计算机技术在生命科学领域的应用

F1003028 5100309127 贺天行

生命科学导论课感觉上与高中所学的生物学有很大不同，换句话说，我们更像是在学一门“科学”了，作为这门课期末的一个小总结报告，我希望能结合我的专业：计算机，挖掘一下计算机技术和生命科学技术两者的结合点，作为一篇报告。

生命科学技术和信息技术在对于我们这些对生命科学研究不够深入的学生来说，似乎并不能找出有多大的共同点。但找到一些资料后，我惊奇地发现信息产业的巨头们已经开始让生命技术领域注资，我想，这些走在时代最前沿的风口浪尖的巨头们已经发现生命科学的革命已经势在必行，而他们的最强项：计算机技术将在生命科学领域起到至关重要的作用，让我们来看看他们做了什么吧。

微软除了通过比尔及梅琳达•盖茨基金会大规模资助生命科学与技术研究项目外，自身也于2006年正式成立了医疗事业部并收购了号称医院Google的医疗病历软件Azyxxi，涉足医疗信息市场。2007年，微软先后收购了美国的在线医疗公司MedStory和泰国的GCS公司及其软件Hospital2000。同年10月，微软推出了针对患者的Health Vault网络平台，实现个人健康信息的记录、储存、维护、交换和搜索。针对医疗机构，微软推出了Amalga医疗信息管理系统。2009年，微软进军中国医疗信息市场，在青岛大学医学院附属医院部署了Amalga系统。目前，微软在中国从事医疗产品开发的团队有近百人。在生物信息学领域，微软不仅推出了基于微软.NET的生物信息学扩展工具箱MBF(Microsoft Biology Foundation)，也有研究人员从事生物信息学研究。笔者在阅读文献时，也多次看到来自微软研究院总部的基于人工智能技术预测MHC结合肽的研究。

作为微软最大的竞争对手，Google与生物技术也很有缘份。Google与生物技术明星企业基因泰克，总部都位于旧金山湾区，而基因泰克的CEO也是Google的董事会成员。虽然分属IT与BT行业，但两家公司却具有极为相似的企业文化，也常常被《科学》、《财富》等杂志同时评为最佳雇主、最具创新精神公司。最为有趣的是Google的老板娘本身就毕业于耶鲁大学生物学专业。针对微软的Health Vault，谷歌于去年5月19日，正式发布了公众测试版Google Health,提供从分类健康信息到个性化的健康指导的各种服务，包括电子病历导入、检查结果、治疗方法、免疫记录、适应症、药物过敏、药物相互作用等。谷歌技术副总裁说：Google Health目前还在了解人们需求，并积累经验，这会是个漫长的过程，但无论为之花费多少钱，对于Google来说都是值得的。今年4月，Google又拿出1亿美元成立风险投资基金，准备投向生物技术与健康产业。

其实，不提那些产业巨头，就拿我们这些本科生对生命科学粗浅的了解，也已经能感觉到生命科学与信息学之间深刻的关联。将生命的全部特征囊括在极其简单的DNA序列中，这本身就是信息学的一个完美的实例！甚至，生命科学有时还会反作用于计算机技术，就在我们已经学习到的算法中，明显从生命科学中学来的就有遗传算法，免疫算法，电脑病毒，还有机器学习中非常复杂的神经网络系统，都或多或少是基于我们对于生命形式的理解而建立起来的。

所以，难怪1991年，因发明DNA测序方法而获得诺贝尔奖的Walter Gilbert教授在Nature上大声疾呼：生命科学研究的范式正在改变，生物学家必须成为计算机学者。世纪之交，DNA自动测序仪的发明者、系统生物学的提出者Leroy Hood教授在一次会议上进一步指出：生物学在21世纪将愈发成为一门信息科学。

在本文的后半段，我想介绍一下生物信息学，作为交叉学科的典范之一，相信最能从中感受到计算机技术与生命科学的关联性。

生物信息学（bioinformatics）是生物学与计算机科学以及应用数学等学科相互交叉而形成的一门新兴学科。它通过对生物学实验数据的获取、加工、存储、检索与分析，进而达到揭示数据所蕴含的生物学意义的目的。由于当前生物信息学发展的主要推动力来自分子生物学，生物信息学的研究主要集中于核苷酸和氨基酸序列的存储、分类、检索和分析等方面，所以目前生物信息学可以狭义地定义为：将计算机科学和数学应用于生物大分子信息的获取、加工、存储、分类、检索与分析，以达到理解这些生物大分子信息的生物学意义的交叉学科。

生物信息学是一种重要的研究开发工具。它是一门研究生物和生物相关系统中信息内容物和信息流向的综合系统科学，只有通过生物信息学的计算处理，我们才能从众多分散的生物学观测数据中获得对生命运行机制的详细和系统的理解。它是今后几乎进行所有生物医药研究开发所必需的动力机，用于序列分类、同类性检测，DNA序列中蛋白质编码区和非编码区的分离，分子结构和功能的预测，以及进化史重建，只有基于生物信息学通过对大量已有数据资料的分析处理所提供的理论指导和分析，我们才能选择正确的研发方向，同样，只有选择正确的生物信息学分析方法和手段，我们才能正确处理和评价新的观测数据并得到准确的结论。它不仅是我们了解生命和进化所必需的，同时也是发现新药和诊断方法所必不可少的。

生物信息学在今后的无论是生物医药科研还是开发中都具有广泛而关键的应用价值；而且，由于生物信息学是生物科学与计算科学、物理学、化学和计算机网络技术等密切结合的交叉性学科，使其具有非常强的专业性，这就使得专业的生物医药科研或开发机构自身难以胜任它们所必需的生物信息学业务，残酷的市场竞争及其所带来的市场高度专业化分工的趋势，使得专业的生物医药开发机构不可能在自身内部解决对生物信息学服务的迫切需求，学术界内的生物医药科研机构也是如此，而这种需求，仅靠那些高度分支化和学术化的分散的生物信息学科研机构是远远不能满足的。可见，在生命科学的新世纪，生物信息学综合服务将是一个非常重要的也是一个极具挑战性的领域。

生命科学对于信息技术既是机遇也是挑战，因为生命科学的蓬勃发展，对于计算机技术提出了一个又一个新的挑战，现在，全世界的超级计算机所做的事已经不在局限于计算弹道，模拟物理模型，下国际象棋等等，而是被用来匹配DNA序列，研制新的生物药剂等等。而生命科学也有可能为计算机世界带来革命—生物计算机。对于我们这些计算机系的本科生来说，与生命科学的约会可能才刚刚开始。