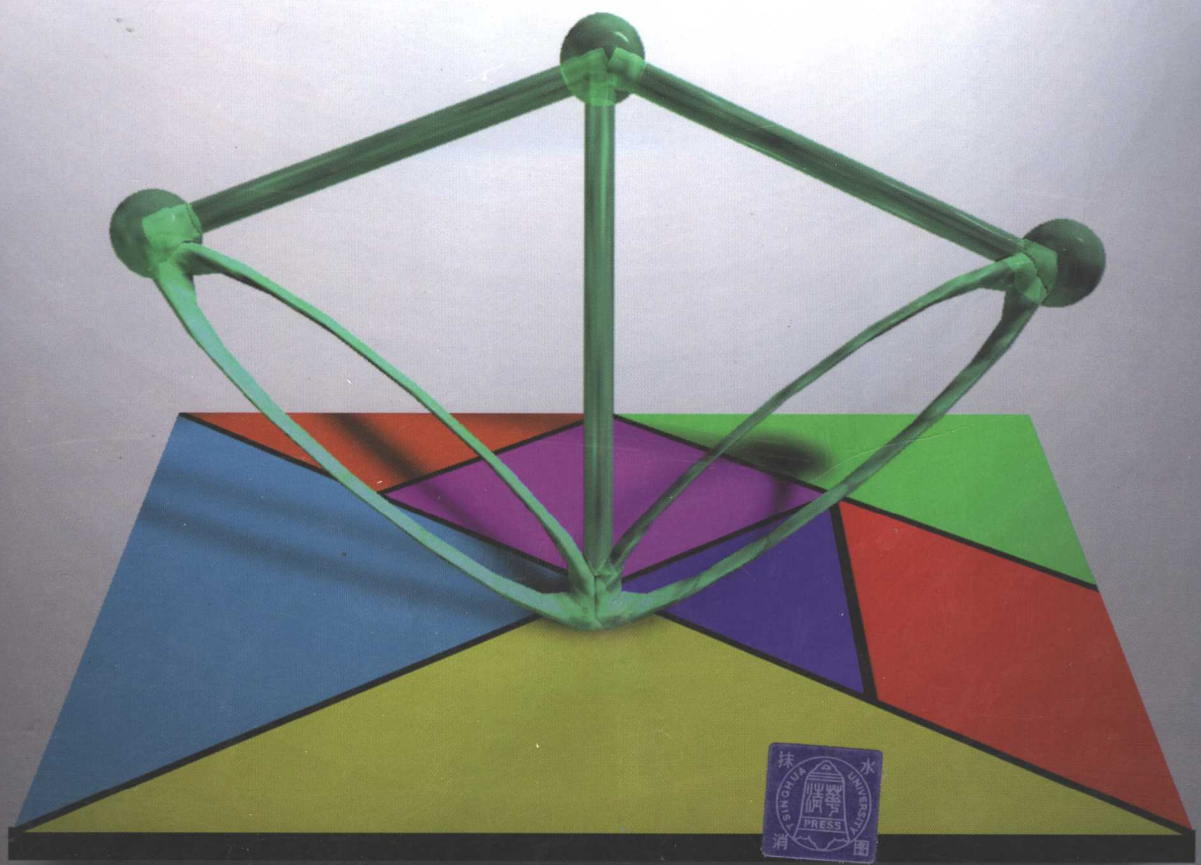
计算机科学组合学丛书

计算机算法导引

——设计与分析

（第2版）

卢开澄 编著



清华大学出版社

计算机科学组合学丛书

计算机算法导引

——设计与分析(第2版)

卢开澄　编著

清华大学出版社
北京

内 容 简 介

本书为《计算机算法导引——设计与分析》的第 2 版。书中内容分 3 部分：第 1 部分是基本算法，按方法论区分，包含优先策略与分治策略、动态规划、概率算法、并行算法、搜索法、数据结构等；第 2 部分是若干专题，包括排序算法、计算几何及计算数论、线性规划；第 3 部分是复杂性理论与智能型算法，其中，智能型算法主要介绍了遗传算法和模拟退火算法。

本书可作为计算机系本科学生及研究生教材，数学系师生和科研工作者也可将其作为参考书。

图书在版编目(CIP)数据

计算机算法导引——设计与分析/卢开澄编著. —2 版. —北京：清华大学出版社，2006.1
(计算机科学组合学丛书)
ISBN 7-302-11501-X

Ⅰ.计… Ⅱ.卢… Ⅲ.①电子计算机—算法设计—高等学校—教材 ②电子计算机—算法分析—高等学校—教材 Ⅳ.TP301.6

中国版本图书馆 CIP 数据核字(2005)第 087654 号

出 版 者：清华大学出版社　　**地　　址**：北京清华大学学研大厦
http://www.tup.com.cn　　**邮　　编**：100084
社 总 机：010-62770175　　**客户服务**：010-62776969
组稿编辑：张　民
文稿编辑：孙建春
封面设计：卢开澄　傅瑞学
印 刷 者：北京鑫丰华彩印有限公司
装 订 者：三河市新茂装订有限公司
发 行 者：新华书店总店北京发行所
开　　本：185×260　**印张**：26.75　**字数**：624 千字
版　　次：2006 年 1 月第 2 版　2006 年 1 月第1 次印刷
书　　号：ISBN 7-302-11501-X/TP·7549
印　　数：1～3000
定　　价：38.00 元

计算机科学组合学丛书序

电子计算机的出现是20世纪的大事，它改变了我们这个世界的面貌。可以毫不夸张地说，它的影响遍及世界的所有角落，几乎无处不感觉到它的存在。数学更不例外。严格地说，电子计算机本身就是近代数学的辉煌成就。将计算机与数学割裂开来，既不合理也不可能。组合学也就是在计算机科学蓬勃发展的刺激下面崛起的，从而成为近若干年来最活跃的数学分支。它研究的问题有的可追溯到欧拉和哈密尔顿等18世纪的数学家，但它成为一新的分支还是近若干年的事。它从与计算机科学相结合中获得了广阔的发展空间，从而也为计算机科学奠定了理论基础。

什么是计算机科学？有的学者将它定义为研究算法的一门学科。研究算法无疑是计算机科学的重要领域，也是本丛书的核心内容，贯穿始终。组合学家在20世纪70年代初建立的算法复杂性的NP理论，至今仍然令无数计算机科学工作者与数学工作者为之折腰。

计算机科学里的组合学内容十分广泛。本丛书涉及组合分析、图论、组合算法、近代密码学、组合优化、编码理论及算法复杂性等七部分。

组合分析是算法的理论基础。组合分析之与组合算法犹如数学分析之与计算数学，众所周知，前者是后者的理论根基。

图论原本是组合数学这个“家族”的主要成员，只因它已成长壮大，故自立门户独立出去。

算法复杂性的NP理论是近30年的一大成就。研究表明对于一类叫做NPC类的困难问题，至今都不存在有效算法，但它们难度相当，只要其中任何一个找到多项式解法，则全体都获得解决；或证明它们根本不存在有效办法。不论是前者还是后者都还看不见露到海平面上的桅杆塔，它吸引了众多的有志之士。密码学是其中十分引人入胜的分支。如若设计好的密码，对它的破译等价于某一NPC类困难问题，无疑这样的密码将是牢不可破的。

在计算机网络深入普及的信息时代，信息本身就是时间，就是财富。信息的传输通过是脆弱的公共信道，信息储存于“不设防”的计算机系统中，如何保护信息的安全使之不被窃取及不至于被篡改或破坏，已成为当今被普遍关注的重大问题。密码是有效而且可行的办法。在计算机网络的刺激下，近代密码学便在算法复杂性理论的基础上建立起来了。密码作为一种技术，自从人类有了战争，不久便有了它。但作为一门学科则是近20多年的事。甚至于它已成为其他学科的基础。密码也从此走出“军营”，进入百姓家。

实际中的“优化”问题是大量的，半个多世纪以来它曾经几度辉煌。近来在计算机科学的影响下，又出现了若干闪光点，十分耀眼，引人注目。

实际上密码也是一种编码。如果说密码学研究的编码是保证通信的保密与安全，则编码理论研究的是通信中如何纠错与检错。计算机纠错码是既实用，理论上又饶有趣味的分支。

本丛书是作者在清华大学计算机科学与技术系长期工作的总结。它不是一部“长篇”记述，而是互相关联又彼此相对独立，因此难免有少量交叉。它们涉及的面如此之广，囿于作者的水平，缺点和错误在所难免，敬请读者不吝指正。谢谢。

前　言

算法设计与分析是一个与计算机科学紧密相关的数学分支，它还很年轻，远未定型。由于它所包含的内容浩如烟海，故不一而足，描述算法的方式也迥异，不像普天之下的微积分学那样都是相似形。

众多的算法彼此独立，读起来往往使人有只见树木不见森林之感。若依方法论来区分，有的又无法概括到。本书为《计算机算法导引——设计与分析》的第 2 版，所讨论的内容分 3 部分：第 1 部分是基本方法，按方法论区分，包含优先策略与分治策略、动态规划、概率算法、并行算法、搜索法等；第 2 部分是若干专题，有排序算法、计算几何及计算数论、线性规划等；第 3 部分是复杂性理论与智能型算法等，智能型算法主要介绍遗传算法和模拟退火算法。

囿于个人的水平，疏漏和缺点在所难免，望读者多多指正。

卢开澄
2004 年 9 月

目　录

第1部分　基本算法

第 2 部分 若 干 专 题

第 3 部分 复杂性理论与智能型算法

第 1 部分

基 本 算 法

第1章 数学准备

“计算机算法”顾名思义讨论的是一类与计算机有关的数学问题，它需要许多数学基础的支持，现在仅就递推关系与母函数作一些介绍。

首先介绍一些常用的符号和术语，例如 R 表示实数集合；R^+ 表示正的实数集合。

又如 $N=\{0,1,2,\cdots\}$ 即包含 0 在内的正整数；$N^+=\{1,2,\cdots\}$ 即正整数。

还经常用到的符号如 $\triangleq$ 表示“定义”，$\exists$ 表示存在，如 $\exists n_0\in N$；$\forall$ 为所有的，如 $\forall n\geqslant n_0$，即对所有 $\geqslant n_0$ 的数 n。

令 f 和 g 是由映射 N 到 R 的两个函数。请读者注意下面三个定义的区别。

定义 令 $f: N\to R^+$，

$O(f)\triangleq\{g: N\to R^+ \mid \exists c\in R^+, \exists n_0\in N, \forall n\geqslant n_0, \text{使 } g(n)\leqslant cf(n)\}$

也就是 $O(f)$ 是函数 $g: N\to R^+$ 的集合，它以 $cf(n)$ 为上界。亦即若存在 $c\in R$，且 $\lim\limits_{n\to\infty}\dfrac{g(n)}{f(n)}=c$，则 $g\in O(f)$。基于上述概念，若 $f(n)=n^3$，$g(n)=n^2$，则 $g\in O(f)$，但 $f\notin O(g)$。

定义 令 $f: N\to R^+$，

$\Omega(f)\triangleq\{g: N\to R \mid \exists n_0\in N, \exists c\in R^+, \forall n\geqslant n_0, \text{使 } cf(n)\leqslant g(n)\}$，

若 $\lim\limits_{n\to\infty}\dfrac{g(n)}{f(n)}=\infty$ 或 $\lim\limits_{n\to\infty}\dfrac{g(n)}{f(n)}=c>0$，则 $g\in\Omega(f)$。

定义 令 $f: N\to R$，

$\Theta(f)=O(f)\cap\Omega(f)$，

若 $\lim\limits_{n\to\infty}\dfrac{g(n)}{f(n)}=c, c\in R^+$，则 $g\in\Theta(f)$。

其中 $c\in R^+$ 表示 $c\neq 0, c\neq\infty$，$g\in\Theta(f)$ 意味着 g 和 f 同阶。

O,Ω,Θ 有以下的性质（假定 $f,g,h: N\to R$）：

(1) 若 $f\in O(g), g\in O(h)$，则 $f\in O(h)$；

(2) $f\in O(g)\Leftrightarrow g\in\Omega(f)$；

(3) 若 $f\in\Theta(g)$，则 $g\in\Theta(f)$；

(4) $O(f+g)=O(\max\{f,g\})$。

1.1 母 函 数

递推关系是计数的一个强有力工具，特别在作算法分析时是必需的。递推关系的求解主要是利用母函数。当然母函数尚有其他用处，但这里主要介绍在解递推关系上的应用。例如

$$(1+a_1x)(1+a_2x)\cdots(1+a_nx)$$
$$=1+(a_1+a_2+\cdots+a_n)x+(a_1a_2+a_1a_3+\cdots+a_{n-1}a_n)x^2+\cdots+a_1a_2\cdots a_na^n \quad (1.1)$$

x^2 项的系数 $a_1a_2+a_1a_3+\cdots+a_{n-1}a_n$ 中所有的项包含了从 n 个元素 $a_1, a_2, \cdots, a_n$ 中

取两个组合的全体；同理，x^3 项系数 $a_1a_2a_3+a_1a_2a_4+\cdots+a_{n-2}a_{n-1}a_n$ 包含了从 n 个元素 $a_1,a_2,\cdots,a_n$ 中取 3 个组合的全体，以此类推。

若令 $a_1=a_2=\cdots=a_n=1$ 在(1.1)式中 x^2 项系数 $a_1a_2+a_1a_3+\cdots+a_{n-1}a_n$ 中每一个组合有 1 个贡献，其他各项以此类推，故有

$$(1+x)^n=1+C(n,1)x+C(n,2)x^2+\cdots+C(n,n)x^n \tag{1.2}$$

另一方面，

$$(1+x)^m(1+x)^n=(1+x)^{m+n}$$

所以 $[C(m,0)+C(m,1)x+\cdots+C(m,m)x^m][C(n,0)+C(n,1)x+\cdots+C(n,n)x^m]$

$=[C(m+n,0)+C(m+n,1)x+\cdots+C(m+n,m+n)x^{m+n}]$

比较等号两端 x 项对应系数，可得一等式

$$C(m+n,r)=C(m,0)C(n,r)+C(m,1)C(n,r-1)+\cdots+C(m,r)C(n,0)$$

这里不过给出另一种比较简单的证明。

同样，对于$(1+x)^n\left(1+\frac{1}{x}\right)^m$，(设 $n\geqslant m$)，用类似方法可得等式

$$C(m+n,m)=C(n,0)C(m,0)+C(n,1)C(m,1)+\cdots+C(n,m)C(m,m)$$

方法如下。

$$(1+x)^n\left(1+\frac{1}{x}\right)^m=x^{-m}(1+x)^{m+n} \tag{1.3}$$

所以 $[C(n,0)+C(n,1)x+\cdots+C(n,n)x^n][C(m,0)+C(m,1)x^{-1}+\cdots+C(m,m)x^{-m}]$

$=x^{-m}[C(m+n,0)+C(m+n,1)x+C(m+n,2)x^2+C(m+n,m+n)x^{m+n}]$

比较等号两端的常数项，即得公式(1.3)。

又如等式

$$(1+x)^m=C(n,0)+C(n,1)x+C(n,2)x^2+\cdots+C(n,n)x^n$$

令 $x=1$，可得

$$C(n,0)+C(n,1)+\cdots+C(n,n)=2n \tag{1.4}$$

(1.2)式等号的两端对 x 求导可得

$$n(1+x)^{n-1}=C(n,1)+2C(n,2)x+3C(n,3)x^2+\cdots+nC(n,n)x^{n-1} \tag{1.5}$$

等式(1.5)两端令 $x=1$，得

$$C(n,1)+2C(n,2)+3C(n,3)+\cdots+nC(n,n)=n2^{n-1} \tag{1.6}$$

类似的办法可以得到

$$C(n,1)x+2C(n,2)x^2+3C(n,3)x^3+\cdots+nC(n,n)x^n$$

$$=nx(1+x)^{n-1}$$

所以

$$C(n,1)+2^2C(n,2)+3^2C(n,3)+\cdots+n^2C(n,n)$$

$$=n(n+1)2^{n-2} \tag{1.7}$$

还可以类似地推出一些等式，但通过上面一些例子已可见函数$(1+x)^n$ 在研究序列 $C(n,0),C(n,1),\cdots,C(n,n)$ 的关系时所起的作用，对于其他序列也有同样结果，现引进母函数概念如下。

定义： 对于序列 $a_0,a_1,a_2,\cdots$，构造一函数

$$G(x)=a_0+a_1x+a_2x^2+\cdots$$