

推动计算产业高质量发展

中国工程院院士王恩东:

从“计算”向“智算”演进面临三大挑战

本报记者 李佳师

“人工智能技术的大爆发,给各行各业带来巨大的赋能作用,同时也对计算产业提出了巨大挑战,从‘计算’向‘智算’演进,目前面临多元化、巨量化、生态化三大挑战。”中国工程院院士王恩东在近日举行的IPF2021浪潮数据中心合作伙伴大会的主题演讲中表示。在当天的演讲中,王恩东对于当前之所以出现“芯片五花八门”“AI计算群雄逐鹿”等产业现象给出了分析,并认为脑机接口最近取得系列重大进展是目前类脑计算发展的一大亮点。

脑机接口取得突破进展

王恩东首先谈及了人类大脑、人工智能大脑、硅基大脑和碳基大脑发展的情况。最近,脑机接口领域有很多突破性进展,比如埃隆·马斯克公布了“脑机接口”的最近实验,科学家在猕猴的大脑皮层中植入了两颗芯片,捕捉大脑活动的信号,实现将大脑信号翻译成游戏的动作,实现猴子动动脑子打乒乓球的游戏。

清华大学教授施路平去年曾经说:“碳基上能够实现的,硅基上一定也能够实现。”而实际情况也显示,碳基生物大脑同硅基系统的结合已经不再是科幻小说或者科幻电影,而是逐步在形成科学和技术上的突破。

王恩东表示,人工智能技术的大爆发才刚刚开始。目前我们已经创造了无数的智慧大脑,包括城市大脑、交通大脑、企业大脑等。人工智能技术,正在给我们的世界带来巨大的改变。例如,基于人工智能技术,可以实现毫秒级的精准识别,准确率达到99%,即使在光线不足的夜晚也可以根据一个尾巴和背影就能够识别出亚洲象。通过对野生大象长期的追踪和观测,为洞察大象物种的习性、种群的迁移轨迹提供了帮助,从而也能够为亚洲象国家公园的规划、栖息地改造、食源地的优化等这些保护措施提供科学的依据。

2020年中国农业无人机的销量超过了1.5万架,一架无人机一天农业喷散面积达到300亩,是人工的20倍。无人机上装了病虫害的识别系统,能够及时发现受病虫害侵蚀的叶子,实现农药的精准喷散,减少了农药的使用量。通过喷散催熟剂可以实现农作物成熟周期的同步,实现规模化的批量收割,成本大大降低,利润率大幅提升。根据普华永道

的预测,到2030年人工智能技术将为全球GDP带来14%的提升,大约16万亿美元。

王恩东认为,人工智能技术的大发展,对产业与经济社会的巨大赋能作用,与计算能力的巨大支撑不无关系。同样也对“计算”向“智算”推进提出了更多的挑战和更高的诉求。智慧计算对计算力的需求呈指数级增长,在2020年以GPU为代表的AI加速芯片所交付的计算力已经超过了同类CPU,预计到2025年加速芯片所提供的计算力可能在整个计算力交付当中超过80%。

智慧计算面临三大挑战

王恩东认为,面对指数级增长的计算需求,计算的技术、产品、产业面临着巨大的挑战,具体来说体现在以下三个方面:一是多元化的挑战。计算场景的复杂、计算架构的多元。二是巨量化的挑战。由巨量模型、巨量数据、巨量算力、巨量应用所引发对现有计算及体系结构的挑战。三是生态化的挑战。现在的智算正处于群雄并起的阶段,自成体系,生态离散,产业链上下游脱节。

第一个挑战是多元化。计算的关键任务是支撑业务的发展,不同的业务类型势必要求不同的计算系统完成,比如针对传统的地震模拟的科学计算,对于数值精度要求很高,而AI训练则可以适用数值范围大、精度相对低的16位浮点运算。

王恩东表示,AI应用引入了新的计算类型,从推理到训练跨度大,同时数据量也从GB到TB、PB不断提升,数据类型从结构化到半结构化、非结构化更加复杂多样。而不同数值精度的计算类型对计算芯片的指令集、架构等要求也不一样,这样就导致之前一直使用的通用CPU没有办法满足多元计算场景的要求,这也是计算芯片种类越来越多的原因。

芯片种类另外的另一个原因就是芯片代工模式的普及。现在芯片设计、制造等关键环节都有开源开放的软件、工具包括代工企业都可以使用。但是芯片从造出来到大规模用起来往往还隔着一个巨大的生态鸿沟,芯片应用一般都面临着开发者学习成本高、用户应用迁移难、芯片制造公司难以规模化的困难和挑战。

人工智能芯片出来之后,大部分都面临着面向开发者的帮助文档、调试工具、交流

回答问题的社区建设不足,开发者学习的时间长、难度大等问题。

对于芯片的最终使用者来说,由于芯片的指令集或者芯片架构的差异,导致编程库、编程模型、算法框架无法有效地横向打通,让大量应用迁移变得困难。

“芯片供应商如果想解决开发者的问题或者是使用者的问题,往往需要投入比芯片的研发成本高数十倍的投入。以往我们认为英特尔就是一个芯片公司、硬件公司,但其在英特尔内部有超过一万人是在为芯片的配套支撑、应用适配、软件优化做工作。”王恩东说。英伟达公司的CUDA软件环境更是通过长期的大规模的投入支撑起GPU在HPC和AI方面的霸主地位。

第二个挑战是巨量化。巨量化首先表现是模型参数多,训练数据量大。以自然语言处理为例,基于此训练模型兴起之后,模型精度随着模型尺寸及训练数据的增加显著提升,在2020年GPT-3模型的参数量首次突破了千亿大关,达到了1750亿。按照当前的发展趋势预测,到2023年模型的参数量会突破百万亿,也就是基本打造人脑的神经突触数量大概是125万亿。

巨量模型需要巨量内存,现有AI芯片的架构已经不足以支撑巨量模型的参数存储需求。

同时,巨量模型依赖海量数据的喂养。目前AI算法本质上还是依赖量变的质变,很难从一个质变跳跃到另外一个质变。例如最新的巨量模型需要万亿级的海量数据,海量数据需要海量存储,在超大规模集群中同时满足几个AI芯片的高精度读取,存储系统也是一个极大挑战。

巨量化的第二个表现是计算力需求的指数级增长。深度学习从2011年兴起到今天,对于算力的需求一直处于指数级增长,每隔三到四个月算力需求翻一倍。

就像我们用千瓦时来度量用电量,在天文学里面,以光年为代表来度量距离,我们以Petaflops-day(简称为PD)来度量算力,2020年GPT-3需要的算力是3640个PD。

GPT-3是以处理英文理解任务为主的,为了满足不同的语言、不同的场景的精度要求,需要训练不同的巨量模型。如此庞大的计算力需求给计算技术和产业带来了机遇和巨大的挑战。解决这样的挑战需要从体系结构、系统软件各个方面开展创新。

到2023年,百万亿参数的巨量模型的算力需要百万PD,现在世界上最快的超算是日本的富岳,在这上面完成百万PD的计算所需要的时间是两年,在不同的领域需要不同类型的巨量的模型。

巨量化的第三个表现是模型应用规模大。AI的应用已经渗透到各个行业,以互联网头部企业为例,他们基于自身的已有的AI技术搭建了开放平台,在这上面承载着各类AI的基础能力(语言识别、图像识别、自然语言处理)等。这些开放平台已经吸引了超百万的AI开发者,开发了各种AI应用和服务,这些AI开放平台每天承载着数亿次的交互量、数百万小时的语音识别、超过百亿张图像识别、超过万亿句自然语言的理解。如此巨量的调用对计算中心的算力、应用能力产生了极大的挑战。

第三个挑战是生态化。AI技术链条和产业链条面临脱节的问题。很多人可能会问一个问题,大家都说人工智能好,但是这个东西与现在的业务、与客户的应用场景怎么结合,用户没有人懂算法和模型,也缺少好的AI开发平台,同时那么多的算法、模型如何与应用进行最优组合,等等。懂AI算法与模型的人往往是集中在科研机构或者头部企业,但缺少对传统行业的需求场景、业务规律的深入理解,也拿不到最关键的业务数据对模型进行训练,这就导致技术无用武之地。

埃森哲等咨询机构的调查报告也表明了这个问题,70%以上有技术的研究机构和科技公司缺少需求场景、领域知识和数据,70%以上的行业用户则缺少技术人才、AI平台和实践能力。

“计算技术与产业生态是离散化,这与我们讲的多元化的挑战是相呼应的,芯片架构五花八门,指令集不同无法兼容,而面向芯片的编程库等又跟芯片紧密绑定,灵活性差。小公司只做了一个环节,这就造成生态的纵向不通;大公司则希望构建封闭系统,就造成了生态横向不通。这些都严重制约了AI技术的应用和发展。”王恩东说。

王恩东进一步表示,通过这些分析我们可以看到,一方面智算需求蓬勃发展的重大机遇,另一方面是多元化、巨量化、生态化的挑战和困难亟须解决,面向新的需求机遇,迎接新的挑战,应该说计算技术和产业需要构建一个新的发展格局。

产业链携手加速破解三大难题

关于如何应对多元化的挑战,王恩东表示,首先要创新智算体系结构,要从硬件重构、软件定义两个方面引领智算技术的体系创新。

王恩东以浪潮的融合架构3.0为例,是通过片内大容量的缓存、片外高带宽内存等存算一体方式实现计算数据结合,突破存储墙。通过高速互联交换网络支持更广泛的多元异构计算,支持新型存储资源置换,提高性能降低功耗,实现更大规模的资源可扩展。进一步通过在互联处理单元中卸载控制平面,实现控制和计算的分离,进而打造更为灵活资源的可重构。在系统软件创新方面,可以重构的硬件资源集组织,通过软件定义的方式实现控制层和基础设施的分离,统一资源组织模型,标准化可编程结构,形成软件定义计算、软件定义存储、软件定义网络。通过微服务化实现集中管理平台的可用性和高扩展。此外,也可以往底层延伸,基于不同的场景来开发定制化的AI加速、网络加速、存储加速产品,最终形成覆盖芯片板卡、系统平台和算法的领先的产品体系,在加速芯片和板卡方面,芯片多元化应该既是挑战也是应对当前算力需求复杂多样的手段。

而面对算法巨量化的挑战,王恩东认为可以通过构建智算中心以服务化的方式来为产业提供服务,面对各行各业以及各类不同的AI算法公司,社会需要AI算力的基础设施,通过构建“智算中心”加速解决巨量化的挑战。一个例子是江苏省建设的智算中心,今年年初江苏按照智能计算中心规划建设指南,在南京投资建设了智算中心,由寒武纪和浪潮联合承建,目前已经投入运营使用。这个智算中心采用了最先进的AI处理芯片和AI计算平台,该中心可以为区域的公共算力服务、创新应用孵化、产业聚集发展、科研创新和人才培养提供助力和支撑。

最后王恩东表示,无论AI产业化还是产业AI化,产业链的拉通和生态的建设都是必需的。在产业AI化的过程当中,浪潮推动元脑计划,希望构建一个开放的AI生态,也希望更多的厂商加入到推动AI生态化发展的进程当中。

(上接第1版)

更好服务构建新发展格局。加快构建以国内大循环为主体、国内国际双循环相互促进的新发展格局,是关系我国发展全局的系统性、深层次变革,是立足当前、着眼长远的战略选择。提升产业链供应链现代化水平,保持制造业比重基本稳定,增强制造业供给体系韧性,促进数字产业化和产业数字化发展,是构建新发展格局的基础和内在要求。要从全局高度准确把握、精心谋划,找准定位和着力点、突破口,在服务构建新发展格局中展现更大作为,促进工业和事业信息化发展行稳致远。

统筹发展和安全。安全是发展的前提,发展是安全的保障。当前和今后一个时期,工业和事业信息化发展形势严峻复杂,各种可以预见和难以预见的风险因素明显增多。必须坚持统筹发展和安全,增强机遇意识和风险意识,树立底线思维,把困难估计得更充分一些,把风险考虑得更深入一些,注重堵漏洞、强弱项,下好“先手棋”、打好“主动仗”,有效防范化解各类风险挑战,牢牢守住安全发展这条底线。

坚持系统观念。系统观念是具有基础

性的思想和工作方法。工业和事业信息化工作点多线长面广、新情况新问题多,更要强调系统观念,加强产业发展统筹,加强资源力量整合,充分发挥改革的突破和先导作用,不断增强工作的前瞻性、整体性、协同性。要把系统观念作为必须遵循的重要原则,加强前瞻性思考、全局性谋划、战略性布局、整体性推进,充分发挥各方面作用和积极性,着力固根基、扬优势、补短板、强弱项,为实现更高质量、更有效率、更加公平、更可持续、更为安全的发展作出应有贡献。

狠抓重点工作任务落实,开创制造强国和网络强国建设新局面

《规划纲要》明确了2035年远景目标和“十四五”时期经济社会发展主要目标。工业和事业信息化系统要认真落实《规划纲要》部署,主动对表对标,坚持目标引领、问题导向,细化完善举措,强化创新驱动、改革推动、融合带动,以更大力度推进制造强国和网络强国建设,为全面建设社会主义现代化国家开好局、起好步提供有力支撑。

加快提升产业创新能力。这是提升产

业核心竞争力、有效应对风险挑战的关键所在。要坚持把科技自立自强作为战略支撑,充分发挥我国超大规模市场优势和新型举国体制优势,深入实施创新驱动发展战略。实施关键核心技术和产品攻关工程,着力突破“卡脖子”技术,推动重点技术和产品生态体系建设。鼓励支持创新,不仅要创造更多单项技术和产品,更要大力提升产业自主创新能力。这就需要深入推进制造业协同创新体系建设,强化基础共性技术供给。加快建立以企业为主体的技术创新体系,充分发挥龙头企业带动作用,推动大中小企业融通创新。

着力提升产业链供应链自主可控能力。这是畅通国民经济循环、保障产业安全的必然要求。要全面系统梳理我国制造业发展状况,探索建立产业链供应链安全风险评估制度,分行业做好战略设计和精准施策,突出有产业集群功能,挖掘产业存量的潜力,布局新兴产业链。实施制造业强链补链行动和产业基础再造工程,聚焦核心基础零部件(元器件)、工业基础软件、关键基础材料、先进基础工艺、产业技术基础等关键瓶颈,全面加大科技创新和

进口替代力度,构建自主可控、安全可靠的国内生产供应体系;持续增强高铁、新能源、通信设备等领域的全产业链优势。

大力推进产业数字化和数字产业化。这是顺应世界经济数字化发展大势、催生发展新动能的必然选择。要推进产业数字化,以制造业数字化转型行动为基础,推进智能制造工程、中小企业数字化赋能行动,对制造业进行全方位、全角度、全链条改造,推动制造业数字化、网络化、智能化发展。推进数字产业化,全面部署新一代通信网络基础设施,有序推动5G、数据中心、工业互联网等的建设部署和创新应用,培育壮大人工智能、大数据、区块链、云计算、网络安全等新兴数字产业,提升通信设备、核心电子器件、关键软件等产业水平。增强网络安全保障能力。

持续推进产业结构优化升级。这是提高产业整体素质、推动高质量发展的内在要求。要坚持深化供给侧结构性改革这条主线,在“巩固、增强、提升、畅通”上狠下功夫,继续完成“三去一降一补”的重要任务,全面优化升级产业结构。深入开展质量提升行动,努力增品种、提品质、创品牌。大

力改造提升传统产业,发展壮大战略性新兴产业,加快培育一批世界级先进制造业集群。认真落实碳达峰碳中和部署,实施工业低碳行动,严控重化工业新增产能规模,加快节能低碳绿色技术创新和应用,提升制造业绿色发展水平。健全优质企业梯度培育体系,大力培育专精特新“小巨人”企业、制造业单项冠军企业和具有生态主导力、核心竞争力的产业链龙头企业。

进一步深化改革开放。这是积极应对外部环境变化、激发产业发展动力活力的重要保障。要坚持市场化改革方向,持续深化“放管服”改革,推动要素市场化改革,建设高标准市场体系,增强制造业对各类资源要素的吸引力,提升制造业国际竞争力。拓展国际化视野,增强国际化思维,更好利用国内国际两个市场两种资源,提升工业和事业信息化领域对外开放水平。依托共建“一带一路”,加强产业链国际合作,构筑互利共赢的产业链供应链利益共同体,支持有实力的企业走出去融入全球产业链、供应链、创新链、价值链。

(本文转自2021年4月19日出版的《人民日报》)

中国电子报 一报在手 行业在握

中国电子报是工业和信息化部主管的传媒机构,创建于1984年。

目前,中国电子报社拥有集报刊、图书、网站、微信、微博、短视频等融媒体传播、会议活动、展览展示、专业大赛、定制服务等会展服务于一体的立体化、多介质产品,成为凝聚行业力量、服务行业发展的重要平台。

《中国电子报》(国内统一连续出版物号:CN 11-0005 邮发代号:1-29)是具有机关报职能的行业报,主要报道内容包括:产业要闻、政策解读、集成电路、新型显示、智能终端、家用电器、5G、人工智能、物联网、工业互联网、移动互联网、大数据、云计算、区块链、VR/AR等。

融媒体服务



- 报纸出版
- 官方网站(电子信息产业网www.cena.com.cn)
- 官方微信(公众号cena1984)
- 官方微博(http://weibo.com/cena1984)
- 视频平台(抖音、快手、央视频、人民视频等)
- 视频服务(视频制作、在线直播、在线会议等)
- 平台推广(学习强国、今日头条、百度百家等)
- 内参专报
- 行业报告
- 图书出版

会展服务



- 会议活动
- 专业大赛
- 展览展示
- 专业培训
- 政府服务
- 企业定制
- 产品评测
- 舆情监测
- 数据营销
- 招商引资



官方微信 官方网站
在这里
让我们一起把握行业脉动
www.cena.com.cn

地址:北京市海淀区紫竹院路66号赛迪大厦18层
电话:010-88558808/8838/9779/8853
传真:010-88558805