

<<基因组信号处理>>

图书基本信息

书名：<<基因组信号处理>>

13位ISBN编号：9787030271235

10位ISBN编号：7030271238

出版时间：2010-4

出版时间：科学出版社

作者：（俄罗斯）什穆列维奇 等编著，刘文斌，高琳 译

页数：214

字数：317000

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>

<<基因组信号处理>>

前言

近年来,各种高通量技术的出现极大地促进了大规模基因表达的研究,人们可以同时确定成千上万个基因的表达水平。

特别是表达微阵列这种复杂的生化-光学系统,它集成了机器点样、计算机成像和分析等技术。由于转录调控是一种多变量输入的复杂处理过程,人们迫切需要一种分析表达谱数据的工具来揭示复杂遗传网络中所发生的各种作用关系。

也就是说,只有对基因组的这些信号进行处理,才能刻画它们的调控结果及其在遗传水平和表观水平所发生的变化。

功能基因组学的两个主要目的是,筛选能够解释特定细胞显型(如疾病)的关键基因和基因组合,并在分子水平利用基因组信号进行疾病的分类。

由于转录信号的传递及相关路径模拟的重要作用,对结构和功能的理解都需要信号处理的理论。

基因组信号处理(GSP)是一个主要研究基因组信号处理的工程学科。

人们可能会由此联想到数字信号处理(DSP),总的来说,GSP和DSP都要应用信号处理的某种数学方法。

然而,这两门学科的名称没有一个与这些数学方法有关,反而是根据它们所处理的物理信号来命名的。

GSP将信号处理的方法、理论应用于功能基因组的研究,其重点是基因组的调控。

因此,GSP涉及与表达谱研究相关的各种方法,如检测、预测、分类、控制、统计以及基因网络的动态模拟等。

总之,GSP是将工程学科中基于模型的结构分析和合成方法引入基因组学研究的一门基础学科。

这一学科主要用于生物组织(切片)的分类及信号转导路径的发现。

实现这些目标需要大量信号处理的方法,其中包括转录信号的表示(如小波分解和随机时间序列分解)和非线性动态系统的建模。

以往两个基因或细胞效应(cellular effect)对之间的相关性分析方法已经不适用于基因组这种多元非线性的复杂网络,调控模型需要各种非线性动力学的方法。

而这些方法已在信号处理、控制,尤其是分布式计算机系统随机数据流网络等领域得到广泛应用。

当然,仅有这些模型系统还不够。

基因组学需要自己的模型系统,而不仅仅是对现有模型的简单应用。

而且,新系统必须能够反映基因组内部的运行规律及分布式调控的机理。

因此,必须发展相应的数学理论,其中包括用于治疗干预的优化理论以及近似理论,并建立一种具有合理复杂度的非线性动态模型。

具体地说,一方面,从诊断和治疗的角度,它能够充分描述基因之间的调控关系;另一方面,相对于现有的实验数据量或者计算机硬件的计算能力,又不过于复杂。

广义上讲,GSP是一门由基因组学研究推动产生的学科。

尽管这种概括性的描述无法反映其中所需的数学知识结构,但它还是强调了数学在基因组科学中的地位。

实际上,这也是数学在所有科学研究中的总体地位的体现。

随着对科学认知论理解的深入,我们将会逐渐体会到这一点。

<<基因组信号处理>>

内容概要

随着各种高通量微阵列技术的飞速发展，基于基因表达谱数据的分析已经成为系统生物学研究中的一个非常重要的领域。

基因表达谱数据的最大特点是高噪声、高维数和小样本，针对这种数据的特点，本书主要介绍了基因调控网络的建模方法及其动态行为分析、分类器的设计及其误差估计、数据和特征的正则化，以及聚类算法及其验证过程等内容。

本书主要面向具有一定数学基础的生物信息学、计算生物学或系统生物学等领域的研究生或科研人员，也可以作为研究基因组学的教科书或参考书。

<<基因组信号处理>>

作者简介

刘文斌，男，工学博士。

温州大学物理与电子信息工程学院副教授，浙江省中青年学科带头人。

2004年在华中科技大学控制系系统工程专业获博士学位，2004-2006年在华中科技大学生物医学工程专业博士后流动站从事博士后研究。

2007年在美国系统生物研究所（Institute for Systems Biology）计算生物学小组做访问学者，主要从事基因调控网络的预测及干预研究。

主要研究方向为计算生物学、DNA计算、数据挖掘、模式识别、智能计算等。

累计在国内外重要期刊和会议发表学术论文40余篇，主持两项国家自然科学基金项目，一项浙江省自然科学基金项目和一项中国博士后科学基金项目（二等）。

获得省部级奖励两项，厅级奖励两项。

高琳，女，工学博士，西安电子科技大学计算机学院教授，学术带头人，博士生导师。

在西安交通大学计算数学专业获理学学士学位，在西北大学计算数学专业获理学硕士学位，在西安电子科技大学电路与系统专业获工学博士学位。

1990～1994年在西电公司计算技术应用研究所从事有限元工程算法设计、计算与分析等研究与应用工作。

1994年10月调入西安电子科技大学从事教学和科研工作至今。

2004～2005年被国家留学基金委选派赴加拿大Guelph大学做访问学者。

在科研方面，主持国家自然科学基金重点项目一项、面上项目两项，教育部博士点学科基金，国家留学回国人员基金及多项省部级项目。

参与了多项国家自然科学基金、国防科技预先研究基金等科研项目，成果获省部级奖励两项，厅局级奖励两项。

在教学方面，荣获校“永新”、“贝尔”、“华为”等奖教金，获西安电子科技大学“十佳师德标兵”称号、“陕西省师德标兵”称号。

近年来的主要研究方向为计算生物信息学、数据挖掘、图论与组合优化算法及应用等，累计在国内外核心期刊和国际会议发表学术论文60余篇。

<<基因组信号处理>>

书籍目录

译者序前言第1章 生物基础 1.1 遗传学 1.1.1 核酸结构 1.1.2 基因 1.1.3 RNA 1.1.4 转录 1.1.5 蛋白质 1.1.6 翻译 1.1.7 转录调控 1.2 基因组学 1.2.1 微阵列技术 1.3 蛋白质组学第2章 基因网络的确定性模型 2.1 图模型 2.2 布尔网络 2.2.1 细胞分化和细胞的功能状态 2.2.2 网络特性及动态行为 2.2.3 网络推理 2.3 布尔网络的推广 2.3.1 异步 2.3.2 多值网络 2.4 微分方程模型 2.4.1 有转录和翻译过程的微分方程模型 2.4.2 连续微分方程模型的离散化第3章 基因网络的随机模型 3.1 贝叶斯网络 3.2 概率布尔网络 3.2.1 定义 3.2.2 推理 3.2.3 PBN的动力学 3.2.4 暂态随机PBN的稳态分析 3.2.5 PBN与贝叶斯网络的关系 3.2.6 基于种子基因的子网络的生长 3.3 干预 3.3.1 基因干预 3.3.2 结构干预 3.3.3 外部控制第4章 分类 4.1 贝叶斯分类器 4.2 分类规则 4.2.1 一致分类器设计 4.2.2 分类规则实例 4.3 有约束的分类器 4.3.1 分散系数 4.3.2 VC维数 4.4 线性分类 4.4.1 Rosenblatt感知器 4.4.2 线性及二次判别分析 4.4.3 基于最小二乘误差的线性判别式 4.4.4 支持向量机 4.4.5 线性判别式的设计误差的表示 4.4.6 基于样本QDA判别式的分布 4.5 神经网络分类器 4.6 分类树 4.6.1 分类与回归树 4.6.2 基于数据划分的强一致规则 4.7 误差估计 4.7.1 重代人法 4.7.2 交叉验证 4.7.3 自举法 4.7.4 支撑 4.7.5 误差估计器性能 4.7.6 特征集排序 4.8 误差校正 4.9 鲁棒分类器 4.9.1 最优鲁棒分类器 4.9.2 鲁棒分类器的性能比较第5章 正则化 5.1 数据正则化 5.1.1 正则化判别分析 5.1.2 噪声注入 5.2 复杂度正则化 5.2.1 误差正则化 5.2.2 结构风险最小化 5.2.3 经验复杂度 5.3 特征选择 5.3.1 峰值现象 5.3.2 特征选择算法 5.3.3 误差估计对特征选择的影响 5.3.4 冗余 5.3.5 并行增量特征选择 5.3.6 贝叶斯变量选择 5.4 特征提取第6章 聚类 6.1 聚类算法的实例 6.1.1 欧氏距离聚类 6.1.2 自组织映射 6.1.3 分层聚类 6.1.4 基于模型的聚类算子 6.2 聚类算子 6.2.1 算法结构 6.2.2 标记算子 6.2.3 贝叶斯聚类器 6.2.4 聚类算子的分布测试 6.3 聚类的验证 6.3.1 外部验证 6.3.2 内部验证 6.3.3 不稳定指数 6.3.4 贝叶斯因子 6.4 聚类算子学习 6.4.1 经验误差聚类算子 6.4.2 最近邻聚类规则索引

<<基因组信号处理>>

章节摘录

所有细胞都含有相同的DNA，那么，如何区分肝细胞和白细胞？细胞的特性（包括结构和活性）由其中的基因活动（特别是表达蛋白）决定。这说明一定有一种内部、外部或者二者都有的调控机制，可以控制与细胞类型或者功能状态有关的蛋白质的表达。

基因的调控过程极其复杂，尤其是在真核生物中。

下面以细菌为例，说明蛋白质对于细胞适应环境变化的重要性。

糖的代谢过程中有大量的酶参与，糖的种类有多种，如乳糖、葡萄糖、半乳糖等，都可以作为细菌能量的来源。

由于糖的种类不同，其分解所需的酶也不同。

一种可能的方法是在细胞中保存所需的各种酶，以便应对出现的各种糖。

显然，这种方式将造成很大的浪费且效率不高，而且合成这些酶需消耗很多能量，其中有很多酶并不需要。

因此，细菌采取的方式是激活所需酶的基因，同时休眠或抑制那些不需要的酶的基因。

这样，就可以根据不同的环境状况来激活或抑制这些基因。

我们已经讨论了转录启动所需的机制：RNA聚合酶必须结合到基因的启动子上。

然而，还有其他的一些DNA结合蛋白能够决定转录是否开始。

例如，在原核生物中，有一种称为激活子（activator）或阻遏子（repressor）的调控蛋白，可以识别启动子附近的一些位点。

激活蛋白必须结合作用位点才能启动转录；阻遏蛋白则相反，它不结合目标位点时才能启动转录，否则，转录将被阻塞。

激活蛋白和阻遏蛋白改变基因转录的一种方式是与RNA聚合酶相互作用：阻遏蛋白的结合可以干扰RNA聚合酶结合DNA，而激活蛋白的结合则可以促进RNA聚合酶结合DNA。

既然调控蛋白的结合能够决定基因转录是否发生，那么，是什么决定这种蛋白质能够结合到目标位点？

大多数情况下，它由蛋白质的DNA结合域（可以直接结合特定DNA序列的区域）和蛋白质的变构位点（allosteric site）[一种称为变构效应因子（allosteric effector）的小分子可以结合并引起构型变化的位点]之间的相互作用决定。

当变构效应因子结合到调控蛋白的变构位点，DNA结合域将发生变化从而改变蛋白质结合目标位点的能力，并最终决定转录是否发生。

例如，对于激活子，变构效应因子的出现可能使得调控蛋白能够结合到DNA上；另外，阻遏子则必须在没有变构效应因子的情况下才能阻止转录的发生。

例如，在大肠杆菌的乳糖（lactose）代谢中，当乳糖（变构效应因子）结合到它的阻遏子时，该阻遏子就会失去结合到目标位点的能力并从DNA上脱落下来。

这样，附近的RNA聚合酶就能够结合到DNA上编码某个乳糖代谢所需的酶的基因附近（图1.10）。

<<基因组信号处理>>

编辑推荐

本书专门探讨基因表达谱数据处理涉及的建模、分类及聚类的基本方法和理论。作者将生物网络的进化、细胞的功能状态等生物学概念与基因调控网络的动态行为有机地结合在一起，揭示了这些抽象数学概念所代表的实际生物学意义。并对PBN这种新的概率模型，作者采取例子驱动的方式，逐步介绍有关的概念、动态行为分析及干预过程，非常浅显易懂。作者还从算子的角度将分类和聚类统一起来，详细讨论了二者涉及的算子的设计、误差估计及鲁棒性。

<<基因组信号处理>>

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>