快速排序算法的数学期望

李江贝 (alex-li@sjtu.edu.cn)

2018 年 6 月 17 日

目录

1 快速排序算法

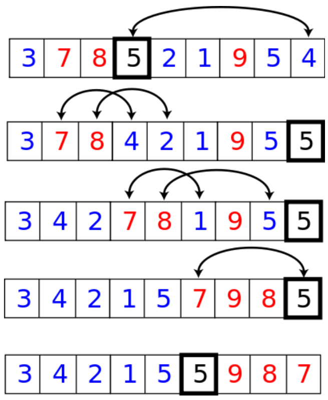

图 1: 快速排序

快速排序由 C. A. R. Hoare 在 1962 年提出。它的基本思想是：通过一趟排序将要排序的数据分割成独立的两部分，其中一部分的所有数据都比另外一部分的所有数据都要小，然后再按此方法对这两部分数据分别进行快速排序，整个排序过程可以递归进行，以此达到整个数据变成有序序列。

快速排序算法的步骤为:

1. 从数列中挑出一个元素，称为" 基准"（pivot）.

2. 重新排序数列，所有比基准值小的元素摆放在基准前面，所有比基准值大的元素摆在基准后面（相同的数可以到任何一边）。在这个分区结束之后，该基准就处于数列的中间位置。这个称为分区（partition）操作.

3. 递归地把小于基准值元素的子数列和大于基准值元素的子数列排序。

快速排序算法在一般情况下的复杂度为 $O(nlogn)$, 我希望通过概率的方法，计算出快速排序 n 个元素所需比较次数的数学期望。

2 快排的数学期望

设 $\bar{x}$ 为第一次比较的基准元素,q_n 表示快速排序 n 个数比较次数的期望, 令 L 表示快速排序 $\bar{x}$ 左边的数所需要的比较次数,R 表示快速排序 $\bar{x}$ 右边的数所需要的比较次数,ξ 表示快速排序所需要的总次数.

$$\xi = L + R + n - 1$$

若 $\bar{x} = x_i$

$$E(L|\bar{x} = x_i) = q_{i-1}, E(R|\bar{x} = x_i) = q_{n-i}$$

$$E(\xi|\bar{x} = x_i) = E(L|\bar{x} = x_i) + E(R|\bar{x} = x_i) + n - 1 = q_{i-1} + q_{n-i} + n - 1$$

又因为 $P(\bar{x} = x_i) = \frac{1}{n}$, 所以由全概率公式

$$\begin{aligned} q_n = E\xi &= \sum_{i=1}^{n} P(\xi|\bar{x} = x_i)P(\bar{x} = x_i) \\ &= \sum_{i=1}^{n}(q_{i-1} + q_{n-i} + n - 1) \cdot \frac{1}{n} \\ &= n - 1 + \sum_{i=1}^{n}(q_{i-1} + q_{n-i}) \\ &= n - 1 + 2\sum_{i=1}^{n} q_{i-1} \end{aligned}$$

$$nq_n - (n-1)q_{n-1} = n(n-1) - (n-1)(n-2) + 2q_{n-1}$$

因此

$$nq_n = 2(n-1) + (n+1)q_{n-1}$$

$$\frac{q_n}{n+1} = \frac{q_{n-1}}{n} + \frac{2(n-1)}{n(n+1)} = \frac{q_{n-1}}{n} + \frac{2}{n} + 4(\frac{1}{n+1} - \frac{1}{n})$$

迭代可得

$$\frac{q_n}{n+1} = 2\sum_{i=1}^{n}\frac{1}{i} + 4(\frac{1}{n+1} - 1)$$

由欧拉公式 $\sum_{i=1}^{n}\frac{1}{n} = ln(n) + \gamma + \frac{1}{2n} + O(n^{-2})$

$$\begin{aligned} q_n &= (n+1)(2(ln(n) + \gamma + \frac{1}{2n} + O(n^{-2})) - \frac{4n}{n+1}) \\ &= 2nln(n) + (2\gamma - 4)n + 2ln(n) + 2\gamma + 1 + O(n^{-1}) \end{aligned}$$

由此也可以推知快速排序的复杂度为 $O(nlogn)$