

Huffman Codes题解

浙江省镇海中学 岑若虚

1 题目大意

若已知文本中每个字母的出现频率，可按如下步骤构造哈夫曼编码：

1. 每个字母建立一个节点，权值为该字母的出现频率，构成一片森林。
2. 新建一个节点，选取森林中根节点权值最小的树作为新节点的左子树，根节点权值次小的树作为右子树（若最小的两棵树权值相等可任意排列）。新节点的权值为两个儿子的权值和。删去原来的两棵树。
3. 重复第2步直到森林中只有一棵树。
4. 对树上每条向左的边标0，向右的边标1。从根走到某个叶子，将经过的边上的标记连起来，就得到了这个叶子代表的字母的哈夫曼编码。

假设每个字母的出现频率都是0.01的正整数倍，和为1。告诉你每个字母的哈夫曼编码，问有多少种字母出现频率的分布情况能得到这样的哈夫曼编码。字母最多有20个。

2 算法分析

先根据哈夫曼编码建立01字母树，这就是构造过程中最后的那棵树，称为哈夫曼树。问题转化为有多少种给哈夫曼树上的节点赋权值的方法。

很容易想到用树形DP或区间DP等方法求解，但是“每次选取权值最小的两个节点”这一操作使得大小关系错综复杂，难以划分出独立的子问题。考虑到问题规模较小，如果能发现一些性质进行剪枝，应该可以通过搜索求解。

我们发现合法的赋值方法具有如下性质：

1. 深度相同的节点中，左边节点的权值不大于右边节点。
2. 深度较大的节点的权值不大于深度较小的节点。

也就是说，将哈夫曼树进行每次先访问右儿子的BFS，BFS序即为按权值降序排序的结果。

简要证明如下。对于性质1，如果两个节点的父亲相同，由于构造过程中将权值较小的节点作为左儿子，结论成立。如果两个节点的父亲不同，设左边节点为 u ，其兄弟为 u' ，父亲为 fu ，右边节点为 v ，其兄弟为 v' ，父亲为 fv 。假设 u 的权值较大。那么在合并 v 和 v' 的时候，由于它们两个是权值最小和次小的节点，所以 v 和 v' 的权值都比 u 和 u' 小。由于父亲的权值等于儿子的权值之和， fv 的权值也比 fu 小。 fu 仍然是 fv 左边的节点，这样可以不断往上推，最后总会出现两个节点的父亲相同的情况。这时左边节点的权值依然较大，这与前面的讨论矛盾。因此假设不成立，结论成立。

性质2可以用类似的方法证明。如果上面节点是下面节点的祖先，结论显然成立。否则，假设上面节点的权值较小，也可推出上面节点的父亲权值比下面节点的父亲小。同样可以不断向上推，最后总会出现上面节点是下面节点祖先的情况，这时上面节点的权值依然较小，与前面的讨论矛盾。因此假设不成立，结论成立。

我们还可以证明，满足这两个性质的赋值方法就是合法的赋值方法。因此，按照从上到下，从右到左的顺序dfs，可以将权值范围缩减到最小。由于合法方案不会很多（最多3万左右），dfs过程中在合法权值范围内枚举即可。

时间复杂度：上限 $O(50^N)$ ，实际远远不到。

空间复杂度： $O(N)$