongdong Zhang, Richang Hong, Qi Tian, "Full-Space Local Topology Extraction for Cross-Modal Retrieval", IEEE Trans. Image Processing, 2015.

CCF B 类:

16. 周建二, Demystifying and Mitigating TCP Stalls at the Server Side, CoNEXT 2015.
17. 周 安 福, Signpost: Scalable MU-MIMO Signaling with Zero CSI Feedback, Mobihoc The ACM international Symposium on mobile Ad Hoc Networking and Computing.
18. Ying Wang, Yinhe Han, Cheng Wang, Huawei Li, Xiaowei Li, "RADAR: A Case for Retention-Aware DRAM Assembly and Repair in Future FGR DRAM Memory", Proc. IEEE/ACM Design, Automation Conference (DAC), 2015.
19. Ying Wang, Yinhe Han, Lei Zhang, Huawei Li, Xiaowei Li, "ProPRAM: Exploiting the Transparent Logic Resources in Non-Volatile Memory for Near Data Processing", Proc. IEEE/ACM Design, Automation Conference (DAC) 2015.



-
20. Jiachao Deng, Yuntan Fang, Zidong Du, Ying Wang, Huawei Li, Olivier Temam, Paolo Ienne, David Novo, Xiaowei Li, Yunji Chen, Chengyong Wu, "Retraining based Timing Error Mitigation for Hardware Neural Networks," Proc. Design Automation and Test in Europe (DATE) 2015.
-
21. Mengran Fan, Haipeng Jia, Yunquan Zhang, Xiaojing An, and Ting Cao. Optimizing image sharpening algorithm on gpu. 44th International Conference on Parallel Processing (ICPP) 2015.
-
22. Yongqing Wang, Huawei Shen, Shenghua Liu, Xueqi Cheng. Learning user-specific latent influence and susceptibility from information cascades. Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI 2015).
-
23. Xinke Chen, Guangfei Zhang, Huandong Wang, Ruiyang Wu, Peng Wu, Longbing Zhang, MRP: Mix Real Cores and Pseudo Cores for FPGA based Chip-multiprocessor Simulation, Proc. Of Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015.
-
24. 姚二林, 谭光明, Bit Flipping Errors in High Performance Linpack at Exascale and Beyond, International Conference on Parallel Processing, 2015.
-
25. 闫洁, 谭光明, 孙凝晖, Study on Partitioning Real-world Directed Graphs of Skewed Degree Distribution, International Conference on Parallel Processing, 2015.
-
26. 谭光明, 张春明, 王文迪, 张佩珩, SuperDragon: A Heterogeneous Parallel System for Accelerating 3D Reconstruction of Cryo-Electron Microscopy Images, ACM Transactions on Reconfigurable Technology and Systems, 2015.
-
27. Zhiguo Liu; Ziyuan Zhu; Jinglin Shi; Jinbao Liu; Shiqiang Li, "A low power buffer-aided vector register file for LTE baseband signal processing," in Computer Design (ICCD) 2015.
-
28. Tong Man, Huawei Shen, Junming Huang, Xueqi Cheng. Context-adaptive matrix factorization for multi-context recommendation. Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM 2015), October 2015, Melbourne, Australia.
-
29. Xiaochun Zhang, Qi Guo, Yunji Chen, Weiwu Hu, HERMES: A Fast Cross-ISA Binary Translator with Post-Optimization, Proc. Of 13th IEEE/ACM International Symposium on Code Generation and Optimization (GGO), 2015, San Francisco, USA, February 07-11, 2015, pp:246-256.
-

CCF C 类:

- 30.侯 陈 达, EasiPCC: Popularity-aware Collaborative Caching for Web Requests in Low-Duty-Cycle Sensor Networks, ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2015.
- 31.谢 开 斌, Low Cost IoT Software Development---Ingredient Transformation and Interconnection, IEEE International Conference on Parallel and Distributed Systems, 2015.
- 32.李 勳, EasiCrawl: A Sleep-aware Schedule Method for Crawling IoT Sensors, IEEE International Conference on Parallel and Distributed Systems, 2015.
- 33.骆裕龙, 谭光明, 莫则尧, 孙凝晖, FAST: A Fast Stencil Auto tuning Framework Based On An Optimal-solution Space Model, International Conference on Supercomputing, 2015.
- 34.赵 喜 全, Implementation of short read mapping algorithm in OpenCL on the Xeon Phi coprocessor, High Performance Computing and Communications, 2015.
- 35.王元戎, 李强强, 谭光明, Application Taxonomy via Algorithmic Commonality for Domain-specific Architecture Design, IEEE International Conference on High Performance Computing, 2015.
- 36.Hongwei Wang; Ziyuan Zhu; Jinglin Shi; Yongtao Su, "An accurate ACOSSO metamodeling technique for processor architecture design space exploration," in Design Automation Conference (ASP-DAC), 2015.
- 37.Jiangnan Lin; Kai Xie; Yongtao Su; Yiqing Zhou; Yanbin Yao; Jinglin Shi, "Parallelized generation of ZC/ZC-DFT sequences in vector DSP," in Wireless Communications and Networking Conference (WCNC), 2015.
- 38.Ling Liu; Garcia, V.; Lin Tian; Zhengang Pan; Jinglin Shi, "Joint clustering and inter-cell resource allocation for CoMP in ultra dense cellular networks," in Communications (ICC), 2015.
- 39.GuoweiZhai; Lin Tian; Yiqing Zhou; Xiaodong Wang; Jinglin Shi, "Real-time guaranteed TDD protocol processing for centralized super base station architecture," in Wireless Communications and Networking Conference (WCNC), 2015.

40. Hang Liu; Yiqing Zhou; Lin Tian; Jinglin Shi, "How can vehicular communication reduce the rear-end collision probability on highway," in Global Telecommunications Conference (GLOBECOM 2015).

41. Chang Liu; Yi Wan; Lin Tian; Yiqing Zhou; Jinglin Shi, "Base station sleeping control with energy-stability tradeoff in centralized radio access networks," in Global Telecommunications Conference (GLOBECOM 2015).

42. 韩 锐, "SARP: Producing Approximate Results with Small Correctness Losses for Cloud Interactive Services," ACM International Conference on Computing Frontiers 2015 (CF 2015).

其它:

43. 刘小丽, 曹政, 臧大伟, 安仲奇, 吴冬冬, 李强, 王展, 安学军, 孙凝晖, 面向移动化部署的高性能计算机设计, 全国高性能计算学术年会, 2015 年

44. 邵恩, 元国军, 邹志轩, 曹政, 面向大规模计算集群的多轨分割网络, 全国高性能计算学术年会, 2015 年

45. 臧大伟, 曹政, 刘小丽, 孙凝晖, 层次化能耗优化的数据中心光电混合网络架构, 全国高性能计算学术年会, 2015 年

46. 肖俊敏, 代俊凯, 刘欢, 尤海航, 一种求解对称正定方程的混沌异步并行迭代算法, 全国高性能计算学术年会, 2015 年

47. 郑华杰, 魏征, 邢晶, 霍志刚, 张佩珩, 面向 DNA 序列数据的压缩技术研究, 全国高性能计算学术年会, 2015 年

48. 王迎瑞, 任江勇, 田荣, 异构系统的高效协同: 基于 SPH 方法的一般性讨论, 全国高性能计算学术年会, 2015 年

49. 李文明, 面向高通量应用的高密度片上网络实现机制, 全国高性能计算学术年会, 2015 年

50. Yanbin Yao; Yongtao Su; Jinglin Shi; Jiangnan Lin, "A low-complexity soft QAM demapper based on first-order linear approximation," in Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on, pp.446-450, Aug. 30 2015-Sept. 2 2015.

期刊论文：

CCF A 类：

1. Tianshi Chen, Shijin Zhang, Shaoli Liu, Zidong Du, Tao Luo, Yuan Gao, Junjie Liu, Dongsheng Wang, Chengyong Wu, Ninghui Sun, Yunji Chen, Olivier Temam: A Small-Footprint Accelerator for Large-Scale Neural Networks. TOCS. 33(2): 6 (2015) (大陆首篇) .
2. Robin WentaoOuyang, Lance Kaplan, Alice Toniolo, Mani Srivastava, and Timothy J. Norman Parallel and streaming truth discovery in large-scale quantitative crowdsourcing, IEEE Transactions on Parallel and Distributed Systems (TPDS).
3. 谭光明, 张春明, 汤文, 孙凝晖, Accelerating Irregular Computation in Massive Short Reads Mapping on FPGA Co-processor, IEEE Transactions on Parallel and Distributed Systems, 2015.
4. 谭光明, 闫洁, Graphine: Programming Graph-Parallel Computation of Large Natural Graphs on Multicore Cluster, IEEE Transactions on Parallel and Distributed Systems, 2015.
5. Guihai Yan, Faqiang Sun, Huawei Li, Xiaowei Li, "CoreRank: Redeeming Imperfect Silicon by Dynamically Quantifying Core-level Healthy Condition of Manycore Processors", TC, 2015.
6. Jun Ma, Guihai Yan, Yinhe Han, Xiaowei Li, An Analytical Framework for Estimating Scale-out and Scale-up Power Efficiency of Heterogeneous Manycores, TC, 03 April 2015.
7. Tianshi Chen, Qi Guo, Olivier Temam, Yue Wu, Yungang Bao, Zhiwei Xu, Yunji Chen: Statistical Performance Comparisons of Computers. TC, 64(5): 1442-1455 (2015).
8. Lei Liu, Yong Li (VMware, US), Chen Ding (Rochester, US), Hao Yang, Chengyong Wu, "Rethinking Memory Management in Modern Operating System: Horizontal, Vertical or Random?", TC, 2015.
9. Chao Wang, Xi Li, Junneng Zhang, Peng Chen, Yunji Chen, Xuehai Zhou, Ray C. C. Cheung: Architecture Support for Task Out-of-Order Execution in MPSoCs. TC, 64(5): 1296-1310 (2015).

-
10. Jiacheng Zhao, Huimin Cui, Jingling Xue, Xiaobing Feng, "Predicting Cross-Core Performance Interference on Multicore Processors with Regression Analysis", TPDS, no.99, pp.1-1. 2015(1).
-
11. Jun Ma, Guihai Yan, Yinhe Han, Xiaowei Li, EcoUp: Towards Economical Datacenter Upgrading, TPDS, 2015.
-
12. Shaoli Liu, Tianshi Chen, Ling Li, Xi Li, Mingzhe Zhang, Chao Wang, Haibo Meng, Xuehai Zhou, Yunji Chen: FreeRider: Non-Local Adaptive Network-on-Chip Routing with Packet-Carried Propagation of Congestion Information. TPDS, 26(8): 2272-2285 (2015).
-
13. Yanfeng Zhang, Shimin Chen, Qiang Wang, Ge Yu, "i2MapReduce: Incremental MapReduce for Mining Evolving Big Data". IEEE Transactions on Knowledge and Data Engineering (TKDE), Volume 27, Number 7, pp.1906-1919, 2015.
-
14. Manos Athanassoulis, Shimin Chen, Anastasia Ailamaki, Phillip B. Gibbons, Radu Stoica, "Online Updates on Data Warehouses via Judicious Use of Solid-State Storage". ACM Transaction on Database Systems (TODS), Volume 40, Number 1, March 2015.
-
15. Jian Wang, Huawei Li, Tao Lv, Tiancheng Wang, Xiaowei Li, and Sandip Kundu, "Abstraction-Guided Simulation Using Markov Analysis for Functional Verification," TCAD, 2015.
-
16. Yanhong Zhou, Tiancheng Wang, Huawei Li, Tao Lv, Xiaowei Li, "Functional Test Generation for Hard-to-reach States Using Path Constraint Solving," TCAD, 2015.
-
17. Ying Wang, Lei Zhang, Yinhe Han, Huawei Li, Xiaowei Li, "Economizing TSV resources in 3D Network-on-Chip design", IEEE Transactions on Very Large Scaled Integration Systems (TVLSI), Vol. 23, No.3, 2015, pp. 493-506.
-
18. Ying Wang, Lei Zhang, Yinhe Han, Huawei Li, Xiaowei Li, "Data Remapping for Static NUCA in Degradable Chip Multiprocessors", IEEE Transactions on Very Large Scaled Integration Systems (TVLSI), Vol. 23, No.5, 2015, pp. 879-892.
-
19. Hang Lu, Bin Zhang, Fu, Ying Wang, Yinhe Han, Guihai Yan, Xiaowei Li, "RISO: Enforce Noninterfered Performance With Relaxed Network-on-Chip Isolation in Many-Core Cloud Processors", IEEE Transactions on Very Large Scaled Integration Systems (TVLSI), Vol. 23, No.12, pp.3053-3064.
-

20. Ying Wang, Yinhe Han, Huawei Li, Xiaowei Li, “VANUCA: Enabling Near-Threshold Voltage Operation in Large-Capacity Cache” , TVLSI, 2015.

21. Ying Wang, Yinhe Han, Huawei Li, Xiaowei Li, “PSI Conscious Write Scheduling: Architectural Support for Reliable Power Delivery in 3D Die-Stacked PCM” , TVLSI, 2015.

22. ShiGang Li, ChangJun Hu, JunChao Zhang, and YunQuan Zhang. Automatic tuning of sparse matrix-vector multiplication on multicore clusters. Science China Information Sciences, 58(9):1 – 14, 2015.

23. Shenghua Liu, Xueqi Cheng, Fuxin Li, Fangtao Li. TASC: Topic-Adaptive Sentiment Classification on Dynamic Tweets. IEEE Transactions on Knowledge and Data Engineering, 27(6):1696-1709, 2015.

24. Zheng Lin, Xiaolong Jin, Xueke Xu, Yuanzhuo Wang, Xueqi Cheng, Weiping Wang, Dan Meng. An Unsupervised Cross-Lingual Topic Model Framework for Sentiment Analysis. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015

25. Robin Wentao Ouyang, Lance Kaplan, Alice Toniolo, Mani Srivastava, and Timothy J. Norman. Truth, ability, and perception: aggregating crowd sourced quantitative claims. IEEE Transactions on Knowledge and Data Engineering (TKDE)

26. Shihong Xia, Congyi Wang, Jinxiang Chai, Jessica Hodgins. Real-time Style Transfer for Unlabeled Heterogeneous Human Motion. ACM SIGGRAPH 2015/ACM Transactions on Graphics 2015. Volume 34, Issue 4, Pages, 119:1-119:10

CCF B 类：

27. 李振宇, Video Delivery Performance of a Large-Scale VoD System and the Implications on Content Delivery, IEEE Transactions on Multimedia, 17, 6, 2015, 880-892.

28. Kun Zhang, Yan Fu, Wen-Feng Zeng, Kun He, Hao Chi, Chao Liu, Yan-Chang Li, Yuan Gao, Ping Xu, Si-Min He, A Note on the False Discovery Rate of Novel Peptides in Proteogenomics, Bioinformatics, 31(20), 2015, 3249-3253.



-
- 29.Jianhui Huang, Shengling Wang, Xiuzhen Cheng, Jingping Bi. Big-Data Routing in D2D Communications with Cognitive Radio Capability, IEEE Wireless communications. (国际顶级杂志) (SCI 一区) (IF: 5.381) (SCI/EI index).
-

CCF C 类

-
- 30.[1] 周 安 福, Cross-layer design with optimal dynamic gateway selection for wireless mesh networks, Computer Communications, 2015, 69 – 79.
-

其它：

-
- 31.王 子 健, Monitoring Algal Blooms using Active Learning Camera Sensor Networks, International Journal of Sensor Networks, 19, 2, 2015, 91-103, 非 ccf 认定类型
-
- 32.王 子 健, Data-Driven Soft Sensor Modeling for Algal Blooms Monitoring, IEEE SENSORS JOURNAL, 15, 1, 2015, 579-590, 非 ccf 认定类型
-
- 33.陈 浩, EasiDSIT: A Two-layer Data Association Method for Multitarget Tracking in Wireless Sensor Networks, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, 62, 1, 2015, 434-443, 非 ccf 认定类型
-
- 34.胡伟武, 靳国杰, 汪文祥, 张晓春, 王焕东, 龙芯指令系统融合技术, 2015, 45 (4) , 459-479
-
- 35.吴瑞阳, 汪文祥, 王焕东, 胡伟武, 龙芯 GS464E 处理器核架构设计, 中国科学: 信息科学, 2015, 45 (4): 480-500
-
- 36.Jianhui Huang, Xiuzhen Cheng, Jingping Bi, Biao Chen. Wireless Relay Selection in Pocket Switched Networks Based on Spatial Regularity of Human Mobility, Sensors, 2016, 16(1), 94. (IF: 2.245) (SCI/EI index)
-
- 37.Fei Wang, Yongjun Xu, Hanwen Zhang, Yujun Zhang and Liehuang Zhu, 2FLIP: A Two-Factor Lightweight Privacy Preserving Authentication Scheme for VANET, IEEE Transactions on Vehicular Technology, vol.99, pp.1-18, 2015. (SCI, IF:1.978)
-

38. Chao Li, YongJunXu, ChaoNongXu, ZhuLin An, BoYuDiao, XiaoWeiLi, "DTMAC: A Delay Tolerant MAC Protocol for Underwater Wireless Sensor Networks". IEEE sensors journal, vol.99, pp.1-11, 2015. (SCI, IF:1.768)
39. BoYuDiao, YongjunXu, ZhuLin An, Fei Wang, Chao Li, Improving Both Energy and Time Efficiency of DepthBased Routing for Underwater, INTERNATIONAL JOURNAL OF DISTRIBUTE Sensor Networks
40. 张秀霞, 刘闯, 谭光明, A Reliable Distributed Convolutional Neural Network for Biology Image Segmentation, Cluster, Cloud and Grid Computing, 印度, 15 届, 2015 年, 777-780
41. 谭光明, 姚二林, Detection of soft errors in LU decomposition with partialpivotingusing algorithm-based fault tolerance, International Journal of High Performance Computing Applications, 2015 年, 422-436
42. 苏勇, 王展, 樊志国, 曹政, HyperFatTree: A Large-scale Tree-based Network with Low-radix Switches, International Journal of Parallel Programming, 30 期, 2015 年, 1-13
43. 臧大伟, 曹政, 王展, 刘小丽, 孙凝晖, 基于 AWGR 的 OCS / EPS 数据中心光电混合网络, 计算机学报, 30 期, 1-13, 2015 年
44. 王展, 曹政, 刘小丽, 苏勇, 李强, 孙凝晖, 基于单根 I/O 虚拟化的多根 I/O 资源池化方法, 计算机研究与发展, 1 期, 52 卷, 2015 年, 83-93
45. 臧大伟, 曹政, 刘小丽, PROP: Using PCIe-based RDMA to Accelerate Rack-Scale Communications in Data Centers, IEEE International Conference on Parallel and Distributed Systems, 2015 年
46. Roberto Proietti, 曹政, A Scalable, Low-latency, High-throughput, Flat Optical Interconnect Architecture based on Arrayed Waveguide Grating Routers, IEEE/OSA Journal of Lightwave Technology (JLT), 4 期, 33 卷, 2015 年
47. 刘炳涛, 王达, 一种缓存数据流信息的处理器前端设计, 计算机研究与发展, 56 卷, 1 期, 1-16
48. 朱亚涛, 王达, E O F D M: 一种面向众核架构的最低能耗搜索方法, 计算机研究与发展, 56 卷, 1 期, 1303-1315



-
- 49.张洋, 王达, 众核片上网络的层次化全局自适应路由机制, 计算机研究与发展, 51 卷, 1 卷, 1-16
-
- 50.Shan Lu Sheng-Bo Fan Bing Yang, Yu-Xin Li, Jia-Ming Meng, Long Wu, Pin Li, Kun Zhang, Mei-Jun Zhang, Yan Fu, Jincailuo, Rui-Xiang Sun, Si-Min He &Meng-Qiu Dong, Mapping native disulfide bonds at a proteome scale, *Nature Methods*,12(4), 2015, 329-331.
-
- 51.Hao Chi, Kun He, Bing Yang, Zhen Chen, Rui-Xiang Sun, Sheng-Bo Fan, Kun Zhang, Chao Liu, Zuo-Fei Yuan, Quan-Hui Wang, Si-Qi Liu, Meng-Qiu Dong, Si-Min He, pFind – Alioth: A novel unrestricted database search algorithm to improve the interpretation of high-resolution MS/MS data, *Journal of Proteomics*, 125, 2015, 89-97
-
- 52.Renmin Han, LiansanWang, Zhiyong Liu, Fei Sun, Fa Zhang. A novel fully automatic scheme for fiducial marker-based alignment in electron tomography. *Journal of Structural Biology*, 2015, 192(3):403-417
-
- 53.Ziying Zhou, Yugang Li, Fa Zhang, Xiaohua Wan. FASART: An iterative reconstruction algorithm with inter-iteration adaptive NAD filter.*Bio-Medical Materials and Engineering*. 26(s1), S1409-S1415, 2015.8
-
- 54.Fa Zhang, Jingrong Zhang, Albert Lawrence, FeiRen, Xuan Wang, Zhiyong Liu, Xiaohua Wan. BSIRT: A Block-Iterative SIRT Parallel Algorithm Using Curvilinear Projection Model. *IEEE Transactions on Nanobioscience*.*IEEE Transactions on Nanobioscience*. 4(2), 229-236, 2015.3
-
- 55.Fa Zhang, Yu Chen, FeiRen, Xuan Wang, Zhiyong Liu and Xiaohua Wan. A Two-phase Improved Correlation Method for Automatic Particle Selection in Cryo-EM. *IEEE-ACM Transactions on Computational Biology and Bioinformatics*. 2015.3.
-
- 56.Hongwei Wang; Ziyuan Zhu; Jinglin Shi; Yongtao Su, "Sensitivity Analysis Based Predictive Modeling for MPSoC Performance and Energy Estimation," in *VLSI Design (VLSID)*, 2015 28th International Conference on , pp.511-516, 3-7 Jan. 2015
-
- 57.Qian Sun; Lin Tian; Yiqing Zhou; Jinglin Shi; Xiaodong Wang, "Energy efficient incentive resource allocation in D2D cooperative communications," in *Communications (ICC)*, 2015 IEEE International Conference on , pp.2632-2637, 8-12 June 2015
-
- 58.Ling Liu; Yiqing Zhou; Lin Tian; Jinglin Shi, "CPC-based backward-compatible network access for LTE cognitive radio cellular networks," in *Communications Magazine*, *IEEE*, vol.53, no.7, pp.93-99, July 2015
-

- 59.ManliQian; Hardjawana, W.; Jinglin Shi; Vucetic, B., "Baseband Processing Units Virtualization for Cloud Radio Access Networks," in Wireless Communications Letters, IEEE , vol.4, no.2, pp.189-192, April 2015
- 60.ManliQian; Yuanyuan Wang; Yiqing Zhou; Lin Tian; Jinglin Shi, " A Super Base Station based Centralized Network Architecture for 5G Mobile Communication Systems," in Digital Communications and Networks,2015
- 61.MingjinGao.; He (Henry) Chen.; Yonghui Li.; zhouyiqing.; Jinglin Shi, " Switch-Delay Aware Computing Resource Allocation in Virtualized Base," in China Communication, 2015
- 62.Yao Yuao; Dalin Zhang; Qing Wang; Jinglin Shi, " Research on the extension of SCTP protocol on the heterogeneous wireless network," in Journal of Communication, 2015
- 63.Yao Yuao; Dalin Zhang; Qing Wang; JinglinShi,"A Multi-flow Network Fairness Selection Scheme Based on Weighted Bigraph Model," in Telecommunications and Networking ,2015
- 64.杜红艳, 周一青, 田霖, 一种基于模糊控制的负载均衡决策机制, 高技术通讯, 已录用, 2015
- 65.孙茜, 田霖, 周一青, 石晶林, 基于 NFV 与 SDN 的未来接入网虚拟化关键技术, 电信通信技, 邀稿, 2015

科研为国分忧

创新与民造福

学位论文

2015 年计算所有 177 人取得硕士学位、81 人取得博士学位。2009 级博士生崔振完成的题为《稀疏与流形表示及其在人脸识别中的应用研究》博士学位论文荣获 2015 年中科院优秀博士学位论文奖。2015 年计算所博士毕业生 53% 进入学术界，其中 33% 毕业生进入科研单位工作。硕士毕业生进入企业工作的比例为 78%，其中去国内企业和外资企业工作的比例为 2.9: 1。

优博论文摘要



作者姓名：崔振

论文题目：稀疏与流形表示及其在人脸识别中的应用研究

作者简介：崔振，男，1981 年 4 月出生，2009 年 9 月师从中国科学院计算技术研究所陈熙霖研究员，于 2014 年 7 月获博士学位。

中文摘要

随着信息电子技术的快速发展，人们需要处理日益增长的多媒体资源数据，例如海量的图片、视频、文本以及声音数据等。如何对这些数据进

行分析和处理，以便从中提取用户所需要的有价值的信息，逐渐成为模式识别、机器学习以及计算机视觉等领域的研究者们所共同关注的问题。不同于传统的少量样本的处理任务，目前的数据处理更多的涉及到同一主题含义下大量样本的学

习, 即从孤立的单张 / 少量样本到数据集合作为整体的处理, 例如对监控摄像机下所拍摄的图像集 (视频片段可以看作它的特例) 进行身份鉴别、场景分析或行为识别等任务。而对图像集数据的信息挖掘和利用涉及到两个基本问题: 1) 如何建模图像集, 以便从图像集中提取有效的特征表示; 2) 如何建模图像集之间的关系, 并设计合理有效的距离度量方法。

本文正是围绕上述这两个基本问题, 从视频人脸识别这一特定任务出发, 一方面利用图像集自身的冗余性、相关性以及自相似性等特点, 研究利用稀疏表示来建模图像集的方法; 另一方面, 利用图像集数据的稠密性和非线性等特点, 从流形的角度探讨和研究图像集之间的关系, 以设计图像集之间的鲁棒距离度量方法。

本文的主要研究工作具体包括:

(1) 提出了一个统一描述静态图像和视频序列的稀疏编码局部描述子框架, 并进一步提出了监督情况下的多尺度度量学习方法。该局部描述子通过统计建模人脸图像 / 视频序列在空间 / 时空体区域上的特征表示, 大幅度缓解了由于人脸图像欠配准问题所引起的识别性能衰退现象。同时通过对原始特征的过完备稀疏表示增强了其相应特征的描述能力。在该框架下, 针对人脸图像包含多个区域描述子的问题, 提出了联合学习多个描述子的马氏距离方法, 该方法整合了人脸图像的全局性信息。在真实环境下的人脸图像数据库 LFW 和人脸视频数据库 YTF 上的实验表明了它们的有效性。

(2) 提出了一种基于稀疏自编码机的局部描述子, 并进一步提出了针对人脸确认任务的自编码预测机方法。基于上述 (1) 中的局部描述子框架, 该方法利用自编码器来前向推理稀疏表

示的近似解, 以加快特征提取的速度。在逐对样本分类的任务中, 自编码预测机尝试学习训练集中成对样本之间的非线性关系。对于任意给定的两个测试样本, 利用该预测机对每个样本的预测值来决定这对样本之间的距离。该方法在 LFW 和 YTF 数据库上严格协议下分别达到了 90.17% 和 80.10% 的识别率。

(3) 提出了一种图像集联合稀疏表示的模型, 将稀疏表示和结构化稀疏表示纳入到一个框架下。该模型从两个层次上建模图像集的稀疏表示: 第一个是采用 $L_{2,1}$ 混合范数的组层次稀疏性, 它假设每一个图像集能用少量若干个注册图像集重构; 第二个是采用 L_1 范数的原子层次稀疏性, 其目标是对每一张图像选取表现相关的原子进行重构。该方法不仅利用了测试图像集的整体性还有效地利用了注册图像集 (或训练图像集) 的分组结构特性。在三个公开的视频人脸数据库上的实验验证了该方法超过了当前已有的视频人脸识别方法。

(4) 提出了一种建模数据集之间关系的无监督流形对齐学习模型。该模型形式化流形对齐为包含流形表现特征匹配、几何结构一致性匹配以及结构保持三个特性的混合整数规划模型。这个模型具有三个特点: i) 完全无监督的流形匹配, 不需要预先指定任何对应性; ii) 同时处理流形结构的发现和对齐; iii) 能使用一个迭代投影算法来发现两流形间的对应点, 且不需要考虑所有的置换情况。本文所提出的迭代投影算法通过沿着增广的凸目标函数的解的路径来发现原目标函数的近似解, 并且在理论上证明了该算法是收敛的。在数据集匹配、人脸识别和域迁移学习等不同任务中的实验表明了该方法的有效性以及应用于实际问题的可行性。



(5) 提出了一种局部仿射凸壳约束的自适应流形距离计算方法。该方法借助于构建非邻接关系的权重矩阵作为正则项约束,进而转化流形距离为流形上局部仿射凸壳之间的距离问题,避免了先前方法在计算流形距离时需要预先划分流形为若干个线性子模型的问题以及由此引起的流形距离计算偏差问题。从建模子空间距离到仿射凸壳最近点的距离的角度,该模型将点、子空间、流形纳入到一个统一的流形距离框架下。在两个挑战性的视频人脸数据库 YTF 和 YTC 上的实验

验证了该方法相比于已有的图像集距离计算方法是更具有竞争力的。

综上所述,本文针对稀疏表示和流形学习理论及其应用开展了广泛而深入的研究。针对同一主题含义下的大量多帧图像这一问题,本文发展和完善了稀疏和流形对图像集表示和建模的理论,提出了一系列基于图像集的建模和分类方法,并将其应用于非可控条件下的人脸识别任务。本文所提出的图像集建模方法不仅适用于人脸识别问题,还可以推广应用于普遍的模式分类问题。

一流人才

一流管理

一流成果

一流效益



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

地址：北京市海淀区中关村
科学院南路 6 号
邮编：100190
电话：(010) 62601116
传真：(101) 62567724
<http://www.ict.ac.cn>