总复习

2023年12月22日　　19:14

第一讲 绪论

香农的信息与信息熵

信息是用来减少不确定性的东西。或者，信息是能够使熵减少的东西。Shannon信息熵是系统不确定性的度量。

布里渊认为：信息就是负熵。由此驱除了麦克斯韦妖(Maxwell's Demon)。麦克斯韦妖不可能不费任何能量地独立获知分子的运动状态信息，从而将不同运动速度的气体分子控制在不同的区域。而获取信息必然需要外界提供额外的能量。

香农信息论

基本问题：信源编码问题；信道编码问题

答案：编码速率大于信源的熵速率时可实现渐进无差错表达：给定失真限界，存在使失真不超过该限界的最小编码速率。当传信率比信道容量高时，必然发生发送差错，否则可以实现渐进无差错传输。

图灵测试

判定机器是否有智能的依据：用测试对象能理解的语言问问题，若干询问后不能区别，则通过。

反向图灵测试

判定交互对象是否是自动机器的依据：测试者为机器，生成人类能回答但机器不能回答的问题。如果A具有足够智能回答，则通过反向测试。

第二讲 信息的基本概念、度量和属性

简单概念

事件的自信息(可正可负可零)；条件自信息；联合自信息；互信息；条件互信息；联合互信息。

熵：平均自信息；疑义度；联合熵

相对熵

熵的性质（P13）

1. $H_K(P)$对概率矢量P的分量对称
2. 非负性
3. 确定性
4. 可扩展性
5. 可加性

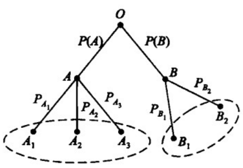


图 2.1.3　熵的可加性

6. 极值性
7. 严格凸函数$H_K(\theta P_1 + (1-\theta)P_2) > \theta H_K(P_1) + (1-\theta)H_K(P_2)$

熵的链式法则：$H(X_1, X_2, \dots, X_n) = H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_{n-1}, \dots, X_1)$

信息的链式法则：$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$

相对熵

相对熵可以衡量两个分布之间的不相似性，即 P 和 Q 两个分布越相似，相对熵越小，否则越大。

$$D(p//q) = \sum p\,(x) log \frac{p(x)}{q(x)} \geq 0$$

疑义度

错误概率(就是不相等的概率之和)

$$P_E = \sum_{\substack{k=0 \\ k\neq j}}^{K-1} \sum_{j=0}^{K-1} P\{X = k, \hat{X} = j\}$$

$H(X|\hat{X}) \leq H(P_E) + P_E log(K-1)$

pf:

$$\begin{aligned}
& H(X|\hat{X}) - H(P_E) - P_E \log(K-1) \\
&= -\sum_{k=0}^{K-1}\sum_{j=0}^{K-1} p(k,j)\log p(k|j) + \sum_{\substack{k=0 \\ k\neq j}}^{K-1}\sum_{j=0}^{K-1} p(k,j)\log P_E + \\
&\quad \sum_{k=0}^{K-1} p(k,k)\log(1-P_E) - \sum_{\substack{k=0 \\ k\neq j}}^{K-1}\sum_{j=0}^{K-1} p(k,j)\log(K-1) \\
&= \sum_{\substack{k=0 \\ k\neq j}}^{K-1}\sum_{j=0}^{K-1} p(k,j)\log\frac{P_E}{(K-1)p(k|j)} + \sum_{k=0}^{K-1} p(k,k)\log\frac{1-P_E}{p(k|k)} \\
&\leq \sum_{\substack{k=0 \\ k\neq j}}^{K-1}\sum_{j=0}^{K-1} p(k,j)\left[\frac{P_E}{(K-1)p(k|j)} - 1\right] + \sum_{k=0}^{K-1} p(k,k)\left[\frac{1-P_E}{p(k|k)} - 1\right] \\
&= \frac{P_E}{K-1}\sum_{\substack{k=0 \\ k\neq j}}^{K-1}\sum_{j=0}^{K-1} p(j) + (1-P_E)\sum_{j=0}^{K-1} p(j) - \sum_{k=0}^{K-1}\sum_{j=0}^{K-1} p(k,j) \\
&= \frac{P_E}{K-1}(K-1)\sum_{j=0}^{K-1} p(j) + (1-P_E) - 1 = 0
\end{aligned}$$

物理意义：已知$\hat{X}$条件下对X还存在的不确定性$H(XI\hat{X})$可分为两部分。第一部分是估计X是否正确，如果是正确的，则从$\hat{X}$可得到X的精确值，这部分不确定性为$H(P_E)$;第二部分,如果估计是不正确的,这时X的可能取值范围是除去估计X已取的值以外的其他$K-1$个值,于是这部分不确定性不会大于$P_E log(K-1)$。

马尔科夫链

性质：条件独立，而不是绝对独立。

$$p(x,z|y) = p(x|y)p(z|y)$$

$$I(X;Z|Y) = 0$$

$then$

$$I(X;Y) \geq I(X;Z)$$

$$I(X;Y) \geq I(X;Y|Z)$$

(数据处理定理:表明对于观察到的数据，增加数据处理的次数，并不会使信息量增加。)

固定转移概率矩阵，熵是输入分布的上凸函数，可以用来计算信道容量；固定输入分布，熵是转移概率矩阵的下凸函数，可以用来计算互信息的最小值。

微分熵

二维正态分布的互信息：$I(X,Y) = -1/2\, ln(1-\rho^2)$

定义微分熵：$H_C(X) = -\int p\,(x) logp(x)dx$可以 < 0

微分熵不反映连续随机变量X的不确定性（一般无穷大）；但能反映相对不确定性；不具备线性变换不变性

极大化

峰值受限：设X取值限于$(-M,M)$，则$H_C(X) \le ln(2M)$，均匀分布时取等。

平均功率受限：σ^2一定，X服从正态分布时，微分熵最大，即$H_C(X) \le ln(\sqrt{2\pi e}\sigma)$

离散平稳源

任意长度的片段的联合概率分布与时间起点无关，只与关联长度有关。

性质：

- $H(X_N|X_1, X_2, \dots, X_{N-1})$随$N$增大单调不增
- $H_N(X)$也随N增大单调不增
- $H_N(X) \ge H(X_N|X_1, \dots, X_{N-1})$
- $\lim_{N\to\infty} H_N(X) = \lim_{N\to\infty} H(X_N|X_1, \dots, X_{N-1})$
- 输出序列的符号熵随序列长度增加单调不增，但总大于熵速率
- 冗余度$R = 1 - \eta = 1 - \frac{H_\infty}{H_0}$

马尔科夫源

马尔科夫信源的熵率等于信源在各状态下的条件熵对状态概率求平均。

$H_\infty = \sum \mu_i\, H(X|s_i)$

时齐(时不变)马尔科夫源：状态转移概率p(n)与时间n无关。

既约（不可约）马尔科夫源：从任一状态出发，经有限步总可以到达任一其它状态。

第三讲 信息感知与压缩

DMS等长编码

绝对无差错等长编码：$N \ge \frac{LlogK}{logD}$，即$R = N/L\, logD \ge logK$

输出的消息序列长L，输出字符表大小K，编码字符表大小D，码字长度N

渐进均分

$\frac{1}{L} I(u^L) \to H(U)$

典型列

平均自信息满足$I_L \in \dot{U}(H(U), \epsilon)$的属于典型列

- $P(u^L) \approx 2^{-LH(U)}$
- 典型列数目满足：$(1-\epsilon)2^{L[H(U)-\epsilon]} \le |A_\epsilon^{(L)}| \le 2^{L[H(U)+\epsilon]}$

- 非典型列的概率不一定低于典型列
- 非典型列的总概率小，但绝对数目不一定小

DMS不等长编码

信源编码的基本要求：唯一可译和即时可译

唯一可译性：码的任意扩展都是非奇异的，则称码是唯一可译的。

判据：没有任何一个后缀分解集包含码字。若S_1是空集，即时可译、唯一可译。(唯一码才分即时不即时)

Kraft不等式：存在长度为$n_1,\ldots,n_k$的D元异字头码(没有一个码字是其他码字的前缀，前缀码)的充要条件是$\sum D^{-n_k}\leq 1$

如果一个码是唯一可译 -> Kraft不等式成立 -> 存在一个具有同样长度的异字头码。

不等长编码定理：$\frac{H(U)}{logD}\leq\bar{n}\leq\frac{H(U)}{logD}+1$

不等长编码编码速率：$R=\bar{n}logD>H(U)$，存在唯一可译不等长码

信源编码的剩余度来自两个方面：信源符号间的相关性；信源符号分布的不均匀性。

编码效率 $=\frac{\bar{n}logD}{H(U)}$

最佳不等长编码(Huffman编码)

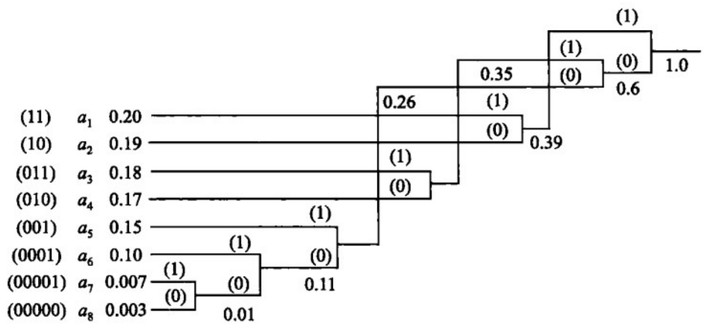


图 3. 3. 1　二元 Huffman 编码

D元Huffman算法：消息总数$K=(D-1)\cdot i+1$，否则需要对K增补

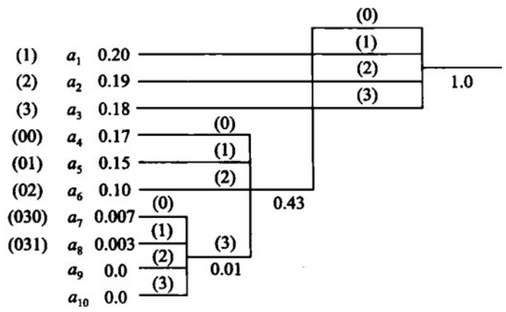


图 3. 3. 2　四元 Huffman 编码

Shannon编码

将概率从大到小排列，定义$P_k=\sum_{i=1}^{k-1}P_i$，用$l_k=\lceil -logp_k\rceil$个比特表示P_k

Fano编码

将消息分成两大组（概率排列不变），每组概率和尽可能相等。一组0一组1；再等概分，以此类推。

$$U=\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.3 & 0.25 & 0.2 & 0.15 & 0.05 & 0.05 \end{pmatrix}$$

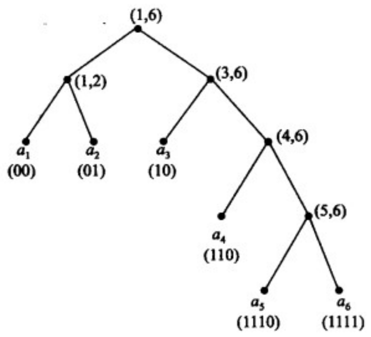


图 3.3.3 Fano 编码的编码树

Fano编码的平均码长 $\le H(U)+1-2p_n$

SFE算法

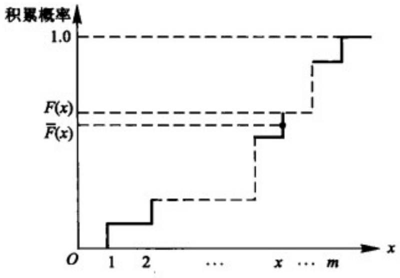


图 3.3.5 积累概率分布 $F(x)$ 和 $\bar{F}(x)$ 图示

截断$\bar{F}(x)$的位数$l(x)=[-log(p(x))]+1$

算数码

1. 置$p(\lambda)=1,\ F(\lambda)=0$
2. $F(x^i)=F(x^{i-1})+p(x^i)c(x)$

   c(x)=(x==1)?p(0):0
3. 截断$F(x^n)$的$[-log(p(x))]$位

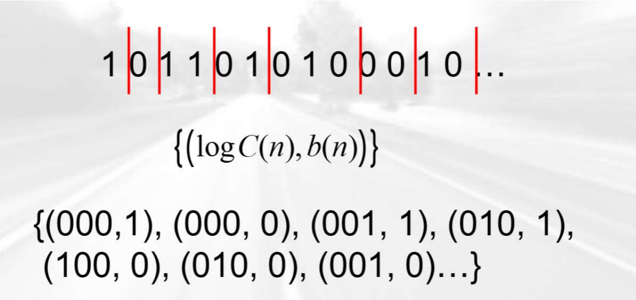
表 3.3.5 例 3.3.7 的编码过程

输入符号	$p(x^i)$	$l(x^i)$	$F(x^i)$	$C(x^i)$
	$p(\lambda)=1$	0	$F(\lambda)=0$	
1	0.11	1	0.01	0.1
1	0.1001	1	0.0111	0.1
1	0.011011	2	0.100101	0.11
0	0.00011011	4	0.100101	0.1010
1	0.0001010001	4	0.1001101011	0.1010
0	0.000001010001	6	0.1001101011	0.100111
1	0.00000011110011	7	0.10011100000001	0.1001111
1	0.0000001011011001	7	0.1001110011110111	0.1001111

Z-L算法

相异分解：C_n表示每个码被相异分解（不重复）的片段数；平均每个信源符号需要的比特数为$\frac{C(n)[logC(n)+1]}{n}$

当序列充足长时，说明二进制序列位的长度由最长码长度决定。



马尔科夫信源编码

算稳态概率和各状态的条件熵。

以Huffman编码为例：

表 3.4.1　图 3.4.1 中 4 状态马尔可夫信源的 Huffman 编码

消息	A	B	C	D
00	0 (0.81)	0 (0.45)	10 (0.25)	111 (0.05)
01	10 (0.09)	111 (0.05)	110 (0.25)	110 (0.05)
10	110 (0.05)	110 (0.25)	111 (0.05)	10 (0.09)
11	111 (0.05)	10 (0.25)	0 (0.45)	0 (0.81)
平均码长 $\bar{n}_L$	1.29	1.85	1.85	1.29

对每个状态单独编码，各自得到平均码长，总体的平均码长等于稳态概率乘平均码长。

第四讲 通信理论

DMC信道容量

$C = maxI(X;Y) = H(Y) - H(Y|X)$

- 无噪信道（一一对应）：$H(X|Y) = 0$，$C = logM$
- 无损信道（一对多）：$H(X|Y) = 0$，$C = logM$
- 确定信道（多对一）：$H(Y|X) = 0$，$C = logm$
- 无用信道：$H(X) = H(X|Y)$，$C = 0$

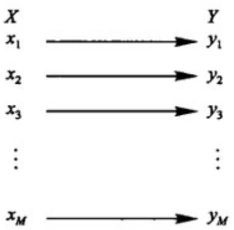


图 4.2.2　无噪信道的概率转移图

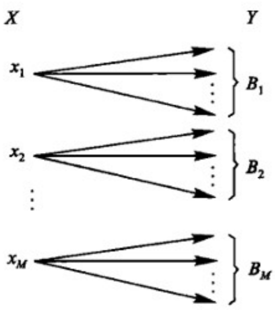


图 4.2.3　无损信道的概率转移图

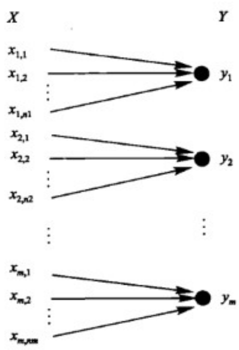


图 4.2.4　确定信道的概率转移图

- BSC信道：$C = 1 - H(p)$
- BEC信道：$C = 1 - \alpha$

- 对称（准对称）信道：$C=I_{等概分布}(X;Y)$

信道容量定理：输入概率矢量Q_k达到DMC的容量的充要条件为$I(X=k;Y)=C \quad \forall k,Q_k>0; I(X=k;Y)\le C \quad \forall k,Q_k=0$

> 说明，达到离散无记忆信道容量时，发送符号集中的每个符号虽然被利用的概率不一定相同，但传送的信息量必然相等。

信道组合

平行信道（积信道）

定义$p(jj'|kk')=p_1(j|k)p_2(j'|k')$，$C=\sum C_i$

和信道

以概率P_A选用A信道，概率$\overline{P_A}$选用B信道，$C=log(\sum 2^{C_i})$

各信道利用概率$P_n=2^{C_n-C}$

级联信道

前一个信道的输出作为下一个信道的输入，构成马尔科夫链$X\to Y\to Z$

先求总的转移概率矩阵，变成普通信道。

DMC编码定理

二元对称信道编码定理

所有低于信道容量C的速率R均是可达的，即当$R<C$时，总存在一系列码$(2^{nR},n)$，当$n\to\infty$，最大误码概率$\lambda^{(n)}\to 0$

> Hamming距离：$r=n(p+\epsilon_2)$

联合典型列

> X、Y独立同分布。

- $P(u^L)\approx 2^{-L[I(X;Y)]-3\epsilon}$
- 典型列数目满足：$(1-\epsilon)2^{L[H(X,Y)-\epsilon]}\le |A_\epsilon^{(L)}|\le 2^{L[H(X,Y)+\epsilon]}$

加性高斯噪声信道

$Y_i=X_i+Z_i \qquad Z_i\sim N(0,N)$
$I(X;Y)=h(Y)-h(Z)\le\frac{1}{2}log(1+\frac{P}{N})$，$P$为最大功率，$N$为$Z$的方差。

当$X\sim N(0,P)$时取等号。

信噪比$=10log(\frac{P}{N})$，最小信噪比为ln2dB

高斯平行信道

k个独立、平行具有共同功率限制的高斯信道

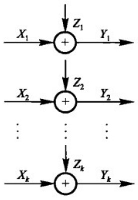


图 4.5.4　平行高斯信道

$C=\frac{1}{2}\sum log(1+\frac{P_i}{N_i})$，$(X_1,X_2,\ldots,X_k)\sim N\left\{0,\begin{pmatrix}P_1&0&0&0\\0&P_2&0&0\\0&0&P_3&0\\0&0&0&P_4\end{pmatrix}\right\}$

采用注水法则确定P_i，

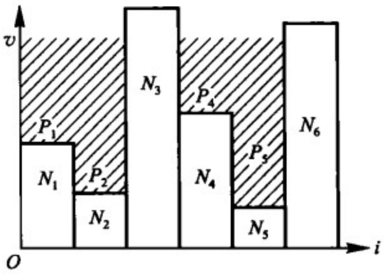
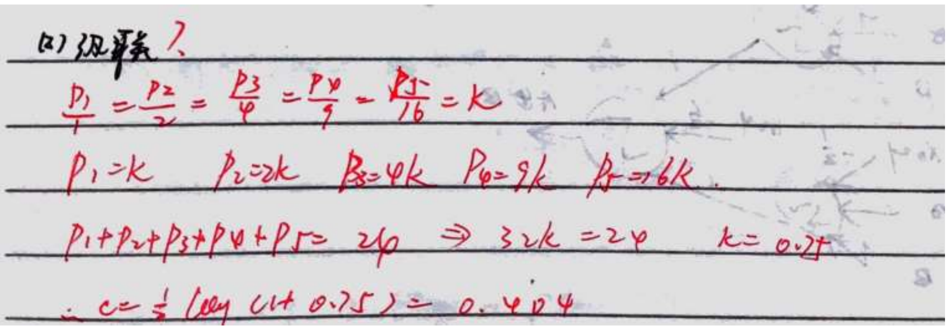


图 4.5.5　求解方程(4.5.31)的灌水法则

在总功率分配给各个子信道时，首先把功率分配给噪声最小的信道，然后分配给第二小噪声信道，就像灌水一样进行，直到水涨到一定的水平 v，使总的灌水量等于总功率 P。如图 4.5.5 所示，有 6 个信道，先把功率分配给噪声最小的第 5 信道，然后分配给第 2 信道。像灌水一样，直到水涨到某一水平 v，这时 $P_1+P_2+P_4+P_5=P$。这时第 3 信道和第 6 信道的噪声水平大于 v，所以不分配任何功率给这两个信道。



(2) 级联

$\frac{P_1}{1}=\frac{P_2}{2}=\frac{P_3}{4}=\frac{P_4}{9}=\frac{P_5}{16}=k$

$P_1=k$　$P_2=2k$　$P_3=4k$　$P_4=9k$　$P_5=16k$

$P_1+P_2+P_3+P_4+P_5=24 \Rightarrow 32k=24$　$k=0.75$

$\therefore C=\frac{1}{2}\log(1+0.75)=0.404$

模拟信道的信道容量

带限、加性白高斯噪声信道

$$C_T=\frac{1}{2}\sum_{i=1}^{2WT} log(1+\frac{P_i}{N_i})=Wlog\{1+\frac{P}{N_0W}\}=\frac{P}{N_0}log\,e(W\to\infty)$$

频带效率$\eta=\frac{每秒传输速率（R）}{传输带宽（W）}\le log(1+\frac{P}{N_0W})$

每比特能量$E_b(\eta)\ge\frac{2^{\eta}-1}{\eta}\cdot N_0\to N_0ln2(\eta\to 0)$，最小比特能量与噪声功率谱密度之比为$ln2$

第五讲 率失真

D_{max}表示无须对信源作任何描述时，可以达到的最小失真，也就是用零比特量化时可达到的最小平均失真。

Hamming失真度量下的贝努利信源

贝努利信源(二元无记忆，输出1的概率为p的率失真函数为

$$R(D)=\begin{cases}H(p)-H(D) & 0\le D\le min(p,1-p)\\ 0 & D\ge min(p,1-p)\end{cases}$$

高斯信源

在平方误差失真度量下，高斯信源$X\sim N(0,\sigma^2)$的率失真函数为

$$R(D)=\begin{cases}\frac{1}{2}log\frac{\sigma^2}{D} & 0\le D\le\sigma^2\\ 0 & D>\sigma^2\end{cases}$$

<u>逆注水法则</u>

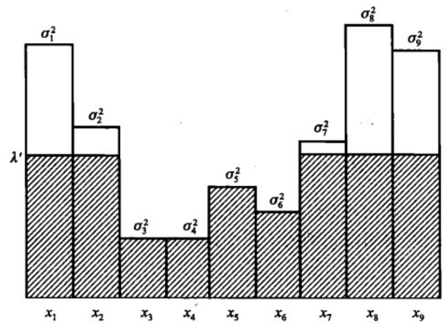


图 5.2.6 分量独立的高斯矢量信源的逆灌水法则

率失真函数的性质

- 向下凸性：$R(D^\theta)\le\theta R(D')+(1-\theta)R(D'')$
- 单调递减、连续

<u>利用信源的对称性计算率失真</u>

1. 转移概率矩阵的分布与失真度量矩阵具有相同的对称性
2. $D=\sum p(x)\cdot q(\hat{x}|x)\cdot d(x,\hat{x})$，计算$R(D)=I(X;Y)$得到率失真曲线

[例 5.3.3] 设二元等概信源$\begin{bmatrix}X\\ p(x)\end{bmatrix}=\begin{bmatrix}x_0 & x_1\\ 0.5 & 0.5\end{bmatrix}$，再生字符表为$\hat{\mathscr{X}}=\{\hat{x}_0,\hat{x}_1,\hat{x}_2\}$，失真度量

矩阵为 $(d)_{2\times3}=\begin{pmatrix}0 & \infty & 1\\ \infty & 0 & 1\end{pmatrix}$，求率失真函数 $R(D)$。

[解] 首先 $D_{\min} = \sum_x p(x) \min_{\hat{x}\in\hat{X}} d(x,\hat{x}) = 0$

$$D_{\max} = \min_{\hat{x}} \sum_x p(x) d(x,\hat{x}) = 1$$

由定理 5.3.1，存在着与失真度量矩阵具有同样对称性的转移概率分布达到率失真 $R(D)$，这个转移概率函数可以写为

$$Q=\begin{pmatrix}\alpha & \beta & \gamma\\ \beta & \alpha & \gamma\end{pmatrix}, \qquad \alpha+\beta+\gamma=1$$

由于 $d(x_0,\hat{x}_1)=d(x_1,\hat{x}_0)=\infty$，所以对于任何有限平均失真，必须 $\beta=0$。于是转移概率矩阵为

$$Q=\begin{pmatrix}\alpha & 0 & 1-\alpha\\ 0 & \alpha & 1-\alpha\end{pmatrix}$$

在失真限制下

$$\begin{aligned}D &= \sum_{x,\hat{x}} p(x)\cdot q(\hat{x}\mid x)\cdot d(x,\hat{x})\\ &=1-\alpha\end{aligned}$$

所以 $\alpha = 1 - D$

相应的检验信道如图 5.3.2 (a)所示。

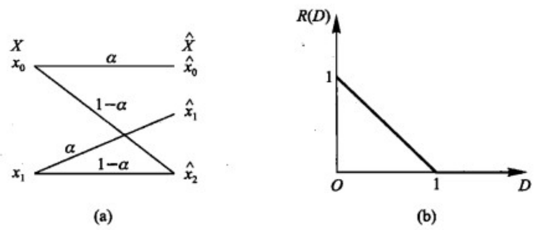


图 5.3.2 例 5.3.3 中相应的检验信道和率失真曲线

容易计算出这时的互信息为

$$\begin{aligned}R(D) &= I(X;Y)\\ &=1-D \qquad 0\leqslant D\leqslant 1\end{aligned}$$

相应的率失真曲线如图 5.3.2 (b)所示。

第六讲 计算理论

计算的性质

- 能行性——明确、有限的求解步骤
- 有穷性——对某些输入在有限步内结束并给出结果
- 通用性——某类问题总能给出正确的答案
- 确定性——每一步都机械、明确

核心问题：自动机理论、可计算性理论、计算复杂性理论

图灵机

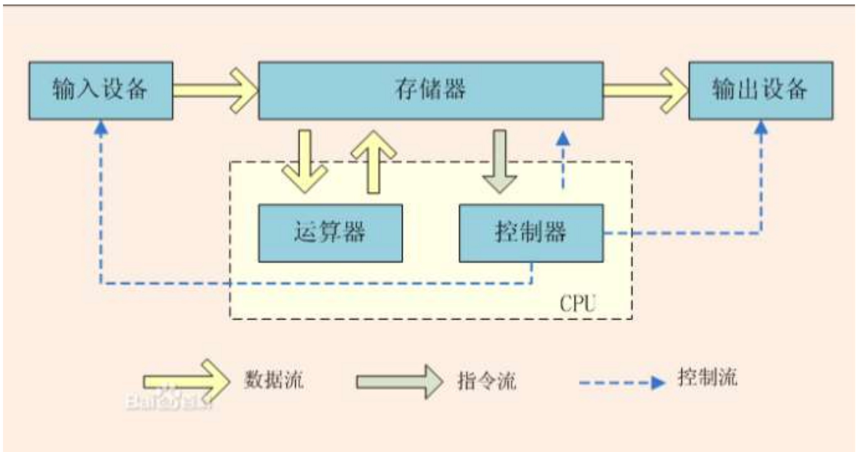
五元指令集：$(q_i, S_j, S_k, R(LN), q_n)$

PPT的图灵机是移动写字头的、保持纸带位置不变；先写入再移动写字头。

理论模型与现代的相同和不同之处

相同之处:程序与数据混合在一起，由控制器控制执行。不同之处:内存无限大!没有考虑输入与输出!(所有信息都在子带上)

冯诺依曼架构



基本原理：存储程序(Stored Program) 并按地址顺序执行。

特点：运算器为中心；存储程序原理；指令由操作码和地址码组成；二进制；硬件、软件分离

局限：存储器与CPU之间的通路狭窄。

可计算性理论

停机问题

假设存在一个测试程序T,能接受任何输入。当输入程序P能终止，输出1；不能终止，输出0。再级联一个程序 `while(x){}` 表示X=1时不终止。则若S终止，则S不终止；若S不终止，S终止。

计算复杂性理论

确定性图灵机：每一步均按由输入和状态确定的方向产生输出并转移到下一个状态。

非确定性图灵机：每一步在多个可能的输出和转移状态的组合中任意选择一个。

设$t(n)$是一个函数，$t(n) \geq n$。则每一个$t(n)$时间的非确定型单带图灵机都与某一个$2^{O(t(n))}$的确定型单带图灵机等价。

多项式时间复杂度：$O(n^k), O(logn)$都是

非多项式时间复杂度：$O(k^n), O(n!)$

P类问题：确定型单带图灵机在多项式时间内可以求解的问题

NP类问题：验证输入是否可被接受。具有多项式时间验证机的问题（一个问题是NP↔它能被某个非确定型多项式时间图灵机验证。）

P=NP？问题：一个非确定性图灵机在多项式时间内能解，是否在确定性图灵机中也能多项式时间可解。

Kolmogorov复杂度

定义：$K_u(x) = min\, l(p)$，能够输出x并停止的所有程序的最小长度。

通用性：如果u是通用计算机，那么对其他任意计算机A，$K_u(x) \leq K_A(x) + c_A$

证明：假定对于计算机$\mathcal{A}$我们有一个输出 x 的程序 p_A。于是$\mathcal{A}(p_A)=x$。我们可以在该程序之前增加一个模拟程序 s_A，它告诉计算机$\mathcal{U}$如何模拟计算机$\mathcal{A}$。然后，计算机$\mathcal{U}$将解释关于$\mathcal{A}$的程序中的指令，执行对应的计算并且输出 x。该程序$\mathcal{U}$是 $p=s_A p_A$，它的长度是

$$l(p)=l(s_A)+l(p_A)=c_A+l(p_A) \qquad (14\text{-}4)$$

其中 c_A是模拟程序的长度。因此，对所有的二元串 x，有

$$K_{\mathcal{U}}(x)=\min_{p:\mathcal{U}(p)=x} l(p) \leq \min_{p:\mathcal{A}(p)=x}(l(p)+c_A)=K_{\mathcal{A}}(x)+c_A \qquad (14\text{-}5)\square$$

条件复杂度：$K(x|l(x)) \leq l(x) + c$

Kolmogorov复杂度的上界：$K(x) \leq K(x|l(x)) + 2\,log\,l(x) + c$

证明： 如果计算机不知道 $l(x)$，定理 14.2.2 的方法就不再适用。我们必须有某种方法来通知计算机什么时候到描述序列的比特串的结尾处。我们来描述一个简单但低效的方法，它使用序列 01 作为一个"逗号"。

假定 $l(x)=n$。为了描述 $l(x)$，将 n 的二进制展开中的每一位重复两次；然后用一个 01 结束这个描述，以便计算机知道已经到了 n 的描述的结尾处。例如，数字 5(二进制表示为 101)将描述为 11001101。这个描述需要 $2\lceil \log n \rceil + 2$ 比特。于是，含有 $l(x)$的二进制表示的程序不会使原有的程序长度增多超过 $2\log l(x) + c$ 比特，由此我们得到定理中的上界。 □

Kolmogorov复杂度的下界：复杂度$K(x) < k$的x的数目满足$|\{x\}| < 2^k$

证明： 短程序并不很多。如果要将所有长度小于 k 的程序列出的话，我们有

$$\underbrace{\Delta}_{1},\underbrace{0,1}_{2},\underbrace{00,01,10,11}_{4},\cdots,\underbrace{\cdots,\overbrace{11\cdots1}^{k-1}}_{2^{k-1}} \tag{14-11}$$

而这样的程序的总数是

$$1+2+4+\cdots+2^{k-1}=2^k-1<2^k \tag{14-12}$$

由于每个程序仅产生一个可能的输出序列，所以复杂度$<k$ 的序列的数目小于 2^k。 □

E.g.

✧例1：n个0的序列

$$K(000\cdots 0|n) = c, \forall n$$

✧例2：π前n位的复杂度

$$K(\pi_1\pi_2\pi_3\cdots\pi_n|n) = c, \forall n$$

✧例3：整数n

$$K(n) \leq c + \log^* n, \forall n$$

✧例4：含k个1的n比特长度序列

$$K(x_1x_2x_3\cdots x_n|n) \leq nH_0\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \frac{1}{2}\log n + c$$

与熵的关系：$E\left[\frac{1}{n}K(X^n|n)\right] \to H(X)$

<u>不可压缩序列</u>

$X_1,\dots,X_n \sim B(0.5)$，$P(K(X_1X_2\dots X_n|n) < n-k) < 2^{-k}$

如果$lim_{n\to\infty}\frac{K(x_1x_2\dots x_n|n)}{n} = 1$，则无限串$x$是无可压缩的。

机器学习

分类：监督学习；非监督学习；强化学习

过拟合解决办法：增加训练样本数、规范化

<u>贝叶斯学习</u>

极大后验（MAP）：给定数据D,在候选假设集合H中可能性最大h，$h_{MAP} = argmaxP(h|D) = argmaxP(D|h)p(h)$

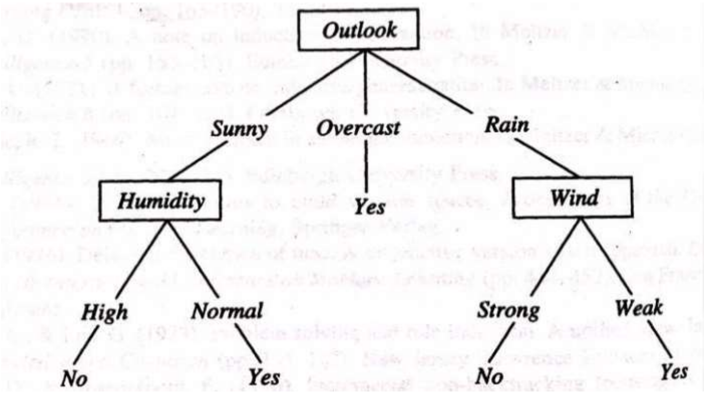
极大似然（ML）：假设先验概率都相等，$h_{ML} = argmaxP(D|h)$

贝叶斯最优分类器：新实例的最可能分类由后验概率加权，合并所有假设的预测得到。

$$v^* = argmax \sum P(v_j|h_i)P(h_i|D)$$

朴素贝叶斯分类：$v_{MAP} = argmaxP(a_1a_2 \dots a_n|v_j)P(v_j) = argmaxP(v_j)\prod P(a_i|v_j)$，假定各属性独立。

决策树



核心问题：什么属性是最佳的分类属性（信息增益）

树的根节点表示分类的开始；叶节点表示一个实例的结束；中间节点表示实例的某种属性；边表示属性的属性值。

$$Entropy(S) = -p_+ logp_+ - p_- logp_-$$

$$Gain(S, A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} H(S_v)$$

表示属性A相对样例集合S的信息增益，$Values(A)$是A所有可能值的集合，S_v是S中属性A的值为v的子集。

决策树构建：计算属性的信息增益；选取增益最大的作为根节点；继续计算基于此时叶节点的信息增益，以此类推直到不可再分。

E.g:

✧按属性Wind分类14个样例得到的信息增益计算：

- $Values(Wind) = Weak，Strong$
- $S = [9+，5-]$
- $S_{Weak} = [6+，2-]$
- $S_{Strong} = [3+，3-]$

$$Gain(S, Wind) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= Entropy(S) - \left(\frac{8}{14}\right) Entropy(S_{Weak}) - \left(\frac{8}{14}\right) Entropy(S_{Strong})$$

$$= 0.94 - \frac{8}{14} * 0.811 - \frac{6}{14} * 1 = 0.048$$

所有四个属性的信息增益为:

$$Gain(S, Outlook) = 0.246$$
$$Gain(S, Humidity) = 0.151$$
$$Gain(S, Wind) = 0.048$$
$$Gain(S, Temperature) = 0.029$$

k-近邻

一个样本在特征空间中k个最相似的样本中的大多数属于某一个类别，则该样本也属于该类别。

定义距离$d(x_i,x_j)=\sqrt{\sum(a_r(x_i)-a_r(x_j))^2}$

问题：维度灾难（坐标伸展法，属性加权；交叉验证法，留一法、k-fold交叉）

多层感知机

实现布尔函数

聚类

- 将一个给定的数据分组成多类，类内对象相似度高，不同类对象差别大。

K-means：

1. 从D中随机取k个元素，作为k个簇的中心
2. 计算D中所有的点到这k个中心的相异度，将其分别划归到相异度最低的簇
3. 根据聚类结果，调整k个聚类中心，即将聚类的中心移动到聚类的几何中心（即平均值）处
4. 重复第2、3步直到聚类的中心不再移动，此时算法收敛

缺点：K值难估计；需要人为确定初始聚类中心；结果只能保证局部最优；噪声敏感。。。

第七章 控制系统

状态空间

$$sX(s)=AX(s)+BU(s)$$
$$Y(s)=CX(s)+DU(s)$$

传递函数$G(s)=\frac{Y(s)}{U(s)}$

线性定常系统的状态方程：$x(k+1)=Ax(k)+Bu(k),\quad y(k)=Cx(k)$

能控性

概念：有限时间内能否转移到要求状态。（支配能力）

判据：线性定常离散系统完全能控的充要条件是能控性矩阵$U_C=[B,AB,\ldots,A^{n-1}B]$满秩。

能观性

概念：有限时间内通过对输出结果判断系统初始状态（反应状态变量）

判据：能观性矩阵$U_O=\begin{bmatrix}C\\CA\\\cdots\\CA^{n-1}\end{bmatrix}$满秩。

稳定性机制

Routh稳定判据

劳斯表中第一列元素全部大于0.若出现小于0的元素则系统不稳定。第一列元素符号改变的次数等于系统正实部根的个数。

$$
\begin{array}{c|ccccc}
s^n & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots \\
s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots \\
s^{n-2} & A_1 & A_2 & A_3 & A_4 & \cdots \\
s^{n-3} & B_1 & B_2 & B_3 & B_4 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \\
s^2 & D_1 & D_2 & & & \\
s^1 & E_1 & & & & \\
s^0 & F_1 & & & &
\end{array}
$$

$$A_1=\frac{a_{n-1}a_{n-2}-a_na_{n-3}}{a_{n-1}},\ A_2=\frac{a_{n-1}a_{n-4}-a_na_{n-5}}{a_{n-1}},\ A_3=\frac{a_{n-1}a_{n-6}-a_na_{n-7}}{a_{n-1}}\cdots\text{全零}$$

$$B_1=\frac{A_1a_{n-3}-a_{n-1}A_2}{A_1},\ B_2=\frac{A_1a_{n-5}-a_{n-1}A_3}{A_1},\ B_3=\frac{A_1a_{n-7}-a_{n-1}A_4}{A_1}\cdots\text{全零}$$

【例】：特征方程：$a_5s^5+a_4s^4+a_3s^3+a_2s^2+a_1s+a_0=0$

Routh表如下：

$$
\begin{array}{c|ccc}
s^5 & a_5 & a_3 & a_1 \\
s^4 & a_4 & a_2 & a_0 \\
s^3 & \dfrac{a_4a_3-a_5a_2}{a_4}=A_1 & \dfrac{a_4a_1-a_5a_0}{a_4}=A_2 & 0 \\
s^2 & \dfrac{A_1a_2-a_4A_2}{A_1}=B_1 & a_0 & \\
s^1 & \dfrac{B_1A_2-A_1a_0}{B_1}=C_1 & 0 & \\
s^0 & a_0 & &
\end{array}
$$

$$
\begin{array}{c|ccc}
s^4 & 1 & 3 & 5 \\
s^3 & 2 & 4 & 0 \\
s^2 & \dfrac{2\times3-1\times4}{2}=1 & \dfrac{2\times5-1\times0}{2}=5 & \\
s^1 & \dfrac{1\times4-5\times2}{1}=-6 & 0 & \\
s^0 & 5 & &
\end{array}
$$

Nyquist稳定判据

s右半平面的闭环 极点个数$Z=P-2N$，P为右半平面开环极点个数，N为开环Nyquist曲线逆时针包围$(-1,j0)$点的圈数，且$N=N_+-N_-$.

起点为$\omega=0$，终点为$\omega=\infty$

李雅普诺夫第一法（间接法）

外部稳定性：零初始条件下，任意一个有界输入，输出也有界。

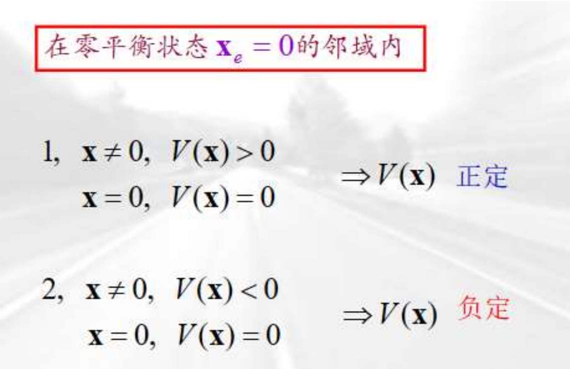
外部稳定性的充要条件：传递函数矩阵中所有元素的极点位于s左半平面。（分式可约）

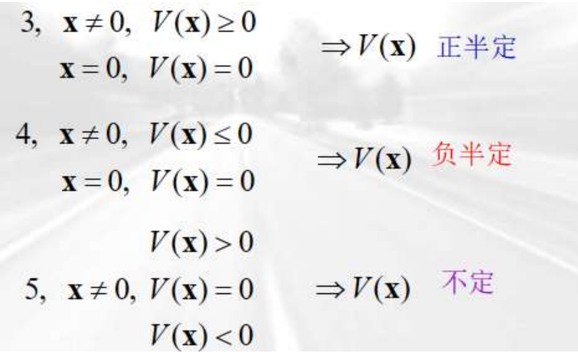
$$\begin{cases}\dot{\mathbf{x}}(t)=\mathbf{Ax}(t)+\mathbf{Bu}(t)\\ \mathbf{y}(t)=\mathbf{Cx}(t)\end{cases}\qquad \mathbf{W}(s)=\mathbf{C}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}=\frac{\mathbf{C}(s\mathbf{I}-\mathbf{A})^{*}\mathbf{B}}{|s\mathbf{I}-\mathbf{A}|}$$

内部稳定性：求$det(\lambda I-A)=0$，A的特征值全小于0（分式约前的所有极点，建议还是算一算），渐近稳定；全小于0，有一个无重根0，李雅普诺夫稳定；其他，不稳定。

李雅普诺夫第二法（直接法）

求出系统的能量函数$V(x,t)$，求出随时间变化率，根据两者的定号性判断平衡状态稳定性。





$V(x)$正定，$\dot{V}(x)$负定，大范围渐近稳定。

$V(x)$正定，$\dot{V}(x)$半负定且不恒等于0，大范围渐近稳定。

$V(x)$正定，$\dot{V}(x)$负半定且恒等于0，李雅普诺夫意义下的稳定。