上海科普创作出版专项资金资助

基因宝库丛书
谈家桢 主编
上海市农业生物基因中心 编

基因计算

钟 扬 张文娟
王 莉 赵佳媛 >>> 编著

上海教育出版社
SHANGHAI EDUCATION PUBLISHING HOUSE

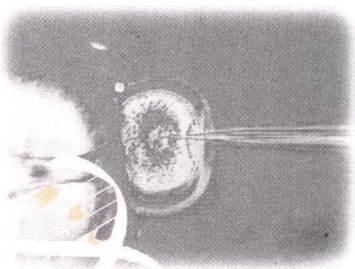
基因宝库丛书

jiyinjisuan

基因计算

上海农业生物基因中心　编

编　著：钟　扬　张文娟
王　莉　赵佳媛



上海教育出版社

图书在版编目（CIP）数据

基因计算 / 钟杨等编著. —上海：上海教育出版社，2005.12
（基因宝库丛书 / 谈家桢主编）
ISBN 7-5444-0525-7

Ⅰ.基… Ⅱ.钟… Ⅲ.计算机应用—基因—遗传工程—青少年读物 Ⅳ.Q78-39

中国版本图书馆CIP数据核字（2005）第154936号

基因宝库丛书

基因计算

谈家桢 主编

上海世纪出版集团
上海教育出版社 出版发行

易文网：www.ewen.cc

（上海永福路123号 邮政编码：200031）

各地新华书店经销 上海中华印刷有限公司印刷

开本 889×1194 1/32 印张 5.75 字数 105,000

2005年12月第1版 2005年12月第1次印刷

印数 1～3,000本

ISBN 7-5444-0525-7/Q·0006 定价：23.00元

（如发生质量问题，读者可向工厂调换）

本书主编 **谈家桢**

1909年9月出生

著名遗传学家

中国科学院院士

第三世界科学院院士

美国科学院外籍院士

责任编辑 吴延恺 肖征波

封面设计 一步设计

主　编：谈家桢

副主编：吴爱忠 罗利军

编　委：沈大棱 林榕辉

袁正守 潘重光

（按姓氏笔划）

编辑策划：肖征波 吴延恺

钟 扬

1964年5月生。1979年考入中国科学技术大学少年班，1984年毕业于该校无线电电子学系。1984至1999年在中国科学院武汉植物研究所工作。曾在美国加州大学柏克莱分校、密西根州立大学、日本文部科学省统计数理研究所工作。现为复旦大学生命科学学院教授、常务副院长、植物学和生物信息学博士生导师，兼任上海生物信息技术研究中心副主任、北京大学理论生物学中心教授、西藏大学教授。

张文娟

1978年出生于湖北武汉。2001年毕业于华中师范大学生命科学学院，现为复旦大学生命科学学院博士研究生，专业方向生物信息学。

王 莉

1977年12月生。2000年毕业于武汉大学信息学部。2001年起为复旦大学生命科学学院进化与生态学系博士研究生，曾作为交流学生赴加拿大女王大学学习，研究方向为生物信息学、植物学。

赵佳媛

1980年生于上海。2003年毕业于复旦大学生命科学学院，现为该院研究生。

序

年初，上海农科院吴爱忠教授和上海农业生物基因中心罗利军教授告诉我，上海市科委和科协将设专项基金资助科技工作者撰写科普书籍。他们打算组织长期从事教育和科技工作的专家编写基因科学丛书，定名为“基因宝库”。我认为科委和科协的决定及两位教授的打算很有意义。向公众传播科学知识，无疑能提高劳动者的科技素质，促进先进生产力的发展。

生命科学自上世纪50年代进入分子生物学时代以来，基因科学突飞猛进，新概念、新名词日新月异，与时俱增。基因也成为运用次数最多的字眼之一。但由于基因科学既包含遗传、变异、个体、群体，分子、细胞，基因、环境，核酸、蛋白质等诸多矛盾的统一，基因科学又与国计民生关系十分密切，丰衣足食、安居乐业、健康长寿、天下太平都离不开基因科学。因此要较全面地了解基因科学知识及基因科学在工业、农业、医学等诸多方面的应用价值，实非易事。组织专家编写普及基因科学的系列丛书，无疑又是先进文化发展的需要，我

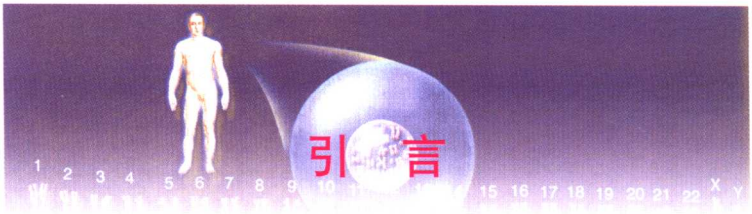
是非常支持的。

自我国取得抗击SARS的初步胜利后，吴爱忠、罗利军两位教授委托上海交大潘重光教授转告我，市科委、科协已正式同意资助“基因宝库”的编写，我很高兴。我因年迈已不能亲自参加丛书的编写，但我很乐意做力所能及的事。我托潘重光同志转告吴、罗两位教授，编写“基因宝库”丛书是一件很有意义的事，希望在编写过程中，特别要重视科学性，在保证科学性的基础上，应该积极探索趣味性和可读性，努力把“基因宝库”编成公众喜欢阅读的丛书。

谈家桢

2003年10月9日

目录



一、推销员该往哪里走?

早在十九世纪中叶，一位名叫汉密尔顿（William Hamilton)的爱尔兰天文学家和数学家发明了一种有趣的棋类游戏——Icosian Game，规则是用最快的速度走遍正二十面体上的所有顶点，最后又能回到起点。汉密尔顿用25英镑的价钱把这个点子卖给了一个玩具制造商，商人又将之发展成了一个以二十个孔代表世界城市的圆盘，用木条代替玩家标记行走路线，并起了一个相当通

Icosian Game 玩具

俗的名字叫“环游世界”。不过，遗憾的是，玩具商似乎并没有从中盈利。

然而这个游戏所具有的意义并不需要在金钱上体现。到了二十世纪三十年代，德国著名数学家门格尔(Karl Menger)把它具体化成了一个很有意思的数学问题：

假定一个推销员要去几个城市售货，在给定起点城市和终点城市的情况下，他是否可以找到恰好经过其他每个城市一次且路程最短的路径？

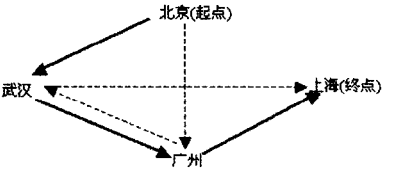
这就是后来名噪一时的数学问题——“旅行推销员问题”(Traveling Salesman Problem)。这个问题一经提出，立刻引起了各国数学家和数学爱好者们的极大兴趣。由于是从汉密尔顿的游戏因袭而来，于是也被称为“汉密

伏案计算的门格尔

符号构成的门格尔像

尔顿路径问题”。

我们先看一个简单的例子。假定推销员是从北京出发，要去的城市是武汉、广州、上海(终点)。从下图中可以找到两条符合要求的路径。当然，如果要找最短路径的话，就只能是北京→武汉→广州→上海，它比另一条路径的距离短1204公里(以目前直达列车里程计算)。



路径1：北京→武汉→广州→上海

路径2：北京→广州→武汉→上海

四个城市间的两条汉密尔顿路径示意图

可以想象，随着城市数目的增加，汉密尔顿路径问题会变得越来越复杂，寻找所有可能路径的工作量也会越来越大，甚至可能像静静屹立在印度某神庙中预兆世界末日的河内塔一样，成为一个人在有生之年都不可能完成的任务。但这个问题可不是一个简单的数学游戏，它在理论和应用两方面都具有重要的价值：

如果我们将包含所有要经过城市的地图看成一个网络，在每个网络中都一定存在着汉密尔顿路径吗？直到二十世纪七十年代初期，人们才终于证明这是一个“NP完备”(Nondeterministic Polynomial的缩写，表示非确定的多项式。)的组合数学问题。目前，NP完备问题在数学

计算上难度极大，通常很难找到一种有效的算法，即便是功能强大的电子计算机，兴许也要花费超出你所能想象的时间，才有可能确定是否真的存在一条这样的路径。比方说，当城市数目达到100个时，一台电子计算机就需要大约好几百年的计算时间！所以，从事计算工作的科学家都想把所有的NP问题找出来，用一些“启发式”算法或并行算法来获得近似解，以避免人力财力的浪费。

寻找最短路径在应用数学中也称为“汉密尔顿路径最小化”，它在工程优化、交通管理、物流成本和进度控制等很多方面都大有用处，在此基础上不断发展的“赋权汉密尔顿最优解方法”已成为上述领域一项不可或缺的关键技术。

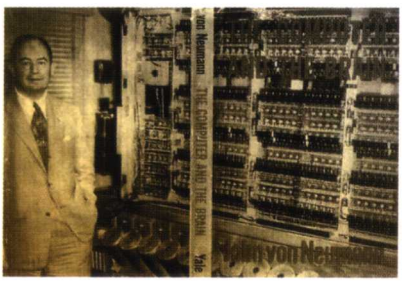
汉密尔顿问题如果到这里就结束的话，也就算不上神奇了。令人意想不到的发展还在后面：1994年底，美国南加州大学一位名叫阿德勒曼 (Leonard M. Adleman) 的计算数学和分子生物学教授在世界著名的《科学》(*Science*) 杂志上发表了一篇颇为轰动的文章。他在自己实验室的试管中采用DNA技术完成了一个能解决7个城市汉密尔顿路径问题的实验。这一革命性的进步，从此拉开了基因计算的帷幕，使之迅速成为生命科学与信息科学交叉领域的前沿，也为解决数学和计算科学中的众多复杂问题带来了新的希望之光！

二、电子计算机还能走多远？

电子计算机的发明和应用无疑是二十世纪中一件了不起的大事，它极大地改变了人类生活方式和社会发展进程。计算机不仅给我们的现实生活提供了种种便利，而且在虚拟的世界中也带给我们无尽的想象和畅游的空间。然而，伟大的事物背后总是隐藏着些许遗憾。电子计算机技术以无与伦比的速度向着顶峰发展，但是巅峰过后，随之而来的将可能是一条漫长的下坡路，在这条路上，电子计算机的不足之处和负面效应也随之日益凸现。

首当其冲的是计算机的存储容量问题。电子计算机之所以能够快速、自动地进行各种复杂的运算，是因为事先已将程序和数据储备在存储器中，运算时只需调用存储器中事先编好的程序，用微处理器进行处理就可以了。这种工作方式自然对存储器设备及其存储能力提出了较高要求。1950年，当冯·诺依曼（John Von Neumann）博士设计出世界上第一台具有内部存储程序功能的计算机EDVAC（Electronic Discrete Variable Automatic Computer，离散变量自动电子计算机）时，使用的是汞延迟线作为存储器。汞在室温时呈液体状态，同时具有导体特性，能使机械波从汞柱的一端开始，让一定厚度的熔融态的金属汞通过一个振动膜片沿着机械波纵向从一端传到另

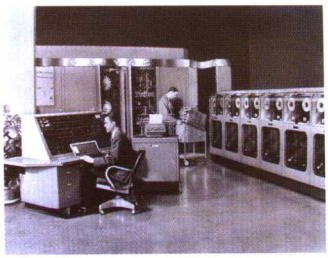
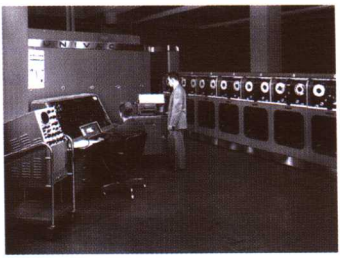
一端，而位于另一端的传感器得到每一比特的信息，就反馈到起点。这一过程是机械和电子的奇妙结合，其设想是通过汞获取并延迟这些数据。由于受环境条件影响，这种存储方式并不精确。



冯·诺依曼和EDVAC

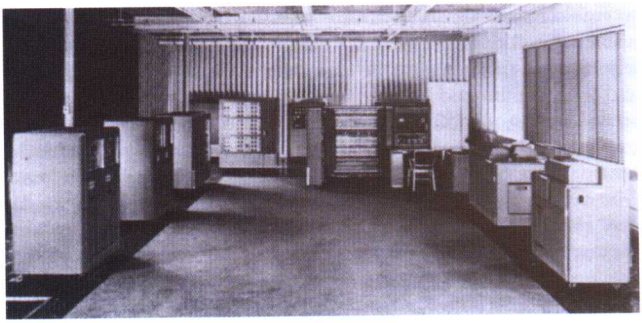
1953年，UNIVAC-I首次采用了磁带机作为外存储器。磁带是所有存储媒体中单位存储信息成本最低、容量最大、标准化程度最高的常用存储介质之一。它互换性好、易于保存。同年，美国国际商用机器（IBM）公司制造的IBM 701计算机开始采用磁鼓作为内存储器，其原理是利

用UNIVAC计算

UNIVAC 全貌

博物馆中的 UNIVAC 主机



IBM 701

用铝鼓筒表面涂覆的磁性材料来存储数据。由于鼓筒旋转速度高，因而存取速度快。也正是凭借这点，IBM 701将UNIVAC甩在了身后。但是磁鼓也有致命的弱点——利用率不高，一个大圆柱体只有表面一层可用于存储。

二十世纪五十年代以来，许多计算机开始采用磁芯存储系统。系统电源关闭后，所存储的数据仍然可以保留，而且磁场能以电子的速度来阅读，这使得交互式计算有了可能。由于采用了电线网格，存储阵列的任何部分都能随机地立即得到访问。直至七十年代初，磁芯存储一直是计算机主存的标准方式。

受磁鼓的启发，磁盘两面都能用来存储，显然利用率比之前的存储材料要高得多。磁盘家族中，最能勾起人们怀旧心理的恐怕就是软盘了，从早期的8英寸软盘、5.25英寸软盘到后来的3.5英寸软盘，都主要是为数据交换和小容量备份之用。其中，3.5英寸1.44MB软盘占据计算机的标准配置地位近20年之久，之后还出现过24MB、100MB、200MB的高密度过渡性软盘和软驱产品。然而，由于USB接口的闪存出现，软盘作为数据交换和小容量备份的统治地位已经动摇，正在慢慢地退出历史舞台。相对软盘而言，世界第一台硬盘存储器是由IBM公司在1956年发明的，总容量只有5MB，共使用了50个直径为24英寸的磁盘。后来，IBM公司提出“温彻斯特”(Winchester)技术，将高速旋转的磁盘、磁头及其寻道机构等全部密封在一个无尘的封闭体中，与外界环境隔绝，避免了灰尘的污染，并采用小型化轻浮力的磁头浮