

<<蛋白质结构预测>>

图书基本信息

书名：<<蛋白质结构预测>>

13位ISBN编号：9787030223876

10位ISBN编号：703022387X

出版时间：2008-9

出版时间：科学出版社

作者：孙向东 等编著

页数：199

字数：251000

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<蛋白质结构预测>>

前言

蛋白质由氨基酸残基线性序列构成，折叠成特定的空间构象后，蛋白质就具有相应生物学活性和功能。

了解氨基酸残基序列与其空间结构的关系，是全面认识蛋白质结构和其生物学功能的关系的重要前提。

近些年来，蛋白质序列数据库的数据积累速度非常快，与之相比，蛋白质结构数据库的数据积累速度远不及序列数据库的数据积累速度。

尽管蛋白质结构测定技术有了较为显著的进展，但是通过实验方法确定蛋白质结构的过程仍然非常复杂，实验周期很长。

另外，随着DNA测序技术的发展，人类基因组及很多模式生物基因组已经或将要完全测序，DNA序列数量将会急增。

由于DNA序列分析技术和基因识别方法的进步，人们可以从DNA序列直接推导出大量的蛋白质序列，这将导致蛋白质序列数据数量急剧增加。

了解了这些序列的结构，可以使它们直接为人类服务。

氨基酸残基序列的结构分析是对生物学家的极大挑战。

20世纪60年代后期，Anfinsen首先发现去折叠蛋白或者说变性蛋白质在允许重新折叠的实验条件下可以重新折叠到原来的结构，这种天然结构对于蛋白质行使生物功能具有重要作用，蛋白质只有在折叠成其天然结构的时候才能具有完全的生物活性。

因此Anfinsen提出了蛋白质折叠信息隐含在蛋白质N—NN构中的观点。

以这种观点为基础，通过对蛋白质一级结构的研究，发现其折叠密码后，仅通过一级结构信息就能预测蛋白质空间结构。

蛋白质结构预测主要有两大类方法。

一类是蛋白质分子特性理论分析方法或从头算方法，通过理论计算(如分子力学、分子动力学计算)进行结构预测。

该类方法假设折叠后的蛋白质取能量最低的构象。

从原则上来说，人们可以根据物理、化学原理，通过计算来进行结构预测。

另一类蛋白质结构预测的方法是统计学方法。

该类方法对已知结构的蛋白质进行统计分析、建立序列到结构的映射模型、进而根据映射模型对未知结构的蛋白质直接从氨基酸序列预测结构。

这是进行蛋白质结构预测较为成功的一类方法。

这类方法包括经验性方法、结构规律提取方法、同源模拟化方法等。

统计学方法本身就是不确定性方法，目前虽然还不能完全替代第一类方法而成为预测蛋白质结构的主要方法，但是发展前景很广阔。

其中以统计学习理论为基础的支持向量机预测蛋白质结构的方法发展非常迅速。

统计学习理论是在20世纪90年代逐渐成熟的机器学习理论，以这种理论为基础的支持向量机与以往的学习机器相比具有支持小样本、不会陷入局部势井、具有很好的鲁棒性以及运算成本低等优势。

<<蛋白质结构预测>>

内容概要

统计学习理论是20世纪90年代逐渐成熟的机器学习理论，以这种理论为基础的支持向量机与以往的学习机器相比具有支持小样本、不会陷入局部势井、鲁棒性好以及运算成本低等优势。

实现这种理论的支持向量机算法已经成为机器学习和知识挖掘的标准工具。

自从2001年支持向量机被首次用于蛋白质二级结构的预测以来，这种算法发展到蛋白质的结构类型、亚细胞结构和膜蛋白的结构等领域的预测中。

本书详细介绍了依据统计学习理论构建支持向量机的方法、各种相关软件原理和使用方法，并以二级结构和结构域为例介绍了以支持向量机为工具预测蛋白质结构的方法。

书中使用了大量的原创性实验结果，理论联系实际，详细阐述了以支持向量机为工具预测蛋白质结构的全过程。

本书适合从事蛋白质结构基础研究的学生和科技工作者阅读。

<<蛋白质结构预测>>

书籍目录

前言第1章 蛋白质结构预测概述 1.1 蛋白质预测基本方法简介 1.2 蛋白质二级结构和结构域预测方法简介第2章 相关知识背景 2.1 生物信息学 2.1.1 生物信息学的定义、目的、内容和发展趋势 2.1.2 基因组学 2.1.3 蛋白质组学 2.1.4 数据库 2.2 蛋白质序列、结构与功能的关系 2.3 机器学习 2.3.1 机器学习的定义和特点 2.3.2 基本的机器学习模型 2.3.3 机器学习方法分类 2.3.4 应用于生物信息学领域的机器学习方法第3章 统计学习理论 3.1 学习问题的表示方法 3.1.1 概述 3.1.2 学习问题的一般表示 3.1.3 学习问题的模型 3.1.4 经验风险最小化原则 3.1.5 复杂性和推广能力 3.1.6 模式识别问题 3.2 统计学习理论的四个部分 3.2.1 学习过程的一致性 3.2.2 学习过程收敛速度的界 3.2.3 控制学习过程推广能力的理论第4章 构造支持向量机 4.1 优化理论 4.1.1 问题公式化 4.1.2 拉格朗日理论 4.1.3 KKT理论 4.2 支持向量机 4.2.1 支持向量机基本原理简介 4.2.2 线性分类 4.2.3 非线性分类 4.2.4 多重分类第5章 应用于支持向量机的主要算法 5.1 支持向量机算法中目前的研究状况 5.2 分解算法 5.3 顺序最小优化算法 5.3.1 顺序最小优化算法的原理 5.3.2 两个拉格朗日乘子的优化问题 5.3.3 选择待优化拉格朗日乘子的启发式方法 5.3.4 每次最小优化后的重置工作 5.3.5 顺序最小优化算法的特点和优势第6章 Libsvm简介 6.1 公式 6.1.1 C-支持向量分类(二元) 6.1.2 V支持向量分类(二元) 6.2 二次规划问题的解决 6.2.1 C—SVC的分解算法 6.2.2 工作集的选择和停止循环的标准 6.2.3 V支持向量分类的分解方法 6.2.4 解析解法 6.2.5 b和P的计算 6.3 压缩和缓存 6.3.1 压缩 6.3.2 缓存 6.4 多元分类 6.5 非平衡数据集 6.6 模型的选择 6.7 预测蛋白质结构中运用Libsvm的基本操作方法第7章 蛋白质二级结构预测 7.1 蛋白质结构 7.1.1 蛋白质的一级结构第8章 蛋白质折叠类型的预测参考文献附表1 RS126数据集附表2 CB513数据集附表3 蛋白质结构域拓扑层预测样本集附表4 蛋白质结构域同源超族层预测本集附表5 蛋白质结构域序列家族层样本集

<<蛋白质结构预测>>

章节摘录

第1章 蛋白质结构预测概述 1.1 蛋白质预测基本方法简介 生物信息学是近年来最有活力的生物学研究领域之一，人们从生物信息的研究中获得了生命本质更丰富的知识和更深刻的理解。核酸序列中蕴含着生命的基本信息，这些信息是自然界留给人类的、解读生命的“天书”。理解这本天书是最终了解自然、了解生命、了解人类自身的重要途径，是人类从必然王国到自由王国飞跃的基本前提之一。

由基因决定的蛋白质执行着生物体内各种重要的功能，如生物化学反应的催化、营养物质的运输、生长和分化控制、生物信号的识别和传递等。

基因确定了组成蛋白质的氨基酸序列。

虽然蛋白质由氨基酸的线性序列组成，但是它们只有折叠成特定的空间构象才能具有相应的活性和相应的生物学功能。

了解蛋白质的空间结构不仅有利于认识氨基酸残基序列与空间结构的关系，也有利于认识蛋白质的结构与其生物学功能的关系。

根据近些年来的经验，蛋白质序列数据库数据积累速度非常快，而且还有加快的趋势。

尽管蛋白质结构测定技术有了较为显著的进展，但是通过实验方法确定蛋白质结构的过程仍然非常复杂，实验周期很长。

另外，随着DNA测序技术的发展，人类基因组及很多的模式生物基因组已经或将要被完全测序，DNA序列数量将会剧增，由于DNA序列分析技术和基因识别方法的进步，人们可以从DNA序列直接推导出大量的蛋白质序列。

这意味着已知序列的蛋白质数量和已测定结构的蛋白质数量（如蛋白质结构数据库PDB中的数据）的差距将会越来越大。

面对这种蛋白质结构信息与DNA序列信息发展速度的不平衡，人们希望找到一些预测方法，通过这些方法加快蛋白质结构产生速度，缩小二者之间的差距。

为了缩小这种差距，要么改进现有的蛋白质测序技术和结构预测方法，要么发展新的理论分析方法，这是对生物学家的极大挑战。

20世纪60年代后期，Anfinsen首先发现去折叠蛋白质或者说变性蛋白质在允许重新折叠的实验条件下可以重新折叠到原来的结构，这种天然结构对于蛋白质行使生物功能具有重要作用，蛋白质只有在折叠成其天然结构的时候才能具有完全的生物活性。

因此Anfinsen提出了蛋白质折叠的信息隐含在蛋白质的一级结构中的观点。

基于这种观点，人们相信通过对蛋白质一级结构的研究，发现其折叠密码后能够仅通过一级结构信息就能预测蛋白质空间结构。

<<蛋白质结构预测>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>