

第 11 章 外排序

教材中练习题及参考答案

1. 外排序中两个相对独立的阶段是什么？

答：外排序中两个相对独立的阶段是产生初始归并段和多路归并排序。

2. 给出一组关键字 $T = (12, 2, 16, 30, 8, 28, 4, 10, 20, 6, 18)$ ，设内存工作区可容纳 4 个记录，给出用置换-选择排序算法得到的全部初始归并段。

答：置换-选择排序算法的执行过程如表 11.1 所示。共产生两个初始归并段，归并段 1 为 $(2, 8, 12, 16, 28, 30)$ ，归并段 2 为 $(4, 6, 10, 18, 20)$ 。

表 11.1 初始归并段的生成过程

读入记录	内存工作区状态	R_{min}	输出之后的初始归并段状态
12, 2, 16, 30	12, 2, 16, 30	$2(i=1)$	归并段 1: {2}
8	12, 8, 16, 30	$8(i=1)$	归并段 1: {2, 8}
28	12, 28, 16, 30	$12(i=1)$	归并段 1: {2, 8, 12}
4	4, 28, 16, 30	$16(i=1)$	归并段 1: {2, 8, 12, 16}
10	4, 28, 10, 30	$28(i=1)$	归并段 1: {2, 8, 12, 16, 28}
20	4, 20, 10, 30	$30(i=1)$	归并段 1: {2, 8, 12, 16, 28, 30}
6	4, 20, 10, 6	4($4 < 30$, 开始新归并段 $i=2$)	归并段 1: {2, 8, 12, 16, 28, 30} 归并段 2: {4}
18	18, 20, 10, 6	$6(i=2)$	归并段 1: {2, 8, 12, 16, 28, 30} 归并段 2: {4, 6}
	18, 20, 10,	$10(i=2)$	归并段 1: {2, 8, 12, 16, 28, 30} 归并段 2: {4, 6, 10}
	18, 20,,	$18(i=2)$	归并段 1: {2, 8, 12, 16, 28, 30} 归并段 2: {4, 6, 10, 18}
	, 20,,	$20(i=2)$	归并段 1: {2, 8, 12, 16, 28, 30} 归并段 2: {4, 6, 10, 18, 20}

3. 设输入的关键字满足 $k_1 > k_2 > \dots > k_n$ ，缓冲区大小为 m ，用置换-选择排序方法可产生多少个初始归并段？

答：可产生 $\lceil n/m \rceil$ 个初始归并段。设记录 R_i 的关键字为 k_i ($1 \leq i \leq n$)，先读入 m 个记录 R_1, R_2, \dots, R_m ，采用败者树选择最小记录 R_m ，将其输出到到归并段 1， $R_{min}=k_m$ ，在该位置上读入 R_{m+1} ，采用败者树选择最小记录 R_{m-1} ，将其输出到到归并段 1， $R_{min}=k_{m-1}$ ，在该位

置上读入 R_{m+2} ，采用败者树选择最小记录 R_{m-2} ，将其输出到归并段 1， $R_{min}=k_{m-2}$ ， \dots ，以此类推，产生归并段 1: $(R_m, R_{m-1}, \dots, R_1)$ 。同样产生其他归并段 $(R_{2m}, R_{2m-1}, \dots, R_{m+1})$ ， $(R_{3m}, R_{3m-1}, \dots, R_{2m+1})$ ， \dots ，一共有 $\lceil n/m \rceil$ 个初始归并段。

4. 什么是多路平衡归并，多路平衡归并的目的是什么？

答：归并过程可以用一棵归并树来表示。多路平衡归并对应的归并树中，每个结点都是平衡的，即每个结点的所有子树的高度相差不超过 1。

k 路平衡归并的过程是：第一趟归并将 m 个初始归并段归并为 $\lceil m/k \rceil$ 个归并段，以后每一趟归并将 l 个初始归并段归并为 $\lceil l/k \rceil$ 个归并段，直到最后形成一个大的归并段为止。

m 个归并段采用 k 路平衡归并，总的归并趟数 $s = \lceil \log_k m \rceil$ 。其趟数是所有归并方案中最少的，所以多路平衡归并的目的是减少归并趟数。

5. 什么是败者树？其主要作用是什么？用于 k 路归并的败者树中共有多少个结点（不计冠军结点）？

答：败者树是一棵有 k 个叶子结点的完全二叉树，从叶子结点开始，两个结点进行比较，将它们中的败者（较大者）上升到双亲结点，胜者（较小者）参加更高一层的比较。

败者树的主要作用是从 k 个记录中选取关键字最小的记录。

败者树中有 k 个叶子结点，且没有度为 1 的结点，即 $n_0 = k$ ， $n_1 = 0$ ， $n_2 = n_0 - 1 = k - 1$ ，所以 $n = n_0 + n_1 + n_2 = 2k - 1$ 。

6. 如果某个文件经内排序得到 80 个初始归并段，试问：

(1) 若使用多路平衡归并执行 3 趟完成排序，那么应取的归并路数至少应为多少？

(2) 如果操作系统要求一个程序同时可用的输入/输出文件的总数不超过 15 个，则按多路平衡归并至少需要几趟可以完成排序？如果限定这个趟数，可取的最低路数是多少？

答：(1) 设归并路数为 k ，初始归并段个数 $m = 80$ ，根据多路平衡归并趟数计算公式 $s = \lceil \log_k m \rceil = \lceil \log_k 80 \rceil = 3$ ，则 $k^3 \geq 80$ ，即 $k \geq 5$ 。也就是说，可取的最低路数是 5。

(2) 设多路平衡归并的归并路数为 k ，需要 k 个输入缓冲区和 1 个输出缓冲区。1 个缓冲区对应一个文件，有 $k + 1 = 15$ ，因此 $k = 14$ ，可做 14 路归并。由 $s = \lceil \log_k m \rceil = \lceil \log_{14} 80 \rceil = 2$ 。即至少需 2 趟归并可完成排序。

若限定这个趟数，由 $s = \lceil \log_k 80 \rceil = 2$ ，有 $80 \geq k^2$ ，可取的最低路数为 9。即要在 2 趟内完成排序，进行 9 路排序即可。

7. 若采用置换选择排序算法得到 8 个初始归并段，它们的记录个数分别为 37、34、300、41、70、120、35 和 43。画出这些磁盘文件进行归并的 4 阶最佳归并树，计算出总的读写记录数。

答： $k = 4$ ， $m = 8$ ， $k - (m - 1) \bmod (k - 1) - 1 = 2$ ，则设两个虚段。4 阶最佳归并树如图 11.2 所示。

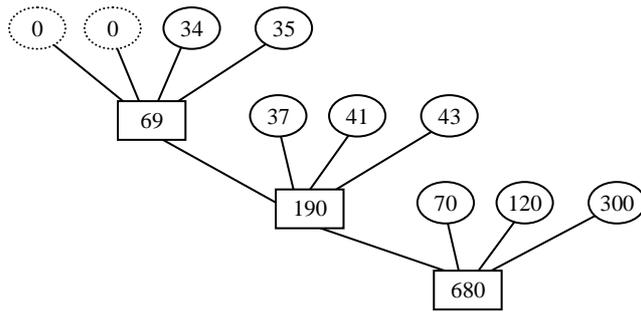


图 11.2 一棵 4 阶最佳归并树

第 1 趟读记录数： $34+35=69$ 。

第 2 趟读记录数： $69+37+41+43=190$ 。

第 3 趟读记录数： $190+70+120+300=680$ 。

总的读记录数= $69+190+680=939$ ，总的读写记录数= $939 \times 2=1878$ 。