AI模型以更少数据精准预测RNA三维结构

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/40678.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模型以更少数据精准预测RNA三维结构。科技日报北京6月30日电 （记者张佳欣）美国弗吉尼亚理工大学科学家团队开发出一种名为“ RNAbpFlow ”的AI模型，在RNA结构预测任务中的表现可媲美谷歌旗下“ 深度思维 ”公司开发的“ 阿尔法折叠3 ”，且无需依赖海量进化序列数据库，有望加速RNA靶向药物研发。相关研究发表于新一期《自然 · 方法》杂志。

RNA是mRNA疫苗等生物技术的核心分子，但其三维结构复杂且不断变化，比蛋白质更难预测。长期以来，由于RNA相关结构数据稀缺，科学家难以准确描绘其空间构象，这也成为RNA药物研发的一大瓶颈。

研究团队介绍，在一项国际通用基准的盲测中，RNAbpFlow成功预测了14个RNA目标中的12个正确结构，而“ 阿尔法折叠3 ”只预测了8个。更重要的是，RNAbpFlow无需依赖大多数主流工具所需的大规模进化序列数据库，仅利用RNA序列及碱基配对信息即可完成预测。

与“ 阿尔法折叠 ”等主要依靠进化信息推断结构的方法不同，RNAbpFlow采用近年来生成式AI领域广泛应用的流匹配技术，通过端到端方式直接生成RNA的全原子三维结构。模型从完全随机的噪声出发，在碱基配对信息引导下逐步折叠形成正确构象，并可一次生成多个不同结构，从而更好地反映RNA分子的动态变化。

该方法最大的优势在于对数据依赖较低。当前多数AI预测模型需要大量不同物种的相关RNA序列作为训练依据，而这类数据往往十分有限。RNAbpFlow无需这些数据库，因此特别适用于缺乏已知近缘序列的RNA分子。团队已利用该方法预测了新冠病毒基因组中的一个保守RNA结构元件以及实验室构建核酶等分子的结构。

不过，对于尺寸更大、结构更加复杂的RNA，依赖丰富进化信息的现有预测平台仍具有一定优势。RNAbpFlow则在数据稀缺的复杂场景中表现更为突出。

研究团队目前正开发升级版本，并计划参加今年夏季举行的CASP国际结构预测竞赛。

作者：张佳欣 来源：科技日报

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