组合算法(选讲)

一.分治策略 (Divide and Conquer)

关键字 key

基本思想：将问题分解成若干子问题，然后求解子问题。 (循逻辑)

"分治策略"可以递归进行。 (recursive) 递推关系

二分查找是分治策略的典型例子(熟悉)。但今天我们介绍两个排序算法。

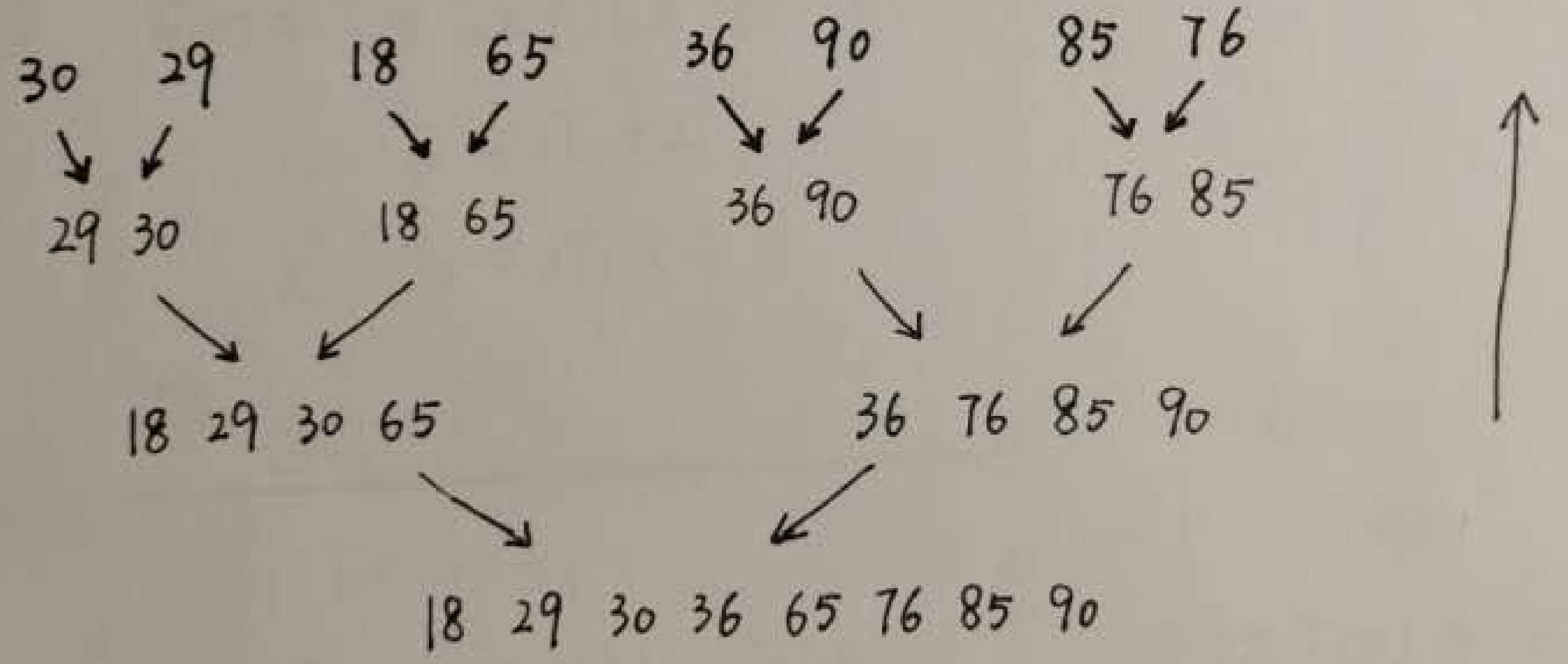
1. 归并排序法 (Merge)

所谓排序(sort)是指已知一序列 $x_1, x_2 \cdots x_n$，将其按从小到大的顺序排列。

排序问题是计算机算法中的重要内容。

分析：将 $x_1, x_2, \cdots x_n$ 一分为二，分成 $x_1, x_2 \cdots x_{\lfloor n/2 \rfloor}$ 和 $x_{\lfloor n/2 \rfloor+1}, x_{\lfloor n/2 \rfloor+2}, \cdots x_n$，对两段分别排序，然后归并为统一的有序列。

算法可递归地进行，例如：8个数 30, 29, 18, 65, 36, 90, 85, 76



关键在于：总是两个有序列合并成一个有序列。

即由 $a_1 < a_2 < a_3 < \cdots < a_m$ 和 $b_1 < b_2 < b_3 < \cdots < b_n$.

$\uparrow i$ $\uparrow j$

$\downarrow k$

$c_1 < c_2 < c_3 < \cdots < c_{m+n}$

简单例示： a={18, 29, 30, 65} b={36, 76, 85, 90} C: 18

$\uparrow i$ $\uparrow j$ $\uparrow k$

18 29 30 65 36 76 85 90 18 29

$\uparrow i$ $\uparrow j$ $\uparrow k$

18 29 30 65 36 76 85 90 18 29 30

$\uparrow i$ $\uparrow j$ $\uparrow k$

a: b: c:

$$\begin{array}{cccc} 18 & 29 & 30 & 65 \\ & & & \uparrow i \end{array} \qquad \begin{array}{cccc} 36 & 76 & 85 & 90 \\ \uparrow j & & & \end{array} \qquad \begin{array}{cccc} 18 & 29 & 30 & 36 \\ & & & \uparrow k \end{array}$$

$$\begin{array}{cccc} 18 & 29 & 30 & 65 \\ & & & \uparrow i \end{array} \qquad \begin{array}{cccc} 36 & 76 & 85 & 90 \\ & \uparrow j & & \end{array} \qquad \begin{array}{ccccc} 18 & 29 & 30 & 36 & 65 \\ & & & & \uparrow k \end{array}$$

$$\begin{array}{cccc} 18 & 29 & 30 & 65 \\ & & & \uparrow i \end{array} \qquad \begin{array}{cccc} 36 & 76 & 85 & 90 \\ & \uparrow j & & \end{array} \qquad \begin{array}{cccccccc} 18 & 29 & 30 & 36 & 65 & 76 & 85 & 90 \end{array}$$

[复杂性分析] 取数据元素个数 $N=2^n$ 情形(有利于分治).

T_n 讨论 2^n个对象的排序在[最好情况下]的比较次数, 则有递归关系

$$T_n = 2T_{n-1} + \underbrace{2^{n-1}}_{\downarrow}, \quad T_1 = 1$$

最后归并所作的比较次数

母函数法

设 $G(x) = T_1 + T_2x + T_3x^2 + \cdots$

x: $T_2 = 2T_1 + 2$

x^2: $T_3 = 2T_2 + 2^2$

$\vdots$

+) $\vdots$

$T_2x + T_3x^2 + \cdots$

$= 2xT_1 + 2x + 2x^2T_2 + 2^2x^2 + \cdots$

$G(x) - T_1$

$= 2xG(x) + 2x(1 + 2x + 2^2x^2 + \cdots)$

则有:

$$G(x) - 1 = 2xG(x) + 2x\frac{1}{1-2x}$$

$$(1-2x)G(x) = 1 + \frac{2x}{1-2x} = \frac{1}{1-2x}$$

$$G(x) = \frac{1}{(1-2x)^2} = (1+2x+2^2x^2+\cdots)(1+2x+2^2x^2+\cdots)$$

$$= 1 + \underbrace{2\cdot2}_{T_2}x + \underbrace{3\cdot2^2}_{T_3}x^2 + \underbrace{4\cdot2^3}_{T_4}x^3 + \cdots$$

故:

$$T_n = n\cdot2^{n-1}$$

$$= \frac{1}{2}n\cdot2^n = \frac{1}{2}n\cdot N = \frac{1}{2}N\log_2 N$$

[最坏情况下的复杂性分析] ✓

估计 $\underbrace{(2^n-1)}$

$$C_n = 2C_{n-1} + \underbrace{2^n-1}, \quad C_1 = 1$$

同样

$$G(x) = C_1 + C_2x + C_3x^2 + \cdots$$

x: $C_2 = 2C_1 + 2^2 - 1$

x^2: $C_3 = 2C_2 + 2^3 - 1$

+) $\vdots$

$$C_2x + C_3x^2 + \cdots$$

$$= 2x(C_1 + C_2x + C_3x^2 + \cdots) +$$

$$2^2x(1 + 2x + 2^2x^2 + \cdots) -$$

$$x(1 + x + x^2 + \cdots)$$

$$G(x) - 1$$

$$= 2xG(x) + 2^2x\frac{1}{1-2x} - \frac{x}{1-x}$$

整理可得:

$$(1-2x)G(x) = 2^2x\frac{1}{1-2x} + 1 - \frac{x}{1-x} = \frac{1+2x}{1-2x} - \frac{x}{1-x}$$

即有, $G(x) = \frac{1+2x}{(1-2x)^2} + \frac{-x}{(1-x)(1-2x)}$

$$= \frac{1}{(1-2x)^2} + \frac{2x}{(1-2x)^2} + \frac{1}{1-x} - \frac{1}{1-2x}$$

展开就有, T4

$$[1 + 2\cdot 2x + 3\cdot 2^2x^2 + 4\cdot 2^3x^3 + \cdots] +$$

$$2x[1 + 2\cdot 2x + 3\cdot 2^2x^2 + 4\cdot 2^3x^3 + \cdots] +$$

$$[1 + x + x^2 + x^3 + \cdots] -$$

$$[1 + 2x + 2^2x^2 + 2^3x^3 + \cdots]$$

$$\frac{-x}{(1-x)(1-2x)} = \frac{A}{(1-x)} + \frac{B}{(1-2x)}$$

$$= \frac{A - 2Ax + B - Bx}{(1-x)(1-2x)}$$

$$= \frac{(A+B) - (2A+B)x}{(1-x)(1-2x)}$$

$$\begin{cases} A + B = 0 \\ 2A + B = 1 \end{cases}$$

解得. $\begin{cases} A = 1 \\ B = -1 \end{cases}$

即有, $C_n = n\cdot 2^{n-1} + (n-1)\cdot 2^{n-1} + 1 - 2^{n-1}$

$$= (n+n-1-1)\cdot 2^{n-1} + 1 = \underbrace{(n-1)\cdot 2^n + 1}$$

$C_n = n \cdot 2^n - 2^n + 1 \qquad (N = 2^n)$

$= N\log_2 N - N + 1$

综合两种情形，时间复杂度 $O(N\log_2 N)$ ✓

与数据元素长度N关系.

2. 快速排序 (Quick Sort)

与归并排序共同之处都是将序列 $a_1, a_2 \cdots a_n$ 分成两个子序列，再分别对子序列进行排序，然后将两个有序子列连接起来.

基本思想：不失一般性，将n个数据对象看作是n个整数(key)的排序。

即取一合适的关键字(key) k，以k为标准把需要排序的n个对象分成两部分，即 (小于k的部分) k (大于k的部分)，

再分别对两部分进行快速排序。算法可以递归进行。

pivot (枢轴)

还是简单例示.

3 9 1 6 5 4 8 2 10 7 取3为K
i↑ ↑j

3 9 1 6 5 4 8 2 10 7 移动
i↑ ↑j

2 9 1 6 5 4 8 3 10 7 交换 (排序)
i↑ ↑j

2 3 1 6 5 4 8 9 10 7 交换
↑i ↑j

2 3 1 6 5 4 8 9 10 7 移动若干次
↑i ↑j

(2 1) 3 (6 5 4 8 9 10 7) 交换、移动

i↑↑j

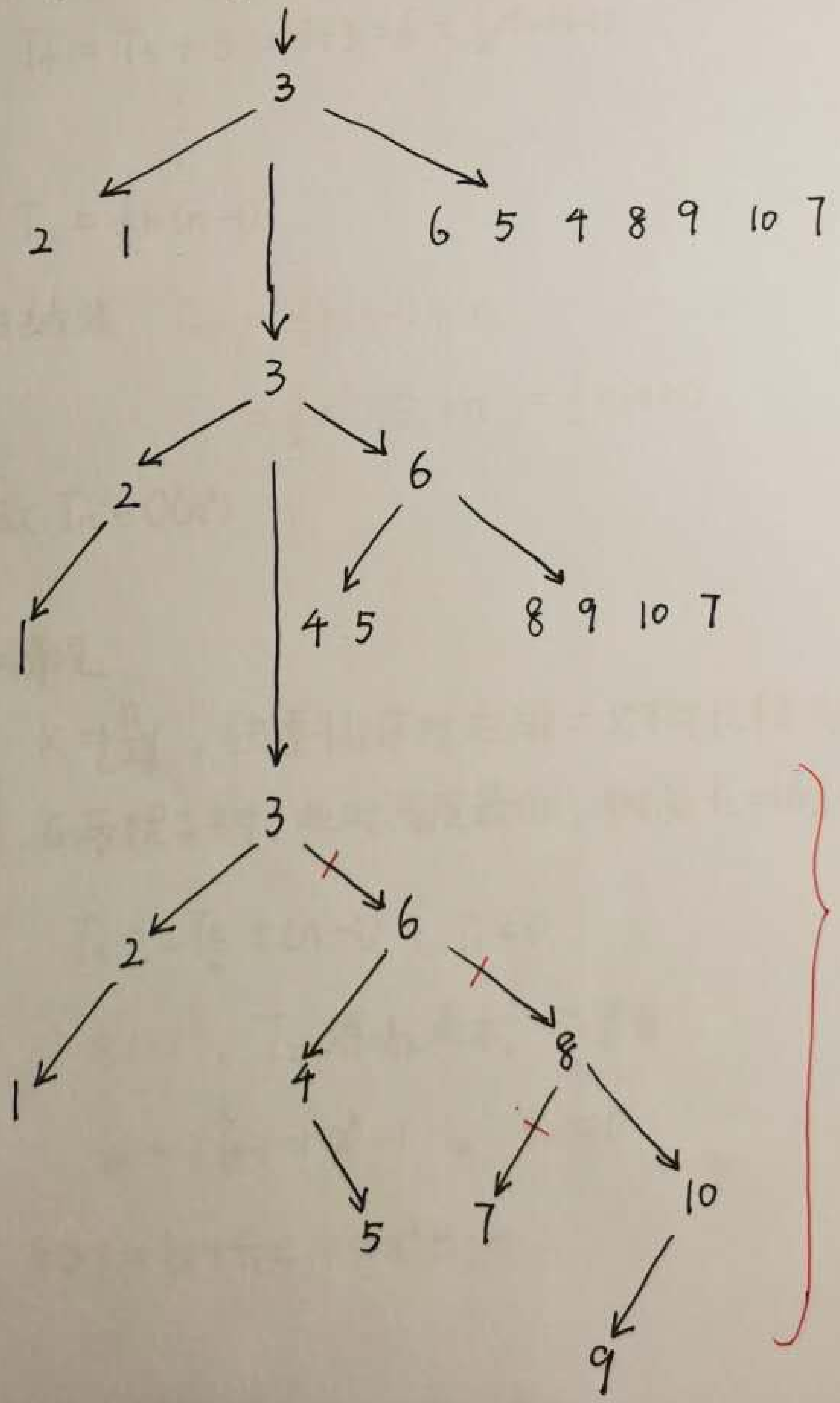
分别对(2, 1) 和(6 5 4 8 9 10 7)进行快速排序.

i_1↑ ↑j_1 i_2↑ ↑j_2

((1) 2) 3 ((4 (5)) 6 (8 9 10 7))

[复杂性分析]

3* 9 ~~4~~1 6 5 4 8 2 10 7



树的深度是比较次数.

例如: 7 三次比较

(1) 最坏情况

假如已是正序，$x_1 < x_2 < \cdots < x_n$，使用快速排序法，有

$$T_n = T_{n-1} + (n-1), \quad T_1 = 0$$

（T_n 是 n 个元素进行排序所需的比较次数）

$$T_{n-1} = T_{n-2} + (n-2)$$

$$\vdots$$

$$T_2 = T_1 + 1 = 1$$

结果有

$$T_3 = T_2 + 2 = 1 + 2 = 3 = \frac{1}{2} \times 3 (3-1)$$

$$T_4 = T_3 + 3 = 3 + 3 = 6 = \frac{1}{2} 4 \times (4-1)$$

$$\vdots$$

$$T_n = \frac{1}{2} n (n-1)$$

数学归纳法. $T_{n+1} = \frac{1}{2} n(n-1) + n$

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

故 $T_n \in O(n^2)$

(2) 最好情况

此时 $k = \lfloor \frac{n}{2} \rfloor$，快速排序对应的二叉树比较均衡，每个非终端结点，都有左、右两棵子树，此时高度最低，树高 $h = \lceil \log_2 n \rceil$.

$$T_n = 2 T_{\frac{n}{2}} + (n-1), \quad T_1 = 0$$

此时 令 $n = 2^k$，T_{2^k} 用 t_k 表示，结果有

$$t_k = 2 t_{k-1} + 2^k - 1, \quad t_0 = 0$$

设 $G(x) = t_0 + t_1 x + t_2 x^2 + \cdots$

$$x: \quad t_1 = 2t_0 + 2^1 - 1$$

$$x^2: \quad t_2 = 2t_1 + 2^2 - 1$$

$$+) \qquad \vdots$$

$$t_1x + t_2x^2 + \cdots$$

$$= 2x(t_0 + t_1x + t_2x^2 + \cdots) + 2x(1 + 2^1x^1 + 2^2x^2 + \cdots) - x(1 + x + x^2 + \cdots)$$

$$G(x) - t_0 = 2xG(x) + 2x\frac{1}{1-2x} - x\cdot\frac{1}{1-x}$$

$$(1-2x)G(x) = \frac{2x}{1-2x} - \frac{x}{1-x}$$

即有： $$G(x) = \frac{2x}{(1-2x)^2} - \frac{x}{(1-x)(1-2x)} = \frac{A}{(1-2x)^2} + \frac{B}{1-2x} + \frac{C}{1-x}$$

$$= A[1 + 2\cdot 2x + 3\cdot 2^2\cdot x^2 + 4\cdot 2^3x^3 + \cdots] +$$

$$B[1 + 2x + 2^2x^2 + 2^3x^3 + \cdots] +$$

$$C[1 + x + x^2 + x^3 + \cdots]$$

$$t_k = A\cdot(k+1)2^k + B\cdot 2^k + C$$

$$= [A(k+1) + B]2^k + C$$

由 $n = 2^k$，则 $k = \log_2 n$

$$t_k = [A(\log_2 n + 1) + B]n + C$$

$$= T_{2^k} = T_n$$

$$T_n \in O(n\log_2 n).$$

(3) 平均复杂度分析.

此时 T_n 取利用快速排序法对 n 个对象进行分类所作的比较的平均数. 由于枢轴取第一个数，n 个数的机率均等，若取 k，则一个序列有 k-1 个数，另一个序列有 n-k 个数，此时有递推关系：

$$T_n = \frac{1}{n}\sum_{k=1}^{n}(n-1+T_{k-1}+T_{n-k})$$

$$= n-1+\frac{2}{n}\sum_{k=0}^{n-1}T_k \text{ ，} \quad T_0 = 0$$

这里，n−1表示第一轮把n个数的序列一分为二时所作的比较次数.

继而可得：

$$nT_n = n(n-1) + 2\sum_{k=0}^{n-1}T_k \text{ ，} \quad T_0=0$$

$$(n+1)T_{n+1} = n(n+1) + 2\sum_{k=0}^{n}T_k$$

故有

$$(n+1)T_{n+1} - nT_n = 2n + 2T_n$$

$$(n+1)T_{n+1} = (n+2)T_n + 2n$$

令 $S_n = T_n/n+1$ ，上式可以化为

$$T_{n+1} = (n+2)\frac{T_n}{n+1} + \frac{2n}{n+1}$$

$$\frac{T_{n+1}}{n+2} = \frac{T_n}{n+1} + \frac{2n}{(n+1)(n+2)}$$

$$S_{n+1} - S_n = \frac{2n}{(n+1)(n+2)} \text{ ，} \qquad S_0 = 0$$

关于S_n的线性常系数非齐次递推关系，

$$S_1 - S_0 = \frac{2\times 0}{1\times 2} = 0$$

$$S_2 - S_1 = \frac{2\times 1}{2\times 3}$$

$$S_3 - S_2 = \frac{2\times 2}{3\times 4}$$

$$\vdots$$

$$+) \qquad S_n - S_{n-1} = \frac{2(n-1)}{n(n+1)}$$

$$S_n = \sum_{k=1}^{n-1}\frac{2k}{(k+1)(k+2)}$$

$$\text{又} \because \frac{2k}{(k+1)(k+2)} = \frac{A}{k+1} + \frac{B}{k+2}$$

$$\frac{A}{k+1}+\frac{B}{k+2}=\frac{Ak+2A+Bk+B}{(k+1)(k+2)}=\frac{(A+B)k+(2A+B)}{(k+1)(k+2)}$$

即有 $\begin{cases} A+B=2 \\ 2A+B=0 \end{cases}$ 解得 $\begin{cases} A=-2 \\ B=4 \end{cases}$

所以，$S_n = 4\sum_{k=1}^{n-1}\frac{1}{k+2} - 2\sum_{k=1}^{n-1}\frac{1}{k+1}$

$$= 4\sum_{k=2}^{n}\frac{1}{k+1} - 2\sum_{k=1}^{n-1}\frac{1}{k+1}$$

$$= \frac{4}{n+1} - \frac{2}{1+1} + 2\sum_{k=2}^{n-1}\frac{1}{k+1}$$

$\because T_n = (n+1)S_n = (n+1)\left[2\sum_{k=2}^{n-1}\frac{1}{k+1} + \frac{4}{n+1} - 2\right]$

但 $\sum_{k=2}^{n-1}\frac{1}{k+1} = \sum_{k=3}^{n}\frac{1}{k} < \int_2^n \frac{1}{x}dx = \ln n - \ln 2$

$\therefore$ ~~S~~ $T_n < \underbrace{(n+1)\left[2\ln n - 2\ln 2 + \frac{4}{n+1} - 2\right]}_{2(n+1)\ln n - 2(n+1)\ln 2 + 4 - 2(n+1)}$

$T_n \in O(n\ln n)$

枢轴应随机选取.

二. 线性常系数递推关系举例.

1. $a_n - 2a_{n-1} = 3^n$， $a_0 = 1$

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

$$x:\quad a_1 - 2a_0 = 3$$
$$x^2:\quad a_2 - 2a_1 = 3^2$$
$$\vdots$$

$+)$

$$G(x) - 1 - 2xG(x) = 3x(1 + 3x + 3^2x^2 + \cdots)$$

$$(1-2x)G(x)=\frac{3x}{1-3x}+1$$

$$=\frac{1}{1-3x}$$

故 $G(x)=\frac{1}{(1-2x)(1-3x)}=\frac{A}{(1-2x)}+\frac{B}{(1-3x)}$

$$=\frac{A-3Ax+B-2Bx}{(1-2x)(1-3x)}=\frac{(A+B)-(3A+2B)x}{(1-2x)(1-3x)}$$

即有，$\begin{cases}A+B=1\\3A+2B=0\end{cases}$ 解得 $\begin{cases}A=-2\\B=3\end{cases}$

$\therefore G(x)=\frac{3}{1-3x}-\frac{2}{1-2x}$

$$=3\left[1+3x+3^2x^2+\cdots\right]-2\left[1+2x+2^2x^2+\cdots\right]$$

$\therefore a_n=3\cdot3^n-2\cdot2^n=3^{n+1}-2^{n+1}$

2. $a_n-2a_{n-1}=(n+5)\cdot3^n$, $a_0=0$, $a_1=18$, $a_2=99$ （非齐次）

解: $\because a_{n-1}-2a_{n-2}=(n+4)\cdot3^{n-1}$

$\therefore 3a_{n-1}-6a_{n-2}=(n+4)\cdot3^n$

所以，$a_n-5a_{n-1}+6a_{n-2}=3^n$ ✓

即. $a_{n-1}-5a_{n-2}+6a_{n-3}=3^{n-1}$

$3a_{n-1}-15a_{n-2}+18a_{n-3}=3^n$ ✓

$a_n-8a_{n-1}+21a_{n-2}-18a_{n-3}=0$ （齐次）

令 $G(x)=a_0+a_1x+a_2x^2+\cdots$

x^3: $a_3 - 8a_2 + 21a_1 - 18a_0 = 0$

x^4: $a_4 - 8a_3 + 21a_2 - 18a_1 = 0$

$\vdots$

$+)$

$$G(x) - a_2x^2 - a_1x - a_0 - 8x(G(x) - a_1x) + 21x^2G(x) - 18x^3G(x) = 0$$

$$(1-8x+21x^2-18x^3)G(x) = 18x + 99x^2 - 144x^2$$

$$= 18x - 45x^2$$

$$G(x) = \frac{18x-45x^2}{1-8x+21x^2-18x^3} = \frac{18x-45x^2}{(1-3x)^2(1-2x)}$$

$$= \frac{A}{(1-3x)^2} + \frac{B}{(1-3x)} + \frac{C}{(1-2x)}$$

$$\begin{cases} A+B+C=0 \\ 2A+5B+6C=-18 \\ 6B+9C=-45 \end{cases} \quad 解得 \quad \begin{cases} A=3 \\ B=6 \\ C=-9 \end{cases}$$

$$\therefore G(x) = \frac{3}{(1-3x)^2} + \frac{6}{1-3x} - \frac{9}{1-2x}$$

$$= 3[1 + 2\cdot 3x + 3\cdot(3x)^2 + \cdots + (n+1)(3x)^n + \cdots] +$$

$$6[1 + 3x + (3x)^2 + \cdots + (3x)^n + \cdots] -$$

$$9[1 + 2x + (2x)^2 + \cdots + (2x)^n + \cdots]$$

$$a_n = (n+1)\cdot 3^{n+1} + 6\cdot 3^n - 9\cdot 2^n$$

又解1.

$a_n - 2a_{n-1} = 3^n$ ✓ $\qquad a_0 = 1$

$a_{n-1} - 2a_{n-2} = 3^{n-1}$

$3a_{n-1} - 6a_{n-2} = 3^n$ ✓

$a_n - 5a_{n-1} + 6a_{n-2} = 0 \qquad (a_1 - 2a_0 = 3 , \quad a_1 = 3 + 2a_0 = 5)$

令 $G(x) = a_0 + a_1 x + a_2 x^2 + \cdots$

$x^2: \quad a_2 - 5a_1 + 6a_0 = 0$

$x^3: \quad a_3 - 5a_2 + 6a_1 = 0$

$\vdots$

，+)

$G(x) - 1 - 5x - 5x(G(x) - 1) + 6x^2 G(x) = 0$

$(1 - 5x + 6x^2) G(x) = 1$

$$G(x) = \frac{1}{1 - 5x + 6x^2} = \frac{1}{(1-2x)(1-3x)}$$

$$= \frac{A}{1-2x} + \frac{B}{1-3x}$$

请同学们回去解. ✓

$$\sum_{i=0}^{\infty} \frac{a_i x^i}{1-x}$$