智能程序开启结构生物学新未来

■本报记者 张双虎 冯丽妃

近日，谷歌旗下DeepMind团队一周之内搞了两件“大事情”。而华盛顿大学戴维·贝克团队的罗塞塔折叠（RosettaFold）也搭载阿尔法折叠（AlphaFold2）的便车风光了一把。

这两款智能程序相继开源昭示着，智能程序正在开启结构生物学的新未来。

大事情

几天前，DeepMind团队在《自然》发表文章，公布了第十四届国际蛋白质结构预测大赛（CASP14）中夺冠的AlphaFold2的源代码。

同一天，华盛顿大学蛋白质设计研究所戴维·贝克团队在《科学》刊文，推出一款名为RosettaFold的人工智能程序。该程序基于深度学习，能根据有限的信息快速、准确地预测出目标蛋白质的结构，“达到与AlphaFold2不相上下的准确度”。

2020年5月至7月，在CASP14上，AlphaFold2以排名第一的准确性轰动一时。一时间，AlphaFold 2“颠覆”“革命性突破”“诺奖级成果”等美誉加身。

很多结构生物学家还未完全从AlphaFold2开源和RosettaFold诞生带来的震撼中回过神来。7月22日，DeepMind团队和欧洲生物信息学研究所（EMBL-EBI）联合在《自然》发表论文，公开AlphaFold2预测的蛋白质结构数据库（AlphaFold DB）。初始的AlphaFold DB涵盖了属于人类以及其他20个重要物种的大多数具有较大价值的蛋白质，包含超过35万个不同的蛋白结构，最终将增加到约1.3亿个三维结构。

“这会让结构生物学乃至整个生命科学上个大台阶。”清华大学结构生物学高精尖创新中心执行主任王宏伟对《中国科学报》说，“原来大家要用很多实验手段去解析单链蛋白质的结构，现在由于高水平结构预测软件的出现，对单链蛋白质实验解析的需求可能没以前那么高了。但另一方面，对多个蛋白质或核酸分子形成的复合体进行结构解析的迫切性会更强，所以对冷冻电镜的技术需求量会更大。”

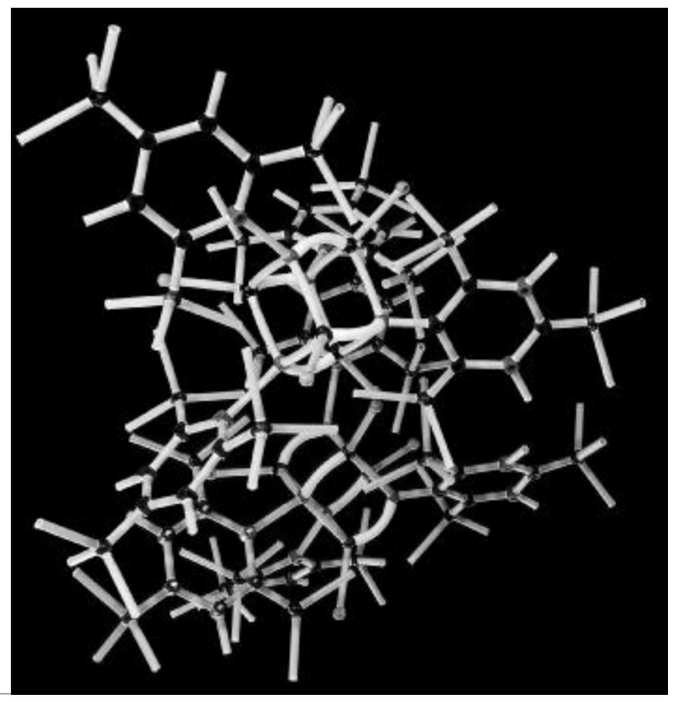
王宏伟认为，这两款软件的开源预示着结构生物学进入新时代，“未来结构生物学的研究对象和研究方式上都会发生较大变化，这实际上是给整个结构生物学领域的升级带来了新的机会”。

“我们已经买新电脑了。”北京大学生命科学学院教授孔道春告诉《中国科学报》。

这两款软件开源后，孔道春团队就迅速配备了显卡更好的电脑。

“我已经让学生用起来了。”孔道春说，“利用传统实验方法解析蛋白结构需要跨越诸多障碍，不仅耗时、费力，还不一定能解析出来。与核磁共振、X射线晶体或冷冻电镜等类似，这些软件是新的、革命性的工具，将极大推动人们对蛋白质/酶的结构和生化作用机理的理解，将对生命科学、医药研究起到极大推动作用，甚至会大大加速人类文明的进程。”

> 软件技术的进步使蛋白质结构变得“唾手可得”，这将对结构生物学的研究范式产生重大影响。



蛋白质分子结构。 图片来源：unsplash

“本尊”和“复现者”

“这两款软件的基本原理都是利用神经网络，依托现有的大数据进行训练，当然也包括很多专业的算法，把这几方面整合到一起，应该说是现在蛋白质结构预测精确度最高的两款软件。”王宏伟说。

“两个软件各有所长，各有自己的特点。”中国科学院大学人工智能学院教授、中国科学院自动化研究所模式识别国家重点实验室研究员杨戈对《中国科学报》说，“可以从三个方面对它们进行比较。”

一是准确度。两者相较而言，AlphaFold2的准确度更高。AlphaFold2预测蛋白质结构的精度已经达到埃（长度单位，1埃相当0.1纳米）级，这是它的最大优势。

二是预测蛋白的复杂程度。这点RosettaFold略胜一筹。AlphaFold2只能预测单个蛋白质，即一个氨基酸链的蛋白，而RosettaFold可以预测蛋白质复合体，即两个乃至数个有相互作用的蛋白质。

三是对计算资源的要求方面，AlphaFold2的要求较高。“AlphaFold2在模型训练阶段对计算资源的要求一般计算中心才能满足，普通的实验室不大可能使用。”而RosettaFold的要求通常单个实验室就能满足，“具备稍好一些的计算机系统就可以‘跑’起来”。

清华大学结构生物学高精尖创新中心研究员龚海鹏介绍说，AlphaFold的第一版和RosettaFold之前的版本，包括其他团队的思路都差不多，比如，先预测氨基酸残基之间的距离，通过一些图像识别算法识别，然后再去折叠蛋白。

“那时候虽然大家的调参能力不同，但相互之间没有本质的区别。”龚海鹏说，“但AlphaFold2采用了全新的架构，从去年参加CASP14开始就崭露头角。”

2020年12月，AlphaFold2的主要研发者John Jumper作了一次报告，简单介绍了一下他们的思路，但很多细节并没有披露出来。

“因为AlphaFold2的准确率非常高，当时几乎所有研究组都想知道他们是怎么做的，有很多人想去复现它。RosettaFold是过去几个月里复现得比较快的，也是复现得最好的，他们根据AlphaFold2释放出来的一些信息，相当于做了一个简化版本。”龚海鹏说，“很多研究组都进行过测试，我觉得在预测精度和准确程度上，RosettaFold离AlphaFold2还有一定的距离，其效果并没有宣称的那样好。”

这两款软件开源完全版后，龚海鹏团队对比发现，两者主体思想虽然差不多，但还是能看出有较大的区别。“有很多细节，AlphaFold2的设计更合理，因此它的效果也更好。”而现在一些自媒体和宣传材料称两者功能相当，甚至RosettaFold的某些方面表现更好，配置要求更低，“这可能会有些误导”。

“AlphaFold2对显卡的要求并不是特别高，预测的时候，如果不是特别长的蛋白链，比如，预测几百个残基、上千个残基，1080Ti这样的显卡就能‘跑’了。但要预测2000多个残基的蛋白链，就需要市面上最好的A100显卡。”龚海鹏说，“在预测方面，RosettaFold并没有太大优势，它在训练上要求的资源少一些。从双方发表的文章来看，AlphaFold2在训练的时候，资源占用大概是RosettaFold的十几倍，但模型训练好后，真正预测的时候两者对资源的要求并没有太大区别。”

坚持“搞事情”

软件技术的进步使蛋白质结构变得“唾手可得”，这将对结构生物学的研究范式产生重大影响。

“预计会有一批实验室转换研究方向，不再做结构预测的方法研究，转而研究下游的一些问题，比如怎么用这个工具去做一些事情。但我们还会沿着这条路走下去。”龚海鹏说，“一是因为AlphaFold2的思路不是唯一的解法。二是受其他因素影响，国内的研究团队不能随时和谷歌合作，很难用上谷歌最新的模型，所以我们需要有一个自己的版本。”

“对这个领域来说，AlphaFold2可以说改变了不少人的理念。以前生物学家可能觉得人工智能只是一个好的工具，但现在，说它将对这个领域带来革命性的影响一点都不过。”杨戈说。

2019年，在美国学习生活了20多年的杨戈回国，到中科院自动化所从事计算生物学方面的研究。回国后他发现，国内的生物技术研究、原创性制药行业远远没有发展起来，甚至有些学生认为生物学是个避之不及的“天坑专业”。

“如果不能很好地抓住发展机会，计算生物学可能就会成为我们的‘卡脖子’问题，其背后的新药开发研制都会被‘卡脖子’。”杨戈说。

龚海鹏认为，DeepMind团队的人才、硬件、软件方面的能力都很强，它能解决的训练问题一般的实验室或小团队很难去复现。我们拿它直接去训练，多半训练不出来，所以我们只能参考它的方法，开发出一些训练代价较小的等价方法。

“达到同一个目的，不会只有一条路。”龚海鹏说，“我们还会一直做下去，包括我了解的几个课题组都是这样，大家会从不同的角度汲取它的优点，融入自己的方法中继续发展。”

相关论文信息：
https://doi.org/10.1038/s41586-021-03819-2
https://doi.org/10.1126/science.abj8754
https://doi.org/10.1038/s41586-021-03828-1

中国首艘智能型无人系统母船在广州开建

本报讯（记者朱汉斌 通讯员吴立坚）近日，记者从南方海洋科学与工程广东省实验室（珠海）（以下简称南方海洋实验室）获悉，我国智能型无人系统母船在广州开工建造。该船有望成为全球首艘具有远程遥控和开阔水域自主航行功能的科考船，将为我国开展海洋科考提供利器。

“智能型无人系统母船是美丽的、全新的‘海洋物种’，将使观测海洋的模式发生革命性的变化。期望参建各方齐心协力、敢为人先，优质高效安全完成好该船的建造工作。”中国科学院院士、南方海洋实验室主任陈大可表示，智能型无人系统母船由中国舰船研究设计中心设计、黄埔文冲船厂建造，贯彻了“未来感”“无人系统保障”“绿色智能”三大设计理念。

作为我国首艘智能型无人系统母船，其所配备的重要设备国产化率较高，“所携带的动力系统、推进系统、智能系统、调查作业支持系统等均为中国制造，核心技术自主可控。”中船黄埔文冲船舶有限公司总建造师樊雷说。

2020年12月，南方海洋实验室与中国舰船研究设计中心、中船黄埔文冲船舶有限公司共同签署了智能型无人系统母船的设计建造合同，将建造中国首艘智能型无人系统母船，项目预估设计及建造周期为18～20个月，预计2022年交付使用。

据了解，智能型无人系统母船长88.5米，型宽14.0米，型深6.1米，设计吃水3.7米，设计排水量约2000吨，最大航速18节，经济航速为13节。该船拥有宽敞的甲板，可搭载数十台配置不同观测仪器的空、海、潜无人系统装备，在目标海区批量化布放，进行面向任务的自适应组网，实现对特定目标的立体动态观测，是南方海洋实验室智能快速机动海洋立体观测系统（IMOSOS）的水面支持平台。

南方海洋实验室海洋智能无人装备创新团队IMOSOS项目于去年正式立项。“利用IMOSOS系统可实时获取和评估不同空间尺度海洋环境信息，预测海洋资源、环境和气候的时空变化，研究和创新海洋多尺度变化及其气候资源效应机理。”相关专家介绍说。

据介绍，IMOSOS系统不仅可以为海洋防灾减灾、海底精细测绘、海洋环境监测、海上风电场维护等提供智慧高效的工具，同时也将为国家海洋事业和地区社会经济发展提供全面、精准的海洋信息服务。

速递

中国学者当选俄罗斯自然科学院外籍院士

本报讯 近日，俄罗斯自然科学院（RAEN）O.L. Kuznetsov院长向东南大学自动化学院教授李新德发来贺信，祝贺他当选为俄罗斯自然科学院外籍院士。

李新德长期从事人工智能、智能机器人、机器视觉感知、学习与理解、人机自然交互、智能信息处理等方面的研究。他首次提出了基于知识图谱推理的场所理解方法，为跨域无人系统提供了必要的解决方案，同时凭借其在自然语言理解的机器人交互式视觉导航领域的成就，为机器人顺利走进家庭、服务场所奠定了基础。他还在国内最早开展似是而非理论（DSmT）和多粒度信度融合理论研究及应用，并开展视觉非接触人体、情绪、精神状态分析的研究，为抑郁症早期筛查创造了条件。

据悉，俄罗斯自然科学院是联合国认可的俄罗斯规模最大的科学院，成员均为自然科学和人文科学领域取得重大成就的科学家和专家，具有重要学术影响力。科学院设立24个学部，有18名诺贝尔奖获得者、270多名俄罗斯科学院院士以及30多名俄罗斯国家医学科学院院士。（张思玮）

新一代超高清AI相机芯片AR9341发布

本报讯 近日，上海酷芯微电子有限公司（以下简称酷芯微电子）发布新一代人工智能（AI）相机芯片AR9341。该芯片是酷芯微电子推出的第二代超高清智能相机芯片。

酷芯微电子联合创始人兼CTO沈泊表示，目前市场对中高端智能相机芯片的需求十分迫切。AR9341芯片适合的应用领域广，包括高端智能IPC、车载辅助驾驶、边缘计算盒子、智能机器人等。

图像信号处理（ISP）画质好、AI算力充足、具有强大的适应能力等，是AR9341芯片的典型特点。

除了芯片硬件性能强悍外，酷芯微电子还提供软件开发工具包、AI工具链、算法等一站式解决方案。其中，“AI工具链提供简洁易用的可视化图形界面，支持Pytorch/Tensorflow/Caffe等多种框架，让用户在很短时间内完成算法到芯片的部署工作。”沈泊说。

据悉，AR9341工程样片将于今年9月提供，12月量产。（秦志伟）

Alphabet孵化公司用机器人训练机器人

本报讯 谷歌的母公司Alphabet成功孵化工业机器人公司Intrinsic。据了解，成立Intrinsic的目的是探索如何将自动感知、深度学习、强化学习、运动规划、力控制和模拟等技术结合起来，使工业机器人更加有用和灵活。

Intrinsic的首个项目是让机械手学会了插插座。通过深度学习和强化学习训练，这类型的机器人可以快速适应新的任务。在另一个示例中，两个Intrinsic的机器人学会了相互协作，高效拼装家具。通过在机器人和生产环境中布置的各类传感器，Intrinsic发明的机器人可以快速感知并适应新的环境。（袁一雪）

耗电大户何以践行“碳中和”？

——探访阿里云张北数据中心

■本报记者 赵广立

数字时代，我们在生产生活中产生的海量数据，它们的存、算、传、用都离不开数据中心这类“新基建”，数据中心也顺理成章地成为当今社会的耗电大户。据“中国IDC圈”统计，2019年数据中心总耗电量超过2045亿千瓦时，占全社会用电总量超过2.4%。

随着我国“碳达峰”“碳中和”目标的提出，数据中心的能耗问题逐渐成为各方关注的焦点。近期，国家发展改革委等4部门更是印发《全国一体化大数据中心协同创新体系算力枢纽实施方案》，明确提出“推动数据中心绿色可持续发展”“加强绿色数据中心建设，强化节能降耗要求”。

把低碳绿色从愿景变成现实，大型、超大型数据中心都有哪些硬招、实招？近日，《中国科学报》实地探访了阿里云张北数据中心。

一年300天，制冷“不花钱”

在北京西北200多公里处的张北县草原天路附近，这座数据中心迎来一年最热的节气——大暑。但即便在三伏天，空调在这里的使用率也并不高。据查，张北坐拥得天独厚的气候优势，年均气温只有2.6℃，年内最低气温更是曾创下零下40℃的纪录。

与其说张北数据中心空调利用率极低，倒不如说这里更多用的是“天然空调”。在园区的气冷机房，运转着的不是空调，而是一种类似于新风系统的设备——AHU风墙。

AHU是Air Handler Unit（风机矩阵空气处理单元）的缩写。当室外温度低于设置值（如25℃）时，AHU设备将室外冷空气经过滤及湿度处理后直接送入数据机房；当室外温度高于设置值时，或通过喷淋降温及过滤后送入数据机房，或启动备用制冷空调为机房服务器降温。

“AHU风墙技术的使用可以大大减少空调机组的运行、实现节能。”阿里云基础设施数据中心总经理高山渊告诉《中国科学报》，在张北数据中心，几乎每年都有300多天可以利用室外冷空气为数据机房降温。

高山渊说，经测算，张北数据中心的电源使用效率（即PUE值，数值越接近1表明效能越高）在冬天最低可达1.09，在夏天也只有1.3左右。

将服务器泡在液体里

在节能降耗方面，除了充分利用自然风冷，阿里云还自研出一套“将服务器泡在水里”的黑科技。

在张北数据中心的液冷机房，可以看到一排排价值不菲的服务器浸泡在液体中，凑近看去，还能看到有些部件在液体中闪着光，活像科幻大片里的桥段。

这就是浸没式液冷技术。浸没式液冷依赖于一种特殊的绝缘冷却液，冷却液与服务器各元器件零距离全方位接触，器件在运行中产生的热量将直接被吸收进入外循环冷却。这种冷却方式不需要开启空调，全程用于散热的能耗几乎为零，整个机房也非常安静。

“浸没式液冷的节能效果超过70%，实现了数据中心100%无机械制冷。”高山渊说，“如果将浸没式液冷向全国推广，那么全国数据中心的PUE都会降低到1.1以下。”

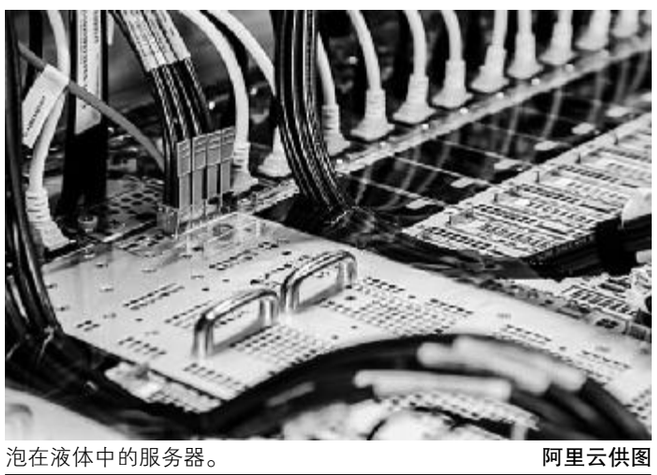
高山渊说，随着未来对能耗密集型服务器（如人工智能服务器）的需求加剧，浸没式液冷或许是唯一解。

液冷的好处不仅体现在散热方面，还在于它能够提升设备的稳定性、降低设备事故率。高山渊告诉《中国科学报》，液冷机房运行3年来，与同等规模的其他机房相比，事故率降低了54%。

但是，浸没式液冷也不是全无死角的“六边形战士”。一方面，液冷虽然从全生命周期来看成本还可以接受，但它的一次性投入成本很高；另一方面，绝缘冷却液跟各类元器件的“磨合”还需要时间给予证明。比如，目前还未校验它与GPU等计算单元的兼容性如何。此外，囿于生产工艺和技术，目前绝缘冷却液距离实现国产化还有一段路程。

“减碳三环”打造“零碳云”

高山渊说，加上模块化设计、AI调温等技术，张北数据中心的全年PUE低于1.2，最低可以达到1.09——这是一个领先行业的数字，这一能效约等于每年可节约标煤8万吨，相当于种植了400万棵树木。



泡在液体中的服务器。 阿里云供图

用大自然的冷风吹、用绝缘冷却液浸泡，把数据中心的PUE值降低到接近1，提高了数据中心的能效。不过，数据中心仍是耗电大户——据“中国IDC圈”统计，2019年数据中心总耗电量超过2045亿千瓦时，占全社会用电总量超过2.4%。

数据中心负荷实在太大了，就算是能耗全部用来支撑计算，它们全年无休地开机运行，用电量也是一个天文数字。当这个天文数字遇到“碳中和”这样的宏伟目标，无疑是一个“需要解决的问题”。

好在张北有“绿电”。早在七八年前，张北就是“广袤的原野上风车林立，数千亩光伏面板如波荡漾”，经过这些年的发展，风车和光伏电板已经成为张家口市的重要电力来源。

使用“绿电”，成为张北数据中心的必然之选。

高山渊透露，张北数据中心无疑是张北绿电的消纳大户。2018年起，阿里就加入张家口“四方协作机制”风电交易，截至今年5月，共交易绿电约4.5亿千瓦时，累计减排二氧化碳近40万吨。

从自身节能减排做起，还只是阿里云数据中心迈向“零碳云”的一环。今年5月，阿里云发布“零碳云”计划，希望在推动自身节能减排的同时，向生态企业输出数字减碳能力、支持绿色技术创新。他们希望向电力能源、钢铁、交通、制造等碳排放大户提供高效云平台支持，为其引入大数据、人工智能技术，帮助上云企业提高效率、节能降耗。

比如，通过向攀钢集团引入阿里云工业大脑，对其炼钢全流程进行工艺优化，帮助攀钢旗下的西昌钢钒公司炼钢厂节省了25%的人工、每生产一吨钢节省1.28公斤铁，生产效益提升2.4倍。在西南某大型垃圾焚烧发电机组上，阿里云利用优化的人工智能算法帮助客户将固废垃圾焚烧效率提升2.6%，相当于燃烧同样的垃圾每年多发电4000多万千瓦时，碳排放相比之前降低约48%。

践行“碳中和”，打造“零碳云”，阿里希望利用数字化能力做好“减碳三环”：自身节能减排的“内环”、推动生态企业脱碳减排的“中环”、公众绿色低碳消费的“外环”。

就像阿里巴巴首席技术官程立说的那样，“碳中和”不仅是环保概念，更是技术路线，在落实“双碳”战略过程中，数字基建会朝着绿色基建迈进。