

本报记者 姬晓婷

在需求增长的同时智能算力供给不足,是当前“算力荒”的核心矛盾。研究机构数据显示,在大模型训练需求的带动下,智能算力的增长速度将远超通用算力。全球智能算力规模将在2027年达到1117.4EFlops,大致相当于2023年的2.7倍、2020年的15倍。那么,如何解决“算力荒”?

统筹智算资源 “集中力量办大事”

当前的算力需求共三种:通用算力、智能算力和超算算力。算力供应不足,更多指代的是智能算力。对应这三类算力需求的,是以不同类型芯片组建的算力基础设施。通用算力以CPU芯片输出的计算能力为主,智能算力以GPU、FPGA、AI加速芯片等输出的人工智能计算能力为主,超算算力以超级计算机输出的计算能力为主。

“通用算力是不缺的,我们现在缺的是能够满足大模型训练需求的智能算力。”鹏博士集团副总裁孙向东在接受《中国电子报》记者采访时表示。模型训练完成后,可以通过压缩、系数化等处理方式放在较低端的算力芯片上做运营和服务,满足这类需求的算力芯片并不少。

“自ChatGPT出现以来,我们能够直观感受到智算需求的增长。今年被认为是大模型真正实现行业应用的一年,金融、医疗等诸多行业都在尝试用大模型解决问题。”孙向东说道。

与满足通用算力的传统数据中心相比,智算中心存在诸多不同。这不仅体现在所需的芯片种类上,还体现在服务器机柜的组件、功耗等诸多方面。传统数据中心的业务以存储为主,其功耗为4~6千瓦时;智算中心的业务则以高性能计算为主,其功耗可能达到12千瓦时甚至更高。不仅需要配备高密度机柜,整个机房的布局也要配合需求进行调整。

为满足市场的强需求,符合人工智能应用需要的智算中心正在建设。但在孙向东看来,建设速度还远赶不上需求增长速度。当前,为了尽可能多地解决智算的供需矛盾,孙向东表示,要由政府或大企业出面,统筹全国的智算资源,“集中力量办大事”。

区别于普通计算,大模型需要的是千卡甚至万卡规模的计算集群。但建设智算中心,一方面,可用服务器供给量有限,且配置成本高。因此,孙向东认为,要想尽可能地提高现有资源的利用效率,就要实现算力中心



的连接,建设成智算网络。这样一来,就可以把可用、符合大模型训练标准的算力基础设施连接起来,搭建起计算网络,当某个计算任务完成后,这部分被占用的资源就能够及时释放出来,提供给其他企业使用,以此满足并行大规模计算的需求。

研制超级计算机 应对大模型训练问题

在超级计算机研制方面,我国已经居于

全球领先地位。当前,国家网络已经接入八个国家超级计算中心,配备有世界上首台峰值运算性能超过每秒十亿亿次浮点运算能力的超级计算机——“神威·太湖之光”,峰值性能达每秒4700万亿次的“天河一号”高效能计算机系统、系统持续性能为每秒796万亿次的神威蓝光超级计算机等诸多性能卓越的超级计算机。

在超算领域,我国有几大优势:具有数十万处理器的搭建经验,具有几大超级计算机供应品牌,具备高速网络搭建和运营能力、绿色液冷技术,具备保证超级计算机运

行可靠性的能力。此外,国内还有经过几十年发展积累的超级计算机领域的丰富专业人才。

由此,中国科学院计算技术研究所研究员张云泉表示,要解决算力荒,可以利用我国在超算领域过去几十年的积累,利用超级计算机技术研制大模型专用超级计算机,以争取时间应对我国因智算资源受限而可能出现的大模型训练落后问题。

超级计算机在计算某种特定任务时具有速度优势。“我们的超级计算机设计的是64位精度,过去主要用于解方程,但大模型

训练主要采用的精度是16位。”张云泉表示,“超级计算机可以把大模型作为具有挑战性的应用来进行针对性研制,从技术实现来说是可行的,通过优化提高效率、性能,是可以实现的。”

张云泉表示,当前的超级计算机不能直接用于大模型训练,需要进行优化和定制。要“向下兼容”,也就是由支持64位为主改为支持16位为主,其软件和算法也要相应地进行优化和改进。

掌握智算资源的企业 将享受算力红利

在智算资源紧缺的情况下,市场上出现了由于掌握算力稀缺资源而颇受市场青睐的企业。例如以印刷为主营业务的鸿博股份,因为搭上英伟达概念,其股价在半年时间内涨幅超过500%。

面对智算资源供应不足的情况,有哪些类型业务将会受到市场关注?

张云泉认为,有三大类公司及其业务将获得发展机遇。第一类是超算基础设施供应商,以曙光、天河、神威为代表,此类公司对算力生态建设比较了解,同时掌握用户需求,对如何搭建大规模计算集群具有技术经验,可为智算提供资产服务。第二类是电信运营商,具有网络建设和调整网络带宽的优势,可通过配备高端资产,依托全国数据资源调配优势,承担算力中心资源调配任务。第三类是能源公司,具有来自光伏等新能源的电力资源,可以在搭建新能源基地的同时配套建设算力中心,具有智算中心所需要的能源优势。

孙向东认为,有两类企业将乘大模型训练之势实现增长。一类是国产人工智能芯片供应商,但此类企业还需要加强技术研发,一方面提升芯片性能,另一方面完善软件生态;另一类是为大模型商业应用提供底层支撑和运营服务的企业。

当问及我国该如何看待当前火热的大模型训练全球竞争时,张云泉表示,在人工智能市场的国际竞争中,我国具有人口基数大、人才资源丰富、新能源产业发展势头强劲等优势。当大模型训练所需的算力越来越大,能源将可能成为制约计算规模继续增长的瓶颈,在那时,我国所具备的绿色能源比较优势便会体现出来。但他也提醒,要警惕发展大模型急功近利的问题:“很多国内公司着急变现,包括投资基金和专注于大模型训练,推理的厂商,很多不愿意追求更高的技术水平,产品稍微能用就去以模型变现、不再继续训练了。这种想法要不得。”



人工智能为湘茶产业添活力

本报记者 张心怡

三湘四水、五彩茶香。在素有“江南茶乡”之称的湖南,以潇湘绿茶、湖南红茶、安化黑茶、桑植白茶、岳阳黄茶为代表的“五彩湘茶”,正在成为湖南茶产业的亮丽名片。如今,在科技的加持下,湘茶产业的发展更具活力。

湖南商务职业技术学院教师阳舜,是一位计算机硬件与软件技术的爱好者与实践者。他主持的课题“五彩湘茶”智能检测与识别典型生产实践项目,不仅将人工智能等新一代信息技术应用于茶叶生产的各个环节,也对校企合作进行了探索和示范。

人工智能技术能否赋能实体经济,推动产业升级和生产力提升,是其从概念走向落地的关键。传统的制茶方式,对制茶师傅的经验和技艺有较高的要求。而人工智能技术的运用,能够使温度、湿度、发酵程度等茶叶加工参数更加可控,也能够提升茶叶采摘、加工、包装、销售等各个环节的效率。在“科技兴茶”的过程中,人工智能正在发挥越来越重要的作用。

在阳舜看来,技术理论知识要运用于实践,服务于生产生活。“五彩湘茶”生产实践项目,实现了对茶叶生产环节的精确监测和数据分析,以提高茶叶品质和生产效率。该项目依托

“楚怡”人工智能应用创新产教融合实训基地,设置了茶园智能监控与管理、茶叶智能采摘机器人、茶叶智能分拣和分级、茶商品销售预测与分析、茶产地可追溯性智能管理五个任务模块,并搭建了模拟操作平台、创新设计平台、数字技术工具、数据分析平台、真实实践平台等项目管理云平台。

在人工智能赋能千行百业的同时,市场对于人工智能人才和专业人员的需求也越发凸显。

自2018年9月至今,阳舜在湖南商务职业技术学院担任人工智能技术应用专业教师。在他眼中,每一位学生都有可能成为未来计算机科研领域的翘楚。他和校方同样意识到,除了课堂上的专业知识和实践操作,技术人才的培养还需要充分结合企业岗位的需求。

湖南省教育厅、湖南省工业和信息化厅联合印发的《关于进一步优化我省高校电子信息类人才培养方案的指导意见》提到,支持高校与行业企业、产业园区深度合作,鼓励高校通过共享企业研发平台、生产基地等方式,建设兼具教学、研发、生产、创新创业等功能的校企一体、产学研用协同的大型实验、实训实习基地。记者看到,在湖南省教育厅于2022年12月发布的湖南省“楚怡”人工智能应用创新产教融合实训基地计划建设单位名单中,就出现了湖南商务职业技术学院的人工智能应用创新产

教融合实训基地。

因此,“五彩湘茶”生产实践项目在规划之初,就以校企联合为重点,并设置了具体的培养目标——人工智能训练师与人工智能应用开发工程师。其中,人工智能训练师要掌握人工智能软件,包括数据库管理、算法参数设置、人机交互设计、性能测试等。而人工智能应用开发工程师要应用相关算法对项目进行设计开发与运维管理。

湖南商务职业技术学院与东软教育科技集团是该项目中校企联合的主体单位,前者负责组织实践性教学活动,后者负责为项目实施提供优质资源。双方形成“三融三通”运行机制,即企业岗位需求与人才培养目标相融合、岗位新技术能力需求与教学内容相融合、生产组织与教学组织培养方式相融合,并围绕人才培养目标定位,研制人才培养方案、构建专业核心课程体系等维度开展合作。

在各方的努力下,“五彩湘茶”智能检测与识别典型生产实践项目将人工智能应用于茶叶的生产过程中,为茶叶生产企业提升生产效率和产品质量提供技术支撑。同时,该项目基于校企合作,以具体的职业岗位能力需求为目标,为人工智能相关专业学生提供了实际操作和训练机会,为助力湘茶产业数字化转型升级提供人才支持。

大力推进现代化产业体系建设 加快发展新质生产力

