信息论与编码

第四章作业解析

第三次作业情况统计

- 总人数：111人；交作业总人数：77
- 交作业人数：　电子1班：21人，总人数30人

  电子2班：27人，总人数32人

  电子3班：22人，总人数32人

  电子2003级：5人，总人数15人

  茅电：2人，总人数2人
- A（A和A-）：共计9人

  电子1班：何国军

  电子2班：王乾刚，李炳兴，钟如秀

  电子3班：刘杨，江燕刚，陈惠卿，曾元(A^{+})

  茅电：顾博川
- D: 共计1人
- 未交作业共计34人

1. 有一信源，它有6个可能的输出，其概率分布如下表所示，表中给出了对应的码A、B、C、D、E和F。

消息	$p(a_i)$	A	B	C	D	E	F
a_1	1/2	000	0	0	0	0	0
a_2	1/4	001	01	10	10	10	100
a_3	1/16	010	011	110	110	1100	101
a_4	1/16	011	0111	1110	1110	1101	110
a_5	1/16	100	01111	11110	1011	1110	110
a_6	1/16	101	011111	111110	1101	1111	011

(1)、求这些码中哪些是唯一可译码；

(2)、求哪些是非延长码（即时码）；

(3)、对所有唯一可译码求出其平均码长和编码效率。

解：

唯一可译码是A， B， C， E,非延长码为A，C，E

A的平均码长：$\bar{n}=\sum_{i=1}^{6}p(s_i)n_i=3(1/2+1/4+1/16+1/16+1/16+1/16)$

$=3$码符号/信源符号

编码效率：$\eta=\dfrac{H(s)}{\bar{n}\log r}=\dfrac{2}{3}*100\%=66.67\%$

B的平均码长：

$\bar{n}=\sum_{i=1}^{6}p(s_i)n_i=1*1/2+2*1/4+3*1/16+4*1/16+5*1/16+6*1/16$

$=17/8=2.125$ 码符号 / 信源符号

编码效率 $\eta=\dfrac{H(s)}{\bar{n}\log r}=\dfrac{2}{2.125}*100\%=94.12\%$

C 的平均码长：

$$\bar{n} = \sum_{i=1}^{6} p(s_i)n_i = 2.125\text{码符号}/\text{信源符号}$$

编码效率：

$$\eta = \frac{H(s)}{\bar{n}\log r} = 94.12\%$$

E 的平均码长：

$$\bar{n} = \sum_{i=1}^{6} p(s_i)n_i = 1*1/2 + 2*1/4 + 4*1/16 + 4*1/16$$

$$+4*1/16 + 4*1/16 = 2\text{码符号}/\text{信源符号}$$

编码效率：

$$\eta = \frac{H(s)}{\bar{n}\log r} = \frac{2}{2}*100\% = 100\%$$

- 判唯一可译码的方法：

  给定一有限长码元序列，可采用下面方法判断是否为唯一可译码:

  – 判断是否为非奇异码;

  – 等长的非奇异码一定是唯一可译码;

  – 对于非等长的非奇异码，先剔除不满足Kraft不等式的码;

  – 用定义判断.

---- 2003级 程华

2. 有一个信源X如下：

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0.32 & 0.22 & 0.18 & 0.16 & 0.08 & 0.04 \end{bmatrix}$$

(1)、求信源熵;

(2)、用Shannon编码法编成二进制变长码，并计算其编码效率；

(3)、用Fano编码法编成二进制变长码，并计算其编码效率；

(4)、用Huffman码编码成二进制变长码，并计算其编码效率；

(5)、用Huffman码编码成三进制变长码，并计算其编码效率；

(6)、比较三种编码方法的优缺点。

解：

(1) $H(X)=\sum_{i=1}^{6}p(x_i)\log\frac{1}{p(x_i)}=2.3522bit/$信源符号

(2) $\log_r\frac{1}{p(x_i)}\le n_i<\log_r\frac{1}{p(x_i)}+1$

由于每个码长肯定是正整数，上式给出了码长选择范围。

符号	$p(x_i)$	累加概率	-log $p(x_i)$	码字长度	码字
x_1	0.32	0	1.6439	2	0 0
x_2	0.22	0.32	2.1844	3	0 1 0
x_3	0.18	0.54	2.4739	3	1 0 0
x_4	0.16	0.72	2.6439	3	1 0 1
x_5	0.08	0.88	3.6439	4	1 1 1 0
x_6	0.04	0.96	4.6439	5	1 1 1 1 0

平均码长 $\bar{n} = 2.84$ 码符号/信源符号

编码效率 $\eta = \dfrac{H(X)}{\bar{n}\log r} = \dfrac{2.3522}{2.84\log 2} \times 100\% = 82.82\%$

FANO编码

符号	$P(x_i)$					码字	码字长度
x_1	0.32	0.54 0	0			00	2
x_2	0.22		1			01	2
x_3	0.18		0.18 0			10	2
x_4	0.16			0.16 0		110	3
x_5	0.08	0.46 1	0.28 1	0.12 1	0	1110	4
x_6	0.04				1	1111	4

平均码字长度$\overline{n} = 2.4$ 码符号/信源符号

编码效率 $$\eta = \frac{H(X)}{\overline{n}\log r} = \frac{2.3522}{2.4\log 2} \times 100\% = 98\%$$

二进制Huffman编码

符号	$p(x_i)$					码字	码字长度
x_1	0.32	0.32	0.32	0.40	0.60	00	2
x_2	0.22	0.22	0.28	0.32	0.40	10	2
x_3	0.18	0.18	0.22	0.28		11	2
x_4	0.16	0.16	0.18			010	3
x_5	0.08	0.12				0110	4
x_6	0.04					0111	4

平均码字长度 $\bar{n} = 2.4$ 码符号/信源符号

编码效率 $\eta = \dfrac{H(X)}{\bar{n}\log r} = \dfrac{2.3522}{2.4\log 2} \times 100\% = 98\%$

三进制Huffman编码

首先，判断$q-(r-1)\alpha \overset{?}{=} r$

$$6-(3-1)\times 2 = 2 < 3$$

选择$m = r-[q-(r-1)\alpha] = 3-2 = 1$个虚假符号

符号	$p(x_i)$			码字	码字长度
x_1	0.32	0.32	0.48 0	1	1
x_2	0.22	0.22	0.32 1	2	1
x_3	0.18	0.18 0	0.22 2	00	2
x_4	0.16	0.16 1		01	2
x_5	0.08 0	0.12 2		020	3
x_6	0.04 1			021	3
x'	0 2			022	舍去

平均码字长度 $\bar{n} = 1.58$ 码符号/信源符号

编码效率 $\eta = \dfrac{H(X)}{\bar{n}\log r} = \dfrac{2.3522}{1.58\log 3} \times 100\% = 93.93\%$

3. 现有一幅已离散量化后的图像，图像的灰度量化分成**8**级，如下表所示。表中数字为相应像素上的灰度级

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	3	3	3
3	3	3	3	3	3	3	4	4	4
4	4	4	4	4	4	4	5	5	5
5	5	5	5	6	6	6	6	6	6
7	7	7	7	7	8	8	8	8	8

另有一无噪无损二元信道，单位时间(秒)内传输100个二元符号。

(1)、现将图像通过给定的信道传输，不考虑图像的任何统计特性，并采用二元等长码，问需要多长时间才能传送完这幅图像？

(2)、若考虑图像的统计特性(不考虑图像的像素之间的依赖性)，求这幅图像的信源熵$H(S)$，并对每个灰度级进行Huffman最佳二元编码，问平均每个像素需用多少二元码符号来表示？这时需多少时间才能传送完这幅图像？

(3)、从理论上简要说明这幅图像还可以压缩，而且平均每个像素所需的二元码符号数可以小于$H(S)$比特。

解：

(1)一幅已离散化后的图象，其灰度划分成 8 级，先不考虑图象的任何统计特性，采用二元等长码，因为q=8,所以要满足

$$2^l \geq q = 8$$

故 $l = 3$ 二元码符号 / 灰度级

即每个灰度等级需采用三位二元符号来传输。

这一幅图象空间离散化后共有N=100个像素，每个像素的灰度需用三个二元符号来编码，所以这幅图象采用二元等长码后共需300个二元符号来描述。所传输的信道是无噪无损信道，其每秒钟传输100个二元符号。因此，需 3 秒钟才能传送完这幅图象。

(2)考虑图象的统计特性(不考虑图象的像素之间的依赖性)时，根据此图象进行统计，把像素的灰度值作为信源S，可得

$$\begin{bmatrix} S \\ P(s_i) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{40}{100} & \frac{17}{100} & \frac{10}{100} & \frac{10}{100} & \frac{7}{100} & \frac{6}{100} & \frac{5}{100} & \frac{5}{100} \end{bmatrix}$$

所以 $H(S) = -\sum_{i=1}^{8} p(s_i)\log p(s_i) \approx 2.572$ 比特 / 灰度级

对此灰度进行哈夫曼最佳二元编码：

码长l_i	码字	灰度级	概率$P(s_i)$					
1	1	1	0.40	0.40	0.40	0.40	0.40	0.60 0
3	001	2	0.17	0.17	0.20	0.23	0.37 0	0.40 1
4	0000	3	0.10	0.13	0.17	0.20 0	0.23 1	
4	0001	4	0.10	0.10	0.13 0	0.17 1		
4	0100	5	0.07 0	0.10 0	0.10 1			
4	0101	6	0.06 1	0.10 1				
4	0110	7	0.05 0					
4	0111	8	0.05 1					

得 $$\bar{L}=\sum_{i=1}^{8}P(s_i)l_i=2.63$$ 二元符号 / 灰度级

通过哈夫曼最佳二元编码后，每个像素平均需要用2.63个二元符号，则此图象平均共需要用263个二元符号来表示。因此，需2.63秒才能传送完这幅图象。

(3)在(2)题中计算时没有考虑图象的像素之间的依赖关系，但实际此图象的像素之间是有依赖的。例如，若考虑像素前后之间灰度的依赖关系，就有灰度“1”后面只可能出现灰度“1”或 “2”；灰度“2”后只可能出现“2” 或“3” ，等等。这时，此图象灰度值信源S可以看成一阶马尔可夫信源。还可以进一步看成为m阶马尔可夫信源。因此，在考虑了这些依赖关系后，像素的灰度值信源S的实际信息熵$H_\infty < H(S)$。根据香农第一理，总可以找到一种编码，使每个灰度级的平均码长$\overline{L} \to H_\infty$(极限熵)。所以，这幅图象还可以进一步压缩，平均每个像素(灰度)所需要的二元码符号数$\overline{L} < H(S)$。