

4.5.3 香农编码

香农(Shannon)编码步骤(以二进制编码为例):

- (1) 将信源符号按概率从大到小的排序;
- (2) 按下式求第 i 个信源符号对应的码字的码长 l_i , 并取整;

$$-\log P(u_i) \leq l_i < -\log P(u_i) + 1$$

- (3) 按下式求 i 个信源符号的累加概率 P_i ;

$$\begin{cases} P_1 = 0 \\ P_i = \sum_{k=1}^{i-1} P(u_k) \quad i = 2, 3, \dots, q \end{cases}$$

- (4) 将累加概率 P_i 转换成二进制数;
- (5) 取 P_i 二进制数小数点后 l_i 个二进制数字作为第 i 个信源符号的码字。

对给定信源 $\begin{bmatrix} U \\ P(u_i) \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$

进行 $r=2$ 进制香农编码。

消息符号 u_i	消息概率 p_i	$-\log p_i$	码长 l_i	累加概率	码字 w_i
u_1	0.20	2.34	3	0	000
u_2	0.19	2.41	3	0.2	001
u_3	0.18	2.48	3	0.39	011
u_4	0.17	2.56	3	0.57	100
u_5	0.15	2.74	3	0.74	101
u_6	0.10	3.34	4	0.89	1110
u_7	0.01	6.66	7	0.99	1111110

$$\bar{l} = 3.14 \text{ 码元/符号}, \quad \eta_c = 83.1\%$$

比较

$$\begin{bmatrix} U \\ P_U \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 \\ 0.20 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$$

	霍夫曼编码	费诺编码	香农编码
平均码长 (码元/符号)	2.72	2.74	3.14
编码效率	95.96%	95.3%	83.1%