



www.most.gov.cn

科学家利用迁移学习建立计算大模型预测基因调控研究

日期：2023年07月10日 17:29 来源：科技部生物中心 【字号：大 中 小】

基因网络绘制需要大量转录组数据用于建立基因之间的联系，这也阻碍了一些数据有限场景（如罕见病）等研究。最近，利用迁移学习的机器学习技术在自然语言和计算机视觉等领域带来了变革性进展，其通过在大规模通用数据集上进行大模型预训练，而后迁移到数据量有限的特定任务进行微调。美国博德研究所等研究团队提出了一个深度学习模型Geneformer可实现特定生物学任务与背景中的预测。该研究成果于近日发表在《Nature》杂志上，题为：Transfer learning enables predictions in network biology。

研究人员开发的深度学习模型Geneformer，在大约3000万个单细胞转录组的大规模数据集上进行了预训练，以便在网络生物学有限数据的情况下进行特异性预测。在预训练期间，Geneformer获得了对网络动态的基本理解，以完全自我监督的方式在模型的注意力权重中编码网络层次。研究人员利用Geneformer基于下游有限数据进行了预测任务，包括“疾病候选靶点预测”“解释拷贝数变异”“关键基因网络调控因子”“基因网络层次编码”“染色质动力学预测”等，并通过实验支持了其预测的结果。

综上，Geneformer代表了一种预训练的深度学习模型，可以对其进行微调，以实现广泛的下游应用，加速发现网络调节关键环节和候选治疗目标。

注：此研究成果摘自《Nature》杂志，文章内容不代表本网站观点和立场，仅供参考。

扫一扫在手机打开当前页



打印本页

关闭窗口

政府网站
找错

版权所有：中华人民共和国科学技术部

办公地址：北京市海淀区复兴路乙15号 | 联系我们

邮政地址：北京市海淀区复兴路乙15号 | 邮政编码：100862

ICP备案序号：京ICP备05022684 | 网站标识码：bm06000001 | 建议使用IE9.0以上浏览器或兼容浏览器