信息论基础

李　莹
liying2009@ecust.edu.cn

第三章：信源及信源熵

一、信源的分类及其数学模型

二、离散单符号信源

三、离散多符号信源

信源的分类及其数学模型

- 信源是产生消息（符号）、消息序列（符号序列）以及时间连续的消息的来源。

- 信源的主要问题：
  - 如何描述信源的输出（信源的建模问题）√
  - 怎样确定信源产生的信息量、产生信息的速率 √
  - 信源编码 （第五、七章）

根据信源输出消息在时间和取值上是离散或连续分类：

时间（空间）	取值	信源种类	举例	消息的数学描述
离散	离散	离散信源（数字信源）	文字、数据、离散化图象	离散随机变量序列 $\boldsymbol{P}(\mathbf{X}) = \boldsymbol{P}(\boldsymbol{X}_1\boldsymbol{X}_2\cdots\boldsymbol{X}_N)$
离散	连续	连续信源		连续随机变量序列 $P(\mathbf{X}) = P(X_1X_2\cdots X_N)$
连续	连续	波形信源（模拟信源）	语音、音乐、热噪声、图形、图象	随机过程 $\{X(e,t)\}$
连续	离散		不常见	

信源的分类

- 根据信源发出的消息序列中的消息，统计特性是否保持不变，信源可分为平稳信源/非平稳信源。
- 根据信源发出的单个消息取值是离散值还是连续值，信源可分为离散信源/连续信源。
- 根据信源发出的消息之间是否有统计依赖关系，信源可分为有记忆信源/无记忆信源。
- 本章重点研究离散平稳无记忆信源，以及较简单的有记忆信源－马尔可夫信源。

信源的分类

- 随机序列
  - 平稳信源
    - 离散平稳信源
      - 离散无记忆信源
      - 离散有记忆信源
        - 记忆长度无限长
        - 记忆长度有限(马尔可夫信源)
    - 连续平稳信源
  - 非平稳信源

随机过程：波形信源

离散单符号信源

- **离散单符号信源：输出离散取值的单个符号的信源。**

离散单符号信源是最简单、最基本的信源，是组成实际信源的基本单元，可以用一个离散随机变量来表示。

- **离散单符号信源X的概率空间：**

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

$$p(x_i) \geq 0 \qquad \sum_{i=1}^{n} p(x_i) = 1$$

离散单符号信源

例： 一个二元无记忆信源，符号集 $A=\{0,1\}$ ，p为$X=0$的概率。写出信源的模型。

解： 信源的模型：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & q \end{bmatrix}$$

离散单符号信源

- 信源输出的所有消息的自信息的 统计平均值，定义为信源的平均自信息（信息熵）：

$$H(X)=E\left[-\log p(x_i)\right]=-\sum_{i=1}^{q}p(x_i)\log p(x_i)$$

- 信息熵表示离散单符号信源的平均不确定性。

第三章：信源及信源熵

一：信源的分类及其数学模型

二：离散单符号信源

三：离散多符号信源

1. 预备知识
2. 离散平稳无记忆信源
3. 离散平稳有记忆信源
4. 马尔可夫信源
5. 信源的相关性和剩余度

1. 预备知识

- 实际信源输出往往是符号序列，称为离散多符号信源。
- 离散多符号信源可以用随机矢量/随机变量序列来描述，即

$$\boldsymbol{X} = X_1 X_2 \cdots X_n \cdots$$

- 一般来说，信源的统计特性随着时间的推移而有所变化。为了便于研究，我们常常假定在一个较短的时间段内，信源是平稳信源。

定义1：对于离散随机变量序列 $X_1X_2\cdots X_n\cdots$ ，若任意两个不同时刻i和j (大于1的任意整数) 信源发出消息的概率分布完全相同，即对于任意的$N=0, 1, 2, \cdots$ ，$X_iX_{i+1}\cdots X_{i+N}$和 $X_jX_{j+1}\cdots X_{j+N}$具有相同的概率分布。也就是

$$P(X_i)=P(X_j)$$

$$P(X_iX_{i+1})=P(X_jX_{j+1})$$

$$\vdots$$

$$P(X_iX_{i+1}\cdots X_{i+N})=P(X_jX_{j+1}\cdots X_{j+N})$$

即各维联合概率分布均与时间起点无关的信源称为离散平稳信源。

对离散平稳信源，由联合概率与条件概率的关系可以推出：

$$P(X_{i+1} | X_i) = P(X_{j+1} | X_j)$$

$$\vdots$$

$$P(X_{i+N} | X_i X_{i+1} \cdots X_{i+N-1}) = P(X_{j+N} | X_j X_{j+1} \cdots X_{j+N-1})$$

因此：

$$H(X_1) = H(X_2) = \cdots = H(X_N)$$

$$H(X_2 | X_1) = H(X_3 | X_2) = \cdots = H(X_N | X_{N-1})$$

$$H(X_3 | X_1 X_2) = H(X_4 | X_2 X_3) = \ldots = H(X_N | X_{N-2} X_{N-1})$$

$$\vdots$$

例：一平稳信源X的符号集$A=\{0,1\}$,产生随机序列$\{x_n\}$，其中

$$P(x_1=0)=p,$$
$$P(x_1=0,x_2=1)=b.$$

求（1） $P(x_n=1),n>1$ （2） $P(x_4=1|x_3=0)$

解（1）由于平稳性 $P(x_n=0)=P(x_1=0)=p$

$$P(x_n=1)=1-p.$$

（2）由于平稳性 $P(x_4=1,x_3=0)=P(x_2=1,x_1=0)=b$

$$P(x_4=1|x_3=0)=P(x_4=1,x_3=0)/P(x_3=0)=b/p.$$

同时 $P(x_2=1|x_1=0)=b/p.$

定义2：随机变量序列中，对前N个随机变量的联合熵求平均称为平均符号熵：

$$H_N(\boldsymbol{X})=\frac{1}{N}H(X_1X_2\cdots X_N)$$

如果当 $N\to\infty$ 时上式极限存在，则 $\lim\limits_{N\to\infty}H_N(\boldsymbol{X})$ 被称为熵率，或极限熵，记为

$$H_\infty=\lim_{N\to\infty}H_N(\boldsymbol{X})$$

2. 离散平稳无记忆信源

- 为了研究离散平稳无记忆信源的极限熵，把信源输出的符号序列看成是一组一组发出的。
  - 例1：电报系统中，可以认为每**2**个二进制数字组成一组。这样信源输出的是由**2**个二进制数字组成的一组组符号。这时可以将它们等效看成一个新的信源，它由四个符号**00**，**01**，**10**，**11**组成，把该信源称为二进制无记忆信源的二次扩展。
  - 例2：如果把每三个二进制数字组成一组，这样长度为**3**的二进制序列就有**8**种不同的符号，可等效成一个具有**8**个符号的信源，把它称为二进制无记忆信源的三次扩展信源。

● 假定信源输出的是N长符号序列，把它看成是一个新信源，称为离散平稳无记忆信源的N次扩展信源，用N维离散随机矢量来表示：

$$\boldsymbol{X} = X_1 X_2 \cdots X_N = X^N$$

● N次扩展信源的概率空间为：

$$\begin{bmatrix} X^N \\ P(\boldsymbol{X}) \end{bmatrix} = \begin{Bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_i & \dots & \alpha_{q^N} \\ p(\alpha_1) & p(\alpha_2) & \dots & p(\alpha_i) & \dots & p(\alpha_{q^N}) \end{Bmatrix}$$

☝ α_i是一个长为N的序列，

$$\alpha_i = x_{i_1} x_{i_2} \cdots x_{i_N}$$

- N次扩展信源的熵：

$$H(\boldsymbol{X}) = H(X^N) = -\sum_{i=1}^{q^N} p(\alpha_i)\log p(\alpha_i)$$

- 离散平稳无记忆信源的N次扩展信源的熵等于离散单符号信源熵的N倍：

$$H(\boldsymbol{X}) = H(X^N) = NH(X)$$

- 离散平稳无记忆信源的熵率：

$$H_\infty = \lim_{N\to\infty} H_N(\mathbf{X}) = \lim_{N\to\infty} \frac{1}{N} \cdot NH(X) = H(X)$$

例1：设有一离散无记忆信源**X**，其概率空间为

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

求该信源的熵率及二次扩展信源的熵。

解：

- 离散单符号信源熵

$$H(X) = -\sum_{i=1}^{3} p(x_i)\log_2 p(x_i) = 1.5 \text{ 比特/符号}$$

- 熵率：　$H_\infty = H(X) = 1.5$ 比特 / 符号

➢ 二次扩展信源的概率空间：

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \begin{Bmatrix} \alpha_1(x_1x_1) & \alpha_2(x_1x_2) & \alpha_3(x_1x_3) & \alpha_4(x_2x_1) & \alpha_5(x_2x_2) & \alpha_6(x_2x_3) & \alpha_7(x_3x_1) & \alpha_8(x_3x_2) & \alpha_9(x_3x_3) \\ 1/4 & 1/8 & 1/8 & 1/8 & 1/16 & 1/16 & 1/8 & 1/16 & 1/16 \end{Bmatrix}$$

➢ 二次扩展信源的熵：

$$H(\boldsymbol{X}) = H(X^2) = -\sum_{i=1}^{9} p(\alpha_i)\log p(\alpha_i) = 3 \quad \text{比特/二个符号}$$

3. 离散平稳有记忆信源

- 实际信源常常是有记忆信源。设信源输出N长的符号序列，则可以用N维随机矢量$\boldsymbol{X} = X_1X_2\cdots X_N$ 来表示信源，其中每个随机变量之间存在统计依赖关系。

- N维随机矢量的联合熵为：

$$H(\boldsymbol{X}) = H(X_1X_2\cdots X_N)$$
$$= H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1X_2) + \cdots + H(X_N \mid X_1X_2\cdots X_{N-1})$$

定理：对于离散平稳信源，有以下几个结论：

（1）条件熵 $H(X_N|X_1X_2\ldots X_{N-1})$ 随着N 的增加而递减；

（2）平均符号熵大于等于条件熵；

$$H_N(\boldsymbol{X}) \geq H(X_N \mid X_1X_2\ldots X_{N-1})$$

（3）平均符号熵 $H_N(\boldsymbol{X})$ 随着N 的增加而递减；

（4）如果 $H_1(X)<\infty$ ，则

$$H_\infty = \lim_{N\to\infty} H_N(\boldsymbol{X}) = \lim_{N\to\infty} \frac{1}{N} H(X_1X_2\ldots X_N)$$
$$= \lim_{N\to\infty} H(X_N \mid X_1X_2\ldots X_{N-1})$$

证明

定理说明：

- 随着序列长度的增加，也就是随着统计约束条件不断增加，平均符号熵和条件熵均随之减少；
- 熵率表示信源输出的符号序列中平均每个符号所携带的信息量。计算方式有两种：

◆求极限平均符号熵

◆求极限条件熵

$$H_\infty = \lim_{N\to\infty} H_N(\boldsymbol{X}) = \lim_{N\to\infty}\frac{1}{N}H(X_1X_2\ldots X_N)$$

$$= \lim_{N\to\infty} H(X_N \mid X_1X_2\ldots X_{N-1})$$

3. 离散平稳有记忆信源（续4）

例：信源X的信源模型为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{4} & \frac{4}{9} & \frac{11}{36} \end{bmatrix}$$

输出符号序列中，只有前后两个符号之间有记忆，条件概率空间见右边的表。求熵率并比较 $H(X)$ 、$H(X_2|X_1)$ 、$1/2H(X_1X_2)$。

条件概率 $P(X_2 \mid X_1)$

X_1 \ X_2	x_1	x_2	x_3
x_1	$\frac{7}{9}$	$\frac{2}{9}$	0
x_2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$
x_3	0	$\frac{2}{11}$	$\frac{9}{11}$

解：

1) $H_\infty = \lim_{N\to\infty} H(X_N \mid X_1X_2\ldots X_{N-1}) = \lim_{N\to\infty} H(X_N \mid X_{N-1})$

$= H(X_2 \mid X_1) = 0.870$ 比特/符号

2) 如果不考虑符号间的相关性，则信源熵为

$H(X) = H(\frac{1}{4},\frac{4}{9},\frac{11}{36}) = 1.542$ 比特/符号

3) 如果把信源发出的符号看成是分组发出的，每两个符号为一组，这个新信源的熵为

$H(X_1X_2) = H(X_1) + H(X_2 \mid X_1) = 2.412$ 比特/两个符号

$$H(X_2 \mid X_1) = \frac{1}{4}H(\frac{2}{9}) + \frac{4}{9}H(\frac{1}{8},\frac{3}{4},\frac{1}{8}) + \frac{11}{36}H(\frac{2}{11},\frac{9}{11})$$

$$H(X_2 \mid X_1) = -\frac{1}{4}\cdot\frac{7}{9}\log\frac{7}{9} - \frac{1}{4}\cdot\frac{2}{9}\log\frac{2}{9} - \frac{4}{9}\cdot\frac{1}{8}\log\frac{1}{8}$$
$$-\frac{4}{9}\cdot\frac{3}{4}\log\frac{3}{4} - \frac{4}{9}\cdot\frac{1}{8}\log\frac{1}{8} - \frac{11}{36}\cdot\frac{2}{11}\log\frac{2}{11} - \frac{11}{36}\cdot\frac{9}{11}\log\frac{9}{11}$$

结论： $$H_\infty < \frac{1}{2}H(X_1 X_2) < H(X)$$

如何从理论上解释这个结果？