第7章 快排序

2020年10月22日 8:45

期望时间为$\theta(n\log n)$

基于比较排序，时间下界：

$\log n! \approx n\log n - 1.44n + O(\log n)$

快排序的平均：

$1.39n\log n + O(n)$

1. 算法描述

（1）方法（分治）

$Divide: A[p \ldots r] \Longrightarrow A[p \ldots q-1] \le A[q] < A[q+1 \ldots r] \ldots (*)$（$A[q]$是划分元）

$Conquer$: 递归排序$A[p \ldots q-1], A[q+1 \ldots r]$

临界问题规模：区间长度为1，空操作

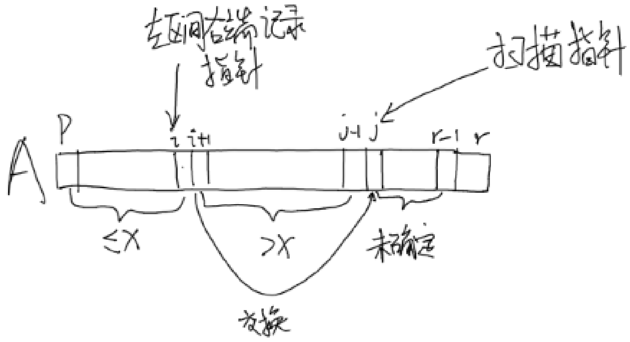
$Combine$: 空操作

（2）算法伪代码

$QuickSort(A, p, r)$
    $if\ p < r$：
        $q = partition(A, p, r)$（返回划分元位置,并完成(∗)式划分）
        $QuickSort(A, p, q-1)$
        $QuickSort(A, q+1, r)$

partition过程：

示意图：



$partition(A, p, r)$
    $x = A[r]$ （得到划分元）
    $i = p - 1$
    $for\ \ j \leftarrow p\ to\ r-1$：
        $if\ A[j] \le x$：
            $i \mathrel{+}= 1$
            $A[i] \leftrightarrow A[j]$
    $A[i+1] \leftrightarrow A[r]$
    $return\ \ i+1$

（3）示例：Fig 7.1

（4）说明：快排不是稳定的，为什么？在代码中什么部分造成的？

代码中的$A[i] \leftrightarrow A[j]$操作导致快排不稳定

1. 性能分析

(1) 最坏划分：$A[p \dots q-1], A[q+1 \dots r]$有一个是空的(或区间长度极不平衡)

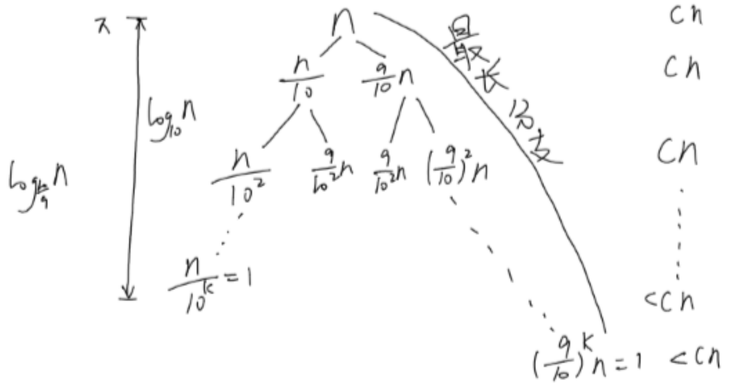
$$T(n) = T(n-1) + T(0) + \theta(n) = T(n-1) + \theta(n) = \sum_{k=1}^{n} \theta(k) = \theta\left(\sum_{k=1}^{n} k\right) = \theta(n^2)$$

(2) 最好划分：一分为二，比较均衡

$$T(n) = 2T\left(\frac{n}{2}\right) + \theta(n) = \theta(n \log n)$$

(3) 固定比例划分：比如区间比为$9:1(1:9)$

如图所示：



$$T(n) \le cn \cdot h = cn \cdot \log_{\frac{10}{9}} n = \theta(n \log n)$$ //但系数要大

(4) 随机划分

1. 随机版本：利用随机数发生器，随机产生划分元

将partition改为：

$RandomPartition(A, p, r)$

$i = Random(p, r)$ //随机产生$[p, \dots, r]$之间的整数i

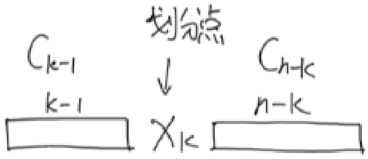
$A[r] \leftrightarrow A[i]$

$return$ Partition(A, p, r)

2. 期望时间分析（书上方法，略）

另解：$x_1, x_2, \dots, x_n$排序

$partition\ node$:



x_k被概率(1/n)选中，c_n表示规模为n的问题的比较次数

$$c_n = (n-1) + \frac{1}{n}\sum_{k=1}^{n}(c_{k-1} + c_{n-k})$$

$$= (n-1) + \frac{1}{n}\sum_{k=1}^{n} c_{k-1} + \frac{1}{n}\sum_{k=1}^{n} c_{n-k} \quad (c_{n-1} + c_{n-2} + \cdots + c_0)$$

$$= (n-1) + \frac{1}{n}\sum_{k=0}^{n-1} c_k + \frac{1}{n}\sum_{k=0}^{n-1} c_k = (n-1) + \frac{2}{n}\sum_{k=0}^{n-1} c_k$$

$$\Longrightarrow nc_n = n(n-1) + 2\sum_{k=0}^{n-1} c_k \quad ①$$

用(n-1)代入n，①式$\Longrightarrow (n-1)c_{n-1} = (n-1)(n-2) + 2\sum_{k=0}^{n-2} c_k$ ②

①-②$\Longrightarrow nc_n - (n-1)c_{n-1} = 2(n-1) + 2c_{n-1}$

$$\Longrightarrow \frac{c_n}{n+1} = \frac{c_{n-1}}{n} + \frac{2(n-1)}{n(n+1)} \quad ③$$

令$D_n = \frac{c_n}{n+1}$, ③式 $\Longrightarrow D_n = D_{n-1} + \frac{2(n-1)}{n(n+1)}$ ④

④$\Longrightarrow D_n = 2\sum_{j=1}^{n}\frac{j-1}{j(j+1)}$ $\quad //\because \frac{2}{j+1} - \frac{1}{j} = \frac{j-1}{j(j+1)}$

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

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

$$= 2\ln n + \theta(1) = \frac{2}{\log e}\log n + \theta(1) \approx 1.44\log n$$

由D_n的定义 $\Longrightarrow \mathrm{C}_n = (n+1)D_n \approx 1.44 n\log n$