基础信息论

香农码

华中科技大学电信学院

学习目标

- 编制香农码
- 评价香农码性能

香农码

设有离散无记忆信源，记作：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} a_1, & a_2, & \cdots, & a_i, & \cdots, & a_n \\ p(a_1), & p(a_2), & \cdots, & p(a_i), & \cdots, & p(a_n) \end{Bmatrix}$$

香农码的编码步骤如下：

(1) 将信源符号按概率从大到小依次排列。设排序后的消息分别记为 $a_1, a_2, \ldots, a_n$，对应概率为

$$p(a_1) \geq p(a_2) \geq \cdots \geq p(a_n)$$

(2) 令$p(a_0) = 0$，并用$p_a(a_j)$表示第j个码字之前的累加概率，即：

$$p_a(a_j) = \sum_{i=0}^{j-1} p(a_i) \quad j = 1, \ldots, n$$

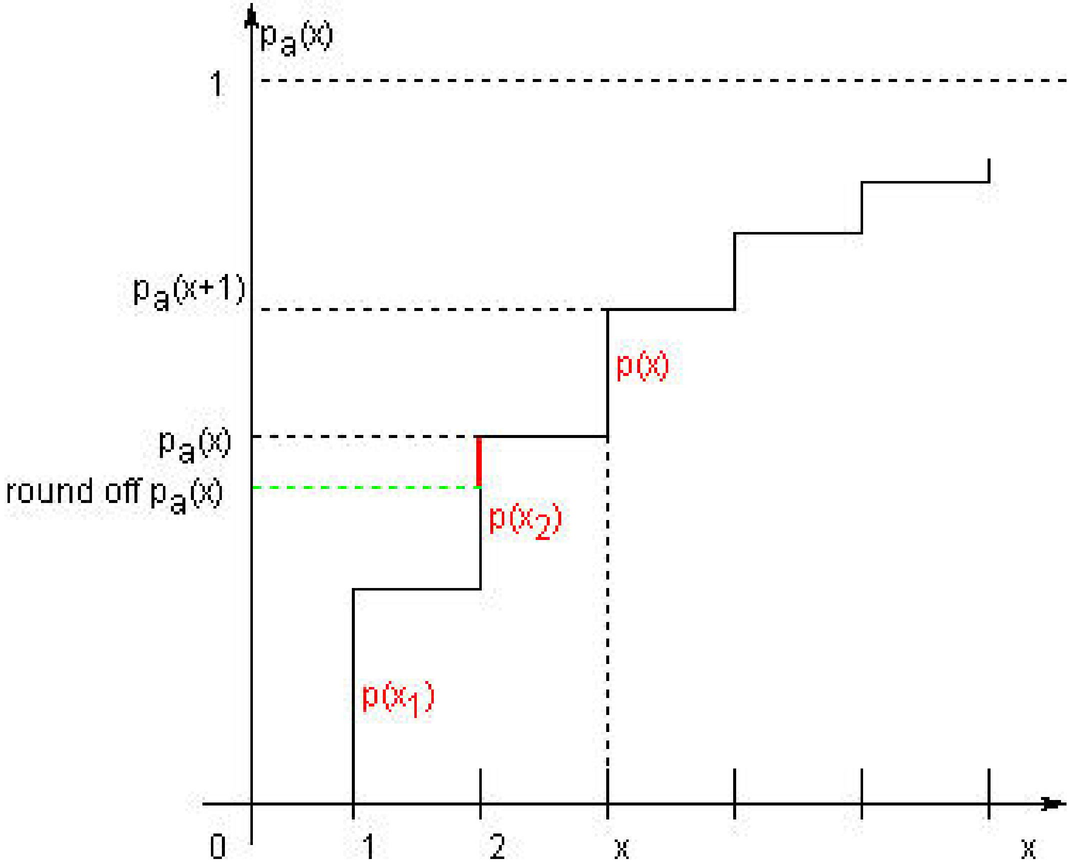
编码步骤

$$p_a(a_j)=\sum_{i=0}^{j-1}p(a_i)\quad j=1,\ldots,n$$

(3) 确定满足下列不等式的整数 k_i, 并令 k_i 为第 i 个码字的长度。

$$-\log_2 p(a_i)\le k_i<1-\log_2 p(a_i)$$

(4) 将累加概率 $p_a(a_j)$ 用二进制表示，去除小数点，根据码长并取小数点后共 k_i 位作为 a_i 的编码。



香农码-例1

例 对信源 $\begin{Bmatrix} a_1', & a_2', & a_3', & a_4', & a_5', & a_6' \\ 0.2, & 0.15, & 0.1, & 0.25, & 0.25, & 0.05 \end{Bmatrix}$ 编香农码。

解：(1) 按概率从大到小依次排列

$$\begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0.25, & 0.25, & 0.2, & 0.15, & 0.1, & 0.05 \end{Bmatrix}$$

(2) 计算累加概率 $p_a(a_j)$

$$\begin{cases} j=1 & a_1 & p_a(a_1)=0 \\ j=2 & a_2 & p_a(a_2)=0.25 \\ j=3 & a_3 & p_a(a_3)=0.5 \\ j=4 & a_4 & p_a(a_4)=0.7 \\ j=5 & a_5 & p_a(a_5)=0.85 \\ j=6 & a_6 & p_a(a_6)=0.95 \end{cases}$$

(3) 计算码字长度 k_i

大于等于自信息量最小整数

$$-\log_2 p(a_i) \le \underline{k_i} < \underline{1-\log_2 p(a_i)}$$

$$\begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0.25, & 0.25, & 0.2, & 0.15, & 0.1, & 0.05 \end{Bmatrix}$$

注意：是对 $p(a_i)$ 算自信息量，而不是对 $p_a(a_j)$算。

$$\begin{matrix} 2 & 2 & 3 & 3 & 4 & 5 \end{matrix} \leftarrow k_i$$

$-\log_2 p(a_1) \le k_1 < 1-\log_2 p(a_1)$	$2 \le k_1 < 3$	$k_1 = 2$
$-\log_2 p(a_2) \le k_2 < 1-\log_2 p(a_2)$	$2 \le k_2 < 3$	$k_2 = 2$
$-\log_2 p(a_3) \le k_3 < 1-\log_2 p(a_3)$	$2.32 \le k_3 < 3.32$	$k_3 = 3$
$-\log_2 p(a_4) \le k_4 < 1-\log_2 p(a_4)$	$2.74 \le k_4 < 3.74$	$k_4 = 3$
$-\log_2 p(a_5) \le k_5 < 1-\log_2 p(a_5)$	$3.32 \le k_5 < 4.32$	$k_5 = 4$
$-\log_2 p(a_6) \le k_6 < 1-\log_2 p(a_6)$	$4.32 \le k_6 < 5.32$	$k_6 = 5$

(3) 计算码字长度 k_i

大于等于自信息量最小整数

$$-\log_2 p(a_i) \le \underline{k_i} < \underline{1-\log_2 p(a_i)}$$

$$\begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0.25, & 0.25, & 0.2, & 0.15, & 0.1, & 0.05 \end{Bmatrix}$$

注意：是对 $p(a_i)$ 算自信息量，而不是对 $p_a(a_j)$算。

$$2 \quad 2 \quad 3 \quad 3 \quad 4 \quad 5 \quad \leftarrow k_i$$

$0.5 \le p(a_i) < 1$	$0 < -\log_2 p(a_i) \le 1$	$k_i = 1$
$0.25 \le p(a_i) < 0.5$	$1 < -\log_2 p(a_i) \le 2$	$k_i = 2$
$0.125 \le p(a_i) < 0.25$	$2 < -\log_2 p(a_i) \le 3$	$k_i = 3$
$0.0625 \le p(a_i) < 0.125$	$3 < -\log_2 p(a_i) \le 4$	$k_i = 4$
$0.03125 \le p(a_i) < 0.0625$	$4 < -\log_2 p(a_i) \le 5$	$k_i = 5$

(4) 根据累加概率 $p_a(a_j)$进行编码

$$p_a(a_j)=\begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0, & 0.25, & 0.5, & 0.7, & 0.85, & 0.95 \end{Bmatrix}$$

注意：此处是按$p_a(a_j)$进行编码，而不是按 $p(a_j)$编码。

十进制整数→二进制整数

	商	余数
2	36	
2	18	0
2	9	0
2	4	1
2	2	0
2	1	0
	0	1

$(36)_{10}=(100100)_2$

十进制小数→二进制小数

$0.6875\times 2=1.375$　　1

$0.375\times 2=0.75$　　0

$0.75\times 2=1.5$　　1

$0.5\times 2=1$　　1

$(0.6875)_{10}=(0.1011)_2$

$$p_a(a_j) = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0, & 0.25, & 0.5, & 0.7, & 0.85, & 0.95 \end{Bmatrix}$$

$$k_i = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 2, & 2, & 3, & 3, & 4, & 5 \end{Bmatrix}$$

a_1 **码字** 00 **注意补零**

a_2 **码字** 01

a_3 **码字** 100 **注意补零**

$0.25 \times 2 = 0.5 \quad 0$

$0.5 \times 2 = 1 \quad 1$

$0.5 \times 2 = 1 \quad 1$

$$p_a(a_j) = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0, & 0.25, & 0.5, & 0.7, & 0.85, & 0.95 \end{Bmatrix}$$

$$k_i = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 2, & 2, & 3, & 3, & 4, & 5 \end{Bmatrix}$$

a_1 **码字** 00 **注意补零**

a_2 **码字** 01

a_3 **码字** 100 **注意补零**

a_4 **码字** 101

$0.7 \times 2 = 1.4$ 1

$0.4 \times 2 = 0.8$ 0

$0.8 \times 2 = 1.6$ 1

$$p_a(a_j) = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0, & 0.25, & 0.5, & 0.7, & 0.85, & 0.95 \end{Bmatrix}$$

$$k_i = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 2, & 2, & 3, & 3, & 4, & 5 \end{Bmatrix}$$

a_1 **码字** 00 **注意补零**

a_2 **码字** 01

a_3 **码字** 100 **注意补零**

a_4 **码字** 101

a_5 **码字** 1101

$0.85 \times 2 = 1.7$ 1

$0.7 \times 2 = 1.4$ 1

$0.4 \times 2 = 0.8$ 0

$0.8 \times 2 = 1.6$ 1

$$p_a(a_j) = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0, & 0.25, & 0.5, & 0.7, & 0.85, & 0.95 \end{Bmatrix}$$

$$k_i = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 2, & 2, & 3, & 3, & 4, & 5 \end{Bmatrix}$$

a_1 码字 00 **注意补零**

a_2 码字 01

a_3 码字 100 **注意补零**

a_4 码字 101

a_5 码字 1101

a_6 码字 11110

$0.95 \times 2 = 1.9$ 1

$0.9 \times 2 = 1.8$ 1

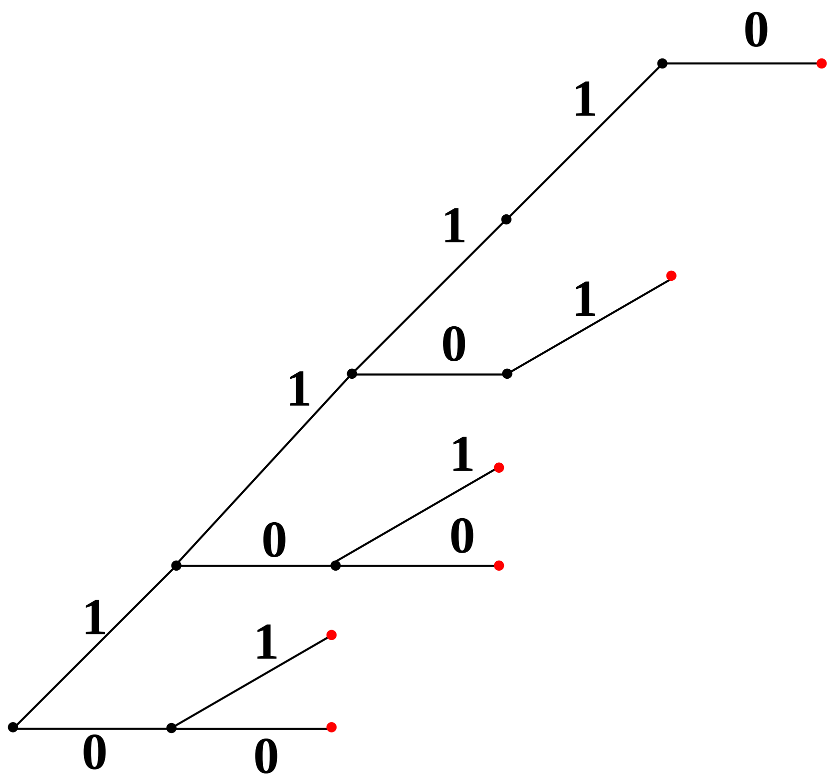
$0.8 \times 2 = 1.6$ 1

$0.6 \times 2 = 1.2$ 1

$0.2 \times 2 = 0.4$ 0

检验是否为即时码?

$$\begin{Bmatrix} X \\ P \\ \text{码字} \end{Bmatrix} = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0.25 & 0.25 & 0.2 & 0.15 & 0.1 & 0.05 \\ 00 & 01 & 100 & 101 & 1101 & 11110 \end{Bmatrix}$$



计算编码效率

$$\eta = \frac{\text{要求平均每个信源符号传递的信息量}}{\text{折算后，平均每个信源符号的最大可能载信量}}$$

$$\eta = \frac{H(X)}{\dfrac{\bar{L} \cdot \log m}{N}}$$

$$\begin{Bmatrix} X \\ P \\ \text{码字} \end{Bmatrix} = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0.25 & 0.25 & 0.2 & 0.15 & 0.1 & 0.05 \\ 00 & 01 & 100 & 101 & 1101 & 11110 \end{Bmatrix}$$

$$\eta = \frac{H(X)}{\frac{\bar{L}\cdot\log m}{N}} = \frac{-0.25\log 0.25 - 0.25\log 0.25 - \ldots - 0.05\log 0.05}{\frac{(0.25\times 2 + 0.25\times 2 + \ldots + 0.05\times 5)}{1}\cdot\log 2}$$

$$= \frac{2.42}{2.7} = 89.63\%$$

通过以上实例发现，香农码的效率并不算高。

当所有消息的概率为 $1/2, 1/4, 1/8, 1/16, \ldots$ **效率可达100%。**

香农码-例2

例　对 $\begin{Bmatrix} a_1', & a_2', & a_3', & a_4', & a_5' \\ \frac{1}{2}, & \frac{1}{8}, & \frac{1}{16}, & \frac{1}{16}, & \frac{1}{4} \end{Bmatrix}$ 编二进制香农码。

解：(1) 按概率从大到小依次排列

$$\begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5 \\ \frac{1}{2}, & \frac{1}{4}, & \frac{1}{8}, & \frac{1}{16}, & \frac{1}{16} \end{Bmatrix}$$

(2) 计算累加概率 $p_a(a_j)$

$$P_a(a_j) \quad \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5 \\ 0, & \frac{1}{2}, & \frac{3}{4}, & \frac{7}{8}, & \frac{15}{16} \end{Bmatrix}$$

(3) 计算各码字长度 k_i

$$\begin{Bmatrix} X \\ P \\ k_i \end{Bmatrix} = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, \\ \frac{1}{2}, & \frac{1}{4}, & \frac{1}{8}, & \frac{1}{16}, & \frac{1}{16}, \\ 1 & 2 & 3 & 4 & 4 \end{Bmatrix}$$

(4) 根据 $p_a(a_j)$**进行编码**

$$p_a(a_j) \quad \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5 \\ 0, & \frac{1}{2}, & \frac{3}{4}, & \frac{7}{8}, & \frac{15}{16} \end{Bmatrix}$$

码字	**0**	**10**	**110**	**1110**	**1111**

$$\begin{Bmatrix} X \\ P \\ \text{码字} \end{Bmatrix} = \begin{Bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, \\ \frac{1}{2}, & \frac{1}{4}, & \frac{1}{8}, & \frac{1}{16}, & \frac{1}{16}, \\ 0 & 10 & 110 & 1110 & 1111 \end{Bmatrix}$$

---

(5) 计算编码效率

$$\eta = \frac{H(X)}{\frac{\bar{L}\cdot \log m}{N}} = \frac{-\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - 2\cdot\frac{1}{16}\log\frac{1}{16}}{(\frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + 2\cdot\frac{1}{16}\cdot 4)\cdot\frac{\log 2}{1}}$$

$$= 100\%$$

洞察：香农码

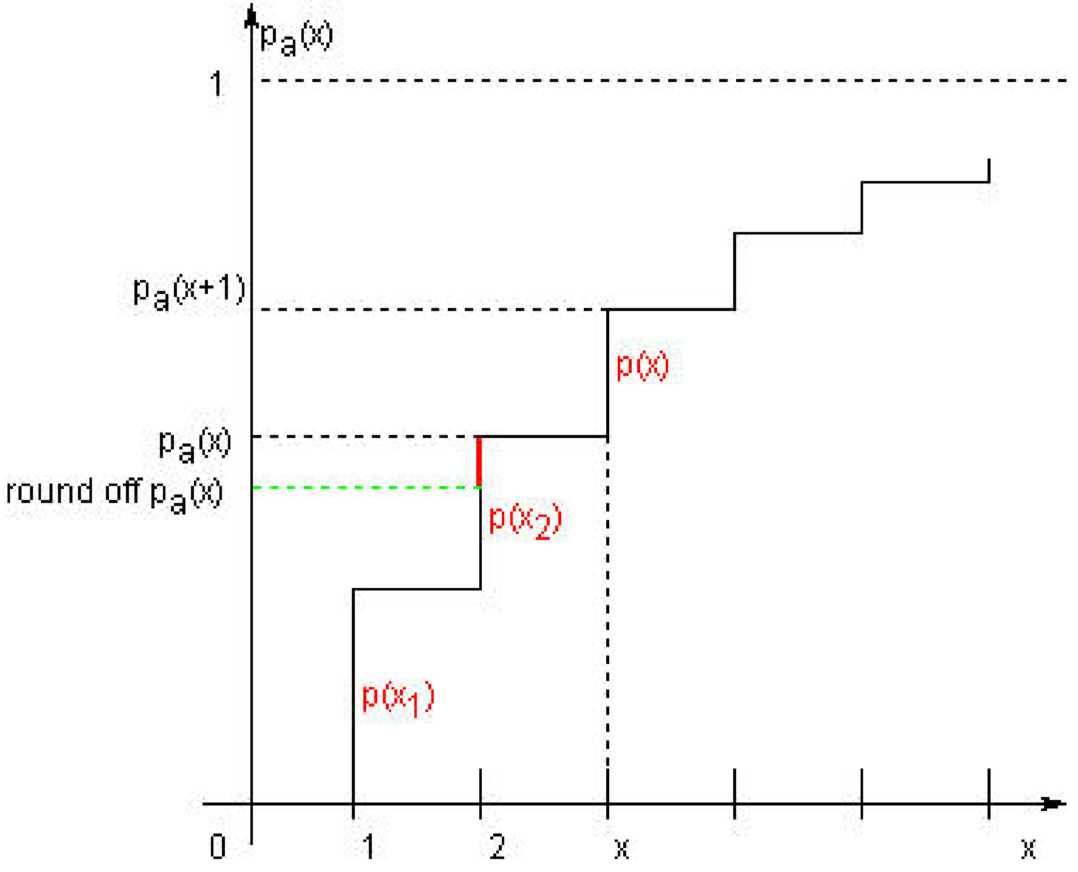
$$p_a(a_j) = \sum_{i=0}^{j-1} p(a_i) \quad j = 1, \ldots, n$$

$$-\log_2 p(a_i) \le k_i < 1 - \log_2 p(a_i)$$

$$\min_{k_1, k_2, \cdots, k_n} L = \sum_{i=1}^{n} p_i k_i$$

$$s.t. \sum_{i=1}^{n} m^{-k_i} \le 1$$

$$k_i^* = -\log_m p_i$$



谢谢！

黑晓军
华中科技大学
电子信息与通信学院
Email: heixj@hust.edu.cn
网址： http://eic.hust.edu.cn/aprofessor/heixiaojun