实验二 基于 Matlab 的信源编码实验

一．实验目的

1. 掌握信源编码的方法，理解香农定理的意义；
2. 能够针对给定的信源进行数据压缩和恢复。

二．实验要求

1. 学习常见的信源编码方案，如等长编码、游程编码、Huffman 编码、LZ 编码等；
2. 自选一个信源，根据信源特性和信道需求选择合适的信源编码方案，利用 matlab 仿真软件进行信源的压缩和恢复；
3. 对比编码前后信源占用空间的大小，计算压缩率和编码效率。

三．实验原理

- 信源编码是一种信息处理技术，主要目的是数据压缩，旨在减少或消除信源的冗余度，提高通信的有效性。信源编码将消息映射成码字，因此当信源是模拟信号时，信源编码也是模/数转换过程。
- 信源编码可以分为无失真（香农第一定理）和率失真（香农第三定理）两种类型。无失真编码能保证码元序列无失真地恢复成信源序列。

四．实验操作步骤

对文本文件的 Huffman 编码及解码

1. 统计文本文件各字符出现的概率
2. 构造 Huffman 树，对各字符进行编码，生成码书文件
3. 对文本文件进行 Huffman 编码
4. 对编码后二进制序列进行 Huffman 解码

五．实验结果记录与分析

1. 统计文本文件中各字符出现频率：

```
dimension = size(w);
word(1:dimension(1), 1) = char(tab(:, 1));
word(dimension(1) + 1, 1) = ' ';
```
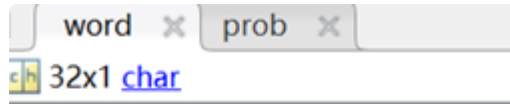
```matlab
for i = 1:dimension(1) + 1
        count(i) = length(find(x == word(i)));
end

prob = count / sum(count); %各字符的概率
```

统计结果：

word	prob

32x1 char

val =

    'I'
    't'

word	prob

1x32 double

	1	2	3
1	0.0083	0.0792	0.0314
2			

2. 基于统计文本文件中的各字符及其出现频率进行大小排序来进行 Huffman 编码：

```matlab
function huff = Code(prob)
    %进行Huffman编码
    n = length(prob);
    p = prob
    sumpro = zeros(n - 1, n);
    %loca用于记录每行最小两概率叠加后概率的排列次序
    for i = 1:n - 1
        [p, index] = sort(p) %进行概率值排序
        sumpro(i, :) = [index(1:n - i + 1), zeros(1, i - 1)] %每一次排序
求和后都会少一个概率值，对有值的赋予排列次序，其余赋0
        p = [p(1) + p(2), p(3:n), 1] %对经过求和处理的概率第一项赋两最小概率的
和值，对3-n项概率照搬，其余概率赋1
    end

    for i = 1:n - 1 %循环生成一个N-1行、n^2列数组C，每行可看作n个段，每段长为
n，记录一个码字（每个码字的长度不会超过n）
        code(i, 1:n * n) = blanks(n * n)
    end

    code(n - 1, n) = '0' % 给code矩阵的N-1行的第一个段赋值0
    code(n - 1, 2 * n) = '1' % 第二个段赋值1。（这两个码字对应编码中最后相加为
一的两个概率）
```

```matlab
    for i = 2:n - 1 %主要的程序，循环N-2次
        code(n - i, 1:n - 1) = code(n - i + 1, n * (find(sumpro(n - i +
1, :) == 1)) - (n - 2):n * (find(sumpro(n - i + 1, :) == 1)))
        code(n - i, n) = '0' %在分支的第一个元素最后补0
        code(n - i, n + 1:2 * n - 1) = code(n - i, 1:n - 1)
        code(n - i, 2 * n) = '1' %在分支的第一个元素最后补1
        %每一行值都从下一行值得到，找到在下一行码字中相加本行最小两个概率得到的概率
的对应码字，
        %本行两个最小概率对应码字分别为此码字最后加"0"，加"1"
        for j = 1:i - 1
            code(n - i, (j + 1) * n + 1:(j + 2) * n) = code(n - i + 1, n
* (find(sumpro(n - i + 1, :) == j + 1) - 1) + 1:n * find(sumpro(n - i +
1, :) == j + 1))
        end

    end

    for i = 1:n
        huff(i, 1:n) = code(1, (find(sumpro(1, :) == i) - 1) * n +
1:find(sumpro(1, :) == i) * n)
    end

end
```
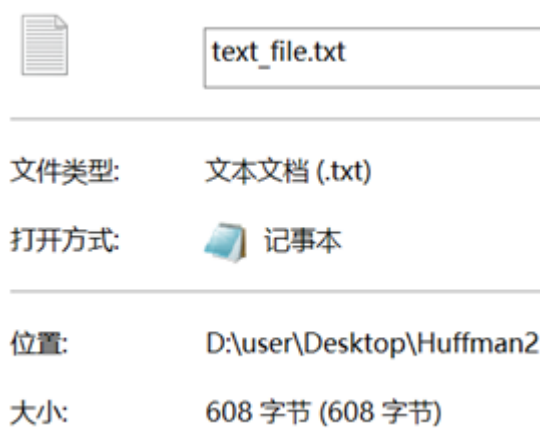
码书文件：

```
📄 codeBook.txt - 记事本
文件(F)  编辑(E)  格式(O)  查看(V)  帮助(H)
|            0101011
t            1101
w            01011
a            0100
s            1010
h            11110
e            011
b            1000110
o            1100
f            10000
i            1001
m            1000111
;            1011011
r            11111
.            1110110
g            111001
d            111010
l            111000
n            10111
p            101100
c            1110111
u            1000100
y            0101010
L            010100110
D            010100111
k            01010000
W            01010001
v            1000101
H            010100100
,            1011010
—            010100101
             00
```
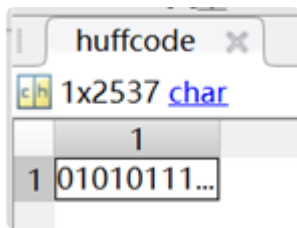
编码前后：

- 编码前：

  📄   text_file.txt

文件类型：	文本文档 (.txt)
打开方式：	📒 记事本
位置：	D:\user\Desktop\Huffman2
大小：	608 字节 (608 字节)

- 编码后：

- 编码效率：

```
ave =

    4.1865


H =

    4.1501


efficiency =

    0.9913
```

压缩率 = 2537/4864 * 100% = 52.2%

3. 译码过程：

- 读取编码后的文件和码书。
- 根据码书中的信息构建字符字典和字符码字典。
- 对编码后的文件进行逐个字符的解码。

译码结果：

- 编码前文本：

text_file.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

It was the best of times; it was the worst of times.It was the age of wisdom; it was the age of foolishness.It was the epoch of belief; it was the epoch of incredulity.It was the season of Light; it was the season of Darkness.It was the spring of hope; it was the winter of despair.We had everything before us; we had nothing before us.We were all going direct to Heaven, we were all going direct the other way —in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.

译码结果:

Decode.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

It was the best of times; it was the worst of times.It was the age of wisdom; it was the age of foolishness.It was the epoch of belief; it was the epoch of incredulity.It was the season of Light; it was the season of Darkness.It was the spring of hope; it was the winter of despair.We had everything before us; we had nothing before us.We were all going direct to Heaven, we were all going direct the other way —in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.

六．思考题回答和心得体会

1. 引入信源编码的目的是什么?
   - 节省带宽和存储空间: 信源编码通过消除或减少数据中的冗余信息来压缩数据,从而减少了数据传输所需的带宽和存储空间。这对于网络通信、存储系统以及传感器数据等应用领域都非常重要。
   - 提高传输效率: 通过信源编码,可以减少数据传输时所需的时间和资源,从而提高了传输的效率。特别是在有限带宽的网络环境下,信源编码可以显著提升数据传输的速度和效率。