

AI 加速破解生命密码

本报记者 李佳师

每一个产业风口的开启,总有一些标志性事件。就像2016年 Google DeepMind 的 AlphaGo 与李世石下棋,2020 年年底 Google DeepMind 的 AlphaFold2 参加 2020 年第 14 届国际蛋白质结构预测竞赛(CASP14)。前者昭示了 AI 第三次浪潮的来临,而后者有可能开启生命科技数字化的黄金十年。

今年7月15日,谷歌 DeepMind 团队与华盛顿大学贝克团队(David Baker)分别开源了 AlphaFold2 与 RoseTTAFold。当这两种创造了前所未有准确度的预测蛋白质结构算法开源后,预示着用 AI 破解生命密码的时代拉开大幕。而在这场用 AI 破解蛋白质结构的全球竞赛中,中国阵营给出了不错的成绩单。日前,来自中国的人工智能企业天壤宣布,其自研的深度学习蛋白质折叠预测平台 TRFold 在基于 CASP14 蛋白质测试集的评估中排名全球第二,仅次于 AlphaFold2。在对 400 个氨基酸的蛋白链预测时,TRFold 仅耗时 16 秒。这是目前国内所有公开蛋白质结构预测模型中取得的最好成绩,标志着我国在计算生物学领域的表现已经处于世界第一梯队。与此同时,深势科技也宣布推出蛋白质结构预测工具 Uni-Fold。它能够复现 AlphaFold2 的全规模训练,并开源训练代码与推理代码。相应解决方案已集成至深势科技自主研发的药物设计平台 Hermite,供用户测试使用。用 AI 预测蛋白质结构,对于加速生命科技的数字化至关重要,在全球生物数字化的竞赛中,中国阵营正采用新思路、新路径,加速推进。



生命数字化开启新风口 蛋白质结构预测是关键

拼多多创始人黄峥和字节跳动创始人张一鸣宣布卸任 CEO 之时,未来规划中有一个共同选项,就是希望投入更多的精力在生命科技的研究与探索上。事实上,其他如百度创始人李彦宏、阿里创始人马云等,都将个人的下一个兴趣目标定在了生命科技上。生物世界的数字化,正在成为新风口。

新晋中国工程院外籍院士张亚勤几天前表示,过去 30 年,信息产业推动了内容的数字化、企业的数字化,而未来的重点是物理世界的数字化和生物世界的数字化,下一个 10 年是 AI 与生物制药融合的大好时机。

“一方面我们的身体从大脑、器官到细胞、蛋白质、基因、分子都在数字化,另一方面人工智能算法、算力和系统的快速发展让大量数据有了使用的场所。”张亚勤说,以前新药研发需要十几年的周期,10 亿美元的投入,AI 正在改变这种状况。

每一个产业的发展都有一些关键基石,而生物数字化、生命数字化的一个关

键是利用数字技术破解蛋白质结构之谜。

蛋白质为什么关键,因为它是生命构成的基本要素,用北京航空航天大学大数据精准医疗高精尖创新中心特聘研究员叶盛的话来说,人体细胞乃至地球上任何一种生命的细胞,都是由蛋白质构成的。可以说,几乎所有的生物学问题都会牵涉蛋白质,而几乎所有的蛋白质功能问题最终都要通过结构研究去回答。不同的蛋白质通过折叠构成形态各异的三维结构,执行多种多样的生理功能。

长期以来,从氨基酸序列到对应蛋白质三维结构的预测问题被认为是生物学领域最具有挑战性的问题之一。蛋白质结构研究为什么难? 是因为蛋白质的尺寸为纳米级,比人类肉眼的可见光波还要小,超越了光学显微镜的观察极限,因此要想研究蛋白质的结构,必须采用某种间接的方式。

而且解蛋白质结构之题,远远难于基因。“蛋白质结构测定与基因组测序最大的区别在于,基因组序列是一维的、线性

的;而蛋白质结构是三维的。因此,基因组测序无论测什么物种的基因组,本质上都是在做同一件事,可以通过机器进行自动化、规模化的操作,而蛋白质结构测定则是不同的课题,科学家始终没有找到批量处理的方法。”叶盛说。

在过去几十年中,结构生物学家为探测蛋白质结构研究出三大实验手段——X 射线晶体学、核磁共振和冷冻电镜,但实验方法成本高、周期长。目前人类有数十亿个已知蛋白质序列,还原出结构的却只有十几个万。生物学发展因此颇受掣肘。

人工智能的发展为解决这一问题带来了一些新的可能。在 2020 年的国际蛋白质结构预测顶级竞赛 CASP14 上,Google DeepMind 推出 AlphaFold2 (简称 AF2) 预测的大部分单体蛋白质结构准确度与实验足够接近,远超此前所有方法。这一结果也标志着蛋白质单体结构的问题在一定意义上得到了解决。中国结构生物学家施一公对此的评价为“这是 AI 对科学领域的最大一次贡献”。结构生物学家 Petr Leiman

未来的重点是物理世界的数字化和生物世界的数字化,下一个 10 年是 AI 与生物制药融合的大好时机。

表示:“我用价值 1000 万美元的电镜努力了好几年,AF2 竟然一下就算出来了。”

除了 AlphaFold2,另外一个被高度关注的 AI 算法是华盛顿大学医学院蛋白质设计研究所的贝克团队,有人说因为 DeepMind 的成功触动贝克团队重视 AI,于是他们联合很多大学与机构研发了一款基于深度学习的蛋白质预测新工具 RoseTTAFold,并在预测蛋白质结构上取得了媲美 AF2 的超高准确率,而且速度更快、所需要的计算机处理能力也较低。

总之,AlphaFold2 与 RoseTTAFold 是 AI 预测蛋白质结构的两个“明星”,破解了出现 50 年之久的蛋白质分子折叠问题,打开了人类预测蛋白质结构的另一扇窗。今年7月,DeepMind 团队与贝克团队分别开源了他们各自的算法,这给全球研究机构进行蛋白质结构预测带来福音,大大降低了研究蛋白质结构的门槛,将帮助科研人员弄清引发某些疾病的机制,并为研发药物、农作物增产,以及可降解塑料的“超级酶”等铺平道路。

利用 AI 解密蛋白质结构之谜的大门才刚刚开启,在这个新赛道中,中国必须参与其中。

模型表达方面,AF2 采用的是端到端模型,而 TRFold 采用的是分段式结构。这些创新和优化,意味着 TRFold 所用的算力资源更少、产生结果的速度更快,在大规模蛋白与蛋白的相互关系分析上,更具有优势,而且能够加速应用在不同场景中。

与此同时,包括中科院、腾讯、复旦大学等 AI 企业、研究机构也加入了这场竞赛。其中,深势科技推出的蛋白质结构预测工具 Uni-Fold,克服了 AF2 未开源训练代码、硬件支持单一、模型不可商用等局限性,复现 AF2 的全规模训练,并开源训练代码与推理代码,为更多人参与推动该领域进一步发展提供了基础。

利用 AI 破解蛋白质结构之谜,将改变人类对生命的认知,有可能重塑生物应用的逻辑

能加速疫苗开发的例子。另外迁移学习用少量原始数据加上动物模型快速发现了对罕见病的药物,几何深度学习找出了广谱、稳定的新冠抗体,对变种株也有效。

李彦宏今年投资了一家生物计算公司——百图生科。他认为,当前生物计算的发展十分迅速,与 20 年前的互联网有诸多相似之处。基因组学研究带来的人体数据、新药研发过程当中所积累的知识和新生的各类机器学习算法三大关键指标在快速增长,会带来巨大的突破和进步。他认为:“生物计算产业的发展,需要生态和产业链的协同,一家乃至 100 家的企业可能都远远不够,我们希望和科学家、企业家一起构建开放的生物计算创新生态,去探索广袤浩瀚的生命科学无人区。”

计算与生物科学的融合,是没有尽头的星辰大海,希望中国企业、研究机构能够在其中贡献更多的中国智慧。

5G 行业虚拟专网正在从 1.0 向 2.0 迈进

本报记者 刘晶

近日,IMT-2020(5G)大会在深圳召开。在此次峰会上,“5G 扬帆计划”的诸多热点问题被与会各方探讨,其中 5G 行业应用是最大的话题。中国通信标准化协会副理事长兼秘书长闻库表示,目前 5G 行业虚拟专网正在从 1.0 向 2.0 迈进,由 To B 通用网络向各行业定制的网络演进。

5G 行业专网超 2300 张 模组价格下探 500 元

会上信息显示,截至今年 10 月底,我国累计建成 5G 基站 129.1 万个,覆盖全国所有地级以上城市市区、97% 以上的县区以及 50% 的乡镇镇区,行业虚拟专网数量超过 2300 张;5G 手机、模组价格分别下探至 1000 元和 500 元,工业网关、车载终端、智能穿戴等新型终端不断涌现;终端连接数超过 4.7 亿,用户渗透率达到 28.7%,5G 应用已覆盖一、二、三产业的诸多领域,部分重点领域已进入商业落地阶段。

闻库说,目前 5G 行业虚拟专网正在从 1.0 向 2.0 迈进,由 To B 通用网络向各行业定制的网络演进,为持续推动 5G 行业虚拟专网的发展,建议如下:

第一,统筹 5G 行业虚拟专网标准。积极发挥中国通信标准化协会、5G 应用产业方阵等行业组织作用,加速推进与电力、钢铁、矿山等行业标准化组织的跨行业合作,形成各行业标准化落地保障体系,率先推动网络模板标准落地。

第二,夯实 5G 行业虚拟专网产业基础。加快行业定制化 UPF、核心网及基站设备的研发及产业化,开展 5G 网络服务能力平台及边缘计算平台的研发,形成 5G 行业虚拟专网的定制化运营运维;加强行业 5G 芯片、模组、终端的技术攻关和产业化。

第三,构建 5G 行业虚拟专网安全体系。建立行业 5G 安全保障模板,以行业为主体形成新型安全体系;开展云管边端各环节安全保障技术的研究,推进行业 5G 安全组网架构及关键设备的开发,提升 5G 行业虚拟专网的安全保障能力。

通过跃变实现 5G 应用纵深发展

中国电信副总经理夏冰认为,新基建承载着 5G 发展新实践的物理载体,但静态新基建不会自动成长为数字经济的新动力,要通过跃变来推动 5G 应用的纵深发展。

要推动 5G 创新,实现从技术创新向应用创新跃变。中国电信正在推动 5G 发展从技术的“低广高”(低时延、广连接和高速率)向应用的“低广高”(低成本、广场景和高可靠)转变。目前,中国电信自主研发的轻量级 UPF、移频 MIMO 系统、一体化边缘云平台、5G 应用加速平台等创新产品,均为广大用户提供了高可靠的应用保障。

要推动 5G 增效,实现从业务复制向能力复用跃变。面对行业客户差异化需求,中国电信将能力原子化,创新推出 5G 能力魔方,通过能力的标准化和模块化,实现行业场景需求挖掘时间缩短 30%,解决方案制定周期降低 20%,商务模式和报价时间缩短 30%,交付成本降低 40%,促进了 5G 规模复制及商业闭环。

要推动 5G 赋能,实现从孤岛驱动向生态驱动跃变。中国电信牵头成立了天翼物联产业联盟和 5G 产业创新联盟,深化构建 NICES 产品生态体系,形成覆盖 20 类行业、200 多个场景的 5G 应用方案。联合合作伙伴共同打造深圳燃气“5G+智慧燃气”数字赋能超大城市公共安全项目、智慧南山“5G 赋能先行示范智能城市治理标杆”等标杆应用。

推动 5G 向行业需求演进

面向未来,在 To C 个人消费市场,随着高清视频及 XR 类业务规模化发展,预计 2025 年热点地区流量将达到 50GB~100GB,2030 年最高可达 250GB,业务体验将对网络保障能力提出新挑战。

在 To B 行业应用市场,5G 在驱动行业向更深层次转型的同时,行业应用也对 5G 提出了“网络无所不达、算力无处不在、能力无所不及”的更高发展要求。因此,亟须产业协作,凝聚共识,共同推进 5G 演进技术 5G-Advanced。

中国移动研究院副院长黄宇红表示,中国移动将围绕三大方向、十大关键技术,构筑 5G 下一站。三大方向包括卓越网络、智生智简和低碳高效。在卓越网络方面,5G 将围绕天地一体、通感一体、多媒体业务体验提升、极致覆盖四大关键技术持续演进,为用户构建一张无所不达、极致体验的 5G 网络。在智生智简方面,将围绕网络自动驾驶、云网融合/算网一体、无源物联网三大关键技术,加速推进网络实现智能化、能力开放化,降低 5G 建设成本和运营成本,提供更为便捷的 5G 服务,加速 5G 在各行各业的普及和渗透。在低碳高效方面,将围绕绿色节能、全双工、弹性小区三大关键技术,提升 5G 设备站点和网络的能效,并通过 5G 赋能各行各业实现全产业链低碳节能,助力国家实现双碳目标。