三个实例：“ 基于动力学的数据科学 ”在生命科学中的应用

作者：writer 来源：爱科学

本文原地址：https://www.iikx.com/news/progress/13570.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！

三个实例：“ 基于动力学的数据科学 ”在生命科学中的应用。21世纪以来，生物大数据在量（多数据种类，海量样本数，多时间点采样等）、质（高时空精度，单细胞测序等）两方面快速发展，大大推动了生命科学的进步，也为生物医学问题的模型建立、数据分析，以及预测和控制，带来了巨大的机遇和挑战。

传统的生物学数据研究方法大多基于数据的静态统计信息，即基于统计学的数据科学(statistics-based data science)，其缺点是，在很多场景下不能准确地解释和预测系统的复杂动态行为。数据分类、数据降维、变量聚类、变量相关性分析等方法都是如此。

然而，即使是静态的数据，往往也蕴含着系统的动力学特征。我们需要利用基于动力学的数据科学（dynamics-based data science），充分建立和利用动力系统的普遍性质（如，稳定平衡点的临界性质、中心流型的低维性、单变量的吸引子的重构性等），对蕴含在数据中的动力学信息进行挖掘和分析。

基于动力学的数据科学将动力系统理论、统计学理论，和数据的实际背景结合在一起，为处理和解释动态生物大数据提供了一种基础坚实、计算高效的理论和方法。

在最近发表于《国家科学评论》（National Science Review，NSR）的观点文章中，中科院生化细胞所的陈洛南研究员(通讯作者)、东京大学的史际帆博士(第一作者)和Aihara教授通过3个具体实例，揭示了如何利用动力系统的普遍性质，由观测的数据对生物学现象进行动力学分析，并解决相关生命科学等实际问题：

1. 利用微分方程的分岔理论，由测量的高维数据，进行健康临界预警和疾病预测。DNB理论利用了系统在临界点附近，复杂网络将表现出有别于非临界点的网络特性，量化临界状态并发现疾病的关键因子，实现疾病预警防病于未然。

2. 利用偏微分方程和diffusion map理论，量化细胞的干性或距离干细胞的远近。LDD方法是，通过建立随机生灭过程的偏微分方程模型，对细胞的分化过程进行了干性量化。利用单细胞测序数据和相关数学方法，可以对每类细胞干性进行估计和分化程度排序，实现量化细胞的干性，并构建干性势能景观。

3. 利用神经网络工具，对基因表达量等的时间序列进行预测。ARNN方法是，利用最新的reservio

r神经网络工具，通过空间-时间信息变换方程STI，即变换高维数据的信息为时间的动态信息，对短序列高维度数据(如基因表达数据)进行学习，可实现复杂系统的短时间序列或动态演化的预测。

基于动力学的数据科学是一个全新交叉领域，相比传统静态的基于统计学的数据科学方法，具有可解释性、可量化性和可拓展性，在今后的生物医学等领域的研究舞台，将扮演不可或缺的重要角色。（来源：科学网）

相关论文信息：https://doi.org/10.1093/nsr/nwab029

作者：陈洛南等 来源：《国家科学评论》

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