

3.1 离散信源的数学模型

信源是产生消息和消息序列的源头,它可以是人、生物、机器或其他事物。信源发出的消息有语音、图像、文字等。信源发出的是消息而不是信息,这是因为信息看不见摸不着,只能通过消息来研究它,消息上带有信息。

离散信源的数学模型是

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$$

其中,共有 n 个信源符号,

$$p(x_i) \geq 0 \quad (i=1,2,\cdots,n) \quad \text{且} \quad \sum_{i=1}^n p(x_i) = 1$$

在信源中,我们不再称 $x_1, x_2, \cdots, x_n$ 为事件,而称为符号。

离散信源的数学模型表明信源总共可以发出 n 个符号,一次(或每一时刻)只能发出这 n 个符号中的一个,发出的这个符号到底是这 n 个符号中的哪一个,这是随机的,但每个符号出现的概率已知。这样信源就能够源源不断地发出一个符号序列,这个符号序列记为 $x_{i1}, x_{i2}, x_{i3}, \cdots$,即第1次(时刻1)发出的符号是 x_{i1} ,第2次(时刻2)发出的符号是 x_{i2} ,第3次(时刻3)发出的符号是 $x_{i3}, \cdots$ 。在不引起混淆的情况下,序列 $x_{i1}, x_{i2}, x_{i3}, \cdots$ 常常简记为 $x_1 x_2 x_3 \cdots$ 。

【例 3-1】 某信源只发出两个符号“0”和“1”,两个符号出现的概率相等,该信源的数学模型是

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

这个信源在不断地按照相等的概率发出 0 和 1 两个符号,形成一个 0-1 序列。“001100010011001100110101”是其中一个可能的序列,其中有 13 个 0、11 个 1。

“010101110001010001100100101111010110100110010001001011001110”也是一个可能的序列,其中有 31 个 0、29 个 1。

【例 3-2】 老师在讲课时不断地发出一个一个的汉字,老师就是一个信源。该信源发

出的符号包括所有的汉字,每个汉字按照一定的概率出现,所有的已经发出的汉字构成了信源序列。

3.2 信源的分类

信源可以分为无记忆信源和有记忆信源。

3.2.1 无记忆信源

如果信源将要发出的符号,与已经发出的符号序列中的任何一个符号都没有关系,或者说已经出现的符号序列对将要出现的符号没有影响,那么这个信源是无记忆信源。

“已经出现的符号序列对将要出现的符号没有影响”用数学的方式表达出来就是

$$p(x_i | x_1 x_2 \cdots x_{i-1}) = p(x_i)$$

即已经出现的符号序列对将要出现的这个符号出现的概率没有影响,其中, $x_1 x_2 \cdots x_{i-1}$ 是信源已经发出的符号序列, x_i 是即将要出现的信源符号。

【例 3-3】 例 3-1 中的信源是一个无记忆信源。这是因为无论信源已经发出的符号是什么,它总是按照各为 1/2 的概率发出 0 或 1 中的一个。

3.2.2 有记忆信源

无记忆信源相对比较简单,但在日常生活和工程实践中并不常见。通常情况下,已经出现的符号对将要出现的符号有影响,这种信源称为有记忆信源。实际问题中的信源往往是有记忆的。下面看两个例子。

【例 3-4】 老师上课的例子中,老师是一个有记忆信源。这是因为汉语是有记忆的,如图 3-1 所示,出现第一个汉字“我”之后,“们”“要”“的”“把”“看”等汉字出现的概率上升,而“碗”“机”“水”“书”“框”等汉字出现的概率下降。即如果假设 $p(\text{们})=0.01$, $p(\text{碗})=0.01$,则出现“我”之后,“们”和“碗”出现的概率有可能变为 $p(\text{们}|\text{我})=0.05$, $p(\text{碗}|\text{我})=0.001$ 。

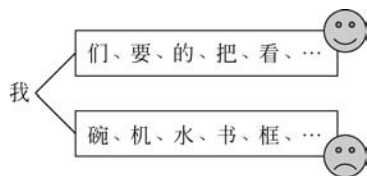


图 3-1 有记忆信源

有记忆信源又可以分为有限记忆信源和无限记忆信源。如果信源将要发出的消息符号只与前面 m 个符号有关,与更前面的符号无关,即

$$p(x_i | x_{i-1} x_{i-2} \cdots x_{i-m} x_{i-m-1} \cdots x_1) = p(x_i | x_{i-1} x_{i-2} \cdots x_{i-m})$$

则该信源为有限记忆信源,即记忆长度有限,为 m 。

如果信源将要发出的消息符号与前面已经发出的所有符号都有关,即

$$p(x_i | x_{i-1} x_{i-2} x_{i-3} \cdots)$$

则该信源为无限记忆信源,即记忆长度无限。

工程实践中遇到的有记忆信源一般都是有限记忆信源。

【例 3-5】 试判断图 3-2 中两个信源的记忆特性,是无记忆信源还是有记忆信源。

无论哪一个信源序列,其中黑色格子和白色格子出现的概率都相等,即信源模型均为



图 3-2 两个信源序列

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} \text{黑} & \text{白} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

对于信源序列 1, 条件概率为

$$p(\text{黑} | \text{黑}) = p(\text{黑} | \text{白}) = p(\text{黑}) = \frac{1}{2}$$

$$p(\text{白} | \text{黑}) = p(\text{白} | \text{白}) = p(\text{白}) = \frac{1}{2}$$

这说明, 无论前一个格子是黑色还是白色, 下一个格子是黑色或者白色的概率均不发生变化, 总是 $\frac{1}{2}$, 因此这是一个无记忆信源。

对于信源序列 2, 条件概率为

$$p(\text{黑} | \text{黑}) = \frac{1}{5}, \quad p(\text{黑} | \text{白}) = \frac{4}{5}$$

$$p(\text{白} | \text{黑}) = \frac{4}{5}, \quad p(\text{白} | \text{白}) = \frac{1}{5}$$

这说明, 前一个格子的颜色对下一个格子的颜色有影响, 由 $\frac{1}{2}$ 变为 $\frac{1}{5}$ 或者 $\frac{4}{5}$, 因此这是一个有记忆信源。

3.3 离散无记忆信源

3.3.1 离散无记忆信源及其熵

离散无记忆信源是最简单, 也是最基本的一类信源。

【定义 3-1】 设信源 X 的符号集为 $\{x_1, x_2, \dots, x_n\}$, n 为信源能发出的符号个数, 每个符号发生的概率为 $p(x_i)$, $i=1, 2, \dots, n$, 这些符号彼此互不相关, 且有

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix} \quad (3-1)$$

其中,

$$p(x_i) \geq 0 \quad (i=1, 2, \dots, n) \quad \text{且} \quad \sum_{i=1}^n p(x_i) = 1$$

则称 X 为离散无记忆信源。

【定义 3-2】信源 X 中消息符号 x_i 的自信息量定义为

$$I(x_i) = -\log p(x_i) \quad (3-2)$$

【定义 3-3】信源 X 的平均自信息量,即信源熵定义为

$$H(X) = -\sum_{i=1}^n p(x_i) \log p(x_i) \quad (3-3)$$

信源中一个消息符号的自信息量表示该符号带有的信息量的多少,而信源熵是所有符号带有的信息量的平均值,因此信源熵表示的是平均每个信源符号带有的信息量的多少,所以信源熵的单位为比特/符号。

【例 3-6】计算机中常见的信源是二元信源,二元信源可以描述为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix} \quad 0 \leq p \leq 1$$

则二元信源的熵为

$$H(X) = -p \log p - (1-p) \log (1-p)$$

可见,二元信源的熵是 p 的函数,函数图形如图 3-3 所示。

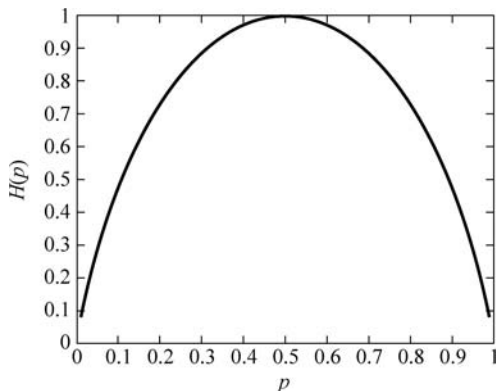


图 3-3 二元信源的熵

例 3-1 中描述的信源的熵为

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1 \text{ 比特 / 符号}$$

即在该信源中,平均每个信源符号上带有 1 比特的信息量。

再如:

【例 3-7】有一个三元信源

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

则该信源的熵为

$$H(X) = -\sum_{i=1}^3 p(x_i) \log p(x_i) = -\frac{1}{2} \log \frac{1}{2} - 2 \times \frac{1}{4} \log \frac{1}{4} = 1.5 \text{ 比特 / 符号}$$

【例 3-8】请从信息论的角度说明,为什么玩扑克牌的时候需要先洗牌?

是否洗牌,得到的信源模型是不同的。假设不考虑大小王,且自己先摸对方后摸,摸到

的第一张牌是 A。

洗牌使得牌的分布是均匀的,则对方也摸到 A 的概率是 $\frac{3}{51}$,摸到其他牌的概率各为 $\frac{4}{51}$,则此时猜测对方的牌,所需的信息量是

$$H_{\text{洗牌}} = -\frac{3}{51} \log \frac{3}{51} - 12 \times \frac{4}{51} \log \frac{4}{51} = 3.6968 \text{ 比特}$$

如果不洗牌,则容易出现几张同样大小的牌连续出现的情况。此时自己摸到 A 之后,假设对方也摸到 A 的概率上升为 $\frac{39}{51}$,摸到其他牌的概率各为 $\frac{1}{51}$,则此时猜测对方的牌,所需的信息量是

$$H_{\text{不洗牌}} = -\frac{39}{51} \log \frac{39}{51} - 12 \times \frac{1}{51} \log \frac{1}{51} = 1.6306 \text{ 比特}$$

$3.6968 > 1.6306$,所以洗过牌之后,在摸牌环节猜测对方的牌将更加困难。

3.3.2 离散无记忆信源的扩展信源及其熵

在 3.3.1 节中,我们给出的信源模型是 $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$,这其实

是在一个符号一个符号地研究信源,但有时这样不能满足实际应用的需要,例如,汉语中更多地考察的是句子,而不是汉字;英语中更多地考察的是单词,而不是字母;图像中更多地考察的是整幅图像,而不是单个像素。因此有必要研究 N 次扩展信源。

【例 3-9】 N 次扩展信源的例子。

(1) 二进制信源: $X = \{00, 01, 10, 11\}$,则 $N=2$

(2) 汉语: $X = \{\text{我们在上课,张三睡着了}, \dots\}$,则 $N=5$

(3) 英语: $X = \{\text{the, car, ear, she, you}, \dots\}$,则 $N=3$

为了给出 N 次扩展信源的数学定义,先来看看离散无记忆二进制信源 $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$ 的

二次扩展信源和三次扩展信源。

1. 二次扩展信源

X 的二次扩展信源记作 X^2 ,其中包含 4 个长度为 2 的序列 $\{00, 01, 10, 11\}$,这 4 个序列分别记为 a_1, a_2, a_3, a_4 。由于原始信源 X 是无记忆的,因此每个序列出现的概率为序列中两个符号出现概率的乘积,因此二次扩展信源为

$$\begin{bmatrix} X^2 \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ p(a_1) & p(a_2) & p(a_3) & p(a_4) \end{bmatrix} = \begin{bmatrix} 00 & 01 & 10 & 11 \\ p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{bmatrix}$$

其中 a_i 是原信源 X 的一个长度为 2 的序列,又是二次扩展信源 X^2 的一个信源符号。

2. 三次扩展信源

X 的三次扩展信源记作 X^3 ,其中包含 8 个长度为 3 的序列 $\{000, 001, 010, \dots, 111\}$,这 8 个序列分别记为 $a_1, a_2, \dots, a_8$,每个序列出现的概率为序列中三个符号出现概率的乘积,因此三次扩展信源为

$$\begin{bmatrix} X^3 \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_8 \\ p(a_1) & p(a_2) & \cdots & p(a_8) \end{bmatrix} = \begin{bmatrix} 000 & 001 & \cdots & 111 \\ p^3 & p^2(1-p) & \cdots & (1-p)^3 \end{bmatrix}$$

其中 a_i 是原信源 X 的一个长度为 3 的序列, 又是三次扩展信源 X^3 的一个信源符号。

3. N 次扩展信源

看过前面的例子, 更容易理解 N 次扩展信源的概念。

【定义 3-4】 离散无记忆信源 $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$ 的 N 次扩展信源

X^N 包含全部 n^N 个长度为 N 的序列, 每一个序列出现的概率为序列中各个符号出现的概率的乘积, 即

$$\begin{bmatrix} X^N \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n^N} \\ p(a_1) & p(a_2) & \cdots & p(a_{n^N}) \end{bmatrix} \quad (3-4)$$

其中,

$$X^N = X \times X \times \cdots \times X, \quad a_i = x_{i_1} x_{i_2} \cdots x_{i_N}$$

$$p(a_i) = \prod_{k=1}^N p(x_{i_k}) = \prod_{k=1}^N p_{i_k}$$

【定理 3-1】 离散无记忆信源 X 的 N 次扩展信源 X^N 的熵等于信源 X 的熵的 N 倍, 即

$$H(X^N) = NH(X) \quad (3-5)$$

证明:

$$\begin{aligned} H(X^N) &= - \sum_{X^N} p(a_i) \log p(a_i) = - \sum_{X^N} p(a_i) \log \prod_{k=1}^N p_{i_k} = - \sum_{X^N} p(a_i) \sum_{k=1}^N \log p_{i_k} \\ &= - \sum_{k=1}^N \sum_{X^N} p(a_i) \log p_{i_k} = - \sum_{k=1}^N \sum_{X^N} p_{i_1} \cdots p_{i_k} \cdots p_{i_N} \log p_{i_k} \\ &= - \sum_{k=1}^N \sum_{X^N} p_{i_1} \cdots (p_{i_k} \log p_{i_k}) \cdots p_{i_N} = - \sum_{k=1}^N \left(\sum_{i_k=1}^n p_{i_k} \log p_{i_k} \sum_{i_1=1}^n p_{i_1} \cdots \sum_{i_N=1}^n p_{i_N} \right) \\ &= - \sum_{k=1}^N \sum_{i_k=1}^n p_{i_k} \log p_{i_k} = \sum_{k=1}^N \left(- \sum_{i_k=1}^n p_{i_k} \log p_{i_k} \right) = \sum_{k=1}^N H(X) = NH(X) \end{aligned}$$

这说明扩展信源中平均一个符号(即长度为 N 的序列)上带有的信息量是原信源中平均一个符号上带有的信息量的 N 倍。

【例 3-10】 接例 3-7, 求 $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$ 的二次扩展信源的熵。

方法一: 二次扩展信源为

$$\begin{bmatrix} X^2 \\ P \end{bmatrix} = \begin{bmatrix} x_1 x_1 & x_1 x_2 & x_1 x_3 & x_2 x_1 & x_2 x_2 & x_2 x_3 & x_3 x_1 & x_3 x_2 & x_3 x_3 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} \end{bmatrix}$$

则

$$H(X^2) = \frac{1}{4} \log 4 + \frac{1}{8} \log 8 + \cdots + \frac{1}{16} \log 16 = 3 \text{ 比特 / 符号}$$

方法二：利用定理 3-1

$$H(X^2) = 2H(X) = 2 \times 1.5 = 3 \text{ 比特 / 符号}$$

可见,两种方法求出的熵是相等的。

3.4 马尔可夫信源(有限记忆信源)

3.4.1 马尔可夫信源的定义

在 3.2.2 节中我们说过,工程实践中遇到的有记忆信源基本上都属于有限记忆信源,这种信源有一个名字,称为马尔可夫信源。

【定义 3-5】 在信源 X 中,如果将要出现的符号仅与刚刚出现的 m 个符号有关,与更早出现的符号无关,则称 X 为 m 阶马尔可夫信源,即

$$p(x_i | x_{i-1}x_{i-2}\cdots x_{i-m}\cdots x_1) = p(x_i | x_{i-1}x_{i-2}\cdots x_{i-m}) \quad (3-6)$$

【例 3-11】 信源 $X = \{0, 1, 2, 3\}$,信源将要发出的符号是前面两个符号的和对 4 的余数,即

$$x_i \equiv (x_{i-1} + x_{i-2}) \bmod 4$$

显然这是一个 2 阶马尔可夫信源,因为将要出现的符号只与前面的两个符号有关。

$$\begin{cases} p(x_i | x_{i-1}x_{i-2}) = 1, & (x_i \equiv (x_{i-1} + x_{i-2}) \bmod 4) \\ p(x_i | x_{i-1}x_{i-2}) = 0, & \text{其他} \end{cases}$$

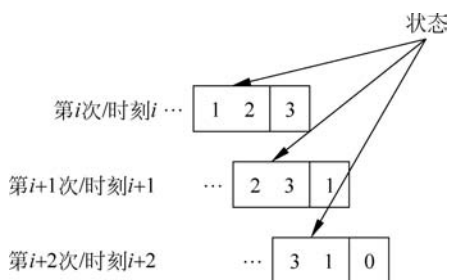


图 3-4 马尔可夫信源

如图 3-4 所示,假设时刻 i 信源已经发出的两个符号是 1、2,则时刻 i 或者说第 i 次发出的符号是

$$(1 + 2) \bmod 4 \equiv 3$$

时刻 $i+1$ 发出的符号与 1 已经无关,只与 2 和 3 有关,时刻 $i+1$ 发出的符号是

$$(2 + 3) \bmod 4 \equiv 1$$

以此类推,信源发出序列是 $\cdots 12310112 \cdots$ 。

在例 3-11 中,每一个将要发出的符号只与前面的两个符号相关,将这两个符号连在一起,看成一个整体,我们把它称为“状态”。时刻 i 发出的符号与状态“12”有关,发出符号 3 之后,状态变为“23”,时刻 $i+1$ 发出的符号与该状态“23”有关,……。

“状态”是马尔可夫信源中的一个重要概念, m 阶马尔可夫信源的状态由 m 个信源符号构成。如果信源有 n 个符号,则 m 阶马尔可夫信源共有 n^m 种状态。假设当前所处的状态为 $S_k = x_{i-1}x_{i-2}\cdots x_{i-m}$,则发出符号 x_i 之后,状态变为 $S_l = x_ix_{i-1}\cdots x_{i-m+1}$ 。因此,式(3-6)所定义的马尔可夫信源中符号对符号序列的依赖关系就变为状态对状态的依赖关系,即

$$\begin{aligned} p(x_i | x_{i-1}x_{i-2}\cdots x_{i-m+1}x_{i-m}) &= p(x_ix_{i-1}\cdots x_{i-m+1} | x_{i-1}x_{i-2}\cdots x_{i-m+1}x_{i-m}) \\ &= p(S_l | S_k) \end{aligned} \quad (3-7)$$

3.4.2 有限状态马尔可夫链

马尔可夫链并不是信源,它体现的是一种状态和状态之间的“一环扣一环”的性质,因此称为“链”。之所以要在介绍信源的时候介绍马尔可夫链,是因为马尔可夫信源是具有马尔可夫链性质的信源,这点将在 3.4.3 节中讲解。

【定义 3-6】 一个状态序列: $S_1, S_2, \dots, S_l, \dots$, 若满足以下条件:

- (1) 有限性: 可能的状态数目 $J < \infty$, 即只有有限个可能的状态;
- (2) 马氏性: 系统将要达到的状态, 只与当前的状态有关, 与更早的状态无关, 即

$$p(S_l | S_{l-1} S_{l-2} \dots S_1) = p(S_l | S_{l-1}) \quad (3-8)$$

则称此随机状态序列为有限状态马尔可夫链。

由式(3-8)能够看出马尔可夫链中将要出现的状态只与它前面的一个状态有关系,体现了状态和状态之间的“一环扣一环”的性质。

描述马尔可夫链的数学工具是状态转移矩阵和状态转移图。已知系统当前处于状态 S_k , 则系统将要到达的状态为 S_l 的概率为

$$p_{kl} = p(S_l | S_k)$$

由于所有可能的状态有 J 个, 因此这样的概率共有 $J \times J$ 个, 所有这些概率组成一个矩阵

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1J} \\ p_{21} & p_{22} & \dots & p_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ p_{J1} & p_{J2} & \dots & p_{JJ} \end{bmatrix} \quad (3-9)$$

由于 p_{kl} 表示由状态 S_k 转移为状态 S_l 的概率, 因此矩阵 $\mathbf{P}$ 称为状态转移矩阵。

状态转移矩阵中每一行元素的和均为 1, 这是因为第 i 行元素的和为

$$\sum_{j=1}^J p_{ij} = \sum_{j=1}^J p(S_j | S_i) = p(S_1, S_2, \dots, S_J | S_i) = 1$$

它的含义是, 当前状态 S_k 必然转移为所有可能状态 $\{S_1, S_2, \dots, S_J\}$ 中的一个。

概率 p_{kl} 还可以用图形化的方式表示, 如图 3-5 所示, 该图称为状态转移图, 表示由状态 S_k 转移为状态 S_l 的概率是 p_{kl} 。

状态转移矩阵 $\mathbf{P}$ 中所有元素都可以表示为图 3-5 的样子, 所有状态之间的转移关系连在一起, 构成了整个系统的状态转移图。

【例 3-12】 一个系统有 2 个状态 $\{S_1, S_2\}$, 状态转移矩阵为

$$\mathbf{P} = \begin{bmatrix} 0.25 & 0.75 \\ 0.5 & 0.5 \end{bmatrix}$$

则状态转移图(见图 3-6)为



图 3-5 状态转移图

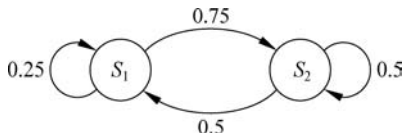


图 3-6 例 3-12 的状态转移图

可见状态转移矩阵和状态转移图之间有一一对应的关系,根据状态转移矩阵可以画出状态转移图,根据状态转移图也能写出状态转移矩阵。

3.4.3 马尔可夫信源的马尔可夫链性质

在式(3-7)中,已经将 m 阶马尔可夫信源中 1 个符号对 m 个符号的依赖关系转换为状态和状态之间的“一环扣一环”的关系,因此马尔可夫信源是具有马尔可夫链性质的信源,这样就能用描述马尔可夫链的工具——状态转移矩阵和状态转移图,来描述马尔可夫信源,而且还可以从另一个角度来定义马尔可夫信源。

【定义 3-7】 若信源输出的符号和状态满足下列条件,则称此信源为 m 阶马尔可夫信源。

(1) 某一时刻,将要出现的信源符号只与当时信源所处的状态有关,与以前的状态无关,即

$$p(x_i | x_{i-1}x_{i-2}\cdots x_{i-m}) = p(x_i | S_k) \quad (3-10)$$

(2) 信源的状态只由当前输出符号和前一时刻信源的状态唯一确定,即

$$p(S_l | x_i, S_k) = \begin{cases} 0, & S_l \neq x_i x_{i-1} \cdots x_{i-m+1} \\ 1, & S_l = x_i x_{i-1} \cdots x_{i-m+1} \end{cases} \quad (3-11)$$

该定义与定义 3-5 是一致的,只不过用了另外一种描述方法。

【例 3-13】 接例 3-11,因为 2 阶马尔可夫信源 $X = \{0, 1, 2, 3\}$ 有 4 个符号,因此该信源共有

$$n^m = 4^2 = 16$$

个状态,其状态转移矩阵为

$$P = \begin{matrix} & \begin{matrix} 00 & 01 & 02 & 03 & 10 & 11 & 12 & 13 & 20 & 21 & 22 & 23 & 30 & 31 & 32 & 33 \end{matrix} \\ \begin{matrix} 00 \\ 01 \\ 02 \\ 03 \\ 10 \\ 11 \\ 12 \\ 13 \\ 20 \\ 21 \\ 22 \\ 23 \\ 30 \\ 31 \\ 32 \\ 33 \end{matrix} & \left[\begin{array}{cccccccccccccccc} 1 & & & & & & & & & & & & & & & \\ & & & & & 1 & & & & & & & & & & \\ & & & & & & & & & & 1 & & & & & \\ & & & & & & & & & & & & & & 1 & \\ & 1 & & & & & & & & & & & & & & \\ & & & & & & 1 & & & & & & & & & \\ & & & & & & & & & & 1 & & & & & \\ & & & & & & & & & & & & 1 & & & \\ & & 1 & & & & & & & & & & & & & \\ & & & & & & & & 1 & & & & & & & \\ & & & & & & & & & 1 & & & & & & \\ & & & & & & & & & & & & 1 & & & \\ & & & & 1 & & & & & & & & & & & \\ & & & & & 1 & & & & & & & & & & \\ & & & & & & & & & 1 & & & & & & \\ & & & & & & & & & & & & & 1 & & \end{array} \right] \end{matrix}$$

状态转移图(见图 3-7)为

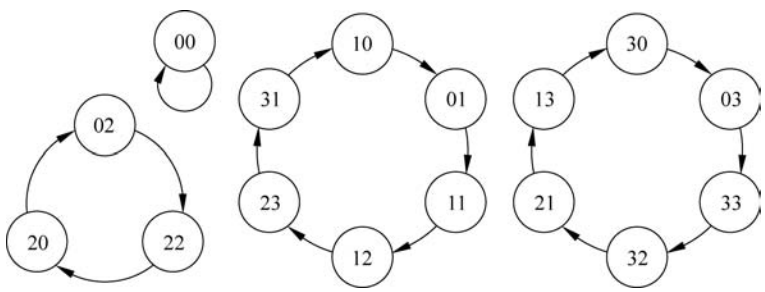


图 3-7 例 3-13 的状态转移图

3.4.4 马尔可夫信源的熵

1. 平稳马尔可夫信源

有的时候,马尔可夫信源输出的状态序列与系统的初始状态有关系,这点可以从下面的例 3-14 看出。

【例 3-14】 接例 3-13,从表 3-1 能够看出,信源的初始状态不同,产生的状态序列是不一样的,这直接导致状态分布和信源分布发生变化,从而使得信源熵发生变化。

表 3-1 初始状态对信源熵的影响

初始状态	状态序列	状态分布	符号序列	信源分布	信源熵 (比特/符号)
00	00	$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} 00 & \text{其他} \\ 1 & 0 \end{bmatrix}$	00000000	$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 0 & 0 \end{bmatrix}$	$H(X)=0$
02	02 22 20	$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} 02/22/20 & \text{其他} \\ \frac{1}{3} & 0 \end{bmatrix}$	022022 022...	$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \end{bmatrix}$	$H(X)=0.9183$
10	10 01 11 12 23 31	$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} 10/01/11 \\ /12/23/31 & \text{其他} \\ \frac{1}{6} & 0 \end{bmatrix}$	101123 101123 ...	$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$	$H(X)=1.7925$
30	30 03 33 32 21 13	$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} 30/03/33 \\ /32/21/13 & \text{其他} \\ \frac{1}{6} & 0 \end{bmatrix}$	303321 303321 ...	$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$	$H(X)=1.7925$

这个例子说明,当信源输出的状态序列受初始状态影响的时候,无法给出信源熵的一个确定的值。因此本课程中,我们只研究状态序列不受初始状态影响的情况。这种马尔可夫信源称为平稳马尔可夫信源。

平稳马尔可夫信源的含义包括两点:

(1) 经过足够长的时间之后,信源的 n^m 个状态出现的概率逐渐稳定下来,不再发生变化。

(2) 状态的稳定分布与初始状态无关,即无论信源一开始处在什么状态,最终都将达到同样的稳定分布。

为了求得平稳马尔可夫信源的熵,首先要判断一个马尔可夫信源是否平稳。我们不加证明给出如下定理。

【定理 3-2】 遍历的马尔可夫信源是平稳马尔可夫信源。

所谓遍历的马尔可夫信源指的是状态转移图中各个状态是互相可达的,即从任何一个状态出发都可以到达其他任何状态,即能将所有状态遍历一遍。像图 3-7 所示的信源就不是遍历的,因为有些状态之间不连通。

在 3.4.2 节介绍马尔可夫链的时候说过,状态转移图和状态转移矩阵是等价的,我们已经知道了如何根据状态转移图判断信源是否为遍历的马尔可夫信源,那如何从状态转移矩阵的角度判断信源是否为遍历的马尔可夫信源呢? 这需要将状态转移矩阵做进一步的划分。

式(3-9)描述的状态转移矩阵表示一个状态经过 1 步转移到另外一个状态的矩阵,即转移 1 次的概率,因此称为一步转移矩阵。如果要描述一个状态经过多次转移,转移到另外一个状态的概率,则需要多步转移矩阵 $\mathbf{P}^{(r)}$,

$$\mathbf{P}^{(r)} = [p_{ij}^{(r)}], \quad i, j = 1, 2, \dots, n^m \quad (3-12)$$

其中, $p_{ij}^{(r)}$ 表示从状态 S_i 经过 r 次转移,转移到状态 S_j 的概率。

多步转移矩阵之间存在下述切普曼-柯尔莫格洛夫方程

$$\mathbf{P}^{(r+t)} = \mathbf{P}^{(r)} \mathbf{P}^{(t)} \quad (r, t \geq 1) \quad (3-13)$$

从式(3-13)不难得到如下结论

$$\mathbf{P}^{(r)} = \mathbf{P}^r \quad (3-14)$$

即 r 步转移矩阵等于一步转移矩阵的 r 次方。

有了多步转移矩阵的概念,可以从状态转移矩阵的角度给出遍历的马尔可夫信源,或者说平稳马尔可夫信源的判定定理。

【定理 3-3】 设 $\mathbf{P}$ 为马尔可夫信源的一步状态转移矩阵,该信源为平稳马尔可夫信源的充要条件是存在一个正整数 N ,使矩阵 $\mathbf{P}^N$ 中的所有元素均大于 0。

定理 3-3 中, $\mathbf{P}^N$ 的含义是最多经过 N 步,“所有元素均大于 0”的含义是各个状态是互相可达的。因此定理 3-3 与定理 3-2 本质上是一致的,区别在于定理 3-3 是从状态转移矩阵角度给出结论,定理 3-2 是从状态转移图角度给出结论。通常一步转移矩阵简称为转移矩阵。

【例 3-15】 设有一个二进制二阶马尔可夫信源,其信源符号集为 $\{0, 1\}$,条件概率为

$$p(0 | 00) = p(1 | 11) = 0.8$$

$$p(1 | 00) = p(0 | 11) = 0.2$$

$$p(0 | 01) = p(0 | 10) = p(1 | 01) = p(1 | 10) = 0.5$$

试判断这个信源是否为平稳马尔可夫信源。

解: 这是一个二进制二阶马尔可夫信源,因此共有 $n^m = 2^2 = 4$ 个状态: 00、01、10、11。状态转移概率为

$$p(00 | 00) = p(0 | 00) = 0.8$$

$$p(10 | 01) = p(0 | 01) = 0.5$$

$$p(01 | 00) = p(1 | 00) = 0.2$$

$$p(11 | 01) = p(1 | 01) = 0.5$$

$$\begin{aligned} p(00 | 10) &= p(0 | 10) = 0.5 & p(10 | 11) &= p(0 | 11) = 0.2 \\ p(01 | 10) &= p(1 | 10) = 0.5 & p(11 | 11) &= p(1 | 11) = 0.8 \end{aligned}$$

其余的为 0。因此状态转移矩阵为

$$\mathbf{P} = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix}$$

存在正整数 $N=2$, 使矩阵 $\mathbf{P}^N$ 中的所有元素均大于 0,

$$\mathbf{P}^2 = \begin{bmatrix} 0.64 & 0.16 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.1 & 0.4 \\ 0.4 & 0.1 & 0.25 & 0.25 \\ 0.1 & 0.1 & 0.16 & 0.64 \end{bmatrix}$$

因此该信源为平稳马尔可夫信源。

从状态转移图, 可以得到同样的结论。从图 3-8 能够看出, 从 4 个状态中的任何一个出发, 最多经过 2 步, 都能够到达其他任何一个状态, 即能够遍历所有状态, 因此该信源为平稳马尔可夫信源。

2. 平稳马尔可夫信源的熵

平稳马尔可夫信源中 n^m 个状态出现的概率能够逐渐稳定下来, 不再发生变化。设达到稳定分布之后, 各个状态的概率为 $p(S_1), p(S_2), \dots, p(S_{n^m})$, 那马

尔可夫信源的熵是 $-\sum_{i=1}^{n^m} p(S_i) \log p(S_i)$ 吗? 当然不是。

这是因为信源的熵指的是信源符号带有的信息量的平均值, 而不是状态带有的信息量的平均值。如何求得信源符号带有的信息量的平均值呢? 是否像离散无记忆信源那样, 求出所有符号出现的概率, 套入熵的公式就可以了呢? 也不是。

这是因为对马尔可夫信源这样的有记忆信源, 信源符号出现的概率与已经出现的符号有关。因此 m 阶平稳马尔可夫信源的熵是在知道了已经出现的 m 个符号的条件下, 信源符号带有的信息量的平均值, 换句话说, 平稳马尔可夫信源的熵实际上是一个条件熵。

$$H_{m+1}(X) = H(X_{m+1} | X_1 X_2 \cdots X_m) \quad (3-15)$$

将该条件熵记作 $H_{m+1}(X)$, 其中 $X, X_1, \dots, X_m, X_{m+1}$ 均来自于集合 X 。

根据条件熵的定义和马尔可夫信源中对状态的定义, 有

$$\begin{aligned} H_{m+1}(X) &= H(X | X_1 X_2 \cdots X_m) \\ &= H(X | S) \\ &= - \sum_{S_j} \sum_i p(x_i, S_j) \log p(x_i | S_j) \\ &= - \sum_{S_j} \sum_i p(S_j) p(x_i | S_j) \log p(x_i | S_j) \end{aligned} \quad (3-16)$$

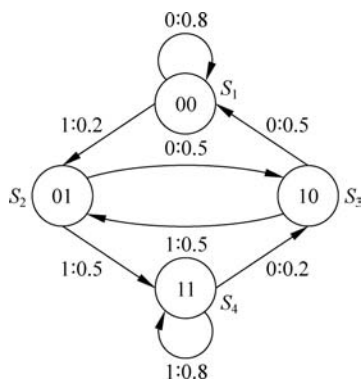


图 3-8 例 3-15 的状态转移图

在式(3-16)中, $p(x_i | S_j)$ 是用来描述马尔可夫信源的转移概率, 是必然要给出的。这样在式(3-16)中, 只有 $p(S_j)$ 是未知量。下面给出求得各个 $p(S_j)$ 的方法。

因为 $p(S_j)$ 表示的是平稳之后各个状态出现的概率, 因此马尔可夫信源再输出一个信源符号之后, 即状态转移一次之后, $p(S_j)$ 是不变的。假设 $\mathbf{S} = [p(S_1) \ p(S_2) \ \cdots \ p(S_{n_m})]$, 因此有

$$\mathbf{SP} = \mathbf{S} \quad (3-17)$$

其中 $\mathbf{P}$ 是状态转移矩阵。而且有

$$p(S_1) + p(S_2) + \cdots + p(S_{n_m}) = 1 \quad (3-18)$$

根据式(3-17)和式(3-18)可以求得各个 $p(S_j)$ 。

【例 3-16】 接例 3-15。试求该平稳马尔可夫信源的熵。

解: 设 $\mathbf{S} = [p(00) \ p(01) \ p(10) \ p(11)]$, 有

$$\left\{ \begin{array}{l} \mathbf{SP} = \mathbf{S} \\ p(00) + p(01) + p(10) + p(11) = 1 \end{array} \right. = \mathbf{S}$$

求得 $p(00) = p(11) = \frac{5}{14}$, $p(01) = p(10) = \frac{1}{7}$, 因此信源熵

$$\begin{aligned} H_{m+1}(X) &= - \sum_{S_j} \sum_i p(S_j) p(x_i | S_j) \log p(x_i | S_j) \\ &= -2 \left[\frac{5}{14} (0.8 \log 0.8 + 0.2 \log 0.2) + \frac{1}{7} (0.5 \log 0.5 + 0.5 \log 0.5) \right] \\ &= 0.8014 \text{ 比特 / 符号} \end{aligned}$$

总结求马尔可夫信源熵的过程, 分为三个步骤:

- (1) 判断是否为平稳马尔可夫信源。
- (2) 对平稳马尔可夫信源, 求出状态的平稳分布。
- (3) 根据式(3-16)求得马尔可夫信源的熵。

求平稳马尔可夫信源的熵, 有很多应用领域, 下面给出两个例子。

【例 3-17】 俄裔美国经济学家列昂惕夫提出了“投入-产出”模型, 并因此获得了 1973 年诺贝尔经济学奖。假设一个经济体系由煤炭、电力和钢铁三个部门组成, 各部门之间的分配如表 3-2 所示。

表 3-2 一个简单的“投入-产出”模型

产 出 部 门	采 购 部 门		
	煤 炭	电 力	钢 铁
煤炭	0.0	0.6	0.4
电力	0.4	0.1	0.5
钢铁	0.6	0.2	0.2

比如表中的第 2 行, 它的意思是电力部门的总产出有 40% 用来购买煤炭、10% 用来满足自己的用电需求、50% 用来购买钢铁。因为所有的产出都分配出去了, 所以每一行的百分比之和为 1。求出平衡价格, 使每个部门的收支平衡。

答：这是一个马尔可夫链，其状态转移矩阵为

$$\mathbf{P} = \begin{pmatrix} 0.0 & 0.6 & 0.4 \\ 0.4 & 0.1 & 0.5 \\ 0.6 & 0.2 & 0.2 \end{pmatrix}$$

则 $\mathbf{P}^2$ 中每一个元素均大于 0，这是一个平稳马尔可夫链。

设 p_C 、 p_E 、 p_S 分别表示煤炭、电力和钢铁部门的平衡价格， $\mathbf{S} = [p_C \quad p_E \quad p_S]$ ，则

$$\mathbf{S}\mathbf{P} = \mathbf{S}$$

解得

$$\mathbf{S} = [0.94p_S \quad 0.85p_S \quad p_S]$$

这意味着，如果钢铁部门的平衡价格是 1 亿元（ $p_S = 1$ 亿元），则煤炭部门的平衡价格是 9400 万元，电力部门的平衡价格是 8500 万元。再进一步，如果煤炭部门生产 1.88 万吨煤，则每吨的价格应该是 5000 元；如果电力部门生产 1.7 亿度电，则每度电的价格应该是 0.5 元。这就是国民生产中，确定每种商品单价的一个基本原则。

【例 3-18】 通常一次网络攻击包括信息收集、权限获取、实施攻击、扩大影响、消除痕迹 5 个步骤。信息收集阶段利用主机扫描、操作系统探测程序等方法收集 IP 地址、操作系统版本、用户标识等信息。权限获取阶段利用系统漏洞或者特洛伊木马等方法获取目标系统的读、写、执行等权限。实施攻击包括窃取信息、破坏系统、安装后门等。扩大影响是以目标系统为“跳板”，对目标所属网络的其他主机进行攻击。消除攻击痕迹是为了防止被识别、追踪。

该过程并不绝对，不同的网络攻击会呈现出不同的形式。有的攻击收集完信息之后直接消除痕迹退出，有的不需要扩大影响……。但是，从统计角度来看，这 5 个步骤仍然有一定规律，可以用转移矩阵来描述。假设某类攻击的转移矩阵为

$$\mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix}$$

矩阵第一行说明这类攻击在执行完信息收集阶段后，有 0.5 的概率进入权限获取阶段，有 0.5 的概率进入扩大影响阶段。攻击的转移矩阵可以作为特征来用，不同的攻击，其转移矩阵是不一样的。有时可以根据转移矩阵的不同，区分不同的攻击。

对上述转移矩阵为 $\mathbf{P}$ 的攻击而言，试计算为了确定攻击所处的步骤，所需要的信息量。

答：可以将该类攻击看作一个一阶马尔可夫信源，状态转移图如图 3-9 所示。

因此这是一个平稳信源。假设各个步骤出现

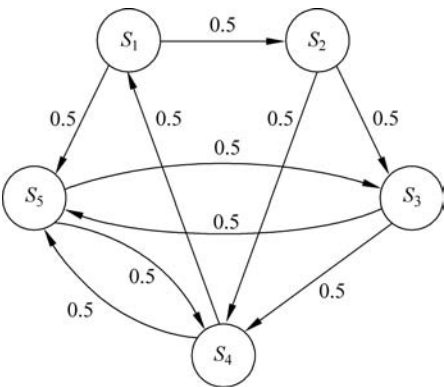


图 3-9 攻击步骤的状态转移图

的概率为 $\mathbf{S}=[p(S_1) \quad p(S_2) \quad \cdots \quad p(S_5)]$, 由 $\mathbf{SP}=\mathbf{S}$, 以及 $p(S_1)+p(S_2)+\cdots+p(S_5)=1$, 得 $\mathbf{S}=[0.1429 \quad 0.0714 \quad 0.1905 \quad 0.2857 \quad 0.3095]$ 。因此信源熵

$$\begin{aligned} H_{m+1}(X) &= - \sum_{S_j} \sum_i p(S_j) p(x_i | S_j) \log p(x_i | S_j) \\ &= 0.1429 + 0.0714 + 0.1905 + 0.2857 + 0.3095 = 1 \text{ 比特 / 步骤} \end{aligned}$$

也就是说, 为了确定攻击所处的步骤, 需要 1 比特的信息量。这与 $\mathbf{P}$ 所表示的攻击矩阵完全相符: 知道当前步骤之后, 下次只可能出现两个步骤之一, 且两个步骤出现的概率各为 0.5。

3. 极限熵

上一节指出, 马尔可夫信源的熵是条件熵 $H_{m+1}(X)$, 这种思想可以扩展到所有平稳有记忆信源, 无论是有限记忆的还是无限记忆的。对无限记忆信源, 记忆长度无限, 此时的熵称为极限熵

$$H_{\infty}(X) = \lim_{m \rightarrow \infty} H_{m+1}(X) = \lim_{m \rightarrow \infty} H(X | X_1 X_2 \cdots X_m) \quad (3-19)$$

其中的“ $m \rightarrow \infty$ ”体现了无限记忆信源记忆长度无限的特点。

实际上, 无记忆信源的熵 $H(X)$ 和马尔可夫信源 (有限记忆信源) 的熵 $H_{m+1}(X)$ 都可以看成极限熵 $H_{\infty}(X)$ 的特殊情况。多数情况下, 马尔可夫信源的熵 $H_{m+1}(X)$ 一般也写为 $H_{\infty}(X)$, 即式 (3-16) 多数情况下写为

$$H_{\infty}(X) = - \sum_{S_j} \sum_i p(S_j) p(x_i | S_j) \log p(x_i | S_j) \quad (3-20)$$

此时极限熵 $H_{\infty}(X)$ 表示平稳马尔可夫信源的熵。

极限熵又称为实际熵, 这与极限熵的含义有很大的关系。从极限熵的式 (3-19) 可以看出, 无论信源是无记忆的、有限记忆的、还是无限记忆的, 极限熵都可以理解为在已知信源已经输出的所有符号的条件下, 再输出一个符号能够获得的信息量, 这与信源的实际输出过程是一致的, 我们就是在已经知道前面的符号条件下才知道当前符号, 因此极限熵又称为实际熵, 即当前符号实际上能够带给我们的信息量。

3.5 离散平稳信源

3.5.1 平稳信源的概念

在讲马尔可夫信源的时候提到了“平稳”的概念, 但是平稳不是马尔可夫信源特有的性质。信源除了可以从记忆的角度分为无记忆信源和有记忆信源之外, 还可以从平稳的角度分为平稳信源和非平稳信源。

【定义 3-8】 如果信源输出各个符号的概率与时间无关, 即

$$P(X_i = x) = P(X_j = x) = p(x) \quad (3-21)$$

则称此信源为一维平稳信源。

【定义 3-9】 如果信源输出长度为 N 的符号序列的概率与时间无关, 即

$$\begin{aligned} &P(X_{i_1} = x_1, X_{i_2} = x_2, \cdots, X_{i_N} = x_N) \\ &= P(X_{j_1} = x_1, X_{j_2} = x_2, \cdots, X_{j_N} = x_N) \end{aligned} \quad (3-22)$$

则称此信源为 N 维平稳信源。

能够看出,一维平稳信源包含在 N 维平稳信源中。平稳信源的实质是无论 i 时刻还是 j 时刻,信源输出同一个信源符号或者符号序列的概率相同,即信源输出信源符号或者符号序列的概率不随时间发生变化。

回忆 3.3 节介绍的离散无记忆信源及其 N 次扩展信源,实际上都是平稳的,因为各个符号或者序列出现的概率均不随时间发生变化。平稳马尔可夫信源也是平稳信源,因为各个状态(即符号序列)出现的概率也会稳定下来,不随时间发生变化。

3.5.2 平稳信源的熵

熵可以分为熵、条件熵、共熵等多种形式,同样,对平稳信源,也有多种度量信息量的方法。

【定义 3-10】 平稳信源输出的长度为 N 的符号序列的联合熵为

$$H(X_1 X_2 \cdots X_N) = - \sum_{X_1 X_2 \cdots X_N} p(x_1 x_2 \cdots x_N) \log p(x_1 x_2 \cdots x_N) \quad (3-23)$$

【定义 3-11】 长度为 N 的信源符号序列中平均每个符号所携带的信息量为平均符号熵,即

$$H_N(X) = \frac{1}{N} H(X_1 X_2 \cdots X_N) \quad (3-24)$$

$H_N(X)$ 的含义是这样理解的: $H(X_1 X_2 \cdots X_N)$ 表示平均每个长度为 N 的序列所携带的信息量,除以 N 之后就是序列中平均每个符号所携带的信息量。

在介绍马尔可夫信源的时候我们定义过 $H_{m+1}(X) = H(X_{m+1} | X_1 X_2 \cdots X_m)$, 即 $H_N(X) = H(X_N | X_1 X_2 \cdots X_{N-1})$, 现在又定义 $H_N(X) = \frac{1}{N} H(X_1 X_2 \cdots X_N)$, 这样定义有问题吗? 后面将会证明,这两个定义从极限熵的角度是等价的。

【定义 3-12】 设信源记忆长度为 $N-1$, 若已知前面 $N-1$ 个符号, 则第 N 个符号所携带的平均信息量为条件熵, 即

$$H(X_N | X_1 \cdots X_{N-1}) = - \sum_{X_1 X_2 \cdots X_N} p(x_1 x_2 \cdots x_N) \log p(x_N | x_1 x_2 \cdots x_{N-1}) \quad (3-25)$$

对于离散平稳信源, 当 $H_1(X) = H(X) < \infty$ 时, 具有以下性质:

性质 1

条件熵 $H(X_N | X_1 X_2 \cdots X_{N-1})$ 随 N 的增加是非递增的, 即

$$\begin{aligned} H(X_1) &\geq H(X_2 | X_1) \geq H(X_3 | X_1 X_2) \geq \cdots \\ &\geq H(X_{N-1} | X_1 X_2 \cdots X_{N-2}) \geq H(X_N | X_1 X_2 \cdots X_{N-1}) \end{aligned} \quad (3-26)$$

证明: 第 2 章中已经给出条件熵和熵之间的关系 $H(X|Y) \leq H(X)$, 该关系表明对于任意的 X 和 Y , 在已知 Y 的条件下关于 X 的平均不确定性不会超过 X 本身的不确定性, 因此当 $N=2$, 即平稳信源输出二维随机矢量 (X_1, X_2) 时, 有

$$H(X_2) \geq H(X_2 | X_1)$$

由于 $X_1, X_2 \in X$ 且具有相同的概率分布, 则

$$H(X_1) = H(X_2) = H(X)$$

因此

$$H(X_1) \geq H(X_2 | X_1)$$

当 $N \geq 3$, 即平稳信源输出 N 维随机矢量 $(X_1, X_2, \dots, X_N)$ 时, 考察

$$\begin{aligned} & H(X_N | X_1 X_2 \cdots X_{N-1}) - H(X_{N-1} | X_1 X_2 \cdots X_{N-2}) \\ &= - \sum_{X_1 \cdots X_{N-1} X_N} p(x_1 x_2 \cdots x_{N-1} x_N) \log p(x_N | x_1 x_2 \cdots x_{N-2} x_{N-1}) \\ &+ \sum_{X_1 \cdots X_{N-2} X_{N-1}} p(x_1 x_2 \cdots x_{N-2} x_{N-1}) \log p(x_{N-1} | x_1 x_2 \cdots x_{N-2}) \end{aligned}$$

由于离散平稳信源的联合概率分布和条件概率分布与时间无关, 有

$$\begin{aligned} & \sum_{X_1 \cdots X_{N-2} X_{N-1}} p(x_1 x_2 \cdots x_{N-2} x_{N-1}) \log p(x_{N-1} | x_1 x_2 \cdots x_{N-2}) \\ &= \sum_{X_2 \cdots X_{N-1} X_N} p(x_2 \cdots x_{N-1} x_N) \log p(x_N | x_2 x_3 \cdots x_{N-1}) \\ &= \sum_{X_1 \cdots X_{N-1} X_N} p(x_1 x_2 \cdots x_{N-1} x_N) \log p(x_N | x_1 x_2 \cdots x_{N-2}) \end{aligned}$$

因此

$$\begin{aligned} & H(X_N | X_1 X_2 \cdots X_{N-1}) - H(X_{N-1} | X_1 X_2 \cdots X_{N-2}) \\ &= \sum_{X_1 \cdots X_{N-1} X_N} p(x_1 x_2 \cdots x_{N-1} x_N) \log \left[\frac{p(x_N | x_2 \cdots x_{N-1})}{p(x_N | x_1 x_2 \cdots x_{N-1})} \times \frac{p(x_1 x_2 \cdots x_{N-1})}{p(x_1 x_2 \cdots x_{N-1})} \right] \\ &= \sum_{X_1 \cdots X_{N-1} X_N} p(x_1 x_2 \cdots x_{N-1} x_N) \log \frac{p(x_N | x_2 \cdots x_{N-1}) \times p(x_1 x_2 \cdots x_{N-1})}{p(x_1 x_2 \cdots x_{N-1} x_N)} \end{aligned}$$

已知对数函数是上凸函数, 于是

$$\begin{aligned} & H(X_N | X_1 X_2 \cdots X_{N-1}) - H(X_{N-1} | X_1 X_2 \cdots X_{N-2}) \\ &\leq \log \left(\sum_{X_1 \cdots X_N} p(x_1 x_2 \cdots x_{N-1} x_N) \frac{p(x_N | x_2 \cdots x_{N-1}) \times p(x_1 x_2 \cdots x_{N-1})}{p(x_1 x_2 \cdots x_{N-1} x_N)} \right) \\ &= \log \left(\sum_{X_1 \cdots X_N} p(x_N | x_2 \cdots x_{N-1}) \times p(x_1 x_2 \cdots x_{N-1}) \right) \\ &= \log \left(\sum_{X_1 \cdots X_{N-1}} p(x_1 x_2 \cdots x_{N-1}) \sum_{X_N} p(x_N | x_2 \cdots x_{N-1}) \right) \\ &= \log \sum_{X_1 \cdots X_{N-1}} p(x_1 x_2 \cdots x_{N-1}) = \log 1 = 0 \end{aligned}$$

因此有

$$H(X_N | X_1 X_2 \cdots X_{N-1}) \leq H(X_{N-1} | X_1 X_2 \cdots X_{N-2})$$

性质 1 的含义是: 我们知道的条件越多, 越容易判断事件的结果, 即该事件包含的信息量越少。这与我们的生活经验是一致的。

性质 2

$$H_N(X) \geq H(X_N | X_1 X_2 \cdots X_{N-1}) \quad (3-27)$$

证明: 由平均符号熵的定义知

$$NH_N(X) = H(X_1 X_2 \cdots X_N) = - \sum_{X_1 \cdots X_N} p(x_1 \cdots x_N) \log p(x_1 \cdots x_N)$$

因为

$$p(x_1 \cdots x_N) = p(x_1) p(x_2 | x_1) p(x_3 | x_1 x_2) \cdots p(x_N | x_1 x_2 \cdots x_{N-1})$$

所以

$$\begin{aligned} NH_N(X) = & - \left[\sum_{X_1 \cdots X_N} p(x_1 \cdots x_N) \log p(x_1) + \sum_{X_1 \cdots X_N} p(x_1 \cdots x_N) \log p(x_2 | x_1) + \cdots \right. \\ & \left. + \sum_{X_1 \cdots X_N} p(x_1 \cdots x_N) \log p(x_N | x_1 x_2 \cdots x_{N-1}) \right] \end{aligned}$$

其中

$$\begin{aligned} \sum_{X_1 \cdots X_N} p(x_1 \cdots x_N) \log p(x_1) &= \sum_{X_1} \left[\sum_{X_2 \cdots X_N} p(x_1 \cdots x_N) \right] \log p(x_1) \\ &= \sum_{X_1} p(x_1) \log p(x_1) = -H(X_1) \\ \sum_{X_1 \cdots X_N} p(x_1 \cdots x_N) \log p(x_2 | x_1) &= \sum_{X_1 X_2} \left[\sum_{X_3 \cdots X_N} p(x_1 \cdots x_N) \right] \log p(x_2 | x_1) \\ &= \sum_{X_1 X_2} p(x_1 x_2) \log p(x_2 | x_1) \\ &= -H(X_2 | X_1) \\ \sum_{X_1 \cdots X_N} p(x_1 \cdots x_N) \log p(x_N | x_1 x_2 \cdots x_{N-1}) &= -H(X_N | X_1 X_2 \cdots X_{N-1}) \end{aligned}$$

于是有

$$H(X_1 X_2 \cdots X_N) = H(X_1) + H(X_2 | X_1) + \cdots + H(X_N | X_1 X_2 \cdots X_{N-1})$$

性质 1 表明

$$H(X_1) \geq H(X_2 | X_1) \geq \cdots \geq H(X_N | X_1 X_2 \cdots X_{N-1})$$

故必有

$$H_N(X) = \frac{1}{N} H(X_1 X_2 \cdots X_N) \geq H(X_N | X_1 X_2 \cdots X_{N-1})$$

即

$$H_N(X) \geq H(X_N | X_1 X_2 \cdots X_{N-1})$$

平均符号熵 $H_N(X)$ 以简单的算术平均表示每一个符号所携带的平均信息量, 而条件熵 $H(X_N | X_1 X_2 \cdots X_{N-1})$ 描述的是已知前面 $N-1$ 个符号时关于符号 X_N 的信息量。可见, 平均符号熵 $H_N(X)$ 和条件熵 $H(X_N | X_1 X_2 \cdots X_{N-1})$ 均为单个符号平均信息量的度量方法。但是条件熵是在已经知道一些条件的前提下, 进一步获得的信息量, 因此条件熵小于或等于平均符号熵。

性质 3

$$H_N(X) \leq H_{N-1}(X) \quad (3-28)$$

证明:

$$\begin{aligned} NH_N(X) &= H(X_1 X_2 \cdots X_N) \\ &= H(X_N | X_1 X_2 \cdots X_{N-1}) + H(X_1 X_2 \cdots X_{N-1}) \\ &= H(X_N | X_1 X_2 \cdots X_{N-1}) + (N-1)H_{N-1}(X) \end{aligned}$$

$$\leq H_N(X) + (N-1)H_{N-1}(X)$$

于是

$$(N-1)H_N(X) \leq (N-1)H_{N-1}(X)$$

因此

$$H_N(X) \leq H_{N-1}(X)$$

性质 3 表明, 平均符号熵 $H_N(X)$ 也是随着 N 的增加而非递增。

进一步分析:

对于熵 $H(X) < \infty$ 的平稳信源 X , 由于熵具有非负性, 即 $H_N(X) \geq 0$, 因此由性质 3 可以看出, 信源符号序列长度不同时, 平稳信源输出的每一个符号所携带的信息量, 即平均符号熵, 满足

$$\infty > H(X) = H_1(X) \geq H_2(X) \geq \cdots \geq H_{N-1}(X) \geq H_N(X) \geq 0 \quad (3-29)$$

可见, 随着信源符号序列长度 N 的增加, $H_N(X)$ 将单调非递增地收敛于某有限值。

对于平稳信源 X , 当其输出的符号序列足够长 ($N \rightarrow \infty$) 时, 每一个信源符号所携带的信息有效地反映出了信源输出的实际信息量。

因此, 定义 $H_\infty = \lim_{N \rightarrow \infty} H_N(X)$ 为离散平稳信源 X 的极限信息量或者极限熵。

性质 4

$H_\infty = \lim_{N \rightarrow \infty} H_N(X)$ 存在, 且

$$H_\infty = \lim_{N \rightarrow \infty} H_N(X) = \lim_{N \rightarrow \infty} H(X_N | X_1 X_2 \cdots X_{N-1}) \quad (3-30)$$

证明: 对于一个任意的正整数 k , 有

$$\begin{aligned} H_{N+k}(X) &= \frac{1}{N+k} H(X_1 X_2 \cdots X_N X_{N+1} \cdots X_{N+k}) \\ &= \frac{1}{N+k} [H(X_1 X_2 \cdots X_{N-1}) + H(X_N | X_1 X_2 \cdots X_{N-1}) + \\ &\quad H(X_{N+1} | X_1 X_2 \cdots X_N) + \cdots + H(X_{N+k} | X_1 X_2 \cdots X_{N+k-1})] \\ &\leq \frac{1}{N+k} [H(X_1 X_2 \cdots X_{N-1}) + (k+1)H(X_N | X_1 X_2 \cdots X_{N-1})] \\ &= \frac{1}{N+k} H(X_1 X_2 \cdots X_{N-1}) + \frac{k+1}{N+k} H(X_N | X_1 X_2 \cdots X_{N-1}) \end{aligned}$$

固定 N , 令 $k \rightarrow \infty$, 得

$$\lim_{k \rightarrow \infty} H_{N+k}(X) \leq 0 + H(X_N | X_1 X_2 \cdots X_{N-1})$$

已知 $N \rightarrow \infty$ 时, $H_N(X)$ 的极限存在且为 H_∞ , 故有

$$\lim_{k \rightarrow \infty} H_{N+k}(X) = H_\infty$$

综合性质 2 给出的关系可知, 条件熵 $H(X_N | X_1 X_2 \cdots X_{N-1})$ 满足

$$H_\infty(X) \leq H(X_N | X_1 X_2 \cdots X_{N-1}) \leq H_N(X)$$

令 $N \rightarrow \infty$, 因为 $\lim_{N \rightarrow \infty} H_N(X) = H_\infty$, 所以

$$\lim_{N \rightarrow \infty} H(X_N | X_1 X_2 \cdots X_{N-1}) = H_\infty$$

于是

$$H_{\infty} = \lim_{N \rightarrow \infty} H_N(X) = \lim_{N \rightarrow \infty} H(X_N | X_1 X_2 \cdots X_{N-1})$$

性质 2、3、4 说明,当平稳信源输出的符号序列长度达到无限大时,平均符号熵 $H_N(X)$ 和条件熵 $H(X_N | X_1 X_2 \cdots X_{N-1})$ 都非递增地收敛于平稳信源的极限熵。因此,对于一个实际的平稳信源,当经过足够长的时间后,信源输出每一个符号所携带的信息量,即信源的极限熵,可以用平均符号熵或者条件熵加以度量。

3.6 信源的相关性和剩余度

前面已经讨论了多类离散信源,包括离散无记忆信源及其 N 次扩展信源、马尔可夫信源、离散平稳信源。由于这些信源的统计特性各不相同,它们输出信息的能力也各不相同。

根据式(3-28)能够得到

$$H(X_N | X_1 X_2 \cdots X_{N-1}) \leq H(X_{N-1} | X_1 X_2 \cdots X_{N-2}) \leq \cdots \leq H(X_2 | X_1) \leq H(X_1)$$

当离散平稳信源输出符号是等概分布时熵最大,记此时的熵为 H_0 ,则 $H_0 = \log n$ 。而且当 $N \rightarrow \infty$ 时, $H(X_N | X_1 X_2 \cdots X_{N-1})$ 收敛于极限熵。于是,有

$$H_0 \geq H_1 \geq H_2 \geq \cdots \geq H_{m+1} \geq \cdots \geq H_{\infty} \quad (3-31)$$

这就是信源的相关性。为什么叫作“相关性”?是因为由于信源符号间的依赖关系,使得信源熵减小。也就是说由于信源符号间存在相关性,知道的信源符号越多,越容易猜测将要出现的符号是什么,即不确定性减小。

为了衡量信源的相关性程度,引入信源剩余度(冗余度)的概念。

【定义 3-13】 信源剩余度定义为

$$R = 1 - \frac{H_{\infty}}{H_0} \quad (3-32)$$

前面我们说过,极限熵 H_{∞} 表示当前符号实际上能够带给我们的信息量,而 H_0 是熵的最大值,因此 H_{∞}/H_0 表示实际值占最大值的比例,该比例称为熵的相对率,记作 η

$$\eta = \frac{H_{\infty}}{H_0} \quad (3-33)$$

可见,信源剩余度

$$R = 1 - \eta$$

信源剩余度越小,说明熵的实际值占最大值的比例越大,即一个信源符号上实际携带的信息量越接近于信源符号能够带有信息量的最大值,即信源符号的利用率越高。利用率越高,当然剩余未用的部分越少,即“剩余度越小”。反之亦然。

怎样才能减小信源剩余度呢?要回答这个问题,首先要明确剩余度的来源。剩余度

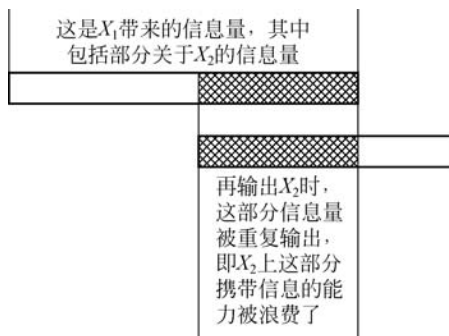


图 3-10 信源相关性和剩余度

来源于信源符号之间的相关性。这是因为如图 3-10 所示,以长度为 2 的序列为例,由于信源符号之间存在相关性,因此输出 X_1 之后,带来的信息量中包括部分关于 X_2 的信息量,再输出 X_2 时,这部分信息量被重复输出,即 X_2 上这部分携带信息的能力被浪费了,或者说这部分携带信息的能力没有用上,因此就产生了剩余。信源符号之间相关性越大,剩余的就越多。究竟剩余了多少,由信源剩余度来衡量。

可见,要减小信源剩余度,就要减小信源相关性。第 6 章和第 7 章要研究的信源编码是减少信源相关性的有效方法。

3.7 本章小结

本书内容主要包括信源、信道、信源编码、信道编码四部分。本章研究了信源,主要是离散信源,主要内容见表 3-3。

信源可以从不同的角度分类。从记忆特性分可以分为无记忆信源和有记忆信源,从平稳特性可以分为平稳信源和非平稳信源。因此信源总共可以分为四类:平稳无记忆信源、平稳有记忆信源、非平稳无记忆信源、非平稳有记忆信源。重点研究的是前两类。

平稳无记忆信源相对简单,给出了信源及扩展信源的模型和熵。

平稳有记忆信源中重点研究平稳有限记忆信源,即平稳马尔可夫信源。包括 m 阶马尔可夫信源的定义、状态转移矩阵、状态转移图、熵(包括熵的求解方法)。

最后是信源的相关性和剩余度。

表 3-3 本章小结

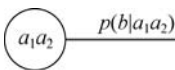
平稳无记忆信源		非平稳无记忆信源
信源	N 次扩展信源	
模型 $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$ 熵： $H(X) = -\sum_{i=1}^n p(x_i) \log p(x_i)$	模型： $\begin{bmatrix} X^N \\ P \end{bmatrix} = \begin{bmatrix} x_1 \cdots x_1 x_1 & x_1 \cdots x_1 x_2 & \cdots & x_n \cdots x_n x_n \\ p(x_1)^N & p(x_1)^{N-1} p(x_2) & \cdots & p(x_n)^N \end{bmatrix}$ 熵： $H(X^N) = NH(X)$	
平稳有记忆信源		非平稳有记忆信源
平稳有限记忆信源(平稳马尔可夫信源) 定义： $p(x_i x_{i-1} x_{i-2} \cdots x_{i-m} \cdots x_1) = p(x_i x_{i-1} x_{i-2} \cdots x_{i-m})$ 状态转移矩阵 状态转移图 熵： $H_\infty(X) = H_{m+1}(X) = -\sum_{S_j} \sum_i p(S_j) p(x_i S_j) \log p(x_i S_j)$ 熵的求解方法： 1. 判断是否为平稳马尔可夫信源； 2. 对平稳马尔可夫信源,求出状态的平稳分布； 3. 根据上式求得马尔可夫信源的熵		
信源相关性： $H_0 \geq H_1 \geq H_2 \geq \cdots \geq H_{m+1} \geq \cdots \geq H_\infty$ 熵的相对率： $\eta = \frac{H_\infty}{H_0}$		平稳无限记忆信源
信源剩余度： $R = 1 - \frac{H_\infty}{H_0} = 1 - \eta$		

3.8 习题

3-1 某信源先后发出两个符号 x_1 和 x_2 , 如果该信源是无记忆的, 则 $p(x_2 | x_1) =$ _____。

3-2 无记忆信源 X 的熵为 1.25 比特, 则它的 4 次扩展信源 X^4 的熵为 _____ 比特。

3-3 假设某一信源的转移矩阵为 $P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$, 则 $\sum_{i=1}^3 p_{2i} =$ _____。

3-4 如果用  表示一个马尔可夫信源, 则该信源是 _____ 阶的。

3-5 设一离散无记忆信源为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3/8 & 1/4 & 1/4 & 1/8 \end{bmatrix}$$

信源输出一条消息为 (202120130213001203210110321010021032011223210), 求

- (1) 此消息的自信息量是多少?
- (2) 在此消息中平均每个符号携带的信息量是多少?

3-6 信源 X 中只有两个消息符号, 画出信源熵的曲线并求解 $H_{\max}(X)$ 。

3-7 某一无记忆信源的符号集为 $\{0, 1\}$, 已知信源的概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

- (1) 求该信源的熵。
- (2) 由 100 个符号构成的序列, 求每一特定序列 (例如有 m 个“0”和 $(100 - m)$ 个“1”构成) 的自信息量的表达式。

3-8 设某离散平稳信源 X , 概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 11/36 & 4/9 & 1/4 \end{bmatrix}$$

并设信源发出的符号只与前一个符号有关, 其联合概率为 $p(a_i, a_j)$, 如题表 3-1 所示。

题表 3-1 联合概率

$a_j \backslash a_i$	0	1	2
0	1/4	1/18	0
1	1/18	1/3	1/18
2	0	1/18	7/36

求信源的信息熵、条件熵与联合熵, 并比较信息熵与条件熵的大小。

3-9 设离散无记忆信源为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.16 & 0.17 \end{bmatrix}$$

求该信源的熵,并解释为什么 $H(x) > \log 6$ 不能满足信源的极值性。

3-10 有一离散无记忆信源

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

设计两个独立实验去观察它,其结果分别为 $Y_1 \in \{0, 1\}, Y_2 \in \{0, 1\}$,已知条件概率如题表 3-2 所示。

题表 3-2 实验结果

$p(y_1 x)$	0	1	$p(y_2 x)$	0	1
0	1	0	0	1	0
1	0	1	1	1	0
2	1/2	1/2	2	0	1

求 $I(X; Y_1)$ 和 $I(X; Y_2)$,并判断哪一个实验好一些。

3-11 二次扩展信源的熵为 $H(X^2)$,而一阶马尔科夫信源的熵为 $H(X_2 | X_1)$,试比较两者的大小,并说明原因。

3-12 一个马尔可夫过程的基本符号为 0, 1, 2,这 3 个符号等概率出现,并且具有相同的转移概率。

(1) 画出一阶马尔可夫过程的状态图,并求稳定状态下的一阶马尔可夫信源熵 H_2 和信源剩余度;

(2) 画出二阶马尔可夫过程的状态图,并求稳定状态下的二阶马尔可夫信源熵 H_3 和信源剩余度。

3-13 设有一个信源,它产生 0-1 序列的消息。该信源在任意时间而且不论以前发生过什么消息符号,均按 $p(0)=0.4, p(1)=0.6$ 的概率发出符号。

(1) 试问这个信源是否是平稳的?

(2) 试计算 $\lim_{N \rightarrow \infty} H_N(X)$ 。

(3) 试计算 $H(X^4)$ 并写出 X^4 信源中可能发出的所有符号。

3-14 有一个马尔可夫信源,已知转移概率为 $p(S_1 | S_1) = \frac{2}{3}, p(S_2 | S_1) = \frac{1}{3}, p(S_1 | S_2) = 1, p(S_2 | S_2) = 0$ 。试画出状态转移图,并求出信源熵。

3-15 一阶马尔可夫信源的状态转移图如题图 3-1 所示,信源 X 的符号集为 $\{0, 1, 2\}$,并定义 $\bar{p} = 1 - p$ 。

(1) 求信源平稳后的概率分布 $p(0)$ 、 $p(1)$ 和 $p(2)$;

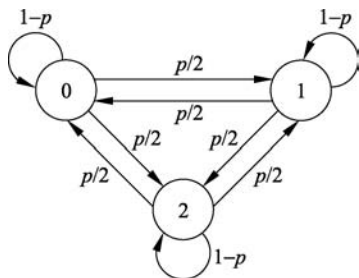
(2) 求此信源的熵;

(3) 近似认为此信源为无记忆时,符号的概率分布等于平稳分布。求近似信源的熵 $H(X)$ 并与 H_∞ 进行比较;

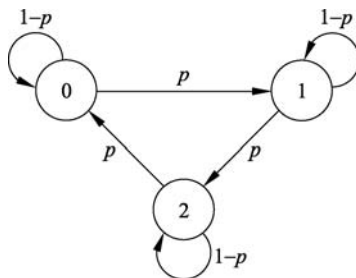
(4) 对一阶马尔科夫信源, p 取何值时 H_∞ 取最大值,又当 $p=0$ 或 $p=1$ 时结果如何?

3-16 一阶马尔可夫信源的状态转移图如题图 3-2 所示,信源 X 的符号集为 $\{0, 1, 2\}$ 。

- (1) 求平稳后的信源的概率分布;
- (2) 求信源熵 H_∞ ;
- (3) 求 $p=0$ 或 $p=1$ 时信源的熵,并说明其理由。



题图 3-1 题 3.15 状态转移图



题图 3-2 题 3.16 状态转移图

3-17 设信源发出两个消息 x_1 和 x_2 , 它们的概率分布为 $p(x_1) = \frac{3}{4}$ 、 $p(x_2) = \frac{1}{4}$ 。求该信源的熵和冗余度。

3-18 设某马尔可夫链的一步转移概率矩阵为

$$\begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{bmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{bmatrix} \end{matrix}$$

试求: (1) 该马尔可夫链的 2 步转移概率矩阵;

(2) 平稳后状态“0”, “1”, “2”的概率分布。

3-19 设有一信源, 它在开始时以 $p(a) = 0.6$ 、 $p(b) = 0.3$ 、 $p(c) = 0.1$ 的概率发出 X_1 , 如果 X_1 为 a 时, 则 X_2 为 a, b, c 的概率为 $1/3$; 如果 X_1 为 b 时, 则 X_2 为 a, b, c 的概率为 $1/3$; 如果 X_1 为 c 时, 则 X_2 为 a, b 的概率为 $1/2$, 为 c 的概率为 0 。而且后面发出 X_i 的概率只与 X_{i-1} 有关。又 $p(X_i | X_{i-1}) = p(X_2 | X_1)$ 。试利用马尔可夫信源的图示法画出状态转移图, 并计算信源熵 H_∞ 。

3-20 黑白传真机的消息元只有黑色和白色两种, 即 $X = \{\text{黑}, \text{白}\}$, 一般气象图上, 黑色出现的概率 $p(\text{黑}) = 0.3$, 白色出现的概率 $p(\text{白}) = 0.7$, 黑白消息前后有关联, 其转移概率为 $p(\text{白}|\text{白}) = 0.9143$ 、 $p(\text{黑}|\text{白}) = 0.0857$ 、 $p(\text{白}|\text{黑}) = 0.2$ 、 $p(\text{黑}|\text{黑}) = 0.8$ 。求该一阶马尔可夫信源的不确定性 $H(X|X)$, 并画出该信源的状态转移图。

3-21 用 α, β, γ 三个字母组字, 设组成的字有以下三种情况:

- (1) 只用 α 一个字母的单字母字;
- (2) 用 α 开头或结尾的两字母字;
- (3) 把 α 夹在中间的三字母字。

假定由这三种字符组成一种简单语言, 试计算当所有字等概出现时, 语言的冗余度。

3-22 令 X 为掷钱币直至其正面第一次向上所需的次数, 求 $H(X)$ 。

3-23 一段代码如下:

```
x = 0;  
for i = 0 to n  
    x = x + 1;  
end
```

不考虑循环次数 n 的限制, x 是越来越大的, 不可能稳定下来。请从信息论的角度证明该循环过程无法达到稳定状态, 即是非平稳过程。

3-24 某地区每年城市和郊区之间的人口移动有一定规律, 大概有 5% 的城市人口流动到郊区, 3% 的郊区人口流动到城市。

(1) 请给出转移矩阵, 该矩阵也称为移民矩阵。

(2) 多年以后, 该地区的城市人口和郊区人口会不会各自达到一个固定的规模? 即城市人口和郊区人口的数量基本稳定, 不再发生大的变化。

(3) 如果能达到一个固定的规模, 并假设该地区共 1000 万人口, 那么城市人口和郊区人口的固定规模各是多少?

3-25 某出租车公司大约有 2000 辆小汽车, 客户可以在同一个营业点取还, 也可以异地取还, 假设取还模型为

	飞机场	市中心	汽车站
飞机场	0.90	0.01	0.09
市中心	0.01	0.90	0.09
汽车站	0.09	0.01	0.90

请问市中心应该准备多少辆车?