
科研人员开发出基于结构自监督学习的深度学习框架

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/31807.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！

科研人员开发出基于结构自监督学习的深度学习框架。

近日，中国科学院微生物研究所吴边和崔颖璐团队在《创新》（The Innovation）上发表了题为Structure-based self-supervised learning enables ultrafast protein stability prediction upon mutation

的研究论文。该研究开发出基于结构自监督学习的深度学习框架Pythia。Pythia可以快速预测蛋白质突变的稳定性变化（ $\Delta\Delta G$ ）。这一研究揭示了蛋白质三维结构蕴含的进化语义信息与热力学稳定性之间的深层关联，并通过开源代码（GitHub链接）与Web服务器技术服务体系为蛋白质工程领域提供便捷的研究工具。

蛋白质突变效应预测是解码分子进化机制、优化蛋白质工程改造的关键物理量。但是，传统预测方法面临挑战。一是基于物理力场的计算方法计算复杂度高，难以满足大规模筛选需求；二是依赖于实验数据的监督学习方法易受训练集偏差的影响，使泛化能力受限。

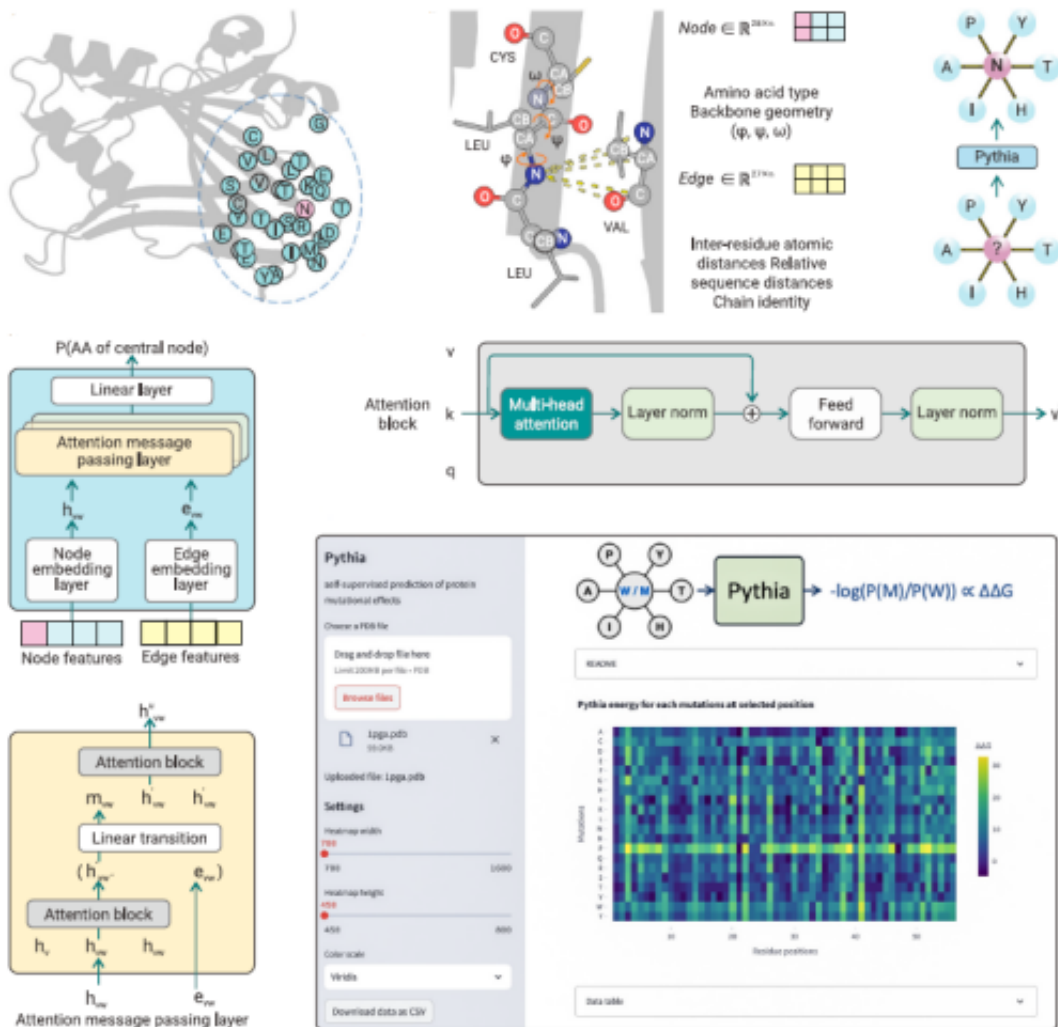
该团队提出了基于三维结构自监督学习的深度学习框架Pythia。Pythia融合了图神经网络与注意力机制，可直接在蛋白质结构中学习氨基酸之间的相互作用。这一“零监督”预训练策略突破了传统方法对标记数据的依赖性，捕获了蛋白质折叠过程中隐藏的物理化学约束规律。

相比于传统的力场函数方法，Pythia单点突变 $\Delta\Delta G$ 预测速度提升5个数量级，单核计算速度达到约50,000个突变/分钟。在标准测试集S2648上，Pythia的Spearman相关系数为0.616，Pearson相关系数为0.598，性能优于现有模型。这一框架为大规模扫描蛋白质序列空间提供了高效工具。在湿实验验证中，Pythia在不依赖专家筛选的情况下，成功率比基于能量函数的方法提高了一倍，证明其在实际应用中的可靠性。

Pythia有望进一步拓展至蛋白质-蛋白质相互作用预测、多突变协同效应分析等复杂场景，将助力推动计算驱动蛋白质设计范式的转型升级。

研究工作得到国家重点研发计划、国家自然科学基金、中国科学院战略性先导科技专项和青年创新促进会会员项目的支持。

[论文链接](#)



Pythia模型框架

研究团队单位：微生物研究所

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](https://www.iikx.com)转发