AI预测蛋白质结构的数据库可能改变生物学

作者：writer 来源：爱科学

本文原地址：https://www.iikx.com/news/progress/14833.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！

AI预测蛋白质结构的数据库可能改变生物学。近日，两个小组公布了计算机科学家、生物学家和物理学家多年工作的成果：先进的建模程序，可以预测蛋白质和一些分子复合物的精确三维原子结构。其中一个研究小组报告说，他们已经使用最新开发的人工智能（AI）程序解决了35万种来自人类和20种模式生物的蛋白质结构——如大肠杆菌、酵母和果蝇。未来几个月里，他们计划将所有已编目的蛋白质列入模型蛋白质名单，大约有1亿个分子。

这是相当惊人的。美国马里兰大学蛋白质折叠专家John Moult说，他每两年举办一次名为关键的蛋白质结构预测方法（CASP）的比赛。Moult表示，几十年来，结构生物学家一直梦想有一天，精确的计算机模型将增加从X射线晶体学等实验方法中获得的极其精确的蛋白质形状的数量。我从没想过这个梦想会成真。Moult说。

这一名为AlphaFold的模型是DeepMind研究人员的成果，DeepMind是一家隶属于谷歌母公司Alphabet的英国AI公司。2020年秋，AlphaFold横扫了CASP竞赛，准确率中位数为92.4（满分100），远远领先于紧随其后的竞争对手。

但是DeepMind的研究人员没有透露他们如何从理论上绘制蛋白质形状的细节，特别是AlphaFold的底层计算机代码。这令其他团队感到沮丧，因为他们无法在这一进展的基础继续前进。

但是这种情况已开始改变。7月15日，西雅图华盛顿大学Minkyung Baek和David Baker研究小组报告说，他们创建了一个名为RoseTTAFold的高度精确的蛋白质结构预测程序，并公开发布了该程序。相关成果在线发表于《科学》。

同一天，《自然》发表了一篇由DeepMind研究人员Demis Hassabis和John Jumper研究小组撰写的论文，公布了AlphaFold的细节。

两个程序都使用AI在庞大的蛋白质结构数据库中识别折叠模式。这些程序通过考虑蛋白质中相邻氨基酸相互作用的基本物理和生物学规则，计算出未知蛋白质最可能的结构。论文显示，Baek和Baker使用RoseTTAFold创建了一个包含数百个G蛋白偶联受体（一类常见的药物靶点）的结构数据库。

而DeepMind的研究人员在《自然》上报告说，他们创造了35万个预测结构，是以前实验方法得到的结果的两倍多。研究人员表示，AlphaFold生产了近44%的人类蛋白质结构，涵盖了人类基因组编码的近60%的氨基酸。AlphaFold确定许多其他人类蛋白质是无序的，这意味着它们的形状不是单一结构。Baker说，当这些无序的蛋白质与蛋白质伴侣结合时，最终可能会采用某种结构。

加州大学旧金山分校结构生物学家 David Agard说，它们也可能会自然地采用多种构造。

DeepMind与欧洲分子生物学实验室（EMBL）合作共建了一个新的蛋白质预测数据库，可以在网上免费访问。能提供这种服务真是太棒了。 Baker说，这真的会加快研究步伐。

因为蛋白质的3D结构在很大程度上决定了它的功能，所以这一数据库便于生物学家厘清数千种未知蛋白质是如何工作的。

DeepMind的合作者表示，AlphaFold已经促进了新型酶的开发，这种酶能更快地在环境中分解塑料，并为被忽视疾病的药物开发带来了新可能性。这是自绘制人类基因组以来最重要的数据集之一。EMBL欧洲生物信息学研究所主任Ewan Birney说。

Baek说，这些预测将有助于实验学家解决结构问题。来自X射线晶体学和冷冻电镜实验的数据可能很难解释，模型将会有所帮助。（来源：中国科学报文乐乐）

相关论文信息：https://doi.org/10.1126/science.abj8754

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。
作者：Minkyung Baek 来源：《科学》

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