《机器智能》：秒级预测蛋白质结构的大模型上线

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/24503.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

《机器智能》：秒级预测蛋白质结构的大模型上线。10月9日，国际学术期刊《自然》旗下子刊《机器智能》发表了飞桨螺旋桨联合百图生科研发的文心生物计算大模型的一项成果：《一种使用蛋白质语言模型的免多序列比对蛋白质结构预测方法》，提出了全球首个开源、并提供在线服务，无需MSA输入的蛋白结构预测大模型HelixFold-Single。据介绍，该预测大模型打破了依赖MSA检索模型的速度瓶颈，将蛋白结构预测速度平均提高数百倍，可实现秒级预测。

据介绍，该工作是在生物计算领域继HelixGEM和Linear Design之后，在蛋白领域的又一突破性成果。近年来，AI一直致力于突破蛋白质的结构预测问题，并在预测精度方面取得了重大进展。然而，以 AlphaFold2 模型为代表的主流蛋白质结构预测方法比较依赖于多序列比对（MSAs）和模板提取的协同进化信息。该项研究打破了依赖MSA检索模型的速度瓶颈，相比AlphaFold2，HelixFold-Single模型推理速度平均提升数百倍，实现了秒级预测。以蛋白长度697的门蛋白7et2_H为例，用AlphaFold2预测其结构需要21分钟，而HelixFold-Single只需要11秒，速度提高了115倍。

该工作的发表为产学研各界带来了使用门槛更低、适用范围更广的蛋白结构预测解决方案，有望促进我国生命科学、生物医药、蛋白研究等领域的发展。

据了解，HelixFold-Single目前已经在国家超算成都中心部署，通过超算平台赋能蛋白领域的科学研究机构。在大分子药物的应用场景上，HelixFold-Single也已经整合进入百图生科AIGP平台，为百图提供更高效的蛋白分析能力，助力其探索大分子创新药。

另据研发团队介绍，针对更具挑战性的抗原抗体、多肽蛋白的相互作用场景，他们还研发了更具通用性和鲁棒性的复合体结构预测算法HelixFold-Multimer，相比业界同类方法，精度提升了数倍，该工作也将在近期上线。（来源：中国科学报 赵广立）

相关论文信息：https://www.nature.com/articles/s42256-023-00721-6

作者：方晓敏等 来源：《机器智能》

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发