

数据结构与算法 II 作业 (10.20)

中国人民大学 信息学院 崔冠宇 2018202147

注：由于 10.25 发现之前上传的作业看错了题目，所以重传。

T8.3-2 下面的排序算法中哪些是稳定的：插入排序、归并排序、堆排序和快速排序？给出一个能使任何排序算法都稳定的方法。这个方法带来的额外时间和空间开销是多少？

解：插入排序和归并排序是稳定排序，堆排序和快速排序是不稳定的。

可以将待排序的值换为 (Index, Value) 对，在比较时需要 $\text{ValueA} < \text{ValueB} \ || \ (\text{ValueA} == \text{ValueB} \ \&\& \ \text{IndexA} < \text{IndexB})$ 。

显然带来的额外空间开销是 n 个待排序数组的下标，每个下标需要 $\log n$ 位二进制数存储，故总额外空间开销是 $O(n \log n)$ ；由于仅是改变了比较条件，而没有改变比较次数，故没有额外时间开销。

T8.3-3 利用归纳法来证明基数排序是正确的。在你所给出的证明中，在哪里需要假设所用的底层排序算法是稳定的？

证明：用归纳法。

① $d = 1$ 时，待排序的数仅有一位，基数排序显然是正确的。
② 假设基数排序对 $d - 1$ 位的数成立，考虑 d 位的数。根据归纳假设，运行第 d 轮排序时，各数已经按低 $d - 1$ 位排好序。在对第 d 位排序时，有下列三种情况：

1. a 的最高位 $< b$ 的最高位，无论底层排序为何， a 排在 b 前面。根据正整数大小比较方法（从左到右一位位比较）， $a < b$ ，所以是正确的。
2. a 的最高位 $> b$ 的最高位，与上面类似，也是正确的。
3. a 的最高位 $= b$ 的最高位，**此处要假设底层排序是稳定的**。因为底层排序是稳定的，这将会保持它们的相对顺序。而根据归纳假设，它们已经按照低 $d - 1$ 位排好序，根据正整数大小比较方法， $a < b$ ，所以是正确的。

综上，基数排序是正确的。□

(如果基数排序的底层排序是不稳定的，那么比如对于 22、23，程序可能会交换它们的位置，这是不正确的。)

P125, T9.3-9 Olay 教授是一家石油公司的顾问。这家公司正在计划建造一条从东向西的大型输油管道，这一管道将穿越一个有 n 口油井的油田。公司希望有一条管道支线沿着最短路径从每口油井连接到主管道（方向或南或北），如图 9-2 所示（图略）。给定每口油井的 x 和 y 坐标，教授应该如何选择主管道的最优位置，使得各直线的总长度最小？
证明：该最优位置可以在线性时间内确定。

证明：

设主管道的位置是 $y = y^*$ 。本题目标是找寻 y_0 ，使得最小化 $\sum_{i=1}^n |y_i - y^*|$ ，即

$$y_0 = \arg \min_{y^*} \sum_{i=1}^n |y_i - y^*|$$

可以看出, 当 y^* 为诸 y_i 的中位数时, $\sum |y_i - y^*|$ 最小。

1. 若 n 为奇数, 根据课堂内容, **Select** 函数可以在线性时间内找到中位数 ($i = (n + 1)/2$)。
2. 若 n 为偶数, 两次调用 **Select** 函数, 分别找到第 $n/2$ 和第 $n/2 + 1$ 小数, 求平均即可。由于 **Select** 是 $O(n)$ 的, 所以两次调用再求平均也是线性时间的。

所以, 该最优位置可以在线性时间内确定。□