Homework05 2021.11.12

1

假定一个数据文件由 8 位字符组成，其中所有 256 个字符出现的频率大致相同：最高的频率也低于最低频率的 2 倍。证明：在此情况下，赫夫曼编码并不比 8 位固定长度编码更有效。

按照赫夫曼编码的算法，由字符出现的频率从小到大将字符进行组合，由于最高的频率低于最低频率的 2 倍，因此频率最小的两个字符组合的频率也大于原来频率最大的字符。

所以赫夫曼树的第二层序号为i的结点由底层序号为$2i$和$2i+1$的节点组合而成，而将这128个结点视为128个字符，设最低的两个频率为$p_1 \leq p_2$,最高的则为$q_1 \geq q_2$，那么显然第二层中，有$q_1+q_2 \leq 2p_1+2p_1 \leq 2(p_1+p_2)$。仍有最高的频率低于最低频率的 2 倍。

同理，第三层至根节点均有如上结论。

故，由赫夫曼算法构建的编码系统每一个字符的编码也是8位01串，即就是说赫夫曼编码并不比 8 位固定长度编码更有效。

2

令 S是一个有限集，$S_1, S_2, \ldots, S_k$是S的一个划分，这些集合都是非空且不相交的。定义结构(S, ϕ)满足条件$\phi = \{A : |A \cap S_i| \leq 1, i = 1, \ldots, k\}$。证明：$(S, \phi)$是一个拟阵。也就是说，与划分中所有子集都最多有一个共同元素的集合A组成的集合构成了拟阵的独立集。

假设$A \subseteq B$且$B \in \phi$，则对于任意$i, i = i, \ldots, k$有$(A \bigcap S_i) \subseteq (B \bigcap S_i)$，则$|A \bigcap S_i| \leq |B \bigcap S_i| \leq 1$，故$A \in \phi$

假设$A \in \phi$且$B \in \phi$，同时$|A| < |B|$，那么必然存在j使得$|B \bigcap S_j| = 1$而$|A \bigcap S_j| = 0$，令$a = B \bigcap S_j$，则$a \notin A$且$|(A \bigcup \{a\}) \bigcap S_j| = 1$，即$A \bigcup \{a\} \in \phi$

综上所述(S, ϕ)是一个拟阵。

3

$A = a_1, \ldots, a_n$表示一个正整数集合。 A中的元素之和为N。设计一个 $O(n \cdot N)$ 的算法来确定是否存在一个 A 的子集 B，使得 $\sum_{a_i \in B} a_i = \sum_{a_i \in A-B} a_i$

算法思想：使用动态规划方法，建立一个数组$reachable[N+1]$，其中$reachable[i] = true$表示存在集合A的子集S，使得$\sum_{a_i \in S} a_i = i$

算法伪代码：

```
Create array reachable[N + 1]
Create array group[n]

GROUP-PARTITION
    sum = 0
    for i = 1 to n do
        sum += group[i]
    if(sum % 2 != 0)
        return false
```

```
    else
        DP-REACHABLE(sum / 2)

DP-REACHABLE(int target)
    reachable[0] = true
    for i = 1 to N do
        reachable[i] = false
    for i = 1 to n do
        for k = target to i do
            if(reachable[k - group[i]] == true)
                reachable[k] = true
    return reachable[target]
```