高等院校信息技术规划教材

数据与算法

徐士良 编著

清华大学出版社

高等院校信息技术规划教材

数据与算法

徐士良 编著

清华大学出版社
北京

内容简介

本书是作者对长期从事数据结构与数值方法课程的教学经验进行总结和提炼而写成的，涉及计算机软件基础与应用中的主要知识、技术和方法，既包含了数据结构的基本知识，又包含了数值方法的内容。具体内容有集合、数据结构与算法的基本概念，线性数据结构的存储与运算，非线性数据结构的存储与运算，查找与排序技术，矩阵与线性方程组，插值与逼近，各种数值问题的近似解法，数值问题的连分式解法。每章都配有一定数量的习题。

本书内容丰富、通俗易懂、实用性强，可作为高等学校相应课程的教材，也可作为广大从事计算机应用工作的科技人员的参考书。

图书在版编目(CIP)数据

数据与算法/徐士良编著. —北京：清华大学出版社，2014
高等院校信息技术规划教材
ISBN 978-7-302-36205-0

Ⅰ. ①数…　Ⅱ. ①徐…　Ⅲ. ①数据结构－高等学校－教材 ②算法分析－高等学校－教材
Ⅳ. ①TP311.12

中国版本图书馆 CIP 数据核字(2014)第 076252 号

责任编辑：焦　虹
封面设计：常雪影
责任校对：梁　毅
责任印制：何　芊

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954
印 装 者：北京密云胶印厂
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：33　　**字　　数**：823 千字
版　　次：2014 年 12 月第 1 版　　**印　　次**：2014 年 12 月第 1 次印刷
印　　数：1～2000
定　　价：54.50 元

产品编号：058873-01

前言 Foreword

作者长期从事数值方法与数据结构方面的教学工作，讲授的主要内容是数值计算与非数值计算的问题，同时也编写了有关的教材。这两方面的知识内容都可以也应该归属于计算机软件的基础与应用领域。随着教学改革的深入，从学生计算机知识的结构以及实际应用考虑，在计算机课程的教学中尝试将这两方面的知识内容作为有机的整体设置成一门课程，即数据与算法。这也是一门计算机软件基础与应用的重要课程。

实际上，不管是离散数据还是数值型数据，都属于数据，其不同点只在于二者的背景以及使用的领域有所不同，在计算机中存储的方式有所不同，对它们的处理方法也有所不同，但它们确实是人们认识事物的同一类对象。另一方面，现代科学技术发展到现在，特别是计算机技术的飞速发展，不论是数值计算还是非数值计算都以计算机作为工具，大量的数据处理与复杂的数值计算都需要算法的支持。因此，不论是数值型数据还是非数值型数据，它们都应该作为计算机处理的对象与算法一起考虑，而算法不仅是离散数据运算的基础，也是数值计算的基础。数据与算法是一个整体，应该作为计算机教学中的一门重要课程。

数值方法与数据结构原来是两门不同的课程，前者属于数学范畴，后者属于计算机软件基础的范畴。这两个原本属于不同范畴的课程内容如何合并在一起？作者认为，既然是不同类型的数据，又采用了不同的运算方式，就应该以数据类型为基础，以“数据＋算法”与“问题＋算法”的方式展开。这样，就不会觉得原本属于数据结构的内容与原本属于数值方法的内容被混杂在一起了。

作为一门新的课程，单纯考虑内容的有机结合是不够的，更重要的是要创新、精简、实用，体现改革与发展。在数值方法中，求解一个非线性方程（即方程求根）有很多经典方法，数值积分也有很多经典方法。一般来说，求解一类数值问题，往往有很多经典方法。在作者看来，绝大部分学生学习数值方法，并不是为了了解各种求解方法，而是为了解决实际问题，因此，不需要介绍所有的求解方法，

只要介绍一种行之有效的方法就行了。作者在长期的教学过程中，对连分式方法进行了一定的研究，认为利用连分式方法解决数值计算问题，具有精确、计算复杂度低、简单实用等优点。因此，在编写本书时就强调了这种方法。对于数值计算的问题，不是介绍解决这类问题的所有方法，而是只介绍解决这类问题的一个基本方法，然后利用连分式的特点对这个基本方法进行校正，以求对不同性态的问题都能得到精确的结果。

本书内容丰富、通俗易懂、实用性强，既包含了数据结构的基本知识，又包含了数值方法的主要内容。书中所有算法均用 C++ 语言描述，这些算法程序均在 Visual C++ 6.0 环境下调试通过。本书可作为高等学校本科生或研究生相关课程的教材，也可作为广大从事计算机应用工作的科技人员的参考书。

如果读者还需要更多的算法，则可以将参考文献中的[1]作为与本书配套的工具书，在那本书中给出了解决各种问题的多种算法。

由于作者水平有限，书中难免有错误或不妥之处，恳请读者批评指正。

作　者

目录 Contents

第1章 chapter 1

预备知识

1.1 集　　合

1.1.1 集合及其基本运算

1. 集合的基本概念

所谓集合，是指若干个或无穷多个具有相同属性的元(元素)的集体。通常，一个集合名称用大写字母表示，而集合中的某个元素用小写字母表示。

如果集合 M 由 $n(n\geqslant 0)$ 个元素 $a_1,a_2,\cdots,a_n$ 组成，则称集合 M 为有限集。例如，大于1而小于100的所有整数构成的集合 A 为有限集。如果一个集合中有无穷多个元素，则称此集合为无限集。例如，所有整数构成的集合 $\mathbf{Z}$，所有实数构成的集合 $\mathbf{R}$，大于0而小于1的所有实数构成的集合 B 等均为无限集。不包括任何元素的集合称为空集。例如，大于1而小于2的整数构成的集合为空集。空集通常用 $\varnothing$ 表示。

如果 M 是一个集合，a 是集合 M 中的一个元素，则记作 $a\in M$，称元素 a 属于集合 M；如果 a 不是集合 M 中的元素，则记作 $a\notin M$，称元素 a 不属于集合 M。

对于一个集合，通常用以下两种方法表示。

(1) 列举法　用列举法表示一个集合是将此集合中的元素全部列出来，或者列出若干项但能根据规律可知其所有的元素。例如，上述举出的几个集合的例子可以用如下列举法表示：

大于1而小于100的所有整数的集合 A 可以表示为

$$A=\{2,3,4,\cdots,99\},\quad \text{有限集}$$

所有整数构成的集合 $\mathbf{Z}$ 可以表示为

$$\mathbf{Z}=\{0,\pm 1,\pm 2,\pm 3,\cdots,\},\quad \text{无限集}$$

空集表示为

$$\varnothing=\{\},\quad \text{空集}$$

(2) 性质叙述法　用性质叙述法表示一个集合是将集合中的元素所具有的属性描述出来。例如：

大于1而小于100的所有整数的集合 A 可以表示为

$$A = \{a \mid 1 < a < 100 \text{ 的所有整数}\}$$

所有整数构成的集合 **Z** 可以表示为

$$\mathbf{Z} = \{z \mid z \text{ 为一切整数}\}$$

大于 0 而小于 1 的所有实数构成的集合 B 可以表示为

$$B = \{b \mid 0 < b < 1 \text{ 的所有实数}\}$$

所有实数构成的集合 **R** 可以表示为

$$\mathbf{R} = \{r \mid r \text{ 为一切实数}\}$$

设 M 与 N 为两个集合。如果集合 M 中的每一个元素也都为集合 N 的元素，则称集合 M 为 N 的子集，记作 $M \subseteq N$ 或 $N \supseteq M$。如果 $M \subseteq N$，且 N 中至少有一个元素 $a \notin M$，则称 M 为 N 的真子集，记作 $M \subset N$ 或 $N \supset M$。如果 $M \subseteq N$ 且 $N \subseteq M$，则称集合 M 和集合 N 相等，记作 $M = N$。

2. 集合的基本运算

(1) 两个集合的并(union)　设有两个集合 M 和 N，它们的并集记作 $M \cup N$，其定义如下：

$$M \cup N = \{\alpha \mid \alpha \in M \text{ 或 } \alpha \in N\}$$

即两个集合 M 与 N 的并集是指 M 与 N 中所有元素(去掉重复的元素)组成的集合。

(2) 两个集合的交(intersection)　设有两个集合 M 和 N，它们的交集记作 $M \cap N$，其定义如下：

$$M \cap N = \{\alpha \mid \alpha \in M \text{ 且 } \alpha \in N\}$$

即两个集合 M 与 N 的交集是指 M 与 N 中所有共同元素组成的集合。

两个集合 M 与 N 的并、交均满足交换律，即

$$M \cup N = N \cup M$$
$$M \cap N = N \cap M$$

(3) 两个集合的差(difference)　设有两个集合 M 和 N，M 和 N 的差集记作 $M - N$，其定义如下：

$$M - N = \{\alpha \mid \alpha \in M \text{ 但 } \alpha \notin N\}$$

两个集合的差不满足交换律，即

$$M - N \neq N - M$$

例 1.1　设集合

$$A = \{a, b, c, d, e\}$$
$$B = \{d, e, f, g, h\}$$

则

$$A \cup B = \{a, b, c, d, e, f, g, h\}$$
$$A \cap B = \{d, e\}$$
$$A - B = \{a, b, c\}$$
$$B - A = \{f, g, h\}$$

3. 集合的并、交、差运算的基本性质

(1) 结合律

$$(A\cap B)\cap C=A\cap(B\cap C)$$
$$(A\cup B)\cup C=A\cup(B\cup C)$$

(2) 分配律

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$$
$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$$

(3) $(A-B)\cup(B-A)=(A\cup B)-(A\cap B)$

(4) $B\cap(A-B)=\varnothing$

$(A\cap B)\cup(A-B)=A$

4. 映射

定义 1.1 设 A、B 是两个非空集。如果根据一定的法则 f,对于每一个 $x\in A$,在 B 中都有唯一确定的元素 y 与之对应,则称 f 是定义在 A 上而在 B 中取值的映射,记作 $f:A\rightarrow B$,并将 x 与 y 的关系记作 $y=f(x)$。x 称为自变元,y 称为在 f 作用下 x 的像。

集合 A 称为 f 的定义域,$f(A)=\{f(x)|x\in A\}$称为 f 的值域。

定义 1.2 设给定映射 $f:A\rightarrow B$,且 $B=f(A)$(即 f 的像充满整个 B)。如果对于每个 $y\in B$,仅有唯一的 $x\in A$ 使 $f(x)=y$,则称 f 有逆映射 f^{-1}(它是定义在 $f(A)$上而取值于 A 的映射)。当映射 $f:A\rightarrow f(A)$有逆映射时,则称 f 是一一映射。

定义 1.3 若 A、B 两集合有一一映射 f 存在,使 $f(A)=B$,则称 A 与 B 成一一对应。

如果集合 A 与 B 为一一对应,则称它们互相对等,并记作 $A\sim B$。当两个集合互相对等时,称它们有相等的浓度(或相等的元素个数)。

例 1.2 设两集合为

$$A=\{1,2,\cdots,10\}=\{x\mid 1\leqslant x\leqslant 10\text{ 的整数}\}$$
$$B=\{1,2,\cdots,20\}=\{y\mid 1\leqslant y\leqslant 20\text{ 的整数}\}$$

若映射 $f:A\rightarrow B$ 为

$$y=2x$$

其中定义域为 A,值域为

$$f(A)=\{2,4,6,\cdots,20\}$$

显然,映射 f 不是一一映射。

例 1.3 设两集合为

$$A=\{x\mid 0\leqslant x\leqslant 4\text{ 的所有实数}\}$$
$$B=\{y\mid 0\leqslant y\leqslant 2\text{ 的所有实数}\}$$

考虑映射 $f:A\rightarrow B$

$$y=\sqrt{x}$$

其中定义域为 A,值域为 $f(A)=B$。

显然，f 为一一映射，集合 A 与 B 互相对等，即 $A \sim B$。

集合的对等满足以下性质：

- 自反性。即 $A \sim A$。
- 对称性。即若 $A \sim B$，则 $B \sim A$。
- 传递性。即若 $A \sim B$ 且 $B \sim C$，则 $A \sim C$。

1.1.2 自然数集与数学归纳法

由所有自然数所组成的集合

$$\{1,2,3,\cdots\}$$

称为自然数集。自然数集是一个无限集。

由自然数组成的集合均是自然数集的子集。自然数集的子集可以是有限集，也可以是无限集。与自然数集对等（即具有相等浓度）的集合称为可列集（或可数集）。任一可列集中的元素排列时可标以正整数下标，即任意可列集 M 均可写成

$$M = \{a_1, a_2, \cdots, a_n, \cdots\}$$

关于自然数集及其子集有以下两个命题成立。

定理 1.1 在自然数集的任一非空子集 M 中，必定有一个最小数。即在集合 M 中有不大于其他任意数的数。

证明：因为 M 非空，所以在 M 中可以取得一自然数 n。

显然，M 中所有不大于 n 的自然数形成的非空子集 N 包含在 M 中，即 $N \subset M$。如果 N 中有最小数，则此最小数就是 M 的最小数。

而在 N 中最多有 n 个自然数（$1 \sim n$），因此，N 中有一个最小数。

综上所述，在自然数集的任一非空子集 M 中，必定有一个最小数。定理得证。

定理 1.2 设 M 是由自然数形成的集合，如果它含有 $1,2,\cdots,k$，并且当它含有数 $n-1, n-2, \cdots, n-k(n>k)$ 时，也含有数 n，那么它含有所有的自然数，即 M 是自然数集。

证明：设 N 是所有不属于 M 的自然数形成的集合，则 $1,2,\cdots,k \notin N$。

现假设 N 不是空集，则由定理 1.1 可知：在 N 中必定有一个最小数。设此最小数为 c。

由于 c 是 N 中的最小数，即 $c \in N$，因此，$c \notin M$，且 $c \neq 1,2,\cdots,k$，同时，$c-1, c-2, \cdots, c-k$ 均为自然数。又由于 c 是 N 中的最小数，所以自然数 $c-1, c-2, \cdots, c-k \notin N$，即 $c-1, c-2, \cdots, c-k \in M$，而根据定理中的条件应有 $c \in M$。

由上所述，一方面有 $c \notin M$，另一方面又有 $c \in M$，这就导致矛盾。这个矛盾是由于一开始假定 N 不是空集所造成的。因此，N 只能为空集，即所有自然数均在 M 中，M 为自然数集。

定理得证。

定理 1.2 是数学归纳法的基础。通常，为了证明一个命题对于所有的自然数是真，采用数学归纳法证明的步骤如下：

(1) 证明命题对于自然数 $1,2,\cdots,k$ 是真的。

(2) 假设命题对于自然数 $n-k, n-k+1, \cdots, n-1(n>k)$ 是真的（这一步称为归纳

假设)。

(3) 证明命题对于自然数 n 也是真的。

上述步骤(1)中 k 值的选取决定于在步骤(3)的证明过程中要用到归纳假设的最小自然数。在步骤(3)中,为了证明命题对于自然数 n 是真,要用到归纳假设的最小自然数。如果它为 $n-k$,则在步骤(1)中要对 $1\sim k$ 中的所有自然数证明命题为真。

例 1.4 证明下列命题:

对于任意的自然数 n,必存在一对非负整数(i,j)有

$$7+n=3i+5j$$

下面用数学归纳法证明这个命题。

证明:

(1) 当 $n=1$ 时,有 $7+1=3+5$,即 $i=1,j=1$。命题成立。

当 $n=2$ 时,有 $7+2=3\times3+5\times0$,即 $i=3,j=0$。命题成立。

当 $n=3$ 时,有 $7+3=3\times0+5\times2$,即 $i=0,j=2$。命题成立。

(2) 假设命题对于自然数 $n-1,n-2,n-3(n>3)$成立(归纳假设)。

(3) 考虑自然数 n,有

$$7+n=[7+(n-3)]+3$$

根据归纳假设,对于自然数 $n-3$ 命题成立,设存在一对非负整数 i_1,j_1 有

$$7+(n-3)=3i_1+5j_1$$

则有

$$7+n=[7+(n-3)]+3=3(i_1+1)+5j_1=3i+5j$$

其中 $i=i_1+1,j=j_1$ 均为非负整数。即对于自然数 n 命题也成立。

由此得出结论,对于所有的自然数 n 命题成立。

在这个例子中,由于步骤(3)的证明过程中,要用到归纳假设的最小自然数为 $n-3$,因此在步骤(1)中取 $k=3$。

1.1.3 笛卡儿积

在 1.1.1 节中介绍了集合的并、交、差运算。对于集合,还有一种很重要的运算——笛卡儿积(Cartesian Product)。

设有 n 个集合 $D_1,D_2,\cdots,D_n$,此 n 个集合的笛卡儿积定义为

$$D_1\times D_2\times\cdots\times D_n=\{(d_1,d_2,\cdots,d_n)\mid d_i\in D_i,i=1,2,\cdots,n\}$$

其中$(d_1,d_2,\cdots,d_n)$称为 n 元组(n-tuple),d_i 称为 n 元组的第 i 个分量。

由笛卡儿积的定义可以看出,n 个集合的笛卡儿积是以 n 元组为元素的集合,而每一个 n 元组中的第 i 个分量取自于第 i 个集合 D_i。

例 1.5 设有三个集合

$$A=\{a_1,a_2,a_3\},\quad B=\{b_1,b_2\},\quad C=\{c_1,c_2\}$$

则它们的笛卡儿积为

$$A\times B\times C=\{(a_1,b_1,c_1),(a_1,b_1,c_2),(a_1,b_2,c_1),(a_1,b_2,c_2),$$

$$(a_2,b_1,c_1)(a_2,b_1,c_2),(a_2,b_2,c_1),(a_2,b_2,c_2),$$
$$(a_3,b_1,c_1),(a_3,b_1,c_2),(a_3,b_2,c_1),(a_3,b_2,c_2)\}$$

如果 n 个集合 $D_1,D_2,\cdots,D_n$ 中的元素个数分别为 $m_1,m_2,\cdots,m_n$，则其笛卡儿积中共有 $m_1\times m_2\times\cdots\times m_n$ 个 n 元组。即 n 个集合的笛卡儿积是所有 n 元组组成的集合。

1.1.4 二元关系

定义 1.4 设 M 和 N 是两个集合，则其笛卡儿积

$$M\times N=\{(x,y)\mid x\in M 且 y\in N\}$$

的每一个子集称为在 $M\times N$ 上的一个二元关系。

如果 $M=N$，则其笛卡儿积

$$M\times M=\{(x,y)\mid x,y\in M\}$$

的每一个子集称为在集合 M 上的一个二元关系，简称为在集合 M 上的一个关系。

例 1.6 设集合 M 为

$$M=\{a,b,c,d,e,f\}$$

则下列每一个二元组的集合是在集合 M 上的一个关系：

$$R_1=\{(a,b),(b,c),(c,d),(d,e),(e,f)\}$$
$$R_2=\{(a,e),(a,a),(c,f),(d,b),(e,a),(f,c),(b,d)\}$$
$$R_3=\{(a,a),(b,b),(c,c),(d,d),(e,e),(f,f),(c,f),(e,a)\}$$
$$R_4=\{(a,b),(b,e),(c,d),(d,f),(a,e),(c,f)\}$$

集合 M 上的一个关系实际上反映了集合 M 中各元素之间的联系。

定义 1.5 设 R 是集合 M 上的一个关系。

(1) 如果 $(a,b)\in R$，则称 a 是 b 的关于 R 的前件(predecessor)，b 是 a 的关于 R 的后件(successor)。

(2) 如果对于每一个 $a\in M$，都有 $(a,a)\in R$，则称关系 R 是自反的(reflexive)；如果对于任何 $a\in M$，$(a,a)\in R$ 均不成立，则称关系 R 是非自反的(antireflexive)。

(3) 如果 $(a,b)\in R$ 时必有 $(b,a)\in R$，则称关系 R 是对称的(symmetric)。

(4) 如果当 $(a,b)\in R$ 且 $(b,c)\in R$ 时必有 $(a,c)\in R$，则称关系 R 是传递的(transitive)。

在例 1.6 中，关系 R_1 是非自反的，但不是对称的，也不是传递的；关系 R_2 是对称的，但不是自反的，也不是非自反的，也不是传递的；关系 R_3 是自反的，但不是对称的，也不是传递的；关系 R_4 是传递的，且是非自反的，但不是对称的。

由此可以看出，集合 M 中的各元素之间的逻辑关系可以由集合 M 上的一个关系来描述。

定义 1.6 设 R 是 M 上的一个传递关系，且 $T\subseteq R$。若对于任何 $(x,y)\in R$，在 M 中有元素 $x_0,x_1,x_2,\cdots,x_n(n\geqslant1)$ 满足：(1) $x_0=x$，(2) $x_n=y$，(3) $(x_{i-1},x_i)\in T(i=1,2,\cdots,n)$；则称关系 T 是关系 R 的基(basis)，又称关系 R 是关系 T 的传递体(transitive hull)。

例 1.7 设集合 M 为

$$M=\{1,2,3,4,5\}$$

集合 M 上的一个关系为

$$R=\{(1,2),(2,3),(3,5),(3,4),(1,3),(1,4),(2,4),(2,5),(1,5)\}$$

可以验证，关系 R 是非自反的，且是传递的。现考虑集合 M 上的另一个关系

$$T=\{(1,2),(2,3),(3,5),(3,4)\}$$

显然 $T\subseteq R$，并且可以验证，对于关系 R 中的每一个二元组 (x,y)，在 M 中存在元素 x_0，$x_1,x_2,\cdots,x_n$，满足定义 1.6 中的三个条件。因此，T 是 R 的具有 4 个元素(二元组)的基。

同样还可以验证关系

$$T_1=\{(1,2),(1,4),(2,3),(3,5),(3,4)\}$$

是 R 的具有 5 个元素(二元组)的基。但关系

$$T_2=\{(1,2),(2,3),(3,4),(4,5)\}$$

不是 R 的基，因为 $(4,5)\notin R$。

1.2 数据结构的基本概念

计算机已被广泛用于数据处理。实际问题中的各数据元素之间总是相互关联的。所谓数据处理，是指对数据集合中的各元素以各种方式进行运算，包括插入、删除、查找、更改等运算，也包括对数据元素进行分析。在数据处理领域中，建立数学模型有时并不十分重要，事实上，许多实际问题是无法表示成数学模型的。人们最感兴趣的是知道数据集合中各数据元素之间存在什么关系；为了提高处理效率，应如何组织它们，即如何表示所需要处理的数据元素。

1.2.1 什么是数据结构

简单地说，数据结构是指相互有关联的数据元素的集合。例如，向量和矩阵就是数据结构，在这两个数据结构中，数据元素之间有着位置上的关系。又如，图书馆中的图书卡片目录，则是一个较为复杂的数据结构，对于列在各卡片上的各种书之间，可能在主题、作者等问题上相互关联，甚至一本书本身也有不同的相关成分。数据元素具有广泛的含义。一般来说，现实世界中客观存在的一切个体都可以是数据元素。例如：

描述一年四季的季节名

春、夏、秋、冬

可以作为季节的数据元素。

表示数值的各个数

18，11，35，23，16，…

可以作为数值的数据元素。

表示家庭成员的各成员名

父亲、儿子、女儿

可以作为家庭成员的数据元素。