

药物设计平台实用指南

徐 峻 中山大学药学院

报告摘要

药物设计平台是在超算支持下的、具有大数据处理能力的药物原始创新引擎的核心装备，它的主要目的是解决药物设计过程中的基本问题：即，发现药物分子的新骨架、预测分子的活性机理、评价分子的生物和化学性质、优化分子的药性性质。本讲座通过实例，介绍在广州超算中心运行的药物设计平台的几个重要模块的原理和使用方法：

- (1) 用于对天然产物进行虚拟筛选的 WEGA 程序；
- (2) 用于对巨量化合物库进行稳定性进行评价的 ChemStable 程序；
- (3) 用于虚拟药物筛选的 LBVS 程序；
- (4) 用于发现药物活性新骨架的活性标注的化学骨架数据库 ASDB

本讲座的目标是：使药物研发人员掌握药物分子设计的基本过程，例如，实现小分子配体与药物靶标的合理对接、根据已知活性化合物知识进行虚拟药物筛选和先导化合物创新或优化、破解中药的活性机理。

个人简历



徐峻，中山大学教授，匹兹堡大学兼职教授、药物分子设计与生物超算研究中心主任、2010 年广东省引进科研团队核心成员、中国化学会计算化学专业委员会委员、国家中医药管理局中药专利资源重点研究室学术委员会副主任、广东省超算学会常务理事、广东省中西医结合学会中药专业委员会常务委员。1988 年毕业于中国科技大学获理学博士学位。

1990-1993 年分别在澳洲国立大学、加拿大麦基尔大学做博士后研究。1993 年以来，在美国生物医药工业界著名公司如 BIO-RAD、Boehringer Ingelheim 制药公司历任资深科学家，主管科学家、研发部总监。研究领域横跨药理学、化学、和信息科学。提出多种药物设计重要算法。发现了多种抗 γ 型人类疱疹病毒、抗 2 型糖尿病、抗耐药菌等先导化合物。主要研究兴趣包括中药现代化、抗癌、抗感染、抗代谢类疾病药物研发、超级计算方法和网络图论应用。

生物信息学实用工具

任 间 中山大学生命科学学院

报告摘要

生物信息学是采用计算机技术和信息论方法研究生命科学中各种生物信息的表达、采集、储存、传递、检索、分析和解读的科学,是现代生命科学与信息科学、计算机科学、数学、统计学、物理学、化学等学科相互渗透和高度交叉形成的学科。本报告简要介绍现代生物信息学的主要研究领域,及实用工具。

个人简历



任间, 男, 博士, 中山大学生命科学大学院、肿瘤防治中心, 教授、博导。中山大学生物信息学中心主任, 高性能计算 2011 协同创新中心应用研究部部长。广东省首批自然科学杰出青年基金获得者, 教育部新世纪优秀人才。

1998 至 2002 年就读于上海交通大学动力与能源工程学院, 获核工程与核技术专业学士学位。2002 至 2007 年在中国科学技术大学近代物理系金革教授实验室攻读博士期间参与国家大科学工程项目“大天区面积多目标光纤光谱望远镜”(LAMOST), 并承担其中的巡天战略系统的研发工作, 在计算机算法和大型工程软件设计开发方面具有丰富的经验。2007 年 7 月获得物理电子学工学博士学位之后进入中国科学技术大学生命科学学院温龙平研究组进行博士后研究工作, 将计算机和天文学算法以及工程学思想成功运用到生物信息学领域, 在蛋白质组生物信息学方面取得了丰富成果。主要包括设计出高效的 GPS (Group-based Prediction System) 预测算法, 并基于该算法开发出十多种蛋白质翻译后修饰位点预测工具, 此外还构建多个数据库及可视化工具。2010 年 3 月获中山大学生命科学学院“百人计划”引进, 2011 年 12 月晋升教授。进入生物领域后, 在包括 Molecular & Cellular Proteomics, Nucleic Acids Research, Cell Research, Briefings in Bioinformatics 等杂志上发表 SCI 论文 30 余篇, 其中通讯或一作 21 篇, 累计影响因子 180。相关文章已被包括 Nature、Science 及 Cell 系列杂志在内的文章引用 1000 余次, 最高单篇引用超 200 次, H-index 14。担任国际杂志 Frontiers in Genetics 编委。主持国家自然科学基金 3 项, 作为骨干参与科技部项目 4 项。发布生物信息学工具及数据库 30 多个, 获颁软件著作权登记证书 14 项。

“天河二号”与大数据处理入门

李家辉 国家超级计算广州中心

报告摘要

本次报告将介绍生物医学数据分析平台 Galaxy，以及大数据分布式处理框架 Hadoop 和 Spark 在天河二号上的部署情况和使用方法。

Galaxy 是一个有广泛应用的开放性平台，其集成了大量的生物信息分析工具，同时为用户提供了一个简单易用的生物信息分析界面。利用 Galaxy 中已安装的分析工具，用户还可创建和调用可重复使用的数据分析流程，并对这些流程进行修改和导入导出，它还有历史记录功能，支持数据的可视化等。除此之外 Galaxy 还支持自定义工具的添加，可按照需求扩展分析工具集。

Hadoop 和 Spark 是目前主流的大数据分析框架，近年随着生物医疗领域数据的积累，数据的有效利用将离不开这些大数据管理分析工具使用。

个人简历



李家辉，浙江大学物理系博士，现任职于中山大学国家超级计算中心应用部，从事高性能计算相关工作。负责“生物计算与个性化精准医疗服务平台”应用程序在天河二号的适配，以及该平台的发展建设。主要使用 C、Fortran 编程，熟悉基于 MPI、OpenMP 的并行程序设计，有丰富高性能平台使用经验。过去主要从事有限差分、有限体积法求解偏微分方程组，目前主要从事大数据处理工具和高性能平台结合的实践。

健康产业中的人工智能

路 川 NVIDIA 公司

报告摘要

深度学习是近十年来人工智能领域取得的重要突破，解决了许多人工智能领域的尽最大努力但仍未能解决的问题。它已经被证明能够擅长发现高维数据中的复杂结构，并应用于科学研究、商业和政府等领域。除在图像识别、语言识别等领域打破了记录，它还在其他领域超越了传统机器学习技术，包括预测潜在药物分子活性、分析粒子加速器数据、重建大脑回路、预测在非编码 DNA 突变对基因表达和基本的影响等。

本讲座主要介绍深度学习解决方案，如何帮助数据科学家和技术人员快速使用深度学习平台，通过简单高效的网页交互式界面实现自己的模型设计和数据训练，快速实现深度学习应用，提升工作效率。

个人简历



路川，Nvidia 机器学习解决方案架构师，在深度学习、高性能计算和云计算领域有着丰富的项目经验，曾作为主要技术负责人员参与国家超算深圳云计算中心、上海超算中心、成都云计算中心等大型超算的技术方案构建、系统建设和支持等工作。