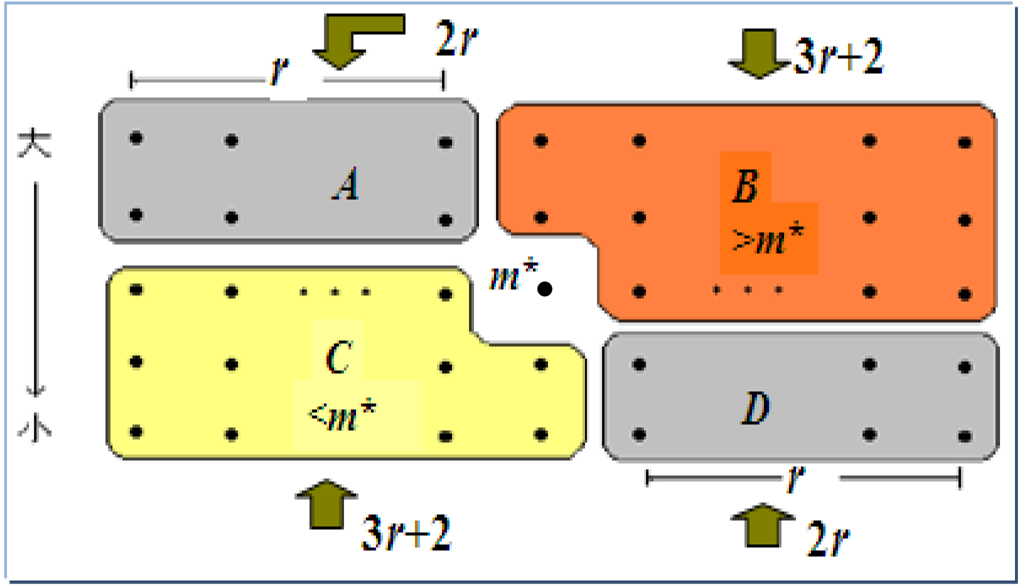
选择问题的
算法分析

伪码

算法 Select (S, k)

1. 将S分5个一组, 共 $n_M = \lceil n/5 \rceil$组
2. 每组排序,中位数放到集合 M
3. $m^* \leftarrow$Select(M,$\lceil |M|/2 \rceil$) //S分A,B,C,D
4. A,D元素小于m^*放 S_1,大于m^*放S_2
5. $S_1 \leftarrow S_1 \cup C$; $S_2 \leftarrow S_2 \cup B$
6. if $k = |S_1| + 1$ then 输出 m^*
7. else if $k \le |S_1|$
8. then Select (S_1, k)
9. else Select (S_2, $k - |S_1| - 1$)

用m*划分



$n = 5\ (2r + 1)$, $|A| = |D| = 2r$

子问题规模至多： $2r+2r+3r+2 = 7r+2$

子问题规模估计

不妨设 $n = 5(2r+1)$，$|A|=|D|=2r$,

$$r = \frac{n/5-1}{2} = \frac{n}{10} - \frac{1}{2}$$

划分后子问题规模至多为

$$7r+2 = 7\left(\frac{n}{10} - \frac{1}{2}\right) + 2$$

$$= \frac{7n}{10} - \frac{3}{2} < \frac{7n}{10}$$

时间复杂度递推方程

算法工作量 $W(n)$

行2: $O(n)$ //每5个数找中位数,构成M

行3: $\underline{W(n/5)}$ // M 中找中位数 m^*

行4: $O(n)$ // 用m^*划分集合 S

行8-9：$\underline{W(7n/10)}$ //递归

$$W(n) \le W(n/5) + W(7n/10) + O(n)$$

递归树

$W(n)=W(n/5)+W(7n/10)+cn$

$$cn \qquad\qquad cn$$

$$\frac{cn}{5} \quad \frac{7cn}{10} \qquad\qquad 0.9cn$$

$$\frac{cn}{25} \quad \frac{7cn}{50} \quad \frac{7cn}{50} \quad \frac{49cn}{100} \qquad\qquad 0.81cn$$

……

$$W(n) \le cn\,(1+0.9+0.9^2+\ldots)=O(n)$$

讨论

分组时为什么5个元素一组？
3个一组或 7个一组行不行？

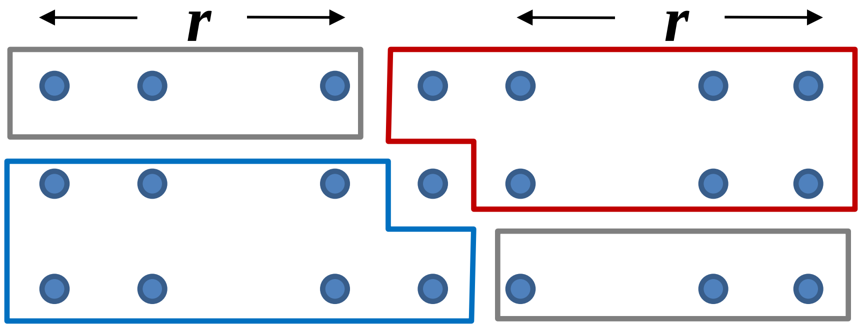
分析：递归调用

1. 求 m^*的工作量与 $|M| = n/t$ 相关，t 为每组元素数. t大, $|M|$小

2. 归约后子问题大小与分组元素数t 有关 . t 大，子问题规模大

3分组时的子问题规模

假设 $t=3$，3个一组：



$$n = 3(2r+1)$$

$$r = (n/3-1)/2 = n/6-1/2$$

子问题规模最多为 $4r+1 = 4n/6-1$

算法的时间复杂度

算法的时间复杂度满足方程

$$W(n) = W(n/3) + W(4n/6) + cn$$

由递归树得 $W(n)=\Theta(n\log n)$

关键：

$|M|$与归约后子问题规模之和小于 n，

递归树每行的工作量构成公比小于 1 的等比级数，算法复杂度才是$O(n)$.

小结

选第 k 小算法的时间分析

- 递推方程
- 分组时每组元素数的多少对时间复杂度的影响