

6.6 哈夫曼树及其应用

6.6.1 哈夫曼树（最优二叉树）

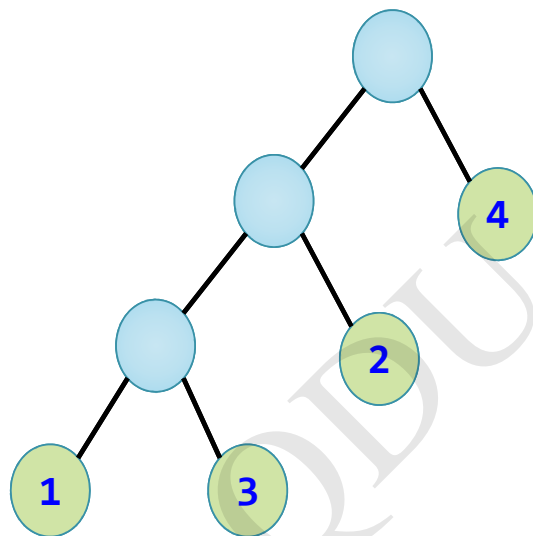
- 设二叉树具有 n 个带权值的叶子结点，从根结点到每个叶子结点都有一个路径长度。
- 从根结点到各个叶子结点的路径长度与相应结点权值的乘积的和称为该二叉树的带权路径长度，记作：

$$WPL = \sum_{i=1}^n w_i l_i$$

- 其中， w_i 为第 i 个叶子结点的权值， l_i 为第 i 个叶子结点的路径长度。

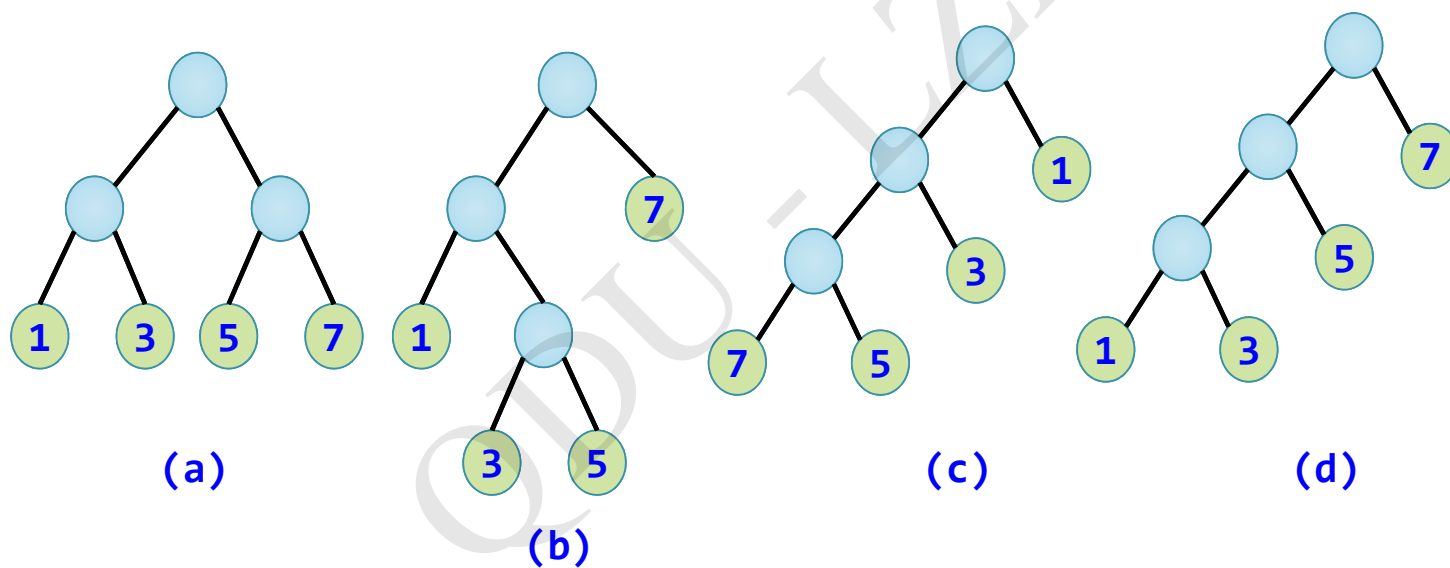
以下二叉树的带权路径长度值：

$$\text{WPL}=1\times 3+3\times 3+2\times 2+4\times 1=20。$$



- 给定一组具有确定权值的叶子结点，可以构造出许多形状的二叉树。
- ✓ 例如，用4个整数2、3、4、11作为4个叶子结点的权值，总共可以构造出120棵不同的二叉树，它们的带权路径长度可能不相同。把其中具有最小带权路径长度的二叉树称为哈夫曼树。

【示例-1】 如下图的4棵二叉树具有相同的叶子结点，计算出它们的带权路径长度。



解：它们的带权路径长度分别为：

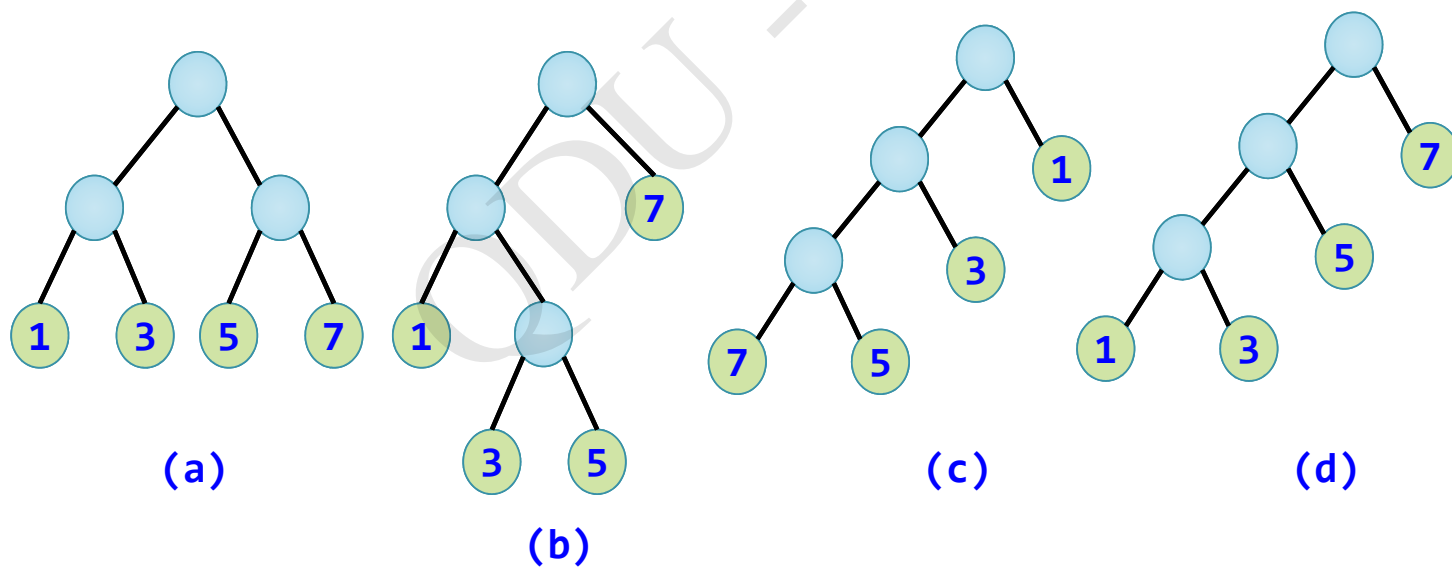
(a) $WPL=1\times 2+3\times 2+5\times 2+7\times 2=32$

(b) $WPL=1\times 2+3\times 3+5\times 3+7\times 1=33$

(c) $WPL=7\times 3+5\times 3+3\times 2+1\times 1=43$

(d) $WPL=1\times 3+3\times 3+5\times 2+7\times 1=29$

其中图(d)所示的二叉树就是一棵哈夫曼树。



6.6.2 构造哈夫曼树

- ✓ 根据哈夫曼树的定义，一棵二叉树要使其WPL值最小，必须使权值越大的叶子结点越靠近根结点。
- ✓ 而权值越小的叶子结点越远离根结点。

哈夫曼算法

构造一棵哈夫曼树的过程如下：

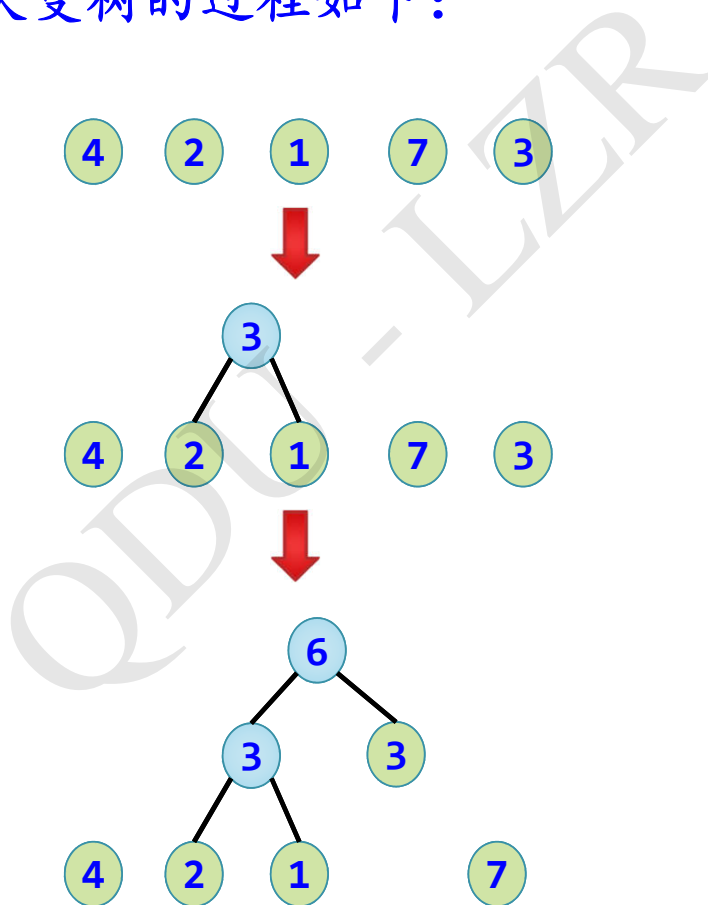
(1) 由给定的 n 个权值 $\{w_1, w_2, \dots, w_n\}$ 构造 n 棵只有一个根结点的二叉树，从而得到一个二叉树的集合 $F=\{T_1, T_2, \dots, T_n\}$ ；

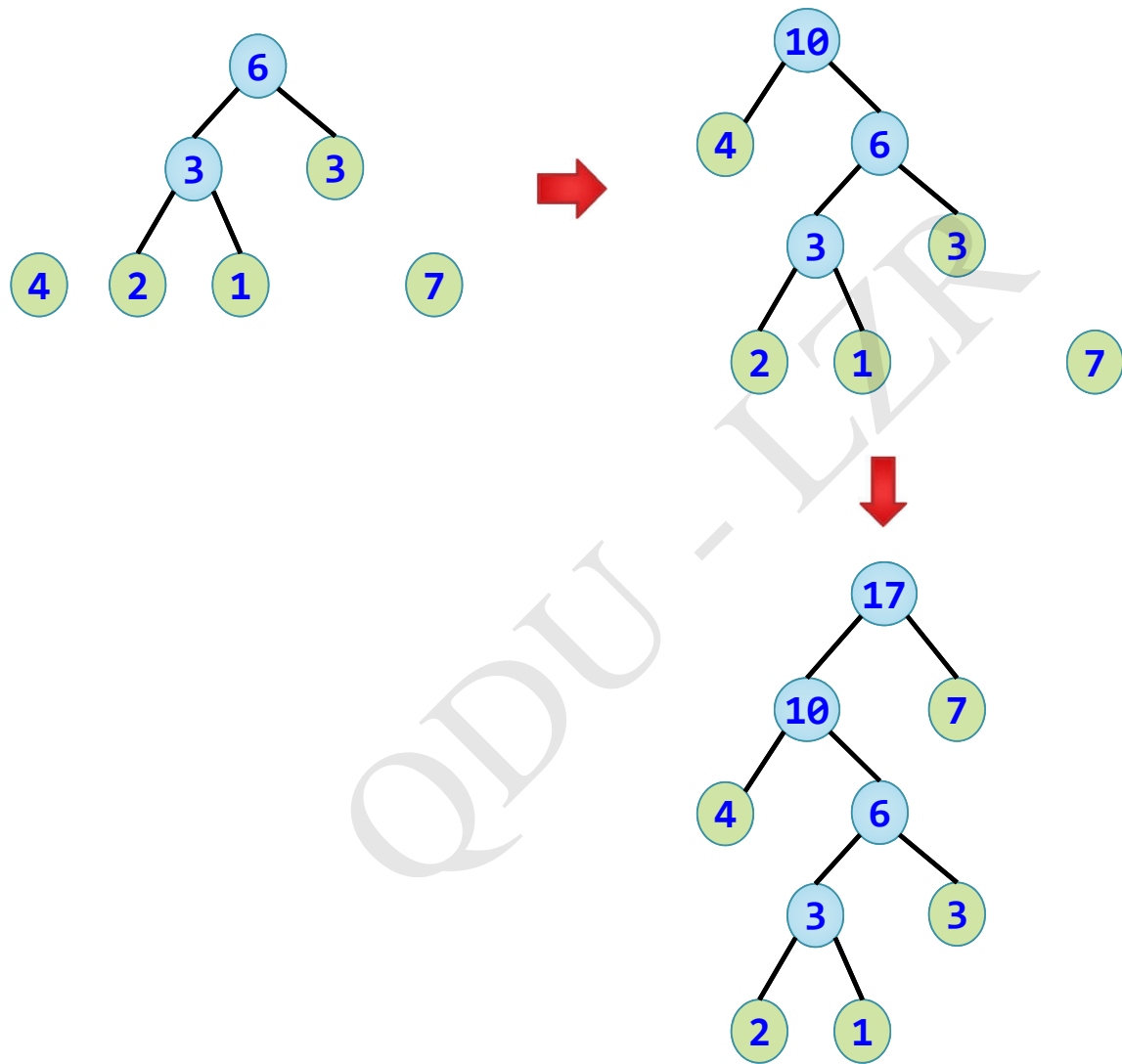
(2) 在 F 中选取根结点的权值最小和次小的两棵二叉树 T_i 、 T_j 进行合并，即增加一个根结点，将 T_i 、 T_j 分别作为它的左、右子树，该根结点的权值为其左、右子树根结点权值之和；

(3) 重复步骤(2)，当 F 中只剩下一棵二叉树时，这棵二叉树便是所要建立的哈夫曼树。

【示例-2】 对于一组给定的叶子结点，它们的权值集合为 $W=\{4,2,1,7,3\}$ ，给出由此集合构造哈夫曼树的过程。

解： 构造哈夫曼树的过程如下：





6.6.3 哈夫曼编码

- ✓ 哈夫曼编码具有广泛的应用，利用哈夫曼树构造的用于通信的二进制编码称为**哈夫曼编码**。
- ✓ 在传送电文时，希望总长尽可能地短。如果对每个字符设计长度不等的编码，且让电文中出现次数较多的字符采用尽可能短的编码，则传送电文的总长便可减少。
- ✓ 若要设计长短不等的编码，则必须是任一个字符的编码都不是另一个字符的编码的前缀，这种编码称做**前缀编码**。

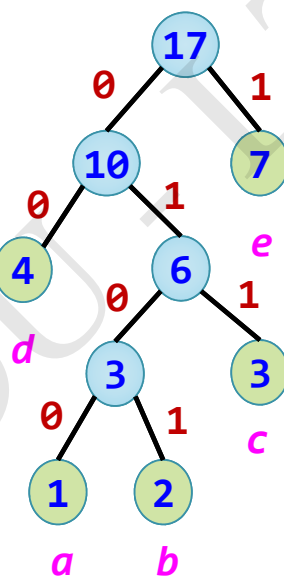
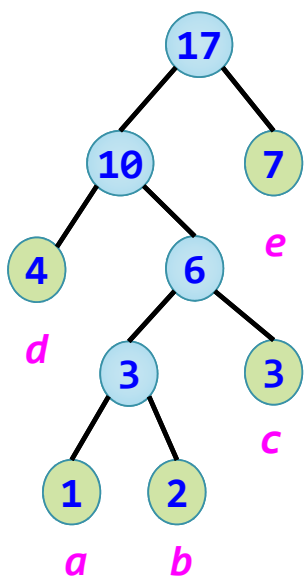
构造哈夫曼编码的方法

- ✓ 用给定的若干字符的频度为权值生成哈夫曼树，并在每个叶子上注明对应的字符。
- ✓ 树中从根到每个叶子都有一条路径，对路径上的各分支约定指向左子树根的分支表示“0”码，指向右子树的分支表示“1”码，取每条路径上的“0”或“1”的序列作为和各个叶子对应的字符的编码，这就是哈夫曼编码。

【示例-3】假设5个字符及其频度：

$a:1$ $b:2$ $c:3$ $d:4$ $e:5$

构造它们的哈夫曼编码。



哈夫曼编码

$a:0100$

$b:0101$

$c:011$

$d:00$

$e:1$

从哈夫曼编码看出，对于 n 个字符，构造它们的哈夫曼编码，没有一个字符的哈夫曼编码是另一个字符的哈夫曼编码的**前缀**。

【练习题-1】 5个字符有如下4种编码方案，不是前缀编码的是_____。

- A. 01, 0000, 0001, 001, 1
- B. 011, 000, 001, 010, 1
- C. 000, 001, 010, 011, 100
- D. 0, 100, 110, 1110, 1100

说明：本题为2014年全国考研题

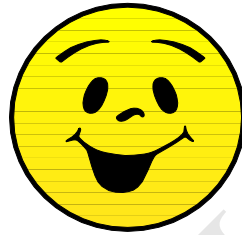
正确答案为：D

【练习题-2】对 n ($n \geq 2$) 个权值均不同的字符构成哈夫曼树，关于该树的叙述中，错误的是_____。

- A. 该树一定是一棵完全二叉树
- B. 该树中一定没有度为1的结点
- C. 树中两个权值最小的结点一定是兄弟结点
- D. 树中任一非叶子结点的权值一定不小于下一层任一结点的权值

说明：本题为2010年全国考研题

正确答案为：A



— END —