数据结构与算法 (十二)

张铭 主讲

采用教材：张铭，王腾蛟，赵海燕 编写
高等教育出版社，2008. 6 （“十一五”国家级规划教材)

http://www.jpk.pku.edu.cn/pkujpk/course/sjjg

第十二章 高级数据结构

- 12.1 多维数组
  - 12.1.1 基本概念
  - 12.1.2 数组的空间结构
  - 12.1.3 数组的存储
  - 12.1.4 数组的声明
  - 12.1.5 用数组表示特殊矩阵
  - 12.1.6 稀疏矩阵
- 12.2 广义表
- 12.3 存储管理
- 12.4 Trie 树
- 12.5 改进的二叉搜索树

基本概念

- 数组（Array）是数量和元素类型固定的有序序列
- 静态数组必须在定义它的时候指定其大小和类型
- 动态数组可以在程序运行才分配内存空间

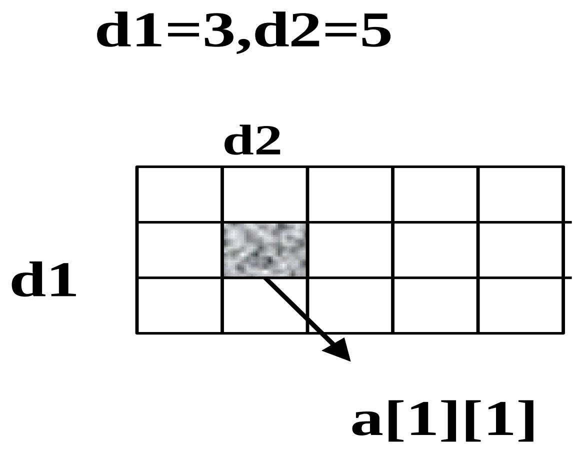
基本概念 (续)

- 多维数组（Multi-array）是向量的扩充
- 向量的向量就组成了多维数组
- 可以表示为：

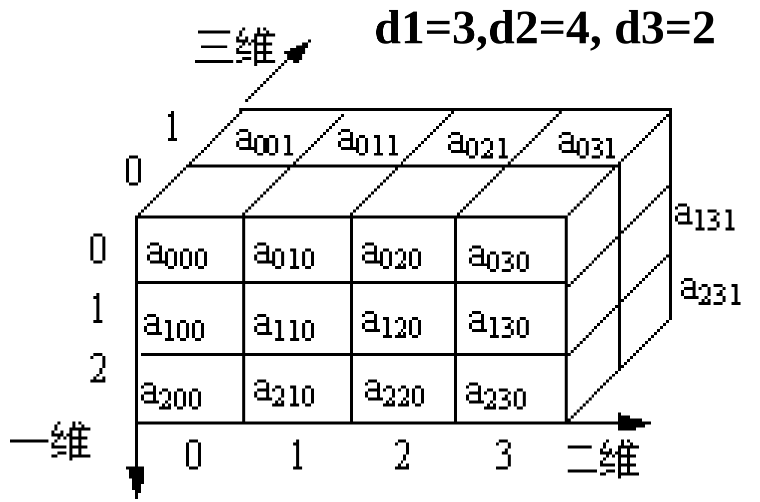
  ELEM $A[c_1..d_1][c_2..d_2]\ldots[c_n..d_n]$
- c_i 和 d_i 是各维下标的下界和上界。所以其元素个数为：

$$\prod_{i=1}^{n}(d_i - c_i + 1)$$

数组的空间结构



二维数组



三维数组

d1[0..2],d2[0..3],d3[0..1]分别为3个维

数组的存储

- 内存是一维的，所以数组的存储也只能是一维的
  - 以行为主序 (也称为 “行优先”)
  - 以列为主序 (也称为 “列优先”)

$$X=\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}$$

12.1 多维数组

Pascal语言的行优先存储 a[1..k,1..m,1..n]

a_{111} a_{112} a_{113} … a_{11n} a_{11*}

a_{121} a_{122} a_{123} … a_{12n} a_{12*}

…………………………

a_{1m1} a_{1m2} a_{1m3} … a_{1mn} a_{1m*}

a_{211} a_{212} a_{213} … a_{21n} a_{21*}

a_{221} a_{222} a_{223} … a_{22n} a_{22*}

…………………………

a_{2m1} a_{2m2} a_{2m3} … a_{2mn} a_{2m*}

⋮

a_{k11} a_{k12} a_{k13} … a_{k1n}

a_{k21} a_{k22} a_{k23} … a_{k2n}

…………………………

a_{km1} a_{km2} a_{km3} … a_{kmn}

FORTRAN的列优先存储

$a[1..k, 1..m, 1..n]$

a_{111} a_{211} a_{311} … a_{k11} a_{*11}

a_{121} a_{221} a_{321} … a_{k21} a_{*21}

…………………………

a_{1m1} a_{2m1} a_{3m1} … a_{km1} a_{*m1}

a_{**1}

a_{112} a_{212} a_{312} … a_{k12}

a_{122} a_{222} a_{322} … a_{k22}

…………………………

a_{1m2} a_{2m2} a_{3m2} … a_{km2}

a_{**2}

a_{11n} a_{21n} a_{31n} … a_{k1n}

a_{12n} a_{22n} a_{32n} … a_{k2n}

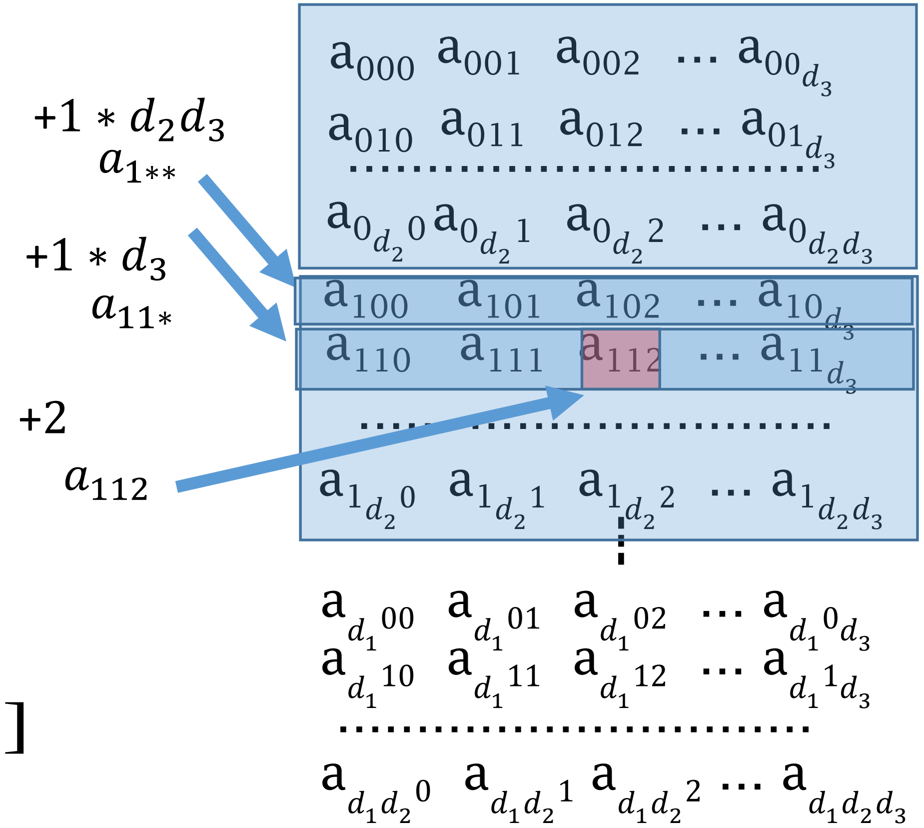
…………………………

a_{1mn} a_{2mn} a_{3mn} … a_{kmn}

12.1 多维数组

- C++ 多维数组ELEM $A[d_1][\ d_2]\ldots[d_n]$;

$$
\begin{aligned}
loc(A[j_1, j_2, \ldots, j_n]) &= loc(A[0,0,\ldots,0]) \\
&\quad + d\cdot[\,j_1\cdot d_2\cdot\ldots\cdot d_n \\
&\quad + j_2\cdot d_3\cdot\ldots\cdot d_n \\
&\quad + \ldots \\
&\quad + j_{n-1}\cdot d_n + j_n\,] \\
&= loc(A[0,0,\ldots,0]) \\
&\quad + d\cdot[\sum_{i=1}^{n-1} j_i \prod_{k=i+1}^{n} d_k + j_n\,]
\end{aligned}
$$



用数组表示特殊矩阵

- 三角矩阵：上三角、下三角
- 对称矩阵
- 对角矩阵
- 稀疏矩阵

下三角矩阵图例

- 一维数组 list[0.. $(n^2+n)/2-1$]
  - 矩阵元素 $a_{i,j}$ 与线性表相应元素的对应位置为 list[$(i^2+i)/2 + j$] ($i>=j$)

$$
\begin{pmatrix}
0 & & & & & \\
0 & 0 & & & & \\
7 & 5 & 0 & & & \\
0 & 0 & 1 & 0 & & \\
9 & 0 & 0 & 1 & 8 & \\
0 & 6 & 2 & 2 & 0 & 7
\end{pmatrix}
$$

对称矩阵

$$\begin{bmatrix} 0 & 3 & 0 & 15 \\ 3 & 0 & 4 & 0 \\ 0 & 4 & 0 & 6 \\ 15 & 0 & 6 & 0 \end{bmatrix}$$

- 元素满足性质 $a_{i,j} = a_{j,i}$，$0 <= (i, j) < n$

  例如，右图的无向图相邻矩阵

- 存储其下三角的值，对称关系映射
- 存储于一维数组 sa[0..n (n+1) /2-1]
  - sa[k] 和矩阵元 $a_{i,j}$ 之间存在着一一对应的关系 ：

$$k = \begin{cases} j(j+1)/2 + i, \text{当} i < j \\ i(i+1)/2 + j, \text{当} i \geq j \end{cases}$$

对角矩阵

- 对角矩阵是指：所有的非零元素都集中在主对角线及以它为中心的其他对角线上。
- 如果 $|i-j| > 1$，那么数组元素 $a[i][j] = 0$。
  - 下面是一个三对角矩阵：

$$
\begin{pmatrix}
a_{0,0} & a_{0,1} & & \cdots\cdots & & 0 \\
a_{1,0} & a_{1,1} & a_{1,2} & & \cdots\cdots & \\
 & & & & & \\
\cdots\cdots & & & & a_{n-2,n-1} & \\
0 & & & a_{n-1,n-2} & a_{n-1,n-1} &
\end{pmatrix}
$$

稀疏矩阵

- 稀疏矩阵中的非零元素非常少，而且分布也不规律

$$\mathbf{A}_{6\times 7} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 11 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

12.1 多维数组

- 稀疏因子
  - 在 m×n 的矩阵中，有 t 个非零元素，则稀疏因子为:

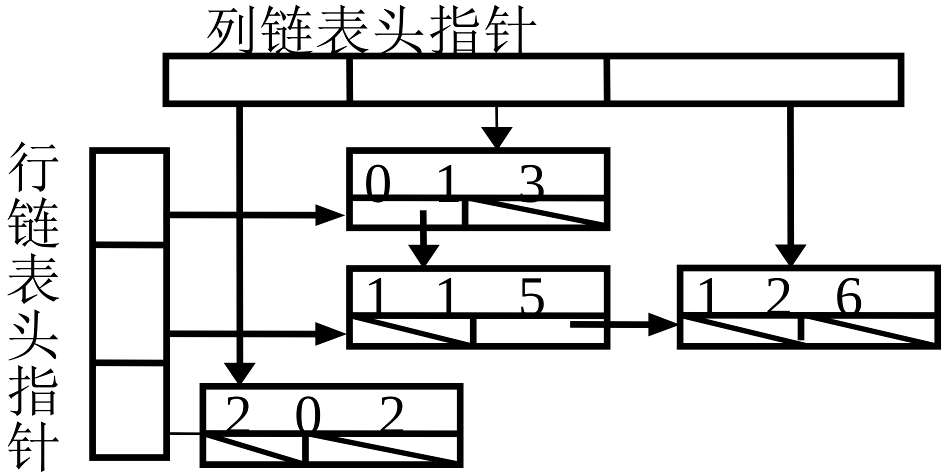
$$\delta = \frac{t}{m \times n}$$

  - 当这个值小于0.05时，可以认为是稀疏矩阵
- 三元组（i, j, a_{ij}）：输入/输出常用
  - i 是该元素的行号
  - j 是该元素的列号
  - a_{ij} 是该元素的值

稀疏矩阵的十字链表

- 十字链表有两组链表组成
  - 行和列的指针序列
  - 每个结点都包含两个指针：同一行的后继，同一列的后继

$$\begin{bmatrix} 0 & 3 & 0 \\ 0 & 5 & 6 \\ 2 & 0 & 0 \end{bmatrix}$$



经典矩阵乘法

- A[c1..d1][c3..d3] , B[c3..d3][c2..d2], C[c1..d1][c2..d2]。

$$C = A \times B \quad (C_{ij} = \sum_{k=c3}^{d3} A_{ik} \cdot B_{kj})$$

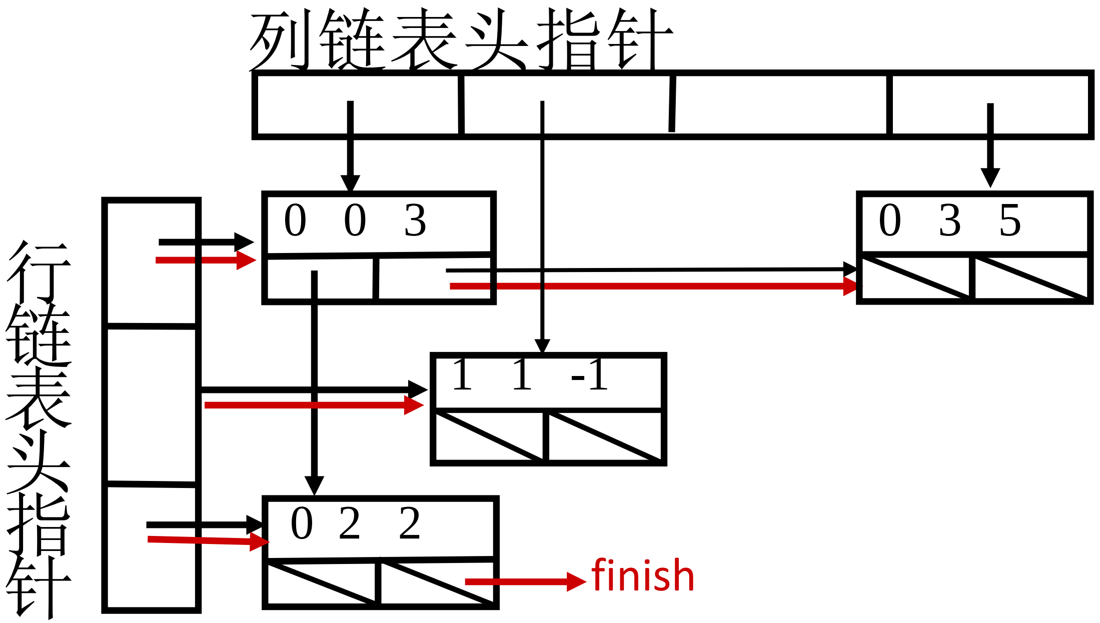
经典矩阵乘法时间代价

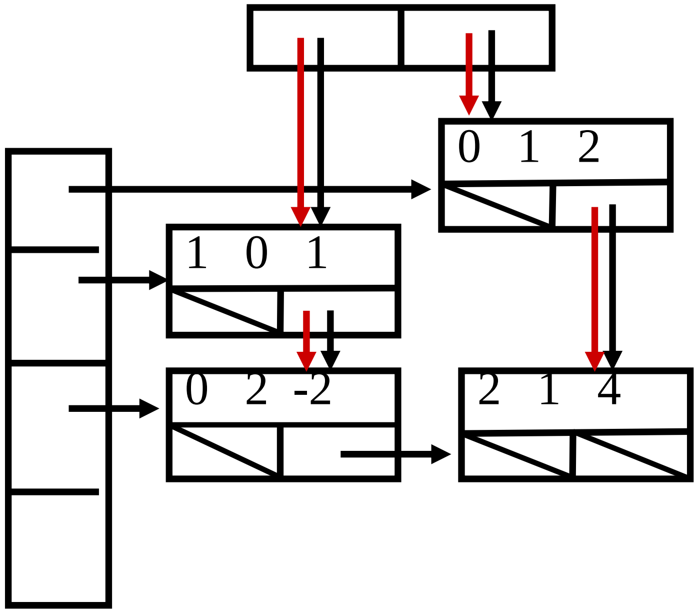
- p=d1-c1+1，m=d3-c3+1，n=d2-c2+1；
- A 为 p×m 的矩阵，B 为 m×n 的矩阵，乘得的结果 C 为 p×n 的矩阵
- 经典矩阵乘法所需要的时间代价为 O (p×m×n)

```
for  (i=c1; i<=d1; i++)
    for  (j=c2; j<=d2; j++){
        sum = 0;
        for  (k=c3; k<=d3; k++)
            sum = sum + A[i,k]*B[k,j];
        C[i , j] = sum;
    }
```

稀疏矩阵乘法

$$\begin{bmatrix}3&0&0&5\\0&-1&0&0\\2&0&0&0\end{bmatrix}\times\begin{bmatrix}0&2\\1&0\\-2&4\\0&0\end{bmatrix}=\begin{bmatrix}0&6\\-1&0\\0&4\end{bmatrix}$$





稀疏矩阵乘法时间代价

- A为 $p \times m$ 的矩阵，B 为 $m \times n$ 的矩阵，乘得的结果 C 为 $p \times n$ 的矩阵
  - 若矩阵 A 中行向量的非零元素个数最多为 t_a
  - 矩阵 B 中列向量的非零元素个数最多为 t_b
- 总执行时间降低为 $O((t_a+t_b) \times p \times n)$
- 经典矩阵乘法所需要的时间代价为 $O(p \times m \times n)$

稀疏矩阵的应用

一元多项式

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$
$$= \sum_{i=0}^{n} a_i x^i$$

	0	1	2		*degree*	*maxDegree*
coef	a_0	a_1	a_2	······	a_n	··········
					↑ n	

思考

- 多元多项式如何使用矩阵表示？如何使用稀疏矩阵进行存储？

一元多项式 $$P_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

- 如何用十字链表结构求解数独(sudoku)问题？

数据结构与算法

谢谢聆听

国家精品课“数据结构与算法”

http://www.jpk.pku.edu.cn/pkujpk/course/sjjg/

张铭，王腾蛟，赵海燕

高等教育出版社，2008. 6。“十一五”国家级规划教材