

本小节内容

2022 年 42 题题目解析

42. (10 分)

现有 n ($n > 100000$) 个数保存在一维数组 M 中, 需要查找 M 中最小的 10 个数, 请回答下列问题。

(1) 设计一个完成上述查找任务的算法, 要求平均情况下的比较次数尽可能少, 简单描述其算法思想, 不需要程序实现。

(2) 说明你所设计的算法平均情况下的时间复杂度和空间复杂度。

答案解析:

这道题非常有代表性, 我们提供 3 种解题思路, 从而对我们前面学习的排序算法能有更深的理解。

方法一: 最小值 (选择排序思想)

(1) 算法思想(高教社官方答案)

定义含 10 个元素的数组 A , 初始时元素值均为该数组类型能表示的最大数 MAX 。

for M 中的每个元素 s

if ($s < A[9]$) 丢弃 $A[9]$ 并将 s 按升序插入到 A 中;

当数据全部扫描完毕, 数组 $A[0] \sim A[9]$ 保存的即是最小的 10 个数。

(2) 时间复杂度: $O(n)$, 需要遍历 10 次数组。

空间复杂度: $O(1)$, 中间过程额外需要常数个变量。

如果这里将 10 修改为 k , 则:

时间复杂度: $O(nk)$, 需要遍历 k 次数组。

空间复杂度: $O(1)$, 中间过程额外需要常数个变量。

方法二: 堆 (堆排序思想)

定义含 10 个元素的大根堆 H , 元素值均为该堆元素类型能表示的最大数 MAX 。

for M 中的每个元素 s

if ($s < H$ 的堆顶元素) 删除堆顶元素并将 s 插入到 H 中;

当数据全部扫描完毕, 堆 H 中保存的即是最小的 10 个数。

2) 算法平均情况下的时间复杂度是 $O(n)$, 空间复杂度是 $O(1)$ 。

进一步解析:

先用 $A[0:9]$ 原地建立大顶堆 (注意: 这里不能用小顶堆), 遍历 $A[10:n]$, 每个元素 $A[i]$ 逐一和堆顶元素 $A[0]$ 进行比较, 其中 $11 \leq i \leq n$, 如果 $A[i]$ 大于等于堆顶元素 $A[0]$, 不进行任何操作, 如果该元素小于堆顶元素 $A[0]$, 那么就删除堆顶元素, 将该元素放入堆顶, 即令 $A[0] = A[i]$, 然后将 $A[0:9]$ 重新调整为大顶堆。

最后堆 $A[0:9]$ 中留存的元素即为最小的 10 个数。

方法三: (这个高教社没给)

通过快速排序，分割思想，第一次知道 $n/2$ 位置，再次 partition 得到 $n/4$ ，最终缩小为 10 个，就拿到了最小的 10 个元素，遍历的平均次数是 $n+n/2+n/4+\dots+1$ ，次数为 $2n$ ，因此时间复杂度为 $O(n)$ 。由于我们不进行快排，只是记录剩余部分的起始和结束，因此空间复杂度是 $O(1)$ 。

现在高教社官方给的答案没有了评分标准，个人认为以上三种方式都可以得满分。