一场学术界和产业界围绕“中国超算服务模式”的大讨论——

中国超算服务，谁主沉浮？

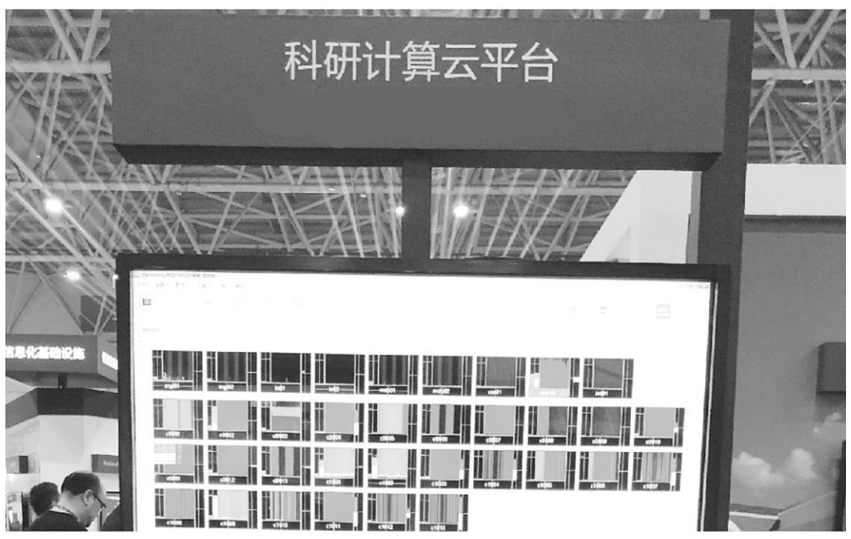
■本报记者 赵广立

最近，超算领域学术界和产业界人士围绕“中国超算服务创新”展开了一场热烈的讨论。

起因是，北京并行科技股份有限公司（以下简称并行科技）总经理陈健从超算建设投入产出比的角度提出，虽然中国超算近年来取得了瞩目的成就，但超算服务的模式还很落后，特别是一些自建院系、单位级和用户组级超算集群（统称为中小微超算），投入产出比远不及大型超算，且年平均利用率低，这种“小农经济式”超算服务形态，急需改进、创新。

陈健提出的解决方案是“超算上云”。作为超算云服务提供商，陈健认为，超算的使用者大都要经历“国内外大超算平台排队上机——自建超算集群——租用超算——超算云服务”的过程，这是一个“从落后的生产力逐渐过渡到更先进的生产力形式”。

以此为发端，超算领域学术界、产业界人士从各个角度给出了他们的理解和思考。



如何为科研提供更好的超算服务，是中国超算的一大命题。

超算“全盘云化”有其局限

之所以以“小农经济”类比，在陈健看来，自建超算集群就好比自种“两亩自留地”，这些自建的中小微超算有些浪费，不如大家都把需求集中起来，把无数的“两亩地”汇集成“万亩良田”，大家通过超算云服务的形式，从大型超算上按需购买计算资源，不但可减少浪费，还能提高应用效率。

相应的，陈健把并行科技比作“制造大型农机具的”——只有大型超算（万亩良田）多了，“现代农机具厂商”才好开张。

但是，超算全部上云（指服务模式，并非云计算），现实吗？

虚拟化和云计算专家麻清刚提出，从技术上来讲，限制超算“上云”的条件是网络带宽。“如果可以以极低成本拉一条400G的网络专线，超算还真有可能全盘云化——但不一定都是公有云，私有云、企业云、行业云等形态都可能存在。”

但是，带宽问题目前很难逾越。“带宽费用是阻碍我们去内蒙古建超算的最主要原因，节约的电费都给运营商买带宽去了。”北京大学高性能计算平台主任工程师、计算中心系统管理室主任樊春这样说。

内蒙古电价为每千瓦时0.26元。近年来，西部地区利用“超低电费”的条件大力投建IT基础设施，这对于一般的云计算中心来说是个利好，但对于超算而言，并不一定可行。

樊春说，对于大部分超算用户来说，速度和价格是排在前两位的。超算中心对外提供计算服务，只有同时满足这两点，用户才愿意买单。

且不说“拉一条400G的网络专线”是个很大的前提，仅超算服务“全盘云化”这一点，中科院计算所研究员、国家超级计算济南中心主任张云泉就不太认同。

他从超算需求的角度出发，认为目前超算云服务只能支持一些中低端计算需求，那些更快、更大规模的需求，超算云服务的形式还提供不了，“就像当年网格计算声称可全面取代超算一样，超算服务全盘云化几乎不可能”。

超算成本下降还需规模效益

樊春还提出一个问题：“超算云技术管理上无法降低成本，从而无法降低对最终用户的价格，这是超算云发展的最大障碍。”

这一观点与陈健提出的观点相悖。陈健此前在接受《中国科学报》采访时称，超算云服务商的角色，就是通过研发软件和方案，以技术输出提高超算集群的易用性和有效利用率，从而降低超算的使用成本——把超算的每核时成本降低至一毛钱。

二者的矛盾点在哪？

一个答案是：用户规模。只有用户规模上去了，两人的观点才能找到平衡点。

一方面，陈健认为樊春所在的北京大学校级高性能计算平台，已经有了足够大的规模应用，规模效益已现，所以北大能够给予用户更低的价格；另一方面，陈健也深知城域网光纤非常贵——他们就投建了从广州超算到北京的城域网光纤专线，之所以这么做，是因为并行科技也实现了规模效益，通过这条光纤专线，并行科技有每年数千万元的超算云服务经营额。

“这（超大带宽网络光纤）是一个极具规模效益的基础设施投入。”陈健感叹。

那么，现在中国超算应用是否到了产生规模效益的节点？

陈健认为“是”。他给出了一组数据：并行科技依托各大超算中心资源提供的超算云服务，现在已经有近1.5万个客户，并行科技给用户提供的超算价格，经核算比自建超算低一半。此外，2018年并行科技超算云服务合同额已达2亿元，预计2019年会达到3亿元。

“这是无数科研工作者用脚投票的结果。”陈健说，长此以往，中小微超算集群建设的越少，中国超大型国家级超算、大型地方超算、校级超算乃至公有云超算就发展得越好。

没谁能“一统江湖”

以云服务的方式提供超算资源，这块市场上并不只有并行科技。阿里云等云计算巨头，也“顺手”做了超算的云服务；一些大型超算中心，也探索了超算云服务的模式，国家超级计算天津中心就是一例。

“我们做超算云，是为了让用户用起来更好用。”国家超级计算天津中心一位要求匿名的专家说，“不管是超算云还是云超算，本质是为用户服务的形式。服务形式确实可以降低用户使用门槛，让用户用起来更简单，但不应过分夸大使用模式（对中国超算实力提高）的作用。”

他认为，并非只有超算云才能解决现在的超算服务问题。

阿里云高性能计算负责人、资深技术专家何万青也持类似观点。他说，超算是个大生态系统，中国是个超算大国，必然有各种模式的超算集群——有国家投资建设的，有诸如并行科技以SaaS和PaaS形式提供增值服务的，也有诸如阿里云等云计算公司提供的商用系统，“只要用户认为他出的价格符合自己所需的价值，就是合理的”。

至于超算服务的价格，何万青认为：“价格永远是市场行为，甲之蜜糖，乙之砒霜。有的用户愿意出高价购买，其他用户也许弃之若敝履，这就是市场。”

何万青举例说，一些科研型和渲染型客户，他只在灵感或者创作来了之后，才会启动计算的点对点模式，这时候资源灵活调度和自动伸缩比线下性能优化更重要——而这其实是云计算比较擅长的领域。

“举例是想说谁也别想一统江湖。大家各有所长，客户、技术需求、业务特点和生态发展也是多样性的。”何万青说。

超算应该部分免费吗

在超算平台建设方面，陈健提出，事实上目前我国的国家级超算一直是不核算建设成本的，对重大应用采取免费或者低价支持的措施；一些高校超算，一般采取免费或低价使用手段，“其实都是亏本在服务校内用户”。

对此，在超算领域从业多年的中科京云总经理贺建海提出一个大胆的想法，“我认为国家级超算应该向基础研究提供免费计算资源。”

曾在三家国家级超算中心工作、现任职于鹏城实验室人工智能研究中心的王丙强认为，该出发点虽好，但在实际操作中，可能会带来混乱：谁会对免费的东西不动心？

但他也提出，有关科技主管机构，确实应该考虑降低使用计算资源的成本和门槛。

张云泉也觉得，免费难免浪费，但面向基础研究领域，超算可以尝试提供尽量低的价格，降低获取成本。

樊春结合他们在北大高性能计算平台提供超算服务的经验提出，面向那些小规模短时长、纳入教学任务的超算使用需求，可以且应该免费，以降低学生学习的门槛；而面向基础科研，比较耗费资源的高性能计算任务是不能免费的，但是应该以极低的价格提供给科学工作者使用。

“收费是为了防止滥用资源，不是为了赚钱。”樊春说，收费的价格要让“挖矿”等无利可图。

而低价格是为了降低用户使用成本，符合“价格低—用户大量使用—计算量大—出更优秀结果的概率增加—超算为人类的贡献也更大”的逻辑。

樊春还提出，如果高性能计算免费的话，就需要一个跨学科的专家委员会来分配资源。不过他也认为，分配资源工作是非常耗费资源和精力的，协调的成本要“远远高于硬件的浪费”。

一位高能物理应用领域的匿名用户认为，国内许多公共服务平台类的大科学装置采用的模式是，国家财政支持装置的运行经费，除企业用户外，装置不得向科研用户收费。用户提出机时申请，用户委员会对申请进行评审筛选。

“事实上科研用户的经费也是从国家财政来的，倒一道手没什么意义。”他表示，“这种只管建设不管运行的模式未必是好模式。”

一个问题是，如果相关主管部门对科研级超算应用实行免费或低价，超算行业还需要并行科技、阿里云超算这类企业吗？

樊春认为，还是需要的。

“这个行业虽然不如传统云计算获利空间大，但也的确有赚钱空间。”他建议，“超算云公司应该提升技术以及管理能力，努力提供更加便宜的资源、更优秀的服务吸引用户，拓展整个高性能计算行业。”

速递

科大讯飞成北京冬奥会智能语音独家供应商

本报讯 9月16日，在位于北京石景山区的北京冬奥组委会驻地，北京冬奥组委会与科大讯飞联合举办“沟通，因爱而能”发布会，宣布科大讯飞正式成为北京2022年冬奥会和冬残奥会官方自动语音转换与翻译独家供应商。

“冬奥会和冬残奥会的举办将给世界一个更深入了解中国的机会，无国界、无障碍沟通事关重大。”冬奥会奥组委技术部部长喻红在发布会上表示，相互的沟通和理解，才能更好地团结，“科大讯飞为此提供了新的解决方案”。

“安徽人许海峰拿到中国奥运会上的第一块金牌，成为民族的骄傲；如今来自安徽的人工智能企业科大讯飞将以先进的语音技术为奥运会添彩。”发布会上，科大讯飞董事长刘庆峰表示，成为北京2022年冬奥会和冬残奥会官方自动语音转换与翻译独家供应商，科大讯飞将以精益求精的智能语音及机器翻译技术为这两项赛事提供服务，助力各国友人在北京2022年冬奥会和冬残奥会期间“沟通无障碍”。

目前科大讯飞的语音转换与翻译技术已应用于消费者领域的多个方面，包括讯飞智能办公本、讯飞输入法、讯飞听见系统、讯飞翻译机等多款集语音识别与转化、翻译技术于一体的产品。今年5月，该公司还发布了讯飞翻译机3.0版本，在新增加的七大行业AI翻译场景中，“体育”位列其中。

值得一提的是，在今年8月底于上海举办的世界人工智能大会上，科大讯飞新一代语音翻译关键技术及系统刚刚摘得SAIL奖唯一一项“应用奖”，该奖项是人工智能领域国内最高规格的官方奖项，被誉为“人工智能领域奥斯卡”。 （赵广立）

2019年中国大数据整体规模超5300亿元

本报讯 近日，大数据产业生态联盟联合赛迪顾问共同编制的《2019中国大数据产业发展白皮书》指出，2019年中国大数据整体规模达到5386.2亿元，预计2020年达到6605.8亿元。大数据产业已经成为数字经济发展“新动能”。

大数据产业日新月异，行业进入应用“深水区”。中科曙光大数据首席科学家、大数据智能研究院院长宋怀明表示，随着大数据基础设施的不断完善，行业已接近完成数据的“原始积累”，需要打造“数字时代”的“智能引擎”才能持续挖掘数据价值。

宋怀明认为，新技术条件下，行业内一些发展趋势特征越加明显。突出表现在：庞大的数据量需要足够敏捷的处理能力；大数据预测和决策支持需要更加智能的大数据分析；大数据、AI、云计算的融合，将推动多学科融合；随着大数据应用越来越广，数据安全变得更加重要。

因此，更敏捷、更智能、更融合、更安全的数据分析和智能工具，将成为企业的主要需求。据介绍，作为中科院计算技术研究所的项目参与方，曙光大数据正在致力打造“1+4+N”业务布局。目前构建的开放城市云数据平台，将为城市管理提供智慧云脑；构建的人工智能公共服务平台，可以为区域的初创企业、科研院所和企业用户提供平台资源支撑。 （辛雨）

华为发布内置5G基带的手机芯片

本报讯 近日，华为消费者业务CEO余承东在2019德国柏林消费电子展（IFA）上，发布了华为新一代芯片麒麟990系列，其中包括麒麟990和麒麟990 5G两款芯片。

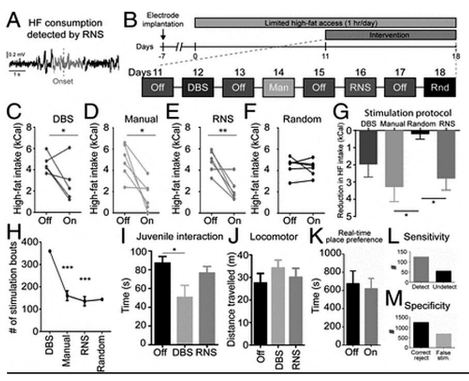
麒麟990是华为第三代AI芯片加持的手机处理器，也是第一代采用华为自研“达芬奇”架构的手机处理器。余承东介绍说，这款芯片将首次内置5G基带，这意味着5G终于走向了实用化。据悉，麒麟990芯片将于月底搭载于Mate30手机上，很快与消费者见面。

余承东称，麒麟990采用台积电第二代7纳米工艺打造，在AI计算方面首次采用了达芬奇架构的自研NPU，而且改用了“两大核一小核”的设计，AI算力是骁龙855近6倍；在图形计算方面，麒麟990的GPU核心数从10核增加到了16核。不过，性能与核心数的提升并没有让芯片体积变大，这一代麒麟990的板级面积相比业界其他方案要小36%。

此外，华为表示，之所以将这一代旗舰手机芯片分成两款推出，是为了兼顾到国外很多还没有发展5G服务的地区。 （赵鲁）

前沿扫描

『减肥脑芯片』电击击退食欲



小鼠每日1小时高脂肪（HF）暴露期间的干预期的示意图，以及不同电刺激方案对HF摄入的结果

想减肥，但又管不住嘴……近日，美国斯坦福大学的一项研究试图通过控制大脑阻截食物的诱惑——在大脑中植入芯片，想吃东西时，芯片就会适量放电，摧毁过度的食欲。

据英国《镜报》报道，六名病态肥胖患者自愿参加了这项临床试验，当他们想暴食时，这种芯片便会对他们的大脑产生刺激。

斯坦福大学Casey Halpern博士介绍，该芯片称为反应性神经刺激系统（RNS），在植入大脑后，它会连续记录分析大脑的活动。

这种芯片最初研制的目的是帮助治疗癫痫患者，它由医疗技术公司NeuroPace开发。该芯片一旦植入大脑，它就会追踪大脑活动，监控大脑运作方式，当芯片检测到癫痫发作前发生的某种模式时，芯片会给患者一个轻微的电击。

最近《美国国家科学院院刊》发表的一项研究表明，同样的技术也可以用来阻止暴食模式。

对小鼠开展的试验表明该芯片可阻止习惯性或不受控制的行为。研究人员记录了来自小鼠“伏隔核”（NAc）的局部场电位，发现1-4Hz振荡的功率可以有效触发神经刺激，从而减少小鼠进食的行为。

此外，在研究过程中，研究人员发现人类的NAc也出现了类似的振荡，因此他们认为该方法可以在人类身上尝试并起到控制一些疾病的作用。

该项研究的重点在于大脑的“快乐中枢”——伏隔核区域，它控制着与饥饿感、饱腹感以及身体新陈代谢有关的激素水平，并且与奖赏和成瘾的感觉紧密相关。

众多研究表明，当涉及大脑对食物的反应时，伏隔核经常被激活，增加与暴饮暴食或肥胖风险相关的因素，也就是说，当肥胖者看到高热量食物时，这一区域的活动会更加活跃。

研究人员表示，对老鼠的研究结果表明，RNS具备治疗顽固性行为障碍的潜力，例如饮食失调、肥胖、成瘾等，而这些行为障碍以前未被认为可以通过神经外科手术方法治疗。

目前，斯坦福大学的科研人员正在测试该方法是否也对例如暴食等“饮食失控”的人起作用。

该试验计划在接下来的五年内进行，RNS芯片将在六位志愿者大脑中植入18个月。

并非所有人都能参与该试验，身高超过170厘米、体重达到130公斤以上的成年人才可以参与这项试验。Casey Halpern表示：“这些患者基本上都会死于肥胖。”

实际上，其他试图治疗肥胖的临床试验也使用了类似的方法，该方法被称为深部脑刺激（DBS）。以治疗帕金森综合征最常使用的DBS为例，它会产生连续的电流，但是，RNS则只在检测到目标活动模式时才会发出电击。

值得关注的是，该研究还面临如何将大脑对高脂肪食物的反应和对健康食物的反应，以及日常其他事物带来的喜悦和快感区分开来的挑战。如果该芯片的刺激会消除大脑产生的所有愉悦和进食快感，是否会使患者失去对事物的兴趣，从而导致精神抑郁等新的问题，还有待进一步研究。

（田瑞颖）

相关论文信息：https://doi.org/10.1073/pnas.1712214114

人工智能让图片秀“千人千面”

■本报记者 计红梅

当你打开一个电商App的时候，如果最近比较关注化妆品，那么一大波化妆品会在页面显著位置推荐给你，而如果你对某品牌的服装感兴趣，那么出现在这一位置的就会是该品牌的服装……

“在人工智能技术的助推下，现在京东商城App已经可以做到‘千人千面’。”近日，在接受《中国科学报》采访时，京东技术总监桂创华告诉记者。

当下，以人工智能（AI）应用为代表的新零售产业正进入增长的上升通道，未来两年也将保持较快的增速。市场调研机构艾瑞发布的报告显示，2018年我国主要零售企业数字化建设的投入为285.1亿元，其中AI投入约为9亿元，占比3.15%。据预测，到2022年这一数字将突破700亿元，而AI投入将超过178亿元，占比达25%。

AI技术何以成为零售企业竞相发力的战略核心？它又是怎样改变传统零售行业面貌的？桂创华以英特尔助力京东打造的AI重塑商品图片系统为例，阐释了人工智能渗透零售行业后所带来的变化。

“当打开京东App的时候，最先吸引我们注意的就是页面的Banner图，即挂在网页最醒目位置的‘广告横幅’。”桂创华告诉记者，以前这个Banner图是由设计师设计的，每个人打开页面时看到的图片都一样。而京东却利用人工智能技术让机器来替代设计师进行设计，让每个人看到的广告图片都是他所感兴趣的。

对于其中的技术挑战，桂创华表示，第一步就是“智能抠图”，即把有背景的图片变成透明的底图，然后才可以做Banner图的设计。

“以前，大家都是用PhotoShop来做这件事情，完成一张图片的处理需要两分钟以上。而现在我们通过机器学习算法，在不到100毫秒的时间里就可以完成这一任务，而且能达到头发丝级别的精细程度。”

而实现这一目标，桂创华坦言不太容易。自动生成Banner图是一个很复杂的过程，将包括图像识别、智能抠图、智能排版等在内的多元计算机视觉算法综合应用，实现多个图像元素的智能组合，进行图像实时合成，还有极致的性能。“最初，我们做了很多上层应用软件的优化，但优化到100毫秒就优化不下去了，再往下就涉及到基础库的优化了。随后，我们和英特尔公司紧密合作，利用英特尔在软件库、ICC编译器及IPP图像库等图片和多媒体处理方面的技术积累做了很多优化工作。”

此外，对于京东、阿里巴巴、苏宁等零售企业而言，如何进行“山寨识别”也是很重要的一项任务。桂创华告诉记者，利用图像识别技术，他们可以对100多个品牌商标进行检测，特别是耐克、阿迪达斯等这些经常被“山寨”的品牌，“如果发现可疑的，我们就立刻揪出来。”此外，他们也会对商品详情页的文字进行识别，比如有些商家在商品详情页的图片上宣传了香奈儿，而它却没有卖香奈儿的资格，就会被认为是违规。“所以‘山寨识别’运用了综合性的技术，既包括文本识别，也包括图像识别。”桂创华说。

他告诉记者，为了在不影响质量的情况下提升山寨检测的性能，京东商城应用了英特尔今年4月发布的第二代至强可扩展处理器。该处理器所具备的Deep Learning Boost深度学习加速库技术，在山寨识别这一场景中极大提升了检测的性能。

人工智能技术在助力精准服务的同时，会不会泄露公众的隐私？对此，桂创华说：“京东对数据保护的要求极为严格，都是利用脱敏的数据构建模型机器去运行，工程师也不直接接触用户隐私数据。”

据悉，即将到来的“双十一”活动中，除了Banner图自动生成、山寨识别等技术外，京东还将推出“视频生成”技术。仅需商家提供原始素材，京东就会利用机器学习方法自动为其生成含丰富转场效果、高品质音乐的视频。“以前制作这样一个短视频，通常需要花几天的时间，支付几百块钱。”桂创华说。