
机器学习提供发现疫情源头新方法

作者：writer 来源：中国科学报

本文原地址：<https://www.iikx.com/news/progress/2797.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！

机器学习提供了一种用基因组序列来快速发现神秘病毒感染的宿主和媒介的方法，它能缩短发现病毒、研究和暴发时反应之间的时间延迟。

已知有200多种RNA病毒能感染人类，它们会引起从普通感冒到埃博拉病毒感染等形形色色的疾病；每年都会发现若干种新的病毒。由未知病毒引起的传染病疫情有可能会快速传播并成为严重的公共卫生危机。了解病毒的天然宿主和传播媒介——即这些病毒最初的动物源头(如鼠类)及病毒传播给人的途径(如通过被感染跳蚤的叮咬)——可帮助确定感染风险最大的群体，并为公共卫生当局作出有效反应提供信息。

然而，要确认某些病原体的动物源头会需要多年的实地和实验室研究，这会极大地限制对疾病快速控防的措施，尤其是在紧急的情况下。尽管未知病毒的生物学可能在多年的时间内仍然晦暗不明，但获取病毒基因组却可快速完成。

Simon Babayan和同事组装了一个含有超过500种单链RNA病毒基因组序列的数据集，并用它创建了一个模型，后者能通过利用机器学习算法来预测病毒的宿主和传播媒介。据Babayan等人披露，关系相近的病毒常常有关系相近的宿主，而病毒基因组组成的偏差会为宿主—病毒关系提供信息。它们的模型可用机器学习在具已知宿主的基因相关病毒的基因组间抽取共演化信号，发现可区分宿主和传播媒介类型的基因组特性。作者通过确认某种潜在的偶蹄类哺乳动物宿主和不为人十分了解的蠓媒下刚果病毒而展示了他们进行预测的能力。

在《科学》文章中，Mark Woolhouse就Mimica等人的模型的局限性进行了讨论，然而，他指出，这项研究是向前迈出的宝贵一步，它有可能预示，我们直接从病毒基因组序列中抽取有价值公共卫生信息的能力会得到进一步的提升。(来源：中国科学报)

相关论文信息：DOI: 10.1126/science.aap9072

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