

第6章 限失真信源编码



问题提出



- 信息编码分类:

无失真编码: 码字与信源符号 (序列) 一一对应。保熵编码

有失真编码: 在信源编码时引入一定失真。

- 引入有失真编码的原因:

(1) 保熵编码并非总是必需的。

(2) 保熵编码并非总是可能的。

(3) 降低信息率有利于传输和处理。

- 限失真编码

在允许的失真范围内把编码后的信息率压缩到最小。

- 信道的信息率R: 信道的平均互信息量:

$$R = I(U; V) = H(U) - H(U | V) = H(V) - H(V | U) \quad \text{bit/符号}$$

各节内容



6.1 失真测度

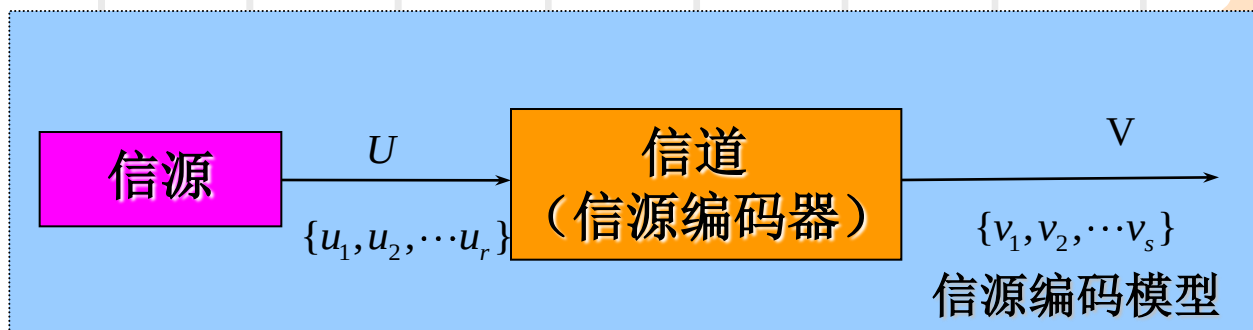
6.2 信息率失真函数及其性质

6.3 限失真信源编码定理

6.4 信息率失真函数的计算



6.1 失真测度



■ 无失真编码

$$H(U|V) = H(V|U) = 0$$

$$R = I(U; V) = H(U) = H(V)$$

■ 限失真编码 ?

1、失真度

- 设信源编码器输入为 $U = \{u_1, u_2, \dots, u_r\}$, 编码后的输出信源为 $V = \{v_1, v_2, \dots, v_s\}$
- 失真度（或失真函数）：编码器输入符号 u_i 与输出符号 v_j 之间的误差或失真，用非负实值函数 $d(u_i, v_j)$ 来描述。
- 常用的失真度有：

误码失真：

$$d(u_i, v_j) = \begin{cases} 0, & u_i = v_j \\ 1, & u_i \neq v_j \end{cases}$$

均方失真：

$$d(u_i, v_j) = (u_i - v_j)^2$$

绝对失真：

$$d(u_i, v_j) = |u_i - v_j|$$

相对失真：

$$d(u_i, v_j) = |u_i - v_j| / |u_i|$$

2、失真矩阵

- 将 $r \times s$ 个 $d(u_i, v_j)$ 排成矩阵形式，称为失真矩阵，记为 $[d]$ ：

$$[d] = \begin{matrix} & \begin{matrix} v_1 & v_2 & \cdots & v_s \end{matrix} \\ \begin{bmatrix} d(u_1, v_1) & d(u_1, v_2) & \cdots & d(u_1, v_s) \\ d(u_2, v_1) & d(u_2, v_2) & \cdots & d(u_2, v_s) \\ \vdots & \vdots & \vdots & \vdots \\ d(u_r, v_1) & d(u_r, v_2) & \cdots & d(u_r, v_s) \end{bmatrix} & \begin{matrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{matrix} \end{matrix}$$

■ 例 设信源 U 取值于 $\{0,1\}$ ，编码器输出取值于 $\{0,1,2\}$ ，规定失真度函数为

$$d(0,0) = d(1,1) = 0$$

$$d(0,1) = d(1,0) = 1$$

$$d(0,2) = d(1,2) = 0.5$$

则失真矩阵为

$$[d] = \begin{bmatrix} 0 & 1 & 0.5 \\ 1 & 0 & 0.5 \end{bmatrix}$$

3、平均失真度

- 对所有符号的失真度 $\{d(u_i, v_j)\}_{i,j}$ 取统计平均，称为平均失真度或平均失真，记为 $\bar{D}$ ：

$$\begin{aligned}\bar{D} &= E\{d(u_i, v_j)\} = \sum_{i=1}^r \sum_{j=1}^s P(u_i, v_j) d(u_i, v_j) \\ &= \sum_{i=1}^r \sum_{j=1}^s P(u_i) P(v_j | u_i) d(u_i, v_j)\end{aligned}$$

4、符号序列的平均失真度和平均失真度

- 对于符号序列，可将失真度或失真函数的定义可推广到矢量形式。设编码器输入 a_h 和输出 b_l 均为 N 长符号序列，即

$$\begin{aligned}\alpha_h &= u_{h_1} u_{h_2} \cdots u_{h_N} & h &= 1, 2, \dots, r^N \\ \beta_l &= v_{l_1} v_{l_2} \cdots v_{l_N} & l &= 1, 2, \dots, s^N\end{aligned}$$

则 N 长符号序列的失真度 $d(a_h, b_l)$ 可定义为，

$$d(a_h, b_l) = \sum_{k=1}^N d(u_{h_k}, v_{l_k})$$

式中 $d(u_{h_k}, v_{l_k})$ 是输入/输出序列第 k 位符号的失真度。

4、符号序列的平均失真度和平均失真度（续）

- N 长符序列的平均失真度为

$$\begin{aligned}\bar{D}(N) &= E\{d(a_h, b_l)\} = \prod_{h=1}^{r^N} \prod_{l=1}^{s^N} P(a_h, b_l) d(a_h, b_l) \\ &= \prod_{h=1}^{r^N} \prod_{l=1}^{s^N} P(a_h, b_l) \prod_{k=1}^N d(u_{h_k}, v_{l_k})\end{aligned}$$

当信源和信道（编码器）均无记忆时，

$$\bar{D}(N) = \prod_{k=1}^N E\{d(u_{h_k}, v_{l_k})\} = \prod_{k=1}^N \bar{D}_k = N\bar{D}$$

其中 $\bar{D}_1 = \bar{D}_2 = \dots = \bar{D}_N = \bar{D}$ 是单符号的平均失真度。

6.2 信息率失真函数及其性质

6.2.1 信息率失真函数定义

1、D允许(试验)信道

- 如果要求平均失真 $\bar{d}$ 小于某个给定值 D , 即要求

$$\bar{d} = E\{d(u_i, v_j)\} = \sum_{i=1}^r \sum_{j=1}^s P(u_i)P(v_j | u_i) d(u_i, v_j) \leq D$$

这意味着对转移概率 $P_{V|U}$ 施加了相应的限制, 上式所给的限制条件称保真度准则。

满足保真度准则 $\bar{d} \leq D$ 的信道称为 D 允许(试验)信道。

所有试验信道的转移概率组成一个集合, 记为:

$$B_D = \{P_{V|U}; \bar{d} \leq D\}$$

- 要使限失真编码满足保真度准则, 则编码器必须是试验信道, 即编码器的转移概率 $P_{V|U} \in B_D$ 。

2、率失真函数的定义

- 在 B_D 中寻求一个 $p_{V|U}$ (即寻求一个特定的编码器) 使 $I(U;V)$ 最小, 这个最小的平均互信息量称为信息率失真函数, 简称率失真函数, 记为 $R(D)$, 即

$$R(D) = \min_{p_{V|U} \in B_D} I(U;V) = \min\{I(U;V); \bar{D} \in D\}$$

- 当最小值不存在时, 可用下界值代替:

$$R(D) = \inf_{p_{V|U} \in B_D} I(U;V)$$

- 对于离散信源, $R(D)$ 可表示成

$$R(D) = \min_{p_{V|U} \in B_D} \sum_{i=1}^r \sum_{j=1}^s P(u_i) P(v_j | u_i) \log \frac{P(v_j | u_i)}{P(v_j)}$$

- $R(D)$ 是保真度准则 ($\bar{D} \in D$) 下所必须传输的信息率, 也是熵压缩编码器输出可能达到的最低熵率。

3、序列的信息率失真函数

- 在序列情形下，可用 N 次扩展信源 U^N 和 N 次扩展信道 $\{U^N, P_{V^N|U^N}, V^N\}$ 讨论。
- 试验信道为所有满足保真度准则 $\bar{D}(N) \in ND$ 的信道的集合，这些试验信道的转移概率组成集合 B_{ND} ：

$$B_{ND} = \{P_{V^N|U^N}; \bar{D}(N) \in ND\}$$

- B_{ND} 必存在一个转移概率（代表某个试验信道）使 $I(U^N; V^N)$ 最小，这个最小值就是 N 次扩展信源 U^N 的信息率失真函数，记为 $R(ND)$ ：

$$R(ND) = \min_{P_{V^N|U^N} \in B_{ND}} I(U^N; V^N) = \min\{I(U^N; V^N); \bar{D}(N) \in ND\}$$

- 若信源和信道均无记忆，则有

$$\begin{aligned} R(ND) &= \min\{I(U^N; V^N); \bar{D}(N) \in ND\} \\ &= \min\{NI(U; V); \bar{D} \in D\} \\ &= NR(D) \end{aligned}$$

- 编码后的信息率 R 就是通过信道的平均互信息量 $I(U; V)$ ，为便于传输和处理，希望将信息率 R 压缩到最小，其最小值就是 $R(D)$ 。
- 若 $R < R(D)$ ，就不能满足保真度准则了

6.2.2 信息率失真函数的性质

1、信息率失真函数的定义域

- $R(D)$ 定义域是 $[0, D_{\max}]$
- 由于 D 是失真度 $d(u_i, v_j)$ 的统计平均，而 $d(u_i, v_j)$ 非负，因此 D 也非负，其下界是零，对应于无失真情况。
- 由 $R(D)$ 的定义式可知其非负，下限值为零，取满足 $R(D)=0$ 的所有 D 中最小者为 $R(D)$ 定义域的上限值 $D_{\max}$ 。
- $R(D)=0$ 意味着 $I(U;V)=0$ ，这时试验信道输入与输出是互相统计独立，即 $P(v_j | u_i) = P(v_j)$ ，这时平均失真为

$$\bar{D} = \sum_{i=1}^r \sum_{j=1}^s P(u_i) P(v_j) d(u_i, v_j)$$

取上式 $\bar{D}$ 的最小值为 $D_{\max}$ ，

$$\begin{aligned} D_{\max} &= \min_{i=1}^r \sum_{j=1}^s P(u_i) P(v_j) d(u_i, v_j) \\ &= \min_{j=1}^s P(v_j) \sum_{i=1}^r P(u_i) d(u_i, v_j) \end{aligned}$$

例 6.2 设输入概率和失真矩阵分别为

$$[P_U] = \begin{bmatrix} 1/3 & 2/3 \end{bmatrix} \quad [d] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{matrix} v_1 & v_2 \\ u_1 & u_2 \end{matrix}$$

求 $D_{\max}$ 。

解 由式(6-2-11)得

$$\begin{aligned} D_{\max} &= \min \sum_{j=1}^2 P(v_j) \sum_{i=1}^2 P(u_i) d(u_i, v_j) \\ &= \min \left\{ P(v_1) \left[\frac{1}{3} \times 0 + \frac{2}{3} \times 1 \right] + P(v_2) \left[\frac{1}{3} \times 1 + \frac{2}{3} \times 0 \right] \right\} \\ &= \min \left\{ \frac{2}{3} P(v_1) + \frac{1}{3} P(v_2) \right\} \end{aligned}$$

根据概率的完备性, 即 $P(v_1) + P(v_2) = 1$, 有

$$D_{\max} = \min \left\{ \frac{1}{3} P(v_1) + \frac{1}{3} \right\}$$

当 $P(v_1) = 0$, 即 $[P_v] = [0 \ 1]$ 时, 得到最小值

$$D_{\max} = \min \left\{ \frac{1}{3} P(v_1) + \frac{1}{3} \right\} = \frac{1}{3} \times 0 + \frac{1}{3} = \frac{1}{3}$$

2、信息率失真函数是 D 的下凸函数

- 若有 I_1 、 I_2 及 D_1 、 D_2 、 D ，满足 $I_1 + I_2 = 1$ 和 $D = I_1 D_1 + I_2 D_2$ ，则有

$$R(D) \leq I_1 R(D_1) + I_2 R(D_2)$$

证明：设 $P_1(v|u)$ 是达到 $R(D_1)$ 的转移概率，设 $P_2(v|u)$ 是达到 $R(D_2)$ 的转移概率，针对同一信源，两种转移概率下的平均互信息量分别设为 $I(U;V_1)$ 和 $I(U;V_2)$ ，则由率失真函数的定义有

$$\begin{aligned} I(U;V_1) &= R(D_1) & E\{d_1(u,v)\} &\leq D_1 \\ I(U;V_2) &= R(D_2) & E\{d_2(u,v)\} &\leq D_2 \end{aligned} \quad (6-2-12)$$

现在定义一个新的转移概率

$$P(v|u) = \lambda_1 P_1(v|u) + \lambda_2 P_2(v|u)$$

在该转移概率下编码器的平均失真为

$$\begin{aligned} E\{d(u_i, v_j)\} &= \sum_{U,V} P(u)P(v | u)d(u, v) \\ &= \sum_{U,V} P(u)[\lambda_1 P_1(v | u) + \lambda_2 P_2(v | u)]d(u, v) \\ &\leq \lambda_1 D_1 + \lambda_2 D_2 = D \end{aligned}$$

设此时编码器的平均互信息量为 $I(U;V)$ 则有

$$R(D) = R(\lambda_1 D_1 + \lambda_2 D_2) \leq I(U;V) \quad (6-2-13)$$

由于平均互信息量是关于转移概率的下凸函数，则得

$$I(U;V) \leq \lambda_1 I(U;V_1) + \lambda_2 I(U;V_2) \quad (6-2-14)$$

综合式(6-2-12)式(6-2-13)和式(6-2-14)即得

$$R(D) \leq \lambda_1 R(D_1) + \lambda_2 R(D_2)$$

3、信息率失真函数是定义域上的非增函数

证明：设 $D_2 \geq D_1$ 则对应试验信道的转移概率有如下包含关系

$$\{P_{V|U}; \bar{D} \leq D_1\} \subset \{P_{V|U}; \bar{D} \leq D_2\}$$

故

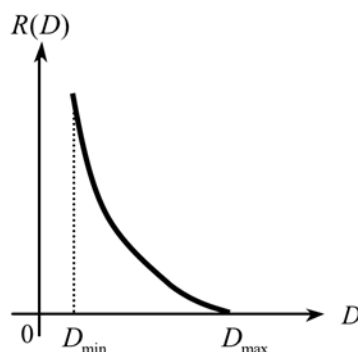
$$\min\{I(U;V), \bar{D} \leq D_2\} \leq \min\{I(U;V), \bar{D} \leq D_1\}$$

即

$$R(D_2) \leq R(D_1)$$

即 $R(D)$ 是定义域上的非增函数。

4、信息率失真函数曲线



由 $R(D)$ 的凸状性质可知 $R(D)$ 在定义域 $[D_{\min}, D_{\max}]$ 上连续又由 $R(D)$ 的非负、下凸、非增及 $R(D_{\max}) = 0$ 这几个性质可知 $R(D)$ 在定义域 $[D_{\min}, D_{\max}]$ 上严格递减，因此 $R(D)$ 是一条连续下凸曲线，由 $R(D_{\min})$ 开始严格递减直至 $R(D_{\max}) = 0$ 为止。

6.3 限失真信源编码定理

- 设离散无记忆平稳信源的信息率失真函数为 $R(D)$ ，只要满足 $R > R(D)$ ，当信源序列 K 足够长时，一定存在一种编码方法，其译码失真小于或等于 $D + \epsilon$ ，其中 ϵ 是任意小的正数；反过来，若 $R < R(D)$ ，则无论采用什么样的编码方法，其译码失真必大于 D 。
- 该定理包含两部分， $R > R(D)$ 的情形称为正定理， $R < R(D)$ 的情形称为逆定理。
- 与香农第二定理（即信道编码定理）一样，限失真信源编码定理只是码的存在性定理。
- 正定理告诉我们 $R > R(D)$ 时，译码失真小于或等于 $D + \epsilon$ 的码肯定存在，但定理本身并未告知码的具体构造方法。

6.4 信息失真函数的计算

6.4.1 离散信源信息率失真函数的参数表示计算方法

1、计算条件

- 已知信息源的概率分布

$$\begin{bmatrix} U \\ P \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_s \\ p(u_1) & p(u_2) & & p(u_s) \end{bmatrix} \quad (6-4-1)$$

失真函数为 $d(u_i, v_j)$, $i=1, \dots, r, j=1, \dots, s$, 选合适的试验信道 $p(v|u)$, 计算平均互信息的极小。

■ 平均互信息表达式

$$I(U;V) = \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) \log \frac{p(v_j | u_i)}{p(v_j)} \quad (6-4-2)$$

其中, $p(v_j)$ 可表示为 $p(v_j) = \sum_{i=1}^r p(u_i) p(v_j | u_i)$

■ 计算的约束条件为

$$\begin{cases} p(v_j | u_i) \geq 0 & (i = 1, 2, \dots, r) (j = 1, 2, \dots, s) \\ \sum_{j=1}^s p(v_j | u_i) = 1 & (i = 1, 2, \dots, r) \\ \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) d(u_i, v_j) = D \end{cases} \quad (6-4-3)$$

2、计算方法

- 应用拉格朗日乘子法，引入乘子 s 和 u_i ，可将求条件极值问题化为无条件极值问题，即求

$$\frac{\partial}{\partial p(v_j | u_i)} [I(U;V) - SD - \mu_i \sum_{j=1}^s p(v_j | u_i)] = 0 \quad (i = 1, 2, \dots, r) (j = 1, 2, \dots, s) \quad (6-4-4)$$

- 分别计算以上3部分的偏导数：

$$\begin{aligned} \frac{\partial I(U;V)}{\partial p(v_j | u_i)} &= \frac{\partial}{\partial p(v_j | u_i)} \left[\sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) \log \frac{p(v_j | u_i)}{p(v_j)} \right] \\ &= \frac{\partial}{\partial p(v_j | u_i)} \left[\sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) \log p(v_j | u_i) - \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) \log p(v_j) \right] \\ &= \frac{\partial}{\partial p(v_j | u_i)} \left[\sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) \log p(v_j | u_i) - \sum_{j=1}^s p(v_j) \log p(v_j) \right] \\ &= p(u_i) \log p(v_j | u_i) + p(u_i) - p(u_i) \log p(v_j) - p(u_i) \\ &= p(u_i) \log \frac{p(v_j | u_i)}{p(v_j)} \end{aligned} \quad (6-4-5)$$

$$\frac{\partial}{\partial p(v_j | u_i)} [SD] = \frac{\partial}{\partial p(v_j | u_i)} \left[S \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) d(u_i, v_j) \right] = S p(u_i) d(u_i, v_j) \quad (6-4-6)$$

$$\frac{\partial}{\partial p(v_j | u_i)} \left[\mu_i \sum_{j=1}^s p(v_j | u_i) \right] = \mu_i \quad (6-4-7)$$



- 将3个偏导的结果代入式（6-4-4）得

$$p(u_i) \log \frac{p(v_j | u_i)}{p(v_j)} - Sp(u_i) d(u_i, v_j) - \mu_i = 0 \quad (i = 1, 2, \dots, r) \quad (j = 1, 2, \dots, s)$$

(6-4-8)

- 解式（6-4-8）得

$$p(v_j | u_i) = p(v_j) \exp\{Sd(u_i, v_j)\} \exp\left\{\frac{\mu_i}{p(u_i)}\right\} \quad (i = 1, 2, \dots, r) \quad (j = 1, 2, \dots, s)$$

(6-4-9)

为计算方便，令 $\lambda_j = \exp\left\{\frac{u_i}{p(u_i)}\right\}$

$$p(v_j | u_i) = \lambda_i p(v_j) \exp\{Sd(u_i, v_j)\} \quad (6-4-10)$$

3、计算 $p(v_j)$ 、 λ_i 和 $p(v_j | u_i)$

利用关系 $\sum_{j=1}^s p(v_j | u_i) = 1$ 和 $p(v_j) = \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i)$ 。

对式 (6-4-10) 求和：

$$1 = \sum_{j=1}^s \lambda_i p(v_j) \exp \{ Sd(u_i, v_j) \} \quad (i = 1, 2, \dots, r) \quad (6-4-11)$$

解得

$$\lambda_i = \frac{1}{\sum_{j=1}^s p(v_j) \exp \{ Sd(u_i, v_j) \}} \quad (i = 1, 2, \dots, r) \quad (6-4-12)$$

对式 (6-4-10) 两边乘以 $p(u_i)$ 求和:

$$p(v_j) = \sum_{i=1}^r p(u_i) \lambda_i p(v_j) \exp \{ S d(u_i, v_j) \} \quad (j = 1, 2, \dots, s)$$

方程两边约去 $p(v_j)$ (假设 $p(v_j) \neq 0$), 得

$$1 = \sum_{i=1}^r p(u_i) \lambda_i \exp \{ S d(u_i, v_j) \} \quad (j = 1, 2, \dots, s) \quad (6-4-13)$$

将式 (6-4-12) 代入式 (6-4-13), 得

$$\sum_{i=1}^r \frac{p(u_i) \exp \{ S d(u_i, v_j) \}}{\sum_{i=1}^s p(v_j) \exp \{ S d(u_i, v_j) \}} = 1 \quad (j = 1, 2, \dots, s) \quad (6-4-14)$$

式 (6-4-14) 是关于 $p(v_j)$ 的有 s 个方程的方程组, 解之可得到以 S 为参量的 s 个 $p(v_j)$ 的值。

将解得的 $p(v_j)$ 代入式 (6-4-12), 得到 r 个 λ_i 的值。

再将 s 个 $p(v_j)$ 和 r 个 λ_i 代入式 (6-4-10), 得到 $s \times r$ 个以 S 为参量的 $p(v_j | u_i)$ 。

4、计算信息率失真函数

平均失真度 $D = \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) d(u_i, v_j)$, 将式 (6-4-10) 代入, 得

$$D(S) = \sum_{i=1}^r \sum_{j=1}^s p(u_i) \lambda_i p(v_j) d(u_i, v_j) \exp \{ S d(u_i, v_j) \} \quad (6-4-15)$$

这里, $p(v_j)$ 和 λ_i 均以 S 为参量, 因此, D 也以 S 为参量。

$$\begin{aligned} R(D) &= \min \{ I(U; V) \} \\ &= \min \left\{ \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) \log \frac{p(v_j | u_i)}{p(v_j)} \right\} \end{aligned}$$

将式 (6-4-10) 代入上式得到以 S 为参量的信息率失真函数的表达式 $R(S)$ 。



$$\begin{aligned} R(S) &= \sum_{i=1}^r \sum_{j=1}^s p(u_i) \lambda_i p(v_j) \exp\{Sd(u_i, v_j)\} \log \frac{\lambda_i p(v_j) \exp\{Sd(u_i, v_j)\}}{p(v_j)} \\ &= \sum_{i=1}^r \sum_{j=1}^s p(u_i) \lambda_i p(v_j) \exp\{Sd(u_i, v_j)\} \log \lambda_i \exp\{Sd(u_i, v_j)\} \\ &= \sum_{i=1}^r \sum_{j=1}^s p(u_i) \lambda_i p(v_j) \exp\{Sd(u_i, v_j)\} [\log \lambda_i + Sd(u_i, v_j)] \\ &= \sum_{i=1}^r p(u_i) \log \lambda_i \left[\sum_{j=1}^s \lambda_i p(v_j) \exp\{Sd(u_i, v_j)\} \right] \\ &\quad + \sum_{i=1}^r \sum_{j=1}^s p(u_i) \lambda_i p(v_j) \exp\{Sd(u_i, v_j)\} Sd(u_i, v_j) \\ &= \sum_{i=1}^r p(u_i) \log \lambda_i \sum_{j=1}^s p(v_j | u_i) + SD(S) \\ &= \sum_{i=1}^r p(u_i) \log \lambda_i + SD(S) \end{aligned}$$

(6-4-16)



5、参量 s 的物理意义

- 除某些特定的情况外，参量 s 一般难以消除，因此很难得到 $R(D)$ 的显式表达式。
- 通过计算 $R(D)$ 的微分，分析参量 s 的物理意义。

$$\frac{dR(D)}{dD} = S \quad (6-4-20)$$



式(6-4-20)表明，参量 S 是信息率失真函数 $R(D)$ 的斜率。由于信息率失真函数 $R(D)$

是 D 的单调递减函数，且是 U 型凸函数，所以 $R(D)$ 的斜率 $S \leq 0$ ， S 随 D 的增加递增，

即 $\frac{dS}{dD} > 0$ 。



斜率 S 与失真度 D ，信息率失真函数 $R(D)$ 与 D 的关系曲线见图 6.3。

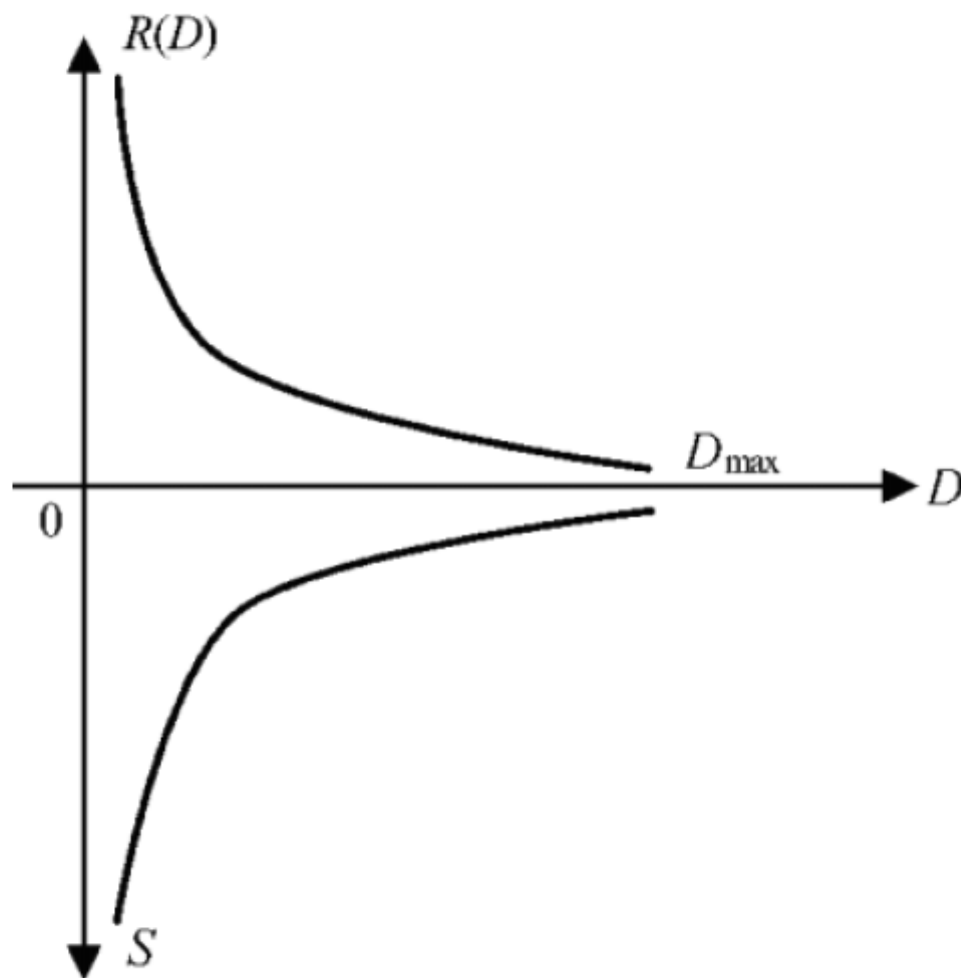


图 6.3 $R(D)$ 及 S 与 D 的关系曲线

当 $D = D_{\min} = 0$ ，式 (6-4-15)，由于失真度 $D(S)$ 是 $r \times s$ 的和，其中 $d(u_i, v_j)$ 、 $p(u_i)$ 、

$p(v_j)$ 和 λ_i 均非负，因此指数项的幂必为负无穷大，即 $S_{\min} \rightarrow -\infty$ 。

当 $D = D_{\max}$ ，参量 S 达到最大值 $S_{\max}$ ，由于 S 的取值为负，所以 $S_{\max} \leq 0$ 。

当 $D > D_{\max}$ ， $R(D) \equiv 0$ ， $\frac{dR(D)}{dD} = 0$ ，因此一般情况下，在 $D = D_{\max}$ 处，参量 S 将从

一个很小的负值跳跃到零！= 在这一点不连续，而在开区间 $(0, D_{\max})$ ， S 是失真度 D 的连续函数。

5、R(D) 参量计算举例

例 6.3 设输入、输出符号集为 $U=V=\{0, 1\}$ ，输入概率密度分布为

$$p(0) = p, p(1) = 1 - p, \text{ 且 } 0 < p \leq \frac{1}{2},$$

失真矩阵为

$$d = \begin{bmatrix} d(u_1, v_1) & d(u_1, v_2) \\ d(u_2, v_1) & d(u_2, v_2) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

求信息率失真函数 $R(D)$ 。



解 (1) 确定 $R(D)$ 的定义域

因为在失真矩阵中, 每行最小值均等于零, 因此最小失真度为

$$D_{\min} = \sum_{i=1}^2 p(u_i) \cdot \left\{ \min_j d(u_i, v_j) \right\} = 0$$

而

$$\begin{aligned} D_{\max} &= \min_j \left\{ \sum_{i=1}^2 p(u_i) \cdot d(u_i, v_j) \right\} \\ &= \min_j \{ p(u_1)d(u_1, v_1) + p(u_2)d(u_2, v_1); p(u_1)d(u_1, v_2) + p(u_2)d(u_2, v_2) \} \\ &= \min_j \{ p(0) \cdot 0 + p(1) \cdot 1; p(0) \cdot 1 + p(1) \cdot 0 \} \\ &= \min_j \{ p(1); p(0) \} \\ &= p \end{aligned}$$

$R(D)$ 的定义域为 $(0, p)$ 。

(2) 求 $\lambda_i (i = 1, 2, \dots, r)$

按式(6-4-11)解方程组

$$\begin{cases} \lambda_1 p(0) \exp \{ Sd(u_1, v_1) \} + \lambda_2 p(1) \exp \{ Sd(u_2, v_1) \} = 1 \\ \lambda_1 p(0) \exp \{ Sd(u_1, v_2) \} + \lambda_2 p(1) \exp \{ Sd(u_2, v_2) \} = 1 \end{cases}$$

方程组简化为

$$\begin{cases} \lambda_1 p + \lambda_2 (1 - p) \exp(S) = 1 \\ \lambda_1 p \exp(S) + \lambda_2 (1 - p) = 1 \end{cases}$$

解得

$$\lambda_1 = \frac{1}{p[1 + \exp(S)]} \quad (6-4-21)$$

$$\lambda_2 = \frac{1}{(1 - p)[1 + \exp(S)]} \quad (6-4-22)$$

(3) 求 $p(v_1), p(v_2)$

按式(6-4-12)有

$$\begin{cases} p(v_1) \exp\{Sd(u_1, v_1)\} + p(v_2) \exp\{Sd(u_1, v_2)\} = \frac{1}{\lambda_1} \\ p(v_1) \exp\{Sd(u_2, v_1)\} + p(v_2) \exp\{Sd(u_2, v_2)\} = \frac{1}{\lambda_2} \end{cases}$$

代入已知条件, 简化为

$$\begin{cases} p(v_1) + p(v_2) \exp(S) = \frac{1}{\lambda_1} \\ p(v_1) \exp(S) + p(v_2) = \frac{1}{\lambda_2} \end{cases}$$

将 λ_1, λ_2 的值代入, 解得

$$p(v_1) = \frac{p - (1 - p) \exp(S)}{1 - \exp(S)} \quad (6-4-23)$$

$$p(v_2) = \frac{(1 - p) - p \exp(S)}{1 - \exp(S)} \quad (6-4-24)$$

(4) 求 $D(S)$

根据(6-4-15), 有

$$\begin{aligned} D(S) = & \lambda_1 p(u_1) p(v_1) d(u_1, v_1) \exp \{ S d(u_1, v_1) \} \\ & + \lambda_1 p(u_1) p(v_2) d(u_1, v_2) \exp \{ S d(u_1, v_2) \} \\ & + \lambda_2 p(u_2) p(v_1) d(u_2, v_1) \exp \{ S d(u_2, v_1) \} \\ & + \lambda_2 p(u_2) p(v_2) d(u_2, v_2) \exp \{ S d(u_2, v_2) \} \end{aligned}$$

代入已知条件

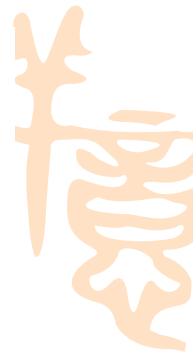
$$D(S) = \lambda_1 p p(v_2) \exp(S) + \lambda_2 (1 - p) p(v_1) \exp(S)$$

代入 $\lambda_1, \lambda_2, p(v_1), p(v_2)$ 的结果, 得

$$D(S) = \frac{\exp(S)}{1 + \exp(S)}$$

从上式可解出参量 S :

$$S = \log \frac{D}{1 - D} \quad (6-4-25)$$



(5) 求 $R(D)$

根据式(6-4-16), 有

$$\begin{aligned} R(D) &= \sum_{i=1}^2 p(u_i) \log \lambda_i + SD \\ &= p(u_1) \log \lambda_1 + p(u_2) \log \lambda_2 + SD \\ &= p \log \lambda_1 + (1-p) \log \lambda_2 + SD \end{aligned}$$

将式(6-4-21)、(6-4-22)和(6-4-25)代入上式得

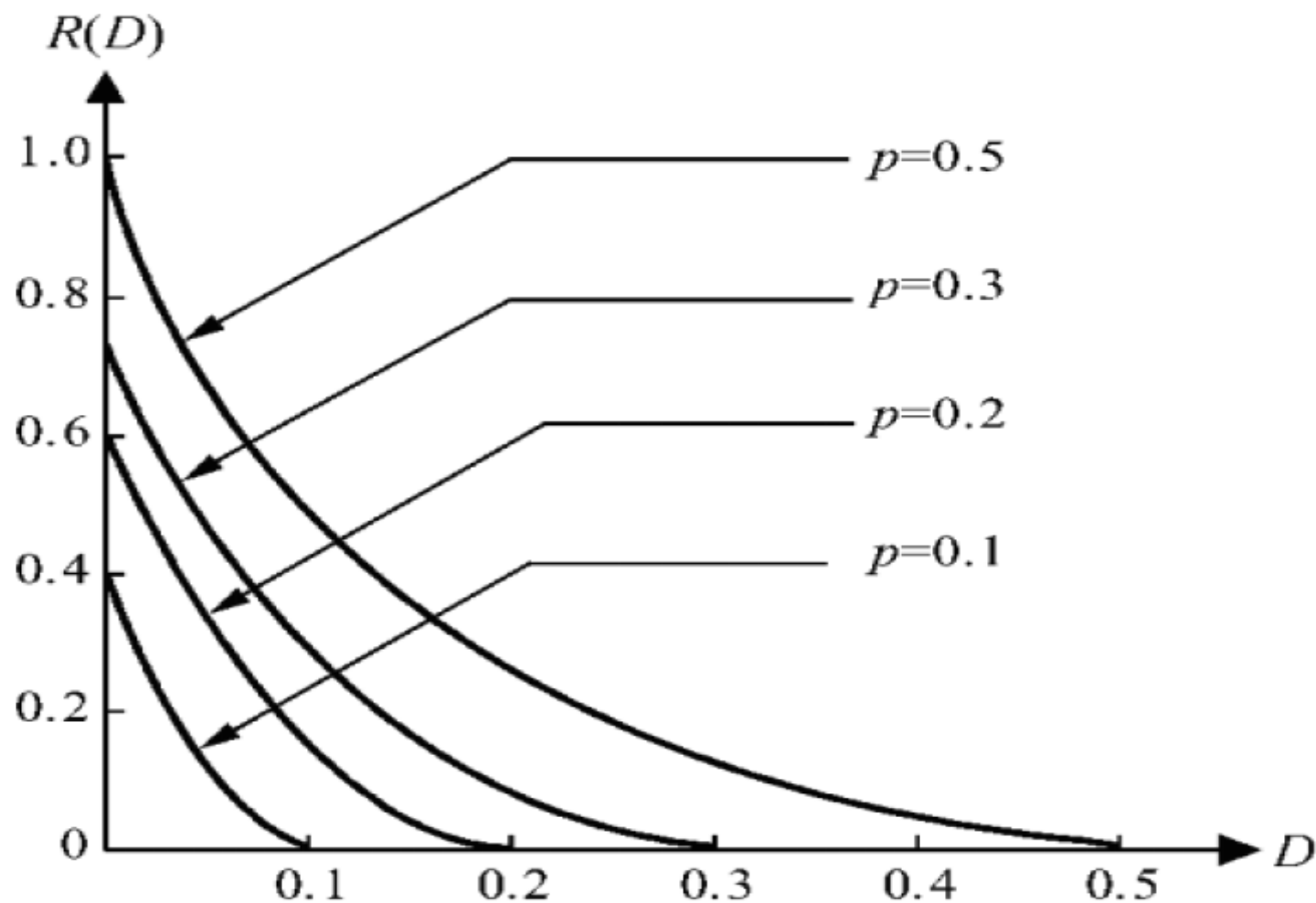
$$\begin{aligned} R(D) &= p \log \left\{ \frac{1}{p[1 + \exp(S)]} \right\} + (1-p) \log \left\{ \frac{1}{(1-p)[1 + \exp(S)]} \right\} + D \log \frac{D}{1-D} \\ &= -p \log p - p \log \frac{1}{1-D} - (1-p) \log(1-p) - (1-p) \log \frac{1}{1-D} + D \log \frac{D}{1-D} \\ &= [-p \log p - (1-p) \log(1-p)] + [D \log D + (1-D) \log(1-D)] \\ &= H(p, 1-p) - H(D, 1-D) \end{aligned}$$

由于 $R(D)$ 的定义域为 $(0, p)$, 因此 $R(D)$ 的完整表达式为

$$R(D) = \begin{cases} H(p, 1-p) - H(D, 1-D) & 0 \leq D \leq p \leq \frac{1}{2} \\ 0 & D \geq p \end{cases}$$



$R(D)$ 的曲线见图 6.4。



通过曲线可知，对于给定的平均失真度 D ，信源分布越均匀（ p 接近 0.5），信源的冗余度越小， $R(D)$ 就越大，信源压缩的可能性就越小。反之，如果信源分布越不均匀，信源的冗余度越大， $R(D)$ 就越小，信源压缩的可能性就越大。

例 6.4 r 无对称信源，输入符号集为 $U = \{u_1, u_2, \dots, u_r\}$ ，信源符号等概分布

$p(u_i) = \frac{1}{r}, (i = 1, 2, \dots, r)$ ，输出符号集为 $V = \{v_1, v_2, \dots, v_r\}$ ，失真函数为

$$d(u_i, v_j) = \begin{cases} 1 & (i \neq j) \\ 0 & (i = j) \end{cases} \quad (i, j = 1, 2, \dots, r) \text{ 求信息率失真函数。}$$

解 (1) $R(D)$ 的定义域

$$D_{\min} = \sum_{i=1}^r p(u_i) \min_j d(u_i, v_j) = 0$$

$$D_{\max} = \min_j \left\{ \sum_{i=1}^r p(u_i) \cdot d(u_i, v_j) \right\} = 1 - \frac{1}{r}$$

(2) 计算 $\lambda_i, (i = 1, 2, \dots, r)$

依题意，失真矩阵为

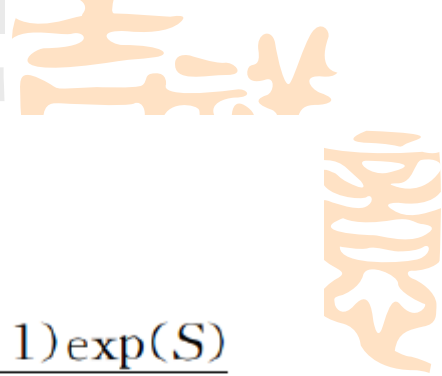
$$d = \begin{bmatrix} d(u_1, v_1) & \cdots & d(u_1, v_r) \\ \vdots & & \vdots \\ d(u_r, v_1) & \cdots & d(u_r, v_r) \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 0 \end{bmatrix}$$

根据式(6-4-11)列方程组

$$\begin{cases} \lambda_1 + \lambda_2 \exp(S) + \cdots + \lambda_r \exp(S) = r \\ \lambda_1 \exp(S) + \lambda_2 + \cdots + \lambda_r \exp(S) = r \\ \vdots \\ \lambda_1 \exp(S) + \lambda_2 \exp(S) + \cdots + \lambda_r = r \end{cases}$$

解得

$$\lambda_i = \frac{r}{1 + (r-1)\exp(S)} \quad (i = 1, 2, \dots, r)$$



(3) 计算 $p(v_j), (j = 1, 2, \dots, r)$

按式(6-4-12)列方程组

$$\begin{cases} p(v_1) + p(v_2)\exp(S) + \dots + p(v_r)\exp(S) = \frac{1 + (r-1)\exp(S)}{r} \\ p(v_1)\exp(S) + p(v_2) + \dots + p(v_r)\exp(S) = \frac{1 + (r-1)\exp(S)}{r} \\ \vdots \\ p(v_1)\exp(S) + p(v_2)\exp(S) + \dots + p(v_r) = \frac{1 + (r-1)\exp(S)}{r} \end{cases}$$

解得

$$p(v_j) = \frac{1}{r} \quad (j = 1, 2, \dots, r)$$



(4) 求 $D(S)$

根据式(6-4-15), 有

$$\begin{aligned} D(S) &= \sum_{i=1}^r \sum_{j=1}^r \lambda_i p(u_i) p(v_j) d(u_i, v_j) \exp \{ S d(u_i, v_j) \} \\ &= \frac{r}{1 + (r-1) \exp(S)} \cdot \frac{1}{r} \cdot \frac{1}{r} \exp(S) \cdot (r^2 - r) \\ &= \frac{(r-1) \exp(S)}{1 + (r-1) \exp(S)} \end{aligned}$$

从上式可解出参量 S

$$S = \log \frac{D}{(r-1)(1-D)}$$



(5) 求 $R(D)$

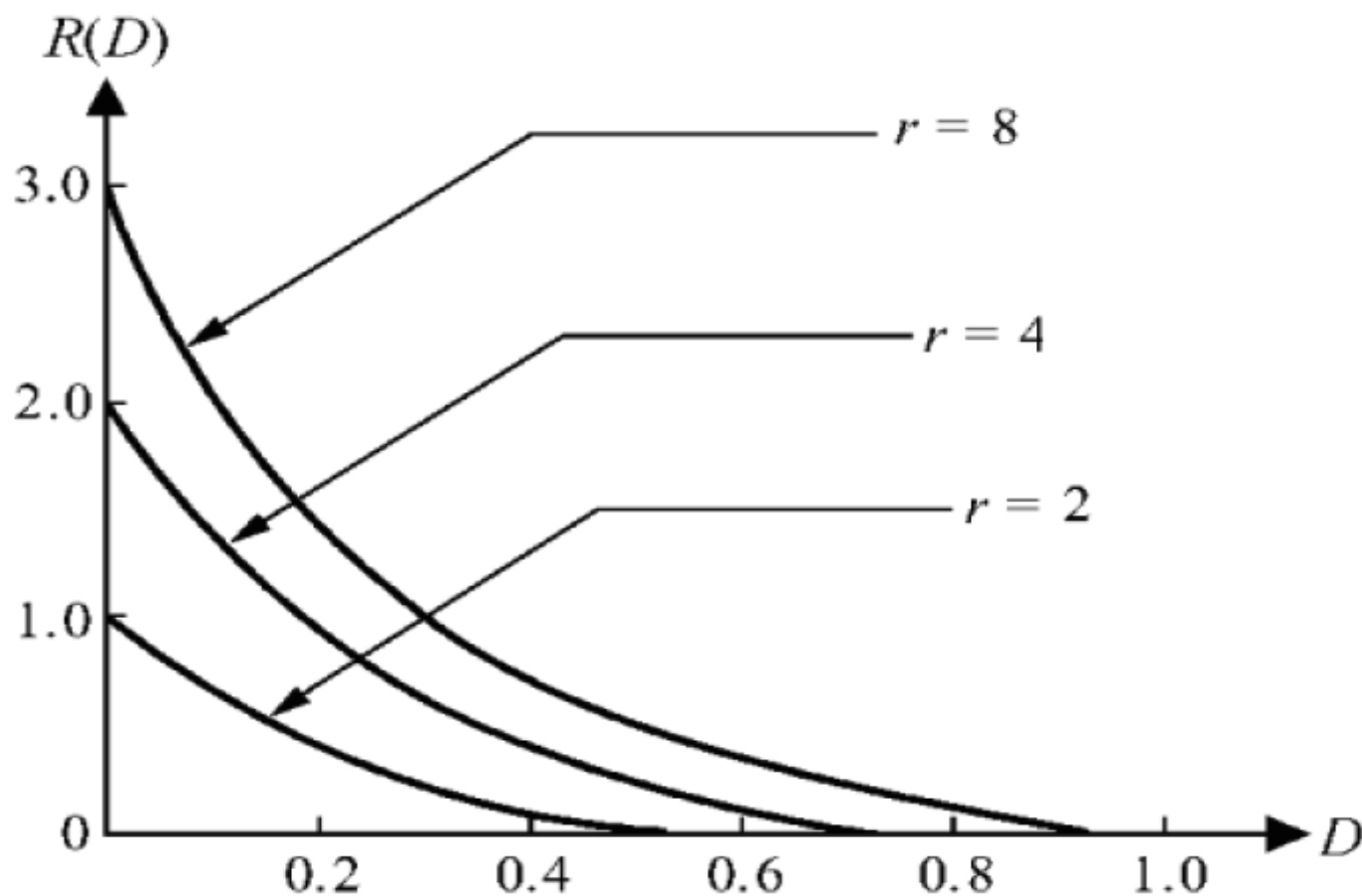
根据式(6-4-16), 有

$$\begin{aligned} R(D) &= \sum_{i=1}^r p(u_i) \log \lambda_i + SD \\ &= \log \left\{ \frac{r}{1 + (r-1) \exp(S)} \right\} + D \log \frac{D}{(r-1)(1-D)} \\ &= \log r - \log \frac{1}{1-D} + D \log D - D \log(r-1) - D \log(1-D) \\ &= \log r - D \log(r-1) + D \log D + (1-D) \log(1-D) \end{aligned}$$





$R(D)$ 的定义域为 $(0, 1 - \frac{1}{r})$ ，值域为 $0 \leq R(D) \leq \log r$ ，其关系曲线见图 6.5。



由图可见，对应于相同的失真度 D ， r 越大，即信源的分层数越多，信源的可压缩性就越小。反之， r 越小，即信源的分层数越少，信源的可压缩性就越大。



6.4.1 离散信源信息率失真函数的迭代计算方法

1、迭代计算条件

设离散信源输入序列为

$$\begin{bmatrix} U \\ P \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_r \\ p(u_1) & p(u_2) & & p(u_r) \end{bmatrix}$$

输出序列为

$$\begin{bmatrix} V \\ P \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \cdots & v_s \\ p(v_1) & p(v_2) & & p(v_s) \end{bmatrix}$$

失真函数为 $d(u_i, v_j), i = 1, 2, \dots, r, j = 1, 2, \dots, s$ 。信息率失真函数 D 是平均互信息 $I(U; V)$ 的极小值。

即在满足约束条件

$$\begin{cases} p(v_j | u_i) \geq 0 & (i = 1, 2, \dots, r)(j = 1, 2, \dots, s) \\ \sum_{j=1}^s p(v_j | u_i) = 1 & (i = 1, 2, \dots, r) \\ \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) d(u_i, v_j) = D \end{cases}$$

下, 计算双重极小化

$$R(D) = \min_{p(v_j)} \min_{p(v_j | u_i)} I(U; V) = \min_{p(v_j)} \min_{p(v_j | u_i)} \sum_{i=1}^r \sum_{j=1}^s p(u_i) p(v_j | u_i) \log \frac{p(v_j | u_i)}{p(v_j)}$$



2、迭代计算公式

(1) 计算 $p^*(v_j | u_i) (i=1,2,\dots,r; j=1,2,\dots,s)$

$p^*(v_j | u_i)$ 为使 $I(U;V)$ 达到极小值的试验信道，计算 $p^*(v_j | u_i)$ 时，假设

$p(v_j) (j=1,2,\dots,s)$ 固定不变，与 $p(v_j | u_i)$ 无关。

$$p(v_j | u_i)^* = \frac{p(v_j) \exp[Sd(u_i, v_j)]}{\sum_{j=1}^s p(v_j) \exp[Sd(u_i, v_j)]} \quad (6-4-31)$$

(2) 计算 $p^*(v_j)(j=1,2,\dots,s)$

$p^*(v_j)$ 为使 $I(U;V)$ 达到极小值的信宿的概率分布, 计算 $p^*(v_j)$ 时, 把 $p^*(v_j)$ 看成独立变量, 即假设 $p(v_j | u_i)(i=1,2,\dots,r; j=1,2,\dots,s)$ 固定不变与 $p(v_j)$ 无关。

$$p(v_j)^* = \sum_{i=1}^r p(u_i) p(v_j | u_i)$$

(6-4-33)

3、迭代计算步骤

(1) 假定 S_1 为一相当大的负值，选定起始传递概率

$p^{(1)}(v_j | u_i) (i=1, 2, \dots, r, j=1, 2, \dots, s)$ ，可取 $p^{(1)}(v_j | u_i) = 1/s$ 。将选定的

$p^{(1)}(v_j | u_i)$ 代入式 (6.4.33)，得到 $p^{(1)}(v_j)$ 。

 (2) 将 $p^{(1)}(v_j)$ 代入式 (6.4.31)，得到 $p^{(2)}(v_j | u_i)$ 。

(3) 再将 $p^{(2)}(v_j | u_i)$ 代入式 6.4.33)，得到 $p^{(2)}(v_j)$ 。



(4) 重复前面的步骤, 计算出第 k 次和第 $k+1$ 次的 $R^{(k)}(D)$ 和 $R^{(k+1)}(D)$ 。

$$R^{(k)}(D) = \sum_{i=1}^r \sum_{j=1}^s p(u_i) p^{(k)}(v_j | u_i) \log \frac{p^{(k)}(v_j | u_i)}{p^{(k)}(v_j)}$$

$$R^{(k+1)}(D) = \sum_{i=1}^r \sum_{j=1}^s p(u_i) p^{(k+1)}(v_j | u_i) \log \frac{p^{(k+1)}(v_j | u_i)}{p^{(k+1)}(v_j)}$$

(5) 当 $R^{(k)}(D)$ 与 $R^{(k+1)}(D)$ 的差别小于预先给定的值, 取 $R^{(k)}(D)$ 或 $R^{(k+1)}(D)$ 作为 $R(S_1)$ 的近似值。

- (6) 选略大一些的负数 S_2 ，重复上述迭代过程，得到对应于 S_2 的 $R(S_2)$ 。
- (7) 重复上述过程，直到 $R(S_{\max})$ 逼近零。
- (8) 根据 S_1 、 S_2 、...、 $S_{\max}$ ，以及与之相对应的 $R(S_1)$ 、 $R(S_2)$ 、...、 $R(S_{\max})$ ，可以作为 $R(S)$ 曲线。根据 $R(D)$ 函数的 S 参量表述理论，进而得到 $R(D)$ 曲线，完成整个迭代过程。

