

随着计算机的不断发展,医学影像的精度、密度等数据信息量会越来越大,生物标本库和生物组学信息的数据也在快速增长,数据的存储甚至人工的智能判读成为重要问题。

精准医疗急等人工智能救场

■本报记者 李惠钰

精准医疗,就是要克服千人一方、万人一药的瓶颈问题,根据患者个体的遗传基因特征,量体裁衣地制定具有个性化的治疗方案。然而,随着生物数据的爆炸式增长,更多的挑战出现在了计算技术领域。

“我们已经进入了一个医学数据大规模整合的时代,要参考基因组学、微生物组、环境、行为、临床检验等各种信息,这个整合就导致数据量出现指数级的增长。”中国国家罕见病注册系统执行总监弓孟春在2017人工智能计算大会上表示,现在精准医疗的知识体系已经超过医生个人承载的范围,知识更新的速度也超出了传统医学教育系统处理的能力。

而人工智能最大的优势就是其碾压性的计算能力,医疗若想更精准,急需人工智能前来救场。

“胖数据”待“瘦身”

“现在病人会拿着基因测序的报告问你,糖尿病会不会遗传给孩子,会不会很早就得眼底病变,会不会很快出现心血管病变,是不是应该用胰岛素,应该先吃什么药,运动会不会对我有效……”精准医疗普及的同时也给医生提出了很多难题,弓孟春就遇到过很多问题不断的病人,这些问题已经远远超出医生原来数据搜集的范围。

“我如果要评估这些问题,就需要对整个环节的其他维度的信息做收集和整理,然后再进行分析,还要参考我的知识体系来判断,最后才能回复我的病人。”弓孟春无奈地说,“医生现在面临的挑战就是,作决策的时候需要参考的信息维度大大增加。”

当前,精准医疗的临床部署对信息技术的依赖程度越来越强,因为精准医学的知识体系和数据量已经远远超过医生个人承载的范围。

中国医学科学院北京协和医学院放射科主任医师惠周光表示,医疗大数据研究的内容包括疾病、药物、医生和患者,其来源从狭义上说,包括医院的信息系统及电子病历数据、检验信息系统数据、医学影像存档和传输系统数据,以及一些医学文献和临床经验等。但是从广义上来说却远远不止这些,还涉及到公共卫生及生命科学领域。

以恶性肿瘤大数据为例,惠周光的体会就是,随着计算机的不断发展,医学影像的精度、密度等数据信息量也会越来越大,生物标本库和生物组学信息的数据也在快速增长,数据的存储甚至人工的智能判读都成为重要问题。

现在,精准医疗本身也有赖于数据的采集,“比如说哮喘患者,有无药物过敏、家族病史、空气的监测和可穿戴设备等数据,都会构成基于患者个人的模型,从而预测该患者的病情未来会不会加重,是否要转移到急诊去。”弓孟春说。



精准医疗的临床部署对信息技术的依赖程度越来越强。

图片来源:百度图片

而回到基因行业领域,英特尔医疗与生命科技集团首席架构师丁华表示,随着全基因组测序成本的不断下降,将来有可能会演进到更多的全基因组的测序。“谈到精准医疗不可回避的一点就是,把组学的这部分内容跟传统临床能够更好地结合,然后共同实现精准医疗的目标。但目前我们还处在怎样把基因数据变成有意义的、有价值的数据,这个数据能够在临床上应用这个阶段,依然面临很多的挑战。”丁华指出。

受益于人工智能浪潮

有业内人士这样比喻:在医疗领域,人工智能分析的作用堪比经验丰富的临床医生。

阿里云此前发布的ET医疗大脑辅助医生判断甲状腺结节的场面,就令不少人记忆犹新:ET医疗大脑通过计算机视觉技术在甲状腺B超影像上圈出结节,并给出良性或者恶性的判断。而在此之前,医生个人判断甲状腺结节的平均准确率仅为60%~70%,当下人工智能算法的准确率已被证明可达85%。

人工智能之所以判断准确率比人类更高,离不开三个关键词:计算、算法、数据。算法的突破及数据洪流的爆发使得几乎所有的“机器辅助功能”都成为可能,深度学习框架的开发和开源,也使得人工智能“算法”的开发越来越便捷。“计算”平台则成为本轮推动人工智能进步的重要因素。

在当天会场的楼下就摆放着数十台样式各

异的计算机、服务器。“这其中一台小小服务器的计算速度,就是20年前最快计算机的60倍。”中国工程院院士、浪潮集团首席科学家王恩东说,由于计算能力的快速发展,结合互联网、物联网带来的海量数据和深度学习等先进算法,共同催生了第三次人工智能浪潮。

精准医疗也将乘着人工智能这波浪潮,在各个领域发挥着超过人类预期的作用。上海交通大学高性能计算中心工程师韦建文团队就准备在白血病诊断方面,想办法解决机器读片的问题。

“当前,血液涂片的人工诊断仍然非常费力,一般专职读片医生需要3至5年的训练,而受限于人眼可阅读的涂片视野,医生根据个人的能力和经验诊断出的结果精确度起伏不定,难以覆盖所有亚型的白血病细胞。”韦建文说。

通过解决超高分辨率、超灵敏度,把基因组学跟医学影像学结合,再加上时间序列上技术突破,研究人员实现了在高性能计算平台上进行血液涂片的解读,并可以让医务人员也非常方便地去使用这个工具。

另外,处理精准医疗的大数据需要极高纬度的分析,在可预料性低的情况下,用传统信息学方法处理难度较大。记者从英特尔方面了解到,英特尔BioX实验室与华大基因在蛋白质组学方面就有成功的合作案例。对于蛋白质飞行质谱实验,在使用深度学习CNN框架和LSTM算法的情况下,较之先前的传统计算,实验人员大幅提升了飞行质谱预测的效率,同时节省了2/3的计算时间。

在美国,微软也宣布将人工智能用于医疗健康计划“Hanover”,他们试图帮助寻找最有效的药物和治疗方案。此外,微软还在研究模拟癌症如何在不同病人身体里扩散,甚至于研究像计算机编程一样创造生物细胞等。

应对计算与数据的挑战

精准医疗可以提供更准确的预测预防、治疗以及康复的解决方案,而支撑精准医疗发展的三驾马车——生命科学、信息技术和临床医学,则在相辅相成中共同推动精准医疗的发展。随着精准医疗的不断提速,其主要难度也逐渐集中在信息技术领域。

有专家表示,目前我国“应用终端的发展”远远走在“硬件架构”的前面,现有计算平台已经不足以完成人工智能对于庞大运算量的需求。而在王恩东看来,尽管人工智能如今蓬勃发展,但仍是挑战重重,其中排在第一位的就是计算性能。

清华大学教授、国家超级计算无锡中心主任杨广文则指出,在过去11年里,内存带宽仅仅提升了15~16倍,而计算能力则提升了30~50倍,这说明内存的性能和计算的性能之间的差距在逐渐拉大,这也是GPU计算今天面临的一个巨大挑战——在相对强悍的计算能力和相对薄弱的内存访问之间,横跨的那一道鸿沟。

除了人工智能本身面临的挑战之外,对于精准医疗来说,数据的深度挖掘也是目前急需解决的问题之一。“机器需要更多的学习次数,以得到更准确的判断。多次学习的基础是数据,医疗影像需要海量医疗影像和医生的结果判断。”一位智能医疗界投资人士表示,精准医疗现阶段最大的困难就是数据积累。

另外,他还指出,影像数据和放射科医生增长速度存在不匹配。相关数据显示,美国、中国的影像数据年增长率将分别达到63.1%和30%。但两国放射科医生年增长率仅有2.2%和4.1%。两者之间的鸿沟一方面促进了人工智能的发展,同时也给人工智能存储及处理海量数据带来巨大难题。

惠周光也以肿瘤大数据为例指出,目前我国尚未建立以医院为基础的癌症筛查网络,缺乏全国肿瘤发病与诊疗状况的权威数据,导致肿瘤防治、临床研究、指南制定等都有一定的盲目性。

惠周光建议要搭建医疗机构的网络平台,建立医疗大数据存储、传输安全和共享机制,并通过标准化解决临床用语与标准数据相割裂的问题。

“另外,现在组学的应用越来越多,大量的计算资源与相应的需求相割裂,都是精准医疗会遇到的问题。我想人工智能相应的计算和分析在这里都是可以大有作用的。”惠周光说。

腾讯过去的业务主要是社交、通讯、数字内容和一些金融服务,其他的主要是投资以及其他生态公司合作。对腾讯来说,有几个基础性的因素是其愿意大力投入的:第一是AI;第二是云计算;第三是大数据。

腾讯董事会主席兼首席执行官马化腾：“不想成为科学家的投资人不是好企业家”

■本报记者 赵广立

“现在全球市值最高的十家企业,前五家在短短的一年内成为了以科技为主的企业,腾讯和阿里巴巴也在其中。随着数字经济的发展,科技越来越融入各行各业,在数字化和智能化的大潮里,科技是产业界必须要去关注的。”

9月8日,“清华—青腾未来科技学堂”开课。在清华大学经济管理学院举行的“洞见论坛——科技·驱动成长”活动上,腾讯董事会主席兼首席执行官马化腾在演讲中打趣说:“不想成为科学家的投资人不是好的企业家”,并表态腾讯不仅与清华大学共同举办“清华—青腾未来科技学堂”,未来还会通过“互联网+”等方式和各行各业、学界和研究界进行合作。

不断尝试促进产研结合创新

记者了解到,首届“清华—青腾未来科技学堂”的48位“同学”是从超过一千家科技企业创办者中挑选出来的,录取率只有4%左右。对此,马化腾直言此次与清华大学合办“清华—青腾未来科技学堂”,是腾讯“希望通过不断地尝试来促进科研和产业界结合与创新”。

“斯坦福大学是硅谷创新的发源地,它很大的特点就是把科技和商业完美地结合在一起,我觉得这是非常值得全球尤其是中国去学习的。”马化腾表示,虽然中国没有像斯坦福大学、哈佛大学这样的私立学校——更多的还是公立学校,但腾讯仍然愿意并且已经作了很多创新的尝试。他披露,腾讯和清华很早就成立了一个联合实验室,近期正转

向研究人工智能(AI)。

“产学研需要有一些创新的生态。”马化腾谈到,一年前他参加了由红杉资本沈南鹏和香港科技大学教授李泽湘发起的产学研结合组织“Hong Kong X”。他们希望依托香港的国际化的高校资源,依托粤港澳大湾区有智能制造、人才和生产基地的大环境“做一些事情”。

马化腾表示,这是因为他们看到了国外在产学研方面的新的变化。“波士顿附近有一个奥林工程学院,这个学院很特别,学生一进来就打破传统院系专业,通过跨界融合的方式与企业合作,让学生通过做项目的方式将理论和实践相结合;麻省理工学院最近有一个叫作“引擎”的计划,通过学校的一些基金,鼓励学生把想法变成现实。”

“这是个新的趋势。”马化腾说,基于这些新模式,马化腾在香港科技大学的推动下,正在探索成立一个新兴的教育机构:“姑且叫作湾区科技大学或者科技中学,目的是从培养创新人才,为整个粤港澳大湾区做服务。”

在人才培养模式的探索上,腾讯的确在不断尝试。今年暑假,腾讯就发起了“腾讯粤港澳大湾区青年营”。

“我们请了粤港澳三地大概一百名青年(主要是高中生)参加夏令营,来深圳到腾讯、大疆等一些企业去接触科技,体验AI、AR/VR、无人机和机器人等技术,以及无现金的消费体验等。”马化腾介绍说,其间还开展一些青年人喜欢的活动,以促进粤港澳三地青年人的融合。此外,腾讯还开设了暑期实习岗位给粤港澳三地的大学生。

声音

从去年开始,无论是在产业界、学术界还是政府层面,人工智能都获得高度关注。这种重视让我们看到,人工智能能够为经济社会发展带来潜在的变化。

人工智能并不是一个新词,自1956年提出至今已有六十多年,并经历了三起三落。大家所讲的人工智能之父就不止一位,其中一个就是图灵,他也同时被称为计算机之父。

图灵发明了计算机的基本原型,又提出人工智能,我们把计算机比作是他的“大儿子”,人工智能是“二儿子”,“大儿子”和“二儿子”的特点不一样。“大儿子”比较稳重、有韧性、不调皮、不捣蛋,一直按照摩尔定律持续发展。“二儿子”则比较活跃,有创新思维、敢冒险,于是人工智能经历了多次起落。

人工智能被第一次提出时,机器就超越了人类,人可以做的事情机器都能做,结果人工智能没有达到预期发展。第二次也是如此,敢说敢干,但发展遇到瓶颈了还得把重心放在计算机领域。

今天,人工智能已经发展到了第三波,似乎比较靠谱了,因为计算发展很快了。今天一台小小的服务器的计算能力就达60万次,是20年前全球最快计算机的60倍。人工智能的发展和支撑,离不开计算、算法和数据,计算是基础。

面对人工智能这样一个蓬勃发展的领域,有众多需求,人工智能计算依然面临着提升计算性能、不同场景计算优化和生态建设三大挑战,我们需要整合各种处理芯片,提供一个先进的计算平台,整合这些框架,提供比较易用的应用开发的平台。

另外也要加强计算芯片创新、可扩展平台架构、系统化设计等基础技术研发和能力构建,同时也要从硬件、软件、应用、人才等多方面综合提升,避免短板效应带来的整体产业发展不良反应。

还希望来自于学术界的、企业界的、应用部门的各方专家,能够从多角度一起来讨论,如何能够打造一个比较易用的计算设备或架构,能够使人工智能的应用发展得更加顺利,而不要再陷入上两次人工智能的窘境里。

从国家角度来讲,希望能够在这一轮的人工智能机会当中站在前沿,企业与学术界也希望能够为人工智能的发展贡献智慧,这就需要人工智能方面的专家人才能够聚在一起,相互交流,形成一个国内平台,形成一些标准,为人工智能的发展扫清障碍、奠定基础,不断推动人工智能的发展。

(作者系中国工程院院士、浪潮集团首席科学家,本报记者李惠钰根据其在2017人工智能计算大会上的发言整理。)

发展人工智能须扫清障碍

■王恩东

按图索“技”



①后摩尔时代面向未来的高效数据中心展示。
②盒马鲜生展展展示了新零售的魅力。
③搭载YunOS Auto操作系统的互联网汽车。图片来源:百度图片

一场科技炫酷的年会

9月8日,阿里巴巴集团迎来公司史上规模最大的一次年会。不仅人数规模创历史新高,科技炫酷程度也前所未有。

18周年庆前一天,阿里巴巴在西溪园区举办了一场“技术互动展”,展出了包括天猫、淘宝、蚂蚁金服、阿里云、菜鸟、闲鱼等在内的22个业务板块的科技成果,现场堪比“科幻片”。

以高精尖技术为例,今年3月提出“NASA计划”后,阿里巴巴在机器学习、芯片、IoT、操作系统、生物识别等方面都进行了广泛的布局。

在今天的技术互动展上,工业级的液冷处理器,人工智能“工业大脑”,能够完成精细的画画、写春联等工作的机械臂,大数据平台等“核武器”尽数亮相。

天猫展区的虚拟试衣镜、试妆镜通过AI技术,能够实现亲身试穿新衣、虚拟化妆的效果,“网红超市”盒马鲜生的展展则展示了新零售的魅力,现场销售的小龙虾月饼也成了会场内人手一只的“抢手货”。

现场最“贵”的展出设备则是一辆货真价实的互联网汽车,这台搭载了YunOS Auto操作系统,由上汽集团生产的互联网汽车,能够实现语音

(贾晓丽编辑)