数字媒体技术

第2章 信号处理技术与信息论基础(6学时)

刘绍辉

计算机科学与技术学院 哈尔滨工业大学

shliu@hit.edu.cn

2021年春季

第2章 信号处理技术与信息论基础

第2章 信息论基础

- 2.5 基本信息论
  - 2.5.1 通信基本模型
  - 2.5.2 信息度量
  - 2.5.3 离散信源熵
  - 2.5.4 连续信源熵
  - 2.5.5 熵速率与信道容量
  - 2.5.6 离散有噪信道中的信道容量
  - 2.5.7 连续有噪信道中的信道容量
- 2.6 信源与信道匹配的编码
  - 2.6.1 编码定理
  - 2.6.2 信源最佳化
  - 2.6.3 符号独立化
  - 2.6.4 概率均匀化
- 2.7 信息率失真函数
- 参考文献和站点

信息论是应用近代概率统计方法研究信息传输、交换、存储和处理的一门学科，也是源于通信实践发展起来的一门新兴应用科学。

信息论

- 物质、能量和信息是构成客观世界的三大要素。信息是物质和能量在空间和时间上分布的不均匀程度，或者说信息是关于事物运动的状态和规律
- **信息论**是研究信息的基本性质及度量方法，研究信息的获取、传输、存储和处理的一般规律的科学。

三个不同的范畴的信息论

- **狭义信息论**：即通信的数学理论，主要研究狭义信息的度量方法，研究各种信源、信道的描述和信源、信道的编码定理，以Shannon信息论为基本内容，也称为基本信息论
- **一般信息论**：研究信息传输和处理问题，也就是狭义信息论方法在调制解调、编码译码以及检测理论等领域的应用，信息传输的基本问题，即通信理论
- **广义信息论：**包括信息论在自然和社会中的新的应用，如模式识别、机器翻译、自学习自组织系统、心理学、生物学、经济学、社会学等一切与信息问题有关的领域

2.5 基本信息论

◆通信基本模型

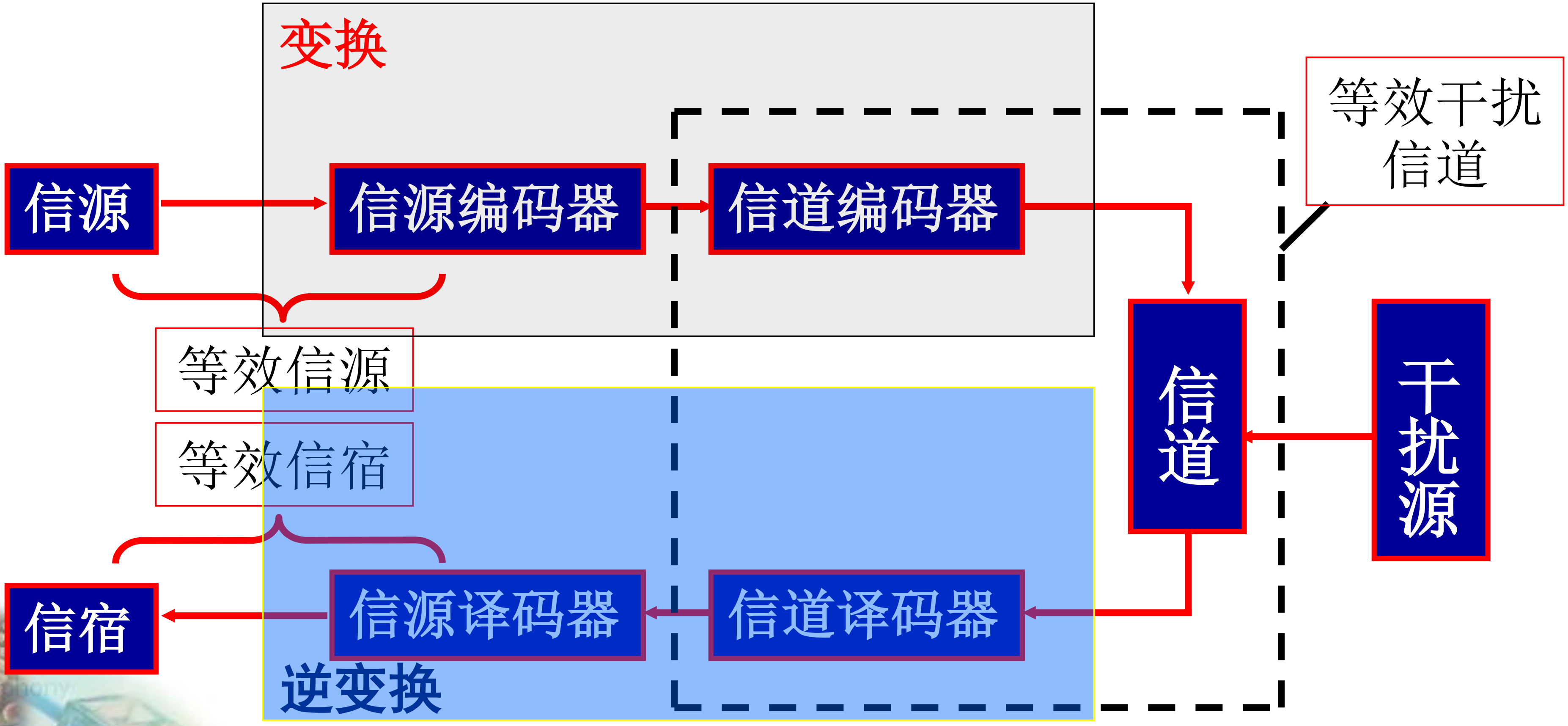


图2-1通信系统的一般模型

2.5 基本信息论(续1)

◆通信基本模型

- 信源，信道，信宿
- 变换
  - ✓编码，译码，调制，解调
  - ✓编码目的
    - 改善信源的传输效率：有效性编码，主要针对信源特性进行处理，也称为信源编码
    - 提高信息传输的可靠性：可靠性编码或抗干扰编码，主要针对信道特性进行编码，也称为信道编码
- 信道分类
  - ✓传输介质：有线，无线
  - ✓信道特性：恒参信道，特性不随时间改变，有线信道，微波接力信道，卫星中继信道；变参信道，短波电离层反射信道，对流层散射信道
  - ✓信道传送信号的形式：离散信道，连续信道

信息论研究范畴中信道的划分方式

2.5 基本信息论(续2)

- **通信-传递信息**
  - 传输信息的速率越大越好，通信的有效性问题
  - 传输的信息受到干扰和噪声越小越好，通信的可靠性问题
  - 矛盾？
- **信息论的研究表明**
  - 在一定条件下，通信系统能以任意小的差错率实现最大的传输率
- **通信就是用传递消息来消除收信人对某事情状态的不定性**
  - 信源要传递的事情一般是属于随机事件
  - 用概率论和随机过程等统计数学工具来描述和分析通信问题------经典信息论的基本研究内容

2.5 基本信息论(续3)

◆ **离散无记忆信源**(Discrete Memoryless Source，简记为DMS)输出的是单个符号的消息，不同时刻发出的符号之间彼此统计独立，而且符号集中的符号数目是有限的或可数的。离散无记忆信源的数学模型为离散型的概率空间，即：

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

$p(x_i)$：信源输出符号消息x_i的先验概率；

满足：$0 \le p(x_i) \le 1$，$1 \le i \le I$　$\sum_{i=1}^{I} p(x_i) = 1$

2.5 基本信息论(续4)

◆**离散平稳有记忆信源**用联合概率空间$\{\boldsymbol{X}, \boldsymbol{P}(\boldsymbol{X})\}$来描述离散有记忆信源的输出。信源在$i$时刻发出什么符号与$i$时刻以前信源所发出的符号有关，即由条件概率$p(x_i | x_{i-1}\, x_{i-2} \cdots)$确定。如果该条件概率分布与时间起点无关，只与关联长度有关，则该信源为**平稳信源**。

对于离散平稳有记忆信源，有：

$$p(x_1 = a_1) = p(x_2 = a_1) = \ldots$$
$$p(x_2 = a_2 | x_1 = a_1) = p(x_3 = a_2 | x_2 = a_1) = \ldots$$
$$p(x_3 | x_2\, x_1) = p(x_4 | x_3\, x_2) = \ldots$$
$$\vdots$$
$$p(x_{i+L} | x_{i+L-1}\, x_{i+L-2} \ldots x_i) = p(x_{j+L} | x_{j+L-1}\, x_{j+L-2} \ldots x_j) = \ldots$$
$$\vdots$$

2.5 基本信息论(续5)

◆离散无记忆信道

➤离散无记忆信道的输入和输出消息都是离散无记忆的单个符号，输入符号$x_i \in \{ a_1 , a_2 , \ldots , a_k \}$，$1 \le i \le I$，输出符号$y_j \in \{ b_1 , b_2 , \ldots , b_D \}$，$1 \le j \le J$，信道的特性可表示为转移概率矩阵：

$$\boldsymbol{P} = \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \cdots & p(y_J|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & \cdots & p(y_J|x_2) \\ & \vdots & & \\ p(y_1|x_I) & p(y_2|x_I) & \cdots & p(y_J|x_I) \end{bmatrix}$$

$p(y_j|x_i)$对应为已知输入符号为x_i，当输出符号为y_j时的信道转移概率，满足$0 \le p(y_j|x_i) \le 1$。

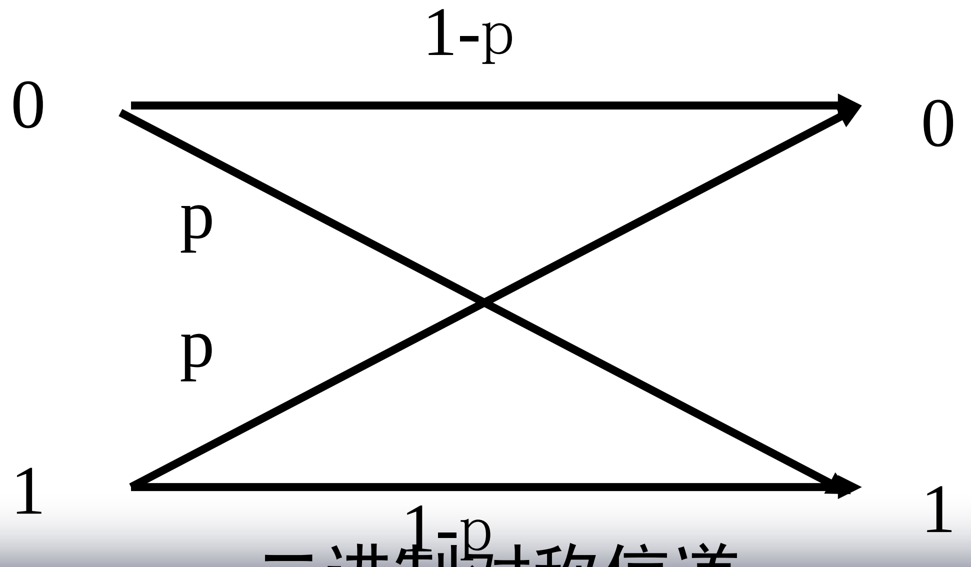
几种常见的离散无记忆信道：

1.**二元对称信道**(Binary Symmetric Channel，简记为BSC)

这是一种很重要的信道，它的输入符号$x \in \{0, 1\}$，输出符号$y \in \{0, 1\}$，转移概率$p(y|x)$如图所示信道特性可表示为信道矩阵

$$\boldsymbol{P} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

其中p称作信道错误概率。



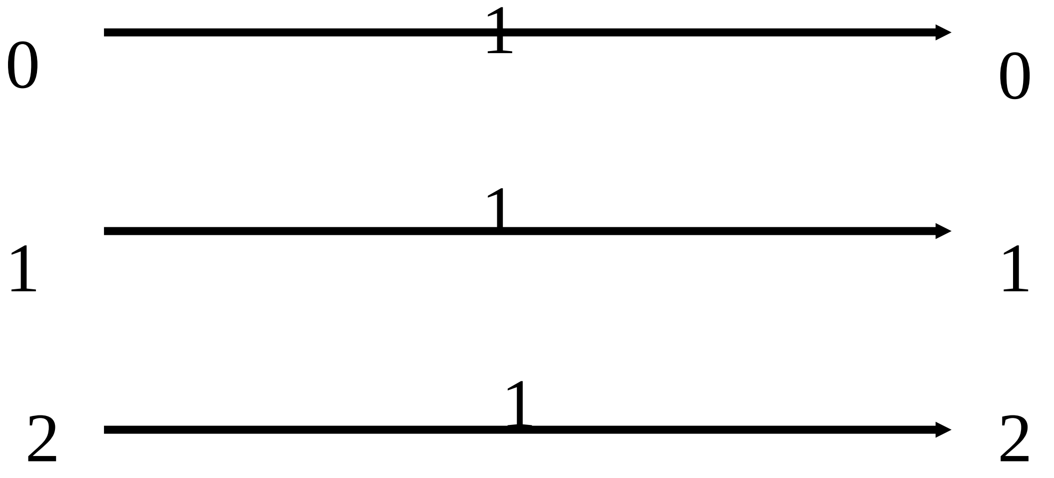
二进制对称信道

2. 无干扰信道

这是一种最理想的信道，也称作无噪信道，信道的输入和输出符号间有确定的一一对应关系，$p(y|x)=\begin{cases}1 & x=y\\ 0 & x\neq y\end{cases}$ 如图三元无干扰信道中，$x,y\in\{0,1,2\}$，

对应信道矩阵是单位矩阵

$$\boldsymbol{P}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$



无干扰信道

3. 二元删除信道

对于接收符号不能作出肯定或否定判决时，引入删除符号，表示对该符号存有疑问，作为有误或等待得到更多信息时再作判决。

二元删除信道如图所示，输入符号$x \in \{0, 1\}$，输出符号$y \in \{0, e, 1\}$，转移概率矩阵为

$$\boldsymbol{P}=\begin{bmatrix} p & 1-p & 0 \\ 0 & 1-p & p \end{bmatrix}$$

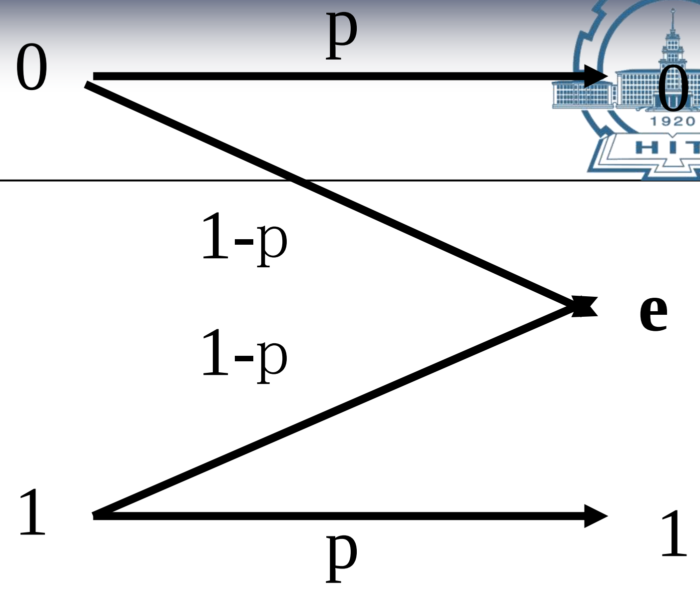


图 二元删除信道

4.二元Z信道

二元Z信道如图所示，信道输入符号$x \in \{0, 1\}$，输出符号$y \in \{0, 1\}$转移概率矩阵为

$$\boldsymbol{P}=\begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix}$$

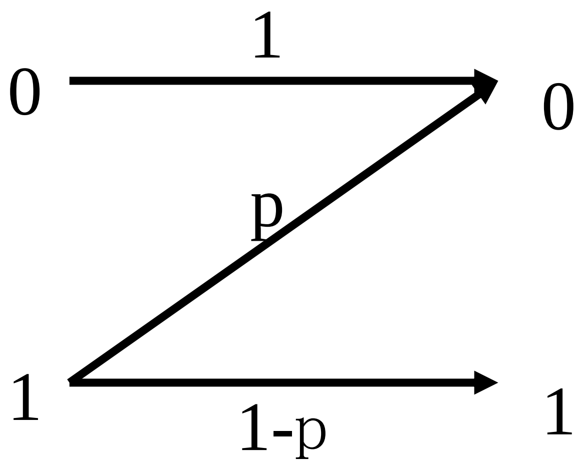


图 二元Z信道

2.5 基本信息论-自信息量和互信息量

- **一个事件的自信息量就是对其不确定性的度量,互信息量则表明了两个随机事件的相互约束程度**
- 对于随机事件集$X=\{x_1,\ x_2,\ \dots,\ x_i,\ \dots,\ x_I\}$中的随机事件$x_i$，其出现概率记为$p(x_i)$，将两个事件$x_i$,y_j同时出现的概率记为$p(x_i\, y_j)$，则$p(x_i)$，$p(x_i\, y_j)$应满足：

$$\begin{cases} p(x_i)\ge 0 \qquad i=1,2,\cdots,I \\ \sum\limits_{i=1}^{I} p(x_i)=1 \end{cases} \qquad \begin{cases} p(x_i y_j)\ge 0 \\ \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J} p(x_i y_j)=1 \end{cases}$$

相应的条件概率为

$$\begin{cases} p(x_i \mid y_j)=\dfrac{p(x_i y_j)}{p(y_j)} \\[2ex] p(y_j \mid x_i)=\dfrac{p(x_i y_j)}{p(x_i)} \end{cases}$$

2.5 基本信息论-自信息量和互信息量(续1)

研究信息度量的必要性

评价通信系统的标准：误码率，接收信噪比，传信率等，传信率：单位时间内所传递的信息量，为了评价它，就必须对消息或信源所含有的信息有一个数量上的度量

信源的不肯定性

不肯定程度与信源概率空间的状态数及其概率分布有关，信源概率空间的概率分布为等概时，不肯定程度最大；等概时，不肯定程度与信源概率空间的可能状态数或相应的概率有关，状态数愈多或相应的概率愈小，不肯定程度愈大

将某事件发生所得到的信息量记为$I(x)$，$I(x)$应该是该事件发生的概率的函数，即

$$I(x)=f[p(x)]$$

100个球的例子：

99 vs 1，50 vs 50 ，25 vs 25 vs 25 vs 25

信息量直观的定义为：

收到某消息获得的信息量 ＝不确定性减少的量

2.5 基本信息论-自信息量和互信息量(续2)

- **直观观察，信息量应满足**
  - $I(x)$ 应该是 $p(x)$ 的单调递减函数：概率小的事件一旦发生赋予的信息量大，概率大的事件如果发生则赋予的信息量小
  - 信息量应具有可加性：对于两个独立事件，其信息量应等于各事件自信息量之和
  - 当 $p(x)=1$ 时，$I(x)=0$：表示确定事件发生得不到任何信息
  - 当 $p(x)=0$ 时，$I(x)\rightarrow\infty$：表示不可能事件一旦发生，信息量将无穷大
- **自信息量** $I(x) \triangleq -\log p(x)$
  - 自信息量的单位与log函数所选用的对数底数有关，如底数分别取 2、 e、 10，则自信息量单位分别为：比特、奈特、哈特
  - 事件 x_i 发生以前，表示事件发生的先验不确定性
  - 当事件 x_i 发生以后，表示事件 x_i 所能提供的最大信息量（在无噪情况下）

2.5 基本信息论-自信息量和互信息量(续3)

◆二维联合集$X\ Y$上元素$x_i\ y_j$的联合自信息量$I(x_i\ y_j)$定义为

$$I(x_i y_j) \triangleq -\log p(x_i y_j)$$

◆条件自信息量，在已知事件y_j条件下,随机事件x_i发生的概率为条件概率$p(x_i \mid y_j)$,条件自信息量定义为

$$I(x_i \mid y_j) \triangleq -\log p(x_i \mid y_j)$$

2.5 基本信息论-自信息量和互信息量(续4)

◆ **从通信的角度考虑，假设不肯定性用H表示**

- 信源X的概率空间为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_I \\ p(x_1) & p(x_2) & \cdots & p(x_I) \end{bmatrix} \xrightarrow{\text{发送消息}} \boldsymbol{H(X)=H(P(X))}$$

- 收信者收到消息Y后，不肯定程度可以描述成:

$$\begin{bmatrix} Y \\ P(X/Y) \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_I \\ p(x_i/y_1) & p(x_i/y_2) & \cdots & p(x_i/y_I) \end{bmatrix} \xrightarrow{\text{收到消息}} \mathbf{H}(X/Y)=H(\mathbf{P}(X/Y))$$

- $P(X)$,$P(X/Y)$分别表示收信者收到信息前后的先验概率和后验概率，这时候收到消息y_j后所获得的信息量:

$$I_j = I(x_i) - I(x_i/y_j)=\log(\text{后验概率}/\text{先验概率})$$

【例】某住宅区共建有若干栋商品房，每栋有5个单元，每个单元住有12户，甲要到该住宅区找他的朋友乙，若：

1. 甲只知道乙住在第5栋，他找到乙的概率有多大？他能得到多少信息？

2. 甲除知道乙住在第5栋外，还知道乙住在第3单元，他找到乙的概率又有多大？他能得到多少信息？

用x_i代表单元数，y_j代表户号：

（1）甲找到乙这一事件是二维联合集$X\ Y$上的等概分布 $p(x_i y_j)=\dfrac{1}{60}$，这一事件提供给甲的信息量为

$$I(x_i y_j) = -\log p(x_i y_j) = \log 60 = 5.907 \text{（比特）}$$

（2）在二维联合集$X\ Y$上的条件分布概率为 $p(y_j|x_i)=\dfrac{1}{12}$，这一事件提供给甲的信息量为条件自信息量

$$I(y_j \mid x_i) = -\log p(y_j \mid x_i) = \log 12 = 3.585 \text{（比特）}$$

2.5 基本信息论-互信息量和条件互信息量

- 互信息量，从通信的角度引出互信息量的概念信源符号$X=\{x_1, x_2, \ldots, x_I\}$，$x_i \in \{a_1,a_2,\ldots,a_k\}$，$i = 1, 2, \ldots, I$，经过信道传输，信宿方接收到符号$Y = \{y_1, y_2, \ldots, y_J\}$，$y_j \in \{b_1,b_2,\ldots,b_D\}$，$j = 1, 2, \ldots, J$
- 事件x_i是否发生具有不确定性，用$I(x_i)$度量；接收到符号y_j后，事件x_i是否发生仍保留有一定的不确定性，用$I(x_i \mid y_j)$度量；观察事件前后，这两者之差就是通信过程中所获得的信息量，则互信息量用$I(x_i\,;y_j)$表示：

$$I(x_i;y_j) = I(x_i) - I(x_i | y_j) = \log \frac{p(x_i | y_j)}{p(x_i)}$$

根据概率互换公式$p(x_i\ y_j) = p(y_j \mid x_i)p(x_i)=p(x_i \mid y_j)p(y_j)$ 互信息量$I(x_i ;y_j)$有多种表达形式:

$$I(x_i;y_j)=\log\frac{p(x_iy_j)}{p(x_i)p(y_j)}=I(x_i)+I(y_j)-I(x_iy_j)$$

$$I(x_i;y_j)=\log\frac{p(y_j|x_i)}{p(y_j)}=I(y_j)-I(y_j|x_i)$$

将事件互信息量的概念推广至多维空间:

在三维$X\ Y\ Z$联合集中，有:

$$I(x_i;y_j\ z_k)=I(x_i;y_j)+I(x_i;z_k \mid y_j)$$

类似，在N维$U_1\ U_2\ \ldots\ U_N$联合空间, 有:

$$I(u_1;u_2u_3\ldots u_N)=I(u_1;u_2)+I(u_1;u_3 \mid u_2)+\ldots+I(u_1;u_i \mid u_2\ldots u_{i-1})+\ldots+I(u_1;u_N \mid u_2\ldots u_{N-1})$$

2.5 基本信息论-互信息量和条件互信息量(续1)

◆条件互信息量

- 三维$X\ Y\ Z$联合集中，在给定条件z_k的情况下, x_i , y_j的互信息量$I(x_i;y_j\mid z_k)$定义为：

$$I(x_i;y_j|z_k)=\log\frac{p(x_i|y_jz_k)}{p(x_i|z_k)}$$

【例】信源包含7个消息$x_0, x_1, x_2, x_3, x_4, x_5, x_6$ 信源编码器将其对应编成7个三位二进制数000，001, …, 110。各消息的先验概率已知，在接收过程中，每收到一个数字，各消息的后验概率都相应地发生变化。考虑在接受100三个数字的过程中，各后验概率的变化，计算信息量$I(x_4;100)$。

表为7个三位二进制数对应的各种概率。

信源消息	码字	消息先验概率	消息后验概率		
			收到1后	收到10后	收到100后
x_0	000	1/16	0	0	0
x_1	001	1/16	0	0	0
x_2	010	1/16	0	0	0
x_3	011	1/16	0	0	0
x_4	100	1/2	2/3	4/5	1
x_5	101	1/8	1/6	1/5	0
x_6	110	1/8	1/6	0	0

根据给定的先验概率，可算出：

$$p(x_4)=\frac{1}{2} \quad p(x_4|1)=\frac{1/2}{1/2+1/8+1/8}=\frac{2}{3} \quad p(x_4|10)=\frac{2/3}{2/3+1/6}=\frac{4}{5} \quad P(x_4 \mid 100)=1$$

将各种后验概率的计算结果列于表中，计算出互信息量：

$$I(x_4;100)=I(x_4;1)+I(x_4;0\mid 1)+I(x_4;0\mid 10)$$

$$=\log\frac{p(x_4|1)}{p(x_4)}+\log\frac{p(x_4|10)}{p(x_4|1)}+\log\frac{p(x_4|100)}{p(x_4|10)}$$

$$=\log\frac{2/3}{1/2}+\log\frac{4/5}{2/3}+\log\frac{1}{4/5}=\log 2=1$$ （比特）

也可直接计算出：

$$I(x_4;100)=\log\frac{p(x_4|100)}{p(x_4)}=\log\frac{1}{1/2}=1$$ （比特）

2.5 基本信息论-熵

◆平均自信息量

- **概率意义下的平均信息量即为熵**
- 人们注意的是整个系统的统计特性，当信源各个消息的出现概率相互统计独立时，这种信源称为无记忆信源，无记忆信源的平均自信息量定义为各消息自信息量的概率加权平均值（统计平均值），即*平均自信息量H(X)*定义为

$$H(X)=\sum p(x_i)I(x_i)=\sum p(x_i)\log[1/p(x_i)]=-\sum p(x_i)\log[p(x_i)]$$

- $H(X)$的表达式与统计物理学中的热熵具有相类似的形式，在概念上二者也有相同之处，故借用熵这个词把$H(X)$称为集合X的信息熵，简称熵
- 条件熵

$$H(X|Y)\triangleq\sum_i\sum_j p(x_iy_j)I(x_i|y_j)=-\sum_i\sum_j p(x_iy_j)\log p(x_i|y_j)$$

2.5 基本信息论-熵(续1)

当X , Y统计独立时，有$p(x_i y_j) = p(x_i)p(y_j)$，$p(x_i \mid y_j) = p(x_i)$，

则 $$H(X|Y) = \sum_i \sum_j p(x_i y_j) I(x_i|y_j)$$

从通信角度来看：

若将$X = \{x_1, x_2, \ldots, x_i, \ldots\}$视为信源输出符号；

$Y = \{y_1, y_2, \ldots, y_j, \ldots\}$视为信宿接收符号；

$I(x_i \mid y_j)$可看作信宿收到y_j后，关于发送的是否为x_i仍然存在的疑义度（不确定性），则

$$H(X|Y) = -\sum_j p(y_j) \sum_i p(x_i) \log p(x_i) = -\sum_i p(x_i) \log p(x_i) = H(X)$$

反映了经过通信后，信宿符号y_j（$j = 1, 2, \ldots$）关于信源符号x_i（$i = 1, 2, \ldots$）的平均不确定性。

2.5 基本信息论-熵(续2)

联合熵$H(XY)$ (共熵)是定义在二维空间XY上，对元素$x_i y_j$的自信息量的统计平均值，若记事件$x_i y_j$出现的概率为$p(x_i y_j)$，其自信息量为$I(x_i y_j)$，则联合熵$H(XY)$ 定义为

$$H(XY) \triangleq \sum_i \sum_j p(x_i y_j) I(x_i y_j)$$

$$= -\sum_i \sum_j p(x_i y_j) \log p(x_i y_j)$$

由熵、条件熵、联合熵的定义式可导出三者的关系式

$$\boldsymbol{H(XY) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)}$$

上式反映了信息的可加性。当X, Y统计独立时，有

$$\boldsymbol{H(XY) = H(X) + H(Y)}$$

2.5 基本信息论-熵(续3)

◆熵函数的性质

设$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$为一个$n$元函数，若对任意$f(\boldsymbol{x}_1), f(\boldsymbol{x}_2) \in f(\boldsymbol{x})$，任意正数$\theta$，$0 \le \theta \le 1$，有$\theta f(\boldsymbol{x}_1) + (1-\theta) f(\boldsymbol{x}_2) \le f[\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2]$，则$f(\boldsymbol{x})$为∩型凸函数，反之为∪型凸函数

◆极大离散熵定理

- **设信源的消息个数为M，则$H(X) \le \log M$，等号当且仅当信源X中各消息等概 (1/M) 时成立，即各消息等概分布时，信源熵最大。**

◆对称；非负；$H(X \mid Y) \le H(X)$，X,Y统计独立时等号成立，$H(XY) \le H(X) + H(Y)$；

$$\begin{cases} H(XY) \ge H(X) \\ H(XY) \ge H(Y) \end{cases}$$

2.5 基本信息论-平均互信息量

◆如果将发送符号与接收符号看成两个不同的“信源”，则通过信道的转移概率来讨论信息的流通问题，一次通信从发送到接收究竟能得到多少信息量呢，这就是平均互信息量

◆定义$x_i \in X$和$y_j \in Y$之间的互信息量为$I(x_i ;y_j)$，在集合X上对$I(x_i ;y_j)$进行概率加权统计平均，可得$I(X;y_j)$为：

$$I(X;y_j) = \sum_i p\left(x_i \middle| y_j\right) \cdot I(x_i;y_j) = \sum_i p\left(x_i \middle| y_j\right) \log \frac{p(x_i|y_j)}{p(x_i)}$$

$$I\left(X;Y\right) = \sum_i \sum_j p(x_i y_j) I(x_i;y_j) = \sum_i \sum_j p(x_i y_j) \log \frac{p(x_i y_j)}{p(x_i) p(y_j)}$$

【例】二元等概信源 $\begin{bmatrix} X \\ p(X) \end{bmatrix} = \begin{bmatrix} x_0 & x_1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$，通过信道转移概率为 $P = \begin{array}{c|cc} & y_0 & y_1 \\ \hline x_0 & \frac{5}{6} & \frac{1}{6} \\ x_1 & \frac{1}{2} & \frac{1}{2} \end{array}$ 的信道传输，信宿接收符号 $Y = \{y_0, y_1\}$，计算信源与信宿间的平均互信息量 I（X；Y）。

（1） 先根据 $p(y_j) = \sum_i p(y_j | x_i) p(x_i)$ 计算出

$$p(\boldsymbol{X}) = \prod_{i-1}^{n} p(X_i)$$

$$p(y_1) = p(y_1 | x_0) p(x_0) + p(y_1 | x_1) p(x_1) = \frac{1}{2}\left(\frac{1}{6} + \frac{1}{2}\right) = \frac{1}{3}$$

（2） 由 $p(x_i | y_j) = \dfrac{p(y_j | x_i) p(x_i)}{p(y_j)}$ 计算后验概率

$$p(x_0 | y_0) = \frac{p(y_0 | x_0) p(x_0)}{p(y_0)} = \frac{5/6 \times 1/2}{2/3} = \frac{5}{8}$$

$$p(x_1|y_0)=\frac{p(y_0|x_1)p(x_1)}{p(y_0)}=\frac{1/2\times 1/2}{2/3}=\frac{3}{8}$$

$$p(x_0|y_1)=\frac{p(y_1|x_0)p(x_0)}{p(y_1)}=\frac{1/6\times 1/2}{1/3}=\frac{1}{4}$$

$$p(x_1|y_1)=\frac{p(y_1|x_1)p(x_1)}{p(y_1)}=\frac{1/2\times 1/2}{1/3}=\frac{3}{4}$$

（3）计算各消息之间的互信息量$I(x_i;y_j)$

$$I(x_0;y_0)=\log\frac{p(x_0|y_0)}{p(x_0)}=\log\frac{5/8}{1/2}=\log\frac{5}{4}=0.322$$ （比特）

$$I(x_0;y_1)=\log\frac{p(x_0|y_1)}{p(x_0)}=\log\frac{1/4}{1/2}=\log\frac{1}{2}=-1$$ （比特）

$$I(x_1;y_0)=\log\frac{p(x_1|y_0)}{p(x_1)}=\log\frac{3/8}{1/2}=\log\frac{3}{4}=-0.415$$ （比特）

$$I(x_1;y_1)=\log\frac{p(x_1|y_1)}{p(x_1)}=\log\frac{3/4}{1/2}=\log\frac{3}{2}=0.585$$ （比特）

（4） 计算平均互信息量

$$I(X;Y)=\sum_{i}\sum_{j}p(x_iy_j)I(x_i;y_j)$$

$$=\sum_{i}\sum_{j}p(x_i)p(y_j|x_i)I(x_i;y_j)$$

$$=\frac{1}{2}\left[\frac{5}{6}\times 0.322+\frac{1}{6}\times(-1)+\frac{1}{2}\times(-0.415)+\frac{1}{2}\times 0.585\right]=0.093$$

（比特）

2.5 基本信息论-条件平均互信息量

平均条件互信息量$I(X;Y\mid Z)$是在联合概率空间$\{XYZ，p(xyz)\}$上定义的物理量。由式（2-11）知道

$$I(x_i;y_j|z_k)=\log\frac{p(x_i|y_jz_k)}{p(x_i|z_k)}=\log\frac{p(x_iy_j|z_k)}{p(x_i|z_k)p(y_j|z_k)}$$

对上式在三维空间XYZ上求概率加权平均值，就得到平均条件互信息量

$$I(X;Y|Z)\triangleq\sum_i\sum_j\sum_k p(x_iy_jz_k)I(x_iy_j|z_k)$$

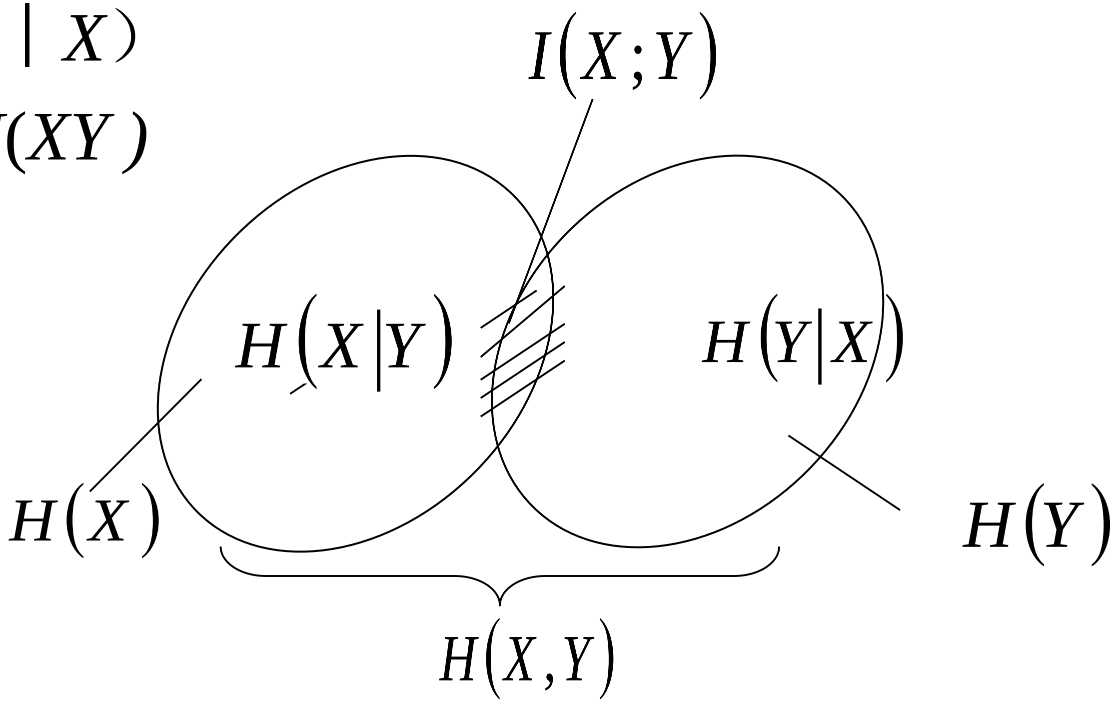
$$=\sum_i\sum_j\sum_k p(x_iy_jz_k)\log\frac{p(x_iy_j|z_k)}{p(x_i|z_k)p(y_j|z_k)} \quad （2-31）$$

式中$p(x_i\ y_j\ z_k)$满足 $\sum\sum\sum p(x_iy_jz_k)=1$

2.5 基本信息论

◆**各种信息量之间的关系：*平均互信息量与信源熵、条件熵的关系***

- $I(X;Y)=H(X)-H(X\mid Y)$
- $I(X;Y)=H(Y)-H(Y\mid X)$
- $I(X;Y)=H(X)+H(Y)-H(XY)$



维拉图

从通信的角度来讨论平均互信息量$I(X;Y)$的物理意义

由第一等式$I(X;Y)=H(X)-H(X|Y)$看$I(X;Y)$的物理意义

设X为发送消息符号集，Y为接收符号集，$H(X)$是输入集的平均不确定性，$H(X|Y)$是观察到Y后，集X还保留的不确定性，二者之差$I(X;Y)$就是在接收过程中得到的关于X, Y的平均互信息量。

对于无扰信道，$I(X;Y)=H(X)$

对于强噪信道，$I(X;Y)=0$

由第二等式$I(X;Y)=H(Y)-H(Y|X)$看$I(X;Y)$的物理意义

$H(Y)$是观察到Y所获得的信息量，$H(Y|X)$是发出确定消息X后，由于干扰而使Y存在的平均不确定性，二者之差$I(X;Y)$就是一次通信所获得的信息量。

对于无扰信道，有$I(X;Y)=H(X)=H(Y)$。

对于强噪信道，有$H(Y|X)=H(Y)$，从而$I(X;Y)=0$。

由第三等式$I(X;Y)=H(X)+H(Y)-H(X,Y)$看$I(X;Y)$的物理意义

通信前，随机变量X和随机变量Y可视为统计独立，其先验不确定性为$H(X)+H(Y)$，

通信后，整个系统的后验不确定性为$H(XY)$，

二者之差$H(X)+H(Y)-H(XY)$就是通信过程中不确定性减少的量，也就是通信过程中获得的平均互信息量$I(X;Y)$。

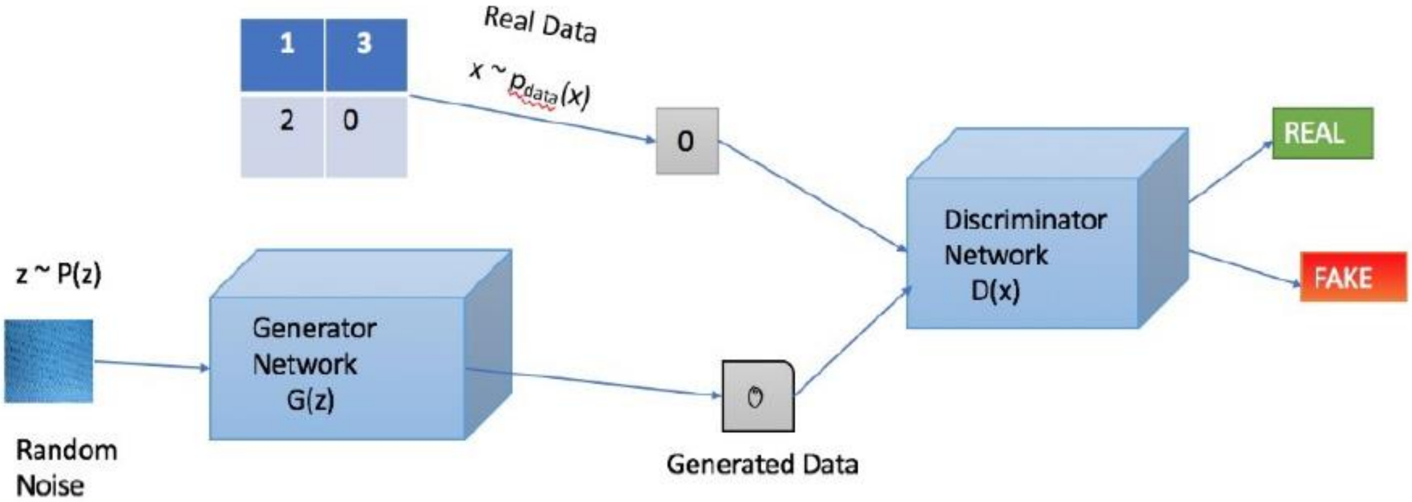
基本描述

- **熵(entropy),互信息(mutual information),相对熵(Relative entropy)**
  - 分别给出了随机变量的不确定性度量以及在消除或减少这一不确定性时所获得信息的度量
    - ✓随机变量不确定性的度量-熵，这从理论和函数形式上来说，香农熵都是最好的一种形式
    - ✓减少不确定可通过两条途径
      - 通过对感兴趣的随机变量X统计关联的变量Y进行观察，从而消除感兴趣随机变量的不确定性($I(X;Y)=H(X)-H(X|Y)$)
      - 通过对感兴趣随机变量本身进行观察，从而获得信息：$D(p||q)=\sum_{x\in X}p(x)\log\left(\frac{p(x)}{q(x)}\right)$:相对熵,交叉熵，**Kullback**熵，**K-L**散度，**K-L**距离，方向散度)
        - 理解为知道真实分布$p(x)$，其平均最短描述的下界为$H(p)$,如果考虑用来描述分布q，则这时候需要的平均最短比特数为：$H(p)+D(p||q)$
- **生成对抗网络(Generative Adversarial Network)**

GAN

◆ 网络结构

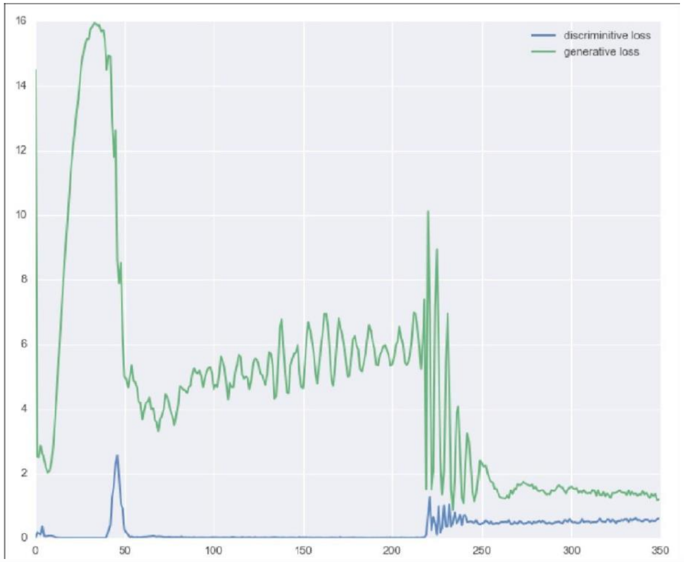
- 生成器G
- 区分器D
- 最小化博弈



- 第一个网络对真实数据建模，当第一个网络的输出等于真实数据时，第二个网络输出概率为**0.5**

$$\min_G \max_D V(D,G)=\mathbb{E}_{x\sim p_{data}(x)}\left[\log D(x)\right]+\mathbb{E}_{x\sim p_z(z)}\left[\log\left(1-D\left(G\left(z\right)\right)\right)\right]$$

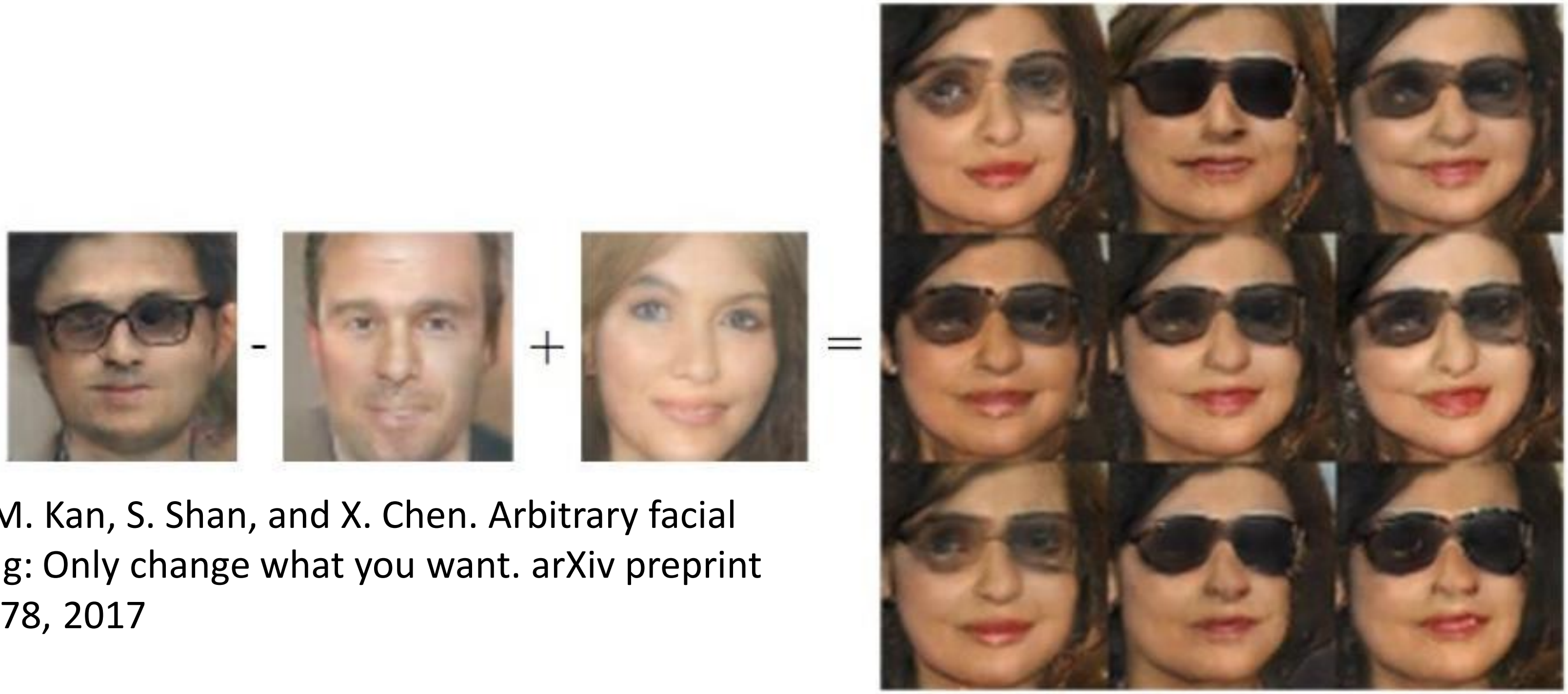
- 有时两个网络同时达到均衡状态；但有时就需要运行更长的时间来达到均衡状态，右图显示**G**，**D**的损失函数随运行次数增多的情况



效果

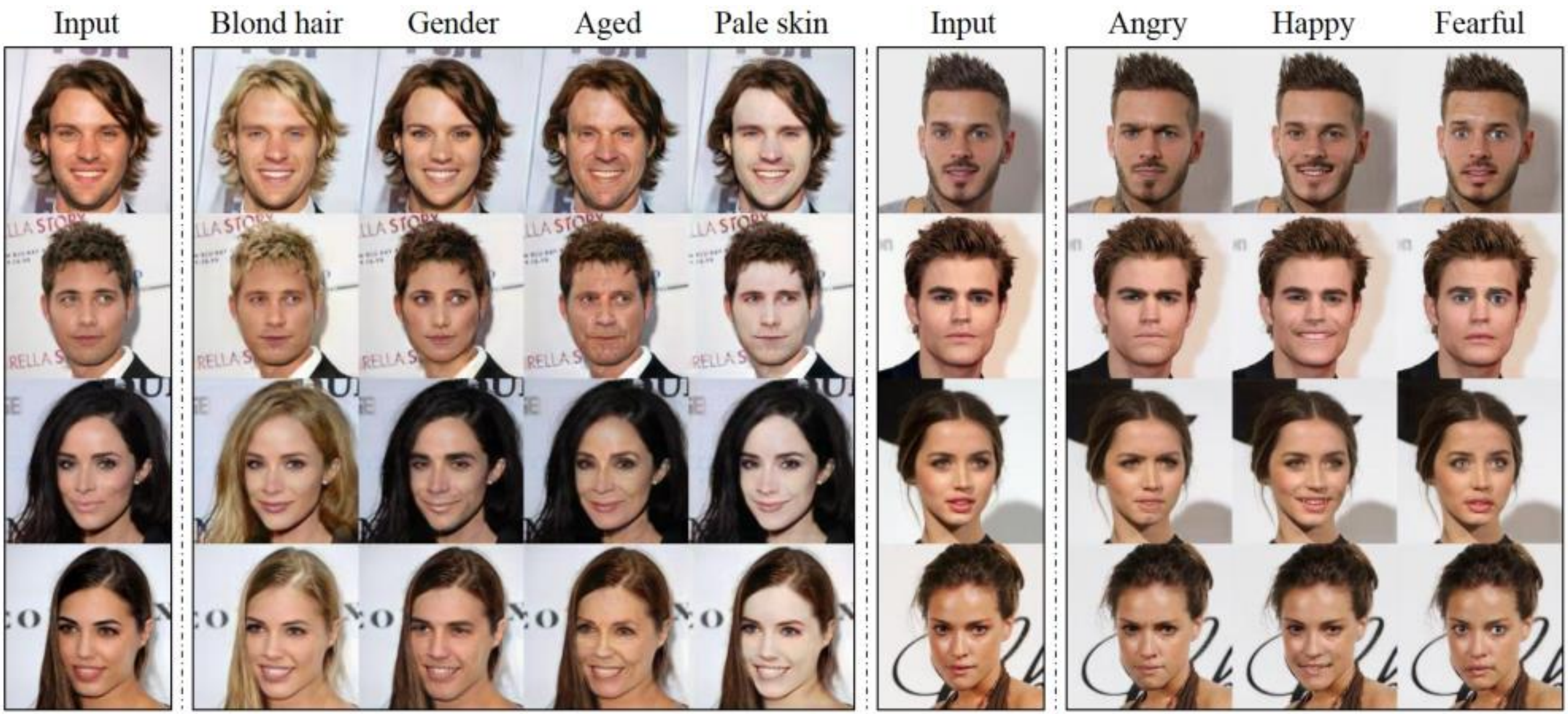
◆ GAN获得很好的效果

- 生成器输入随机噪声，输出样本数据，$G(z)$的输入服从概率分布$p(z)$,生成的数据反馈至区分器网络$D(x)$
- 区分器输入要么来自真实数据，要么来自生成器网络的输出，然后输出一个判断输入是否是真实数据的概率
- 能处理无监督学习，无标签数据远大于有标签的数据
- 能生成很多非常真实的图像



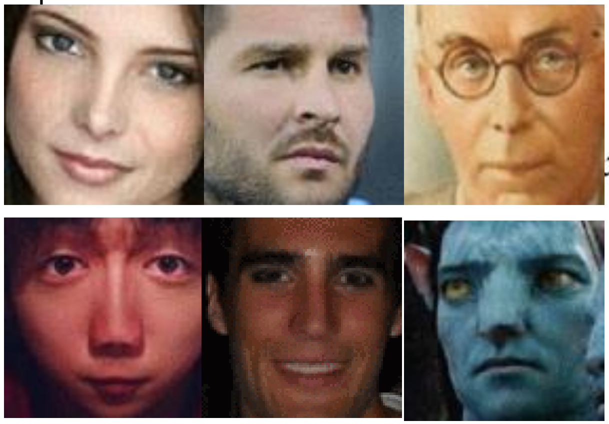
Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen. Arbitrary facial attribute editing: Only change what you want. arXiv preprint arXiv:1711.10678, 2017

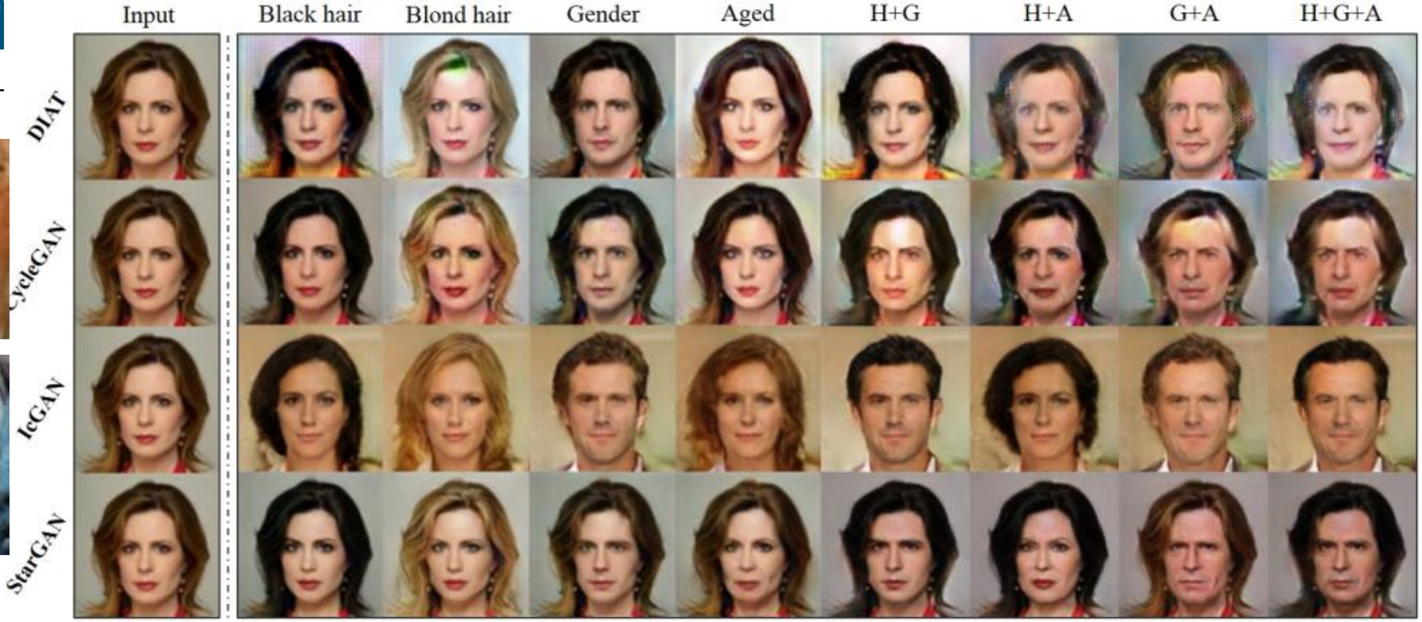
基于GAN的人脸属性编辑

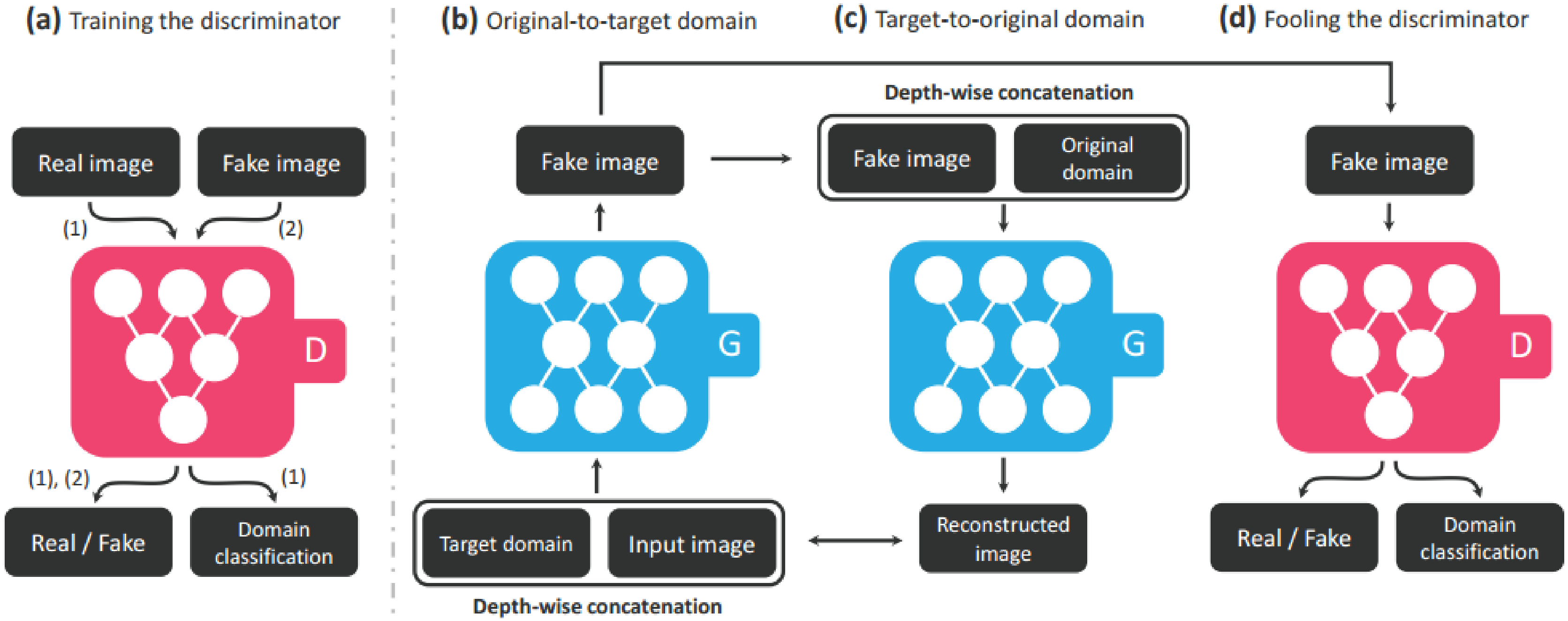


STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing

Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo. Stargan: Unified generative adversarial networks for multidomain image-to-image translation. In IEEE Conference on Computer Vision and Pattern Recognition, pages 8789– 8797, 2018.




Input
Black hair
Blond hair
Gender
Aged
H+G
H+A
G+A
H+G+A
DIAT
CycleGAN
IcGAN
StarGAN


(a) Training the discriminator
(b) Original-to-target domain
(c) Target-to-original domain
(d) Fooling the discriminator
Real image
Fake image
(1)
(2)
D
(1), (2)
(1)
Real / Fake
Domain classification
Depth-wise concatenation
Fake image
Fake image
Original domain
G
G
Target domain
Input image
Reconstructed image
Depth-wise concatenation
Fake image
D
Real / Fake
Domain classification

2.5 基本信息论-连续信源熵

◆**定义：与离散类似，采用概率分布来进行定义，假设分布为$p(v)$**

$$\sum_i p(v_i)dv\log[p(v_i)dv]$$

$$H_{连}=\lim_{dv\to 0}\left\{\sum_i -p(v_i)dv\log[p(v_i)dv]\right\}$$

$$=\lim_{dv\to 0}\left\{\sum_i -p(v_i)\log(p(v_i))dv]\right\}$$

$$-\lim_{dv\to 0}\left\{\sum_i p(v_i)\log(dv)dv\right\}$$

$$=-\int p(v)\log p(v)dv-\lim_{dv\to 0}\left\{\sum_i p(v_i)\underbrace{\log(dv)}_{无穷大}dv\right\}$$

2.5 基本信息论-连续信源熵(续1)

- **连续信源熵的定义**

$$H(x) = -\int p(x)\log p(x)dx$$

$$H_{连} = H(x) - \lim_{dv\to 0}(\log dv)$$

  - 是一个比无穷大大多少的相对量，但离散信源的则不是；但传信率，信道容量都是熵的差，所以不影响计算
- **注意，这种熵有时也称为微分熵(differential entropy).**
- **具有相同方差的随机变量中，高斯随机变量具有最大熵。因此，可以以此为基准，用于测量数据的非高斯性(non-Gaussianity)**
  - **负熵(Negentropy)**$J(y)$
  - $J(y) = H(y_g) - H(y)$，y_g是一个与y具有相同协方差矩阵的高斯随机变量，负熵总是非负的，当且仅当y是高斯变量的时候为0.
  - 计算的时候估计概率密度函数比较困难，一般对数据采用简单的近似来计算，假设随机变量y具有0均值单位方差，则近似可写为：
  - $J(y) \approx (E[G(y)] - E([G(y_g)]))^2$，$y_g$为0均值单位方差的高斯变量，$G(y) = \frac{1}{a}logcosh(ay), 1 \le a \le 2$，$cosh(x) = \frac{e^{-x}+e^x}{2}$，双曲余弦函数
  - 类似的，在随机变量$y = [y_1, y_2, \cdots, y_n]^T$的各个成份之间的互信息：

$$I(y_1, \cdots, y_n) = \sum H(y_i) - H(y)$$

2.5 基本信息论-连续信源熵(续2)

◆连续信源的最大熵

- 假设样值之间相互独立，研究使连续信源熵为最大时信号的一维最佳概率分布

$$\max_{p(x)}\left\{H(x)=-\int p(x)\log p(x)dx\right\}$$

$$subject\ to:$$

$$\int p(x)dx=1\ldots$$

- 一般用变分法解
  - ✓限制条件可写成

$$\int_a^b \varphi_1(x,p)dx=k_1$$

$$\int_a^b \varphi_2(x,p)dx=k_2$$

$$\int_a^b \varphi_m(x,p)dx=k_m$$

  - ✓求积分(泛函)为极值时函数$p(x)$: $g=\int_a^b F(x,p)dx, F(x,p)=-p(x)\log p(x)$

2.5 基本信息论-连续信源熵(续3)

◆变分法求解

$$g=\int_a^b F(x,p)dx, F(x,p)=-p(x)\log p(x)$$

$$\Phi=\int_a^b F(x,p)dx+\lambda_1\int_a^b \varphi_1(x,p)dx+\cdots+\lambda_m\int_a^b \varphi_m(x,p)dx$$

$$\Downarrow$$

$$\frac{\partial\Phi}{\partial p}=\int_a^b(\partial F/\partial p+\lambda_1\partial\varphi_1/\partial p+\cdots+\lambda_m\partial\varphi_m/\partial p)dx=0$$

$$\Downarrow$$

$$\partial F/\partial p+\lambda_1\partial\varphi_1/\partial p+\cdots+\lambda_m\partial\varphi_m/\partial p=0$$

$$\Downarrow$$

$$\lambda_i$$

2.5 基本信息论-连续信源熵(续4)

◆连续信源最大熵

➢ 信源输出的瞬时功率或输出幅度受限的情况

$$-V \le X \le V, solving : H(x) = -\int_{-V} p(x)\log p(x)dx$$

$$subject\ to : \int_{-V}^{V} p(x)dx = 1$$

当幅度限制在-V1和V2之间时：

$$p(x) = \frac{1}{V_1+V_2}$$

$$H(x) = \log(V_1 + V_2)$$

$$\partial F / \partial p = -p(x)\log p(x)$$

$$\partial \varphi_1 / \partial p = 1$$

$$\Rightarrow \log p(x) = \lambda_1 - 1 \Rightarrow p(x) = \frac{1}{2V}$$

➢ 信源输出的平均功率受限的情况

2.5 基本信息论-连续信源熵(续4)

◆连续信源最大熵

➢ 信源输出的平均功率受限的情况

$$solving: H(x) = -\int_{-V} p(x)\log p(x)dx$$

$$subject\ to: \int_{-\infty}^{\infty} p(x)dx = 1, \int_{-\infty}^{\infty} x^2 p(x)dx = \sigma^2$$

$$\partial F / \partial p = 1 + \log p(x)$$

$$\partial \varphi_1 / \partial p = 1$$

$$\partial \varphi_2 / \partial p = x^2$$

$$\Rightarrow \log p(x) = \lambda_1 - 1 + \lambda_2 x^2$$

$$\Rightarrow p(x) = e^{\lambda_1 - 1} e^{\lambda_2 x^2}$$

$$\Rightarrow p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

信号平均功率一定时，信号的最佳概率分布时数学期望为0，方差等于平均功率的高斯分布

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

$$H(x) = -\int_{-\infty}^{\infty} p(x)\log_e p(x)dx$$

$$= \log_e \sqrt{2\pi}\sigma + \frac{1}{2}$$

$$= \log_e \sqrt{2\pi e}\sigma (nat)$$

2.5 基本信息论-连续信源熵(续5)

◆熵功率

- 假如信源输出信号平均功率为N，但幅度不是高斯分布，那么熵比最大熵偏小，用熵功率来表示该信源的剩余
- 定义：就是指与这个平均功率为N的非高斯信源有同样熵的高斯信源的平均功率，若H为熵，则根据熵功率定义得

$$H = \log_2 \sqrt{2\pi e N}$$

$$e^{2H} = 2\pi e \bar{N}$$

$$\text{熵功率}\bar{N} = \frac{e^{2H}}{2\pi e} < N$$

2.5 基本信息论-熵速率和信道容量

◆信源熵速率

- 定义：信源在单位时间内输出的熵(或平均信息量)称为信源的熵速率，也叫信息速率
- 离散信源的熵速率
  - ✓离散信源的熵也就是离散信源每个消息的熵，完全由信源的概率分布决定$H(x)=-\sum P(x)log(P(x))$
  - ✓如果离散信源每秒输出消息数目为n/秒，则离散信源的熵速率$H'(x)=nH(x)$
- 连续信源的熵速率 $H(x)=-\int p(x)\log p(x)dx$
  - ✓输出带宽为W，采样率为$2W$/s
  - ✓连续信源的熵速率

$$H'(x)=2WH(x)=-2W\int p(x)\log p(x)dx$$

2.5 基本信息论-熵速率和信道容量(续1)

◆信道容量

- 信源输出的信息通过信道传送到收信者，需要度量信道的传输能力
- 定义：信道对信源的一切可能的概率分布而言能够传送的最大熵速率
- 离散信道的信道容量
  - ✓信源等概时，输出熵最大，每秒传输**n**个等宽度，有**K**种状态的脉冲

$$C = H'_{max}(x) = nH_{max}(x) = n\log K$$

  - ✓为什么要做信源信道匹配
- 连续信道的信道容量
  - ✓对频带为**W**，平均功率为**N**的受限的连续信源，其幅度分布为高斯分布时，其熵最大，且为

$$H_{max}(x) = \log_e \sqrt{2\pi eN}$$

$$C = 2WH_{max}(x) = W\log_e(2\pi eN)(nat/s)$$

  - ✓要求将信源改造成随机噪声，才能与信道匹配

◆接收熵速率

- 收信者接收一个消息符号所获得的信息量或收到的熵*I=log(后验概率/先验概率)*
- $p(i),p(j),p(i,j),p(i/j)$:发送第i个消息,接收第j个消息的概率情况

$$\log\frac{\text{后验概率}}{\text{先验概率}}=\log\frac{p(i/j)}{p(i)} \xrightarrow{\text{统计平均}} \sum_i p(i/j)\log\frac{p(i/j)}{p(i)}$$

对i进行统计平均

$$I_{cp}=\sum_j p(j)\sum_i p(i/j)\log\frac{p(i/j)}{p(i)}=H(i)-H(i/j)$$

$$I_{cp}=H(x)-H(x/y)$$

- 表示收到Y后获得的关于X的信息，称为他们的互信息。可见，由于干扰或噪声，信道没有把信源熵H(x)全部传输过去，传输过程中损失的信息量称为疑义度、可疑度或含糊度

2.5 基本信息论-离散有噪信道中的熵速率和信道容量(续1)

◆共熵H(x,y)

- $I_{cp}=H(x)-H(x/y)=H(y)-H(y/x)=H(x)+H(y)-H(x,y)$
- $H(y/x)$表示信道输入信号(即信源熵)由于干扰作用在输出端表现的散布范围，称为散布度
- 如果信道每秒传输消息数为n/s,则接收者收到的熵速率R为

$$R=n[H(x)-H(x/y)]=H'(x)-H'(x/y)=H'(y)-H'(y/x)$$

- H’(x/y),H’(y/x)分别表示疑义度熵速率和散布度熵速率

◆疑义度

- H(x/y)的意义：为了克服信道中噪声的影响，从而使接收的信息仍然被正确识别必须额外提供的纠正信息

◆信道容量

- 定义：为信道提供给收信者的最大熵速率(或信息速率)

$$C=\max R=\max[H'(x)-H'(x/y)]=\max[H'(y)-H'(y/x)]$$

2.5 基本信息论-连续有噪信道中的熵速率和信道容量

◆接收熵速率

- 收信者接收的消息y是由发送消息x和信道噪声n所组成

$$y=x+n$$

- 信源x与噪声n的共熵为

$$H(x,n) = H(x)+H(n/x)$$

- 信源和噪声独立：$H(x,n)=H(x)+H(n)$，且$y=x+n$，因此y的概率分布规律与噪声的概率分布规律一样，p(y/x)=p(n/x)
  - ✓ $H(y/x)=H(n/x) \Rightarrow H(x,y)=H(x,n) \Rightarrow H(y)+H(x/y)=H(x)+H(n)$
- 与离散信道一样，定义接收熵为:
  - ✓ $H(x)-H(x/y)=H(y)-H(n)$
- 接收熵速率(信道传信率)
  - ✓ $R=H'(x)-H'(x/y)=H'(y)-H'(n)$(bit/s)

2.5 基本信息论-连续有噪信道中的熵速率和信道容量(续1)

- **信道容量** $C=\max R=\max[H'(y)-H'(n)]$
- ***X*与*n*无关，故*H'*(*n*)与*p*(*x*)无关**

$$C=\max[H'(y)-H'(n)]=\max H'(y)-H'(n)$$

- **假定信道如下特性，干扰为随机噪声，功率谱均匀，幅度为高斯分布，平均功率为N;信道带宽为矩形带宽，宽度为W;信道输入信号平均功率受限，为P，如何计算信道容量?**
- **Shannon公式**

$$\max H'(y)=2W\log_e\sqrt{2\pi e(P+N)}=W\log_e[2\pi e(P+N)](nat/s)$$

$$H'(n)=W\log_e[2\pi eN](nat/s)$$

$$C=\max H'(y)-H'(n)=W\log_e(1+\tfrac{P}{N})(nat/s)$$

$$C=\max H'(y)-H'(n)=W\log_2(1+\tfrac{P}{N})(bit/s)$$

2.5 基本信息论-连续有噪信道中的熵速率和信道容量(续2)

◆结论

- 平均功率受限的高斯白噪声信道，其信道容量*C*与信道带宽*W*和信号噪声功率比*P*/*N*有关
- 平均功率受限的信道，当其输入信号为高斯分布式，该信道的传信率R可以达到传信率理论极限值-容量
- 平均功率受限的信道，高斯白噪声危害最大。因为这时候*H*’(*n*)最大
- 目前还没有实际系统其传信率达到信道容量，但指出了现实系统的潜在能力和理论极限值，可以作为通信系统中带宽与信号噪声比进行互换的理论基础

2.6 信源与信道匹配的编码

- **2.6.1 编码定理**
- **2.6.2 信源最佳化**
- **2.6.3 符号独立化**
- **2.6.4 概率均匀化**

2.6 信源与信道匹配的编码-编码定理

- **当信源输出有剩余的消息时，信道的传信率或熵速率就小于信道容量**
- **Shannon指出:可以通过适当的编码方法使信道的传信率无限接近信道容量**
  - 无噪声离散信道的编码定理
    - ✓ 假设信源熵为H(bit/符号)，无噪声信道容量为C(bit/秒),则总可以找到一种方法对信源的输出进行编码，使得在信道上传输的平均速率为每秒(C/H-e)个符号，其中e为任意小的正数，但要使传输速率大于C/H是不可能的
      - 定理的逆命题，即速率不可能超过C/H可以这样说明，信源每秒送入信道的熵等于信源的熵速率，这个速率不能超过信道容量，也就是说H’<=C,并且每秒钟符号个数H’/H<=C/H
      - 为使信源与信道在统计意义上达到匹配，应当在信源信道之间加入编码器，使得从信道输入端来看，输入信源应当与使信道中熵最大的信源具有同样的统计特性，这就是信源信道之间达到最佳匹配
      - 理想和近似理想的编码系统均有长时间的延迟，实用性受限

2.6 信源与信道匹配的编码-编码定理(续1)

- 有噪声离散信道的编码定理
  - ✓假设离散信源的熵速率为R，噪声信道容量为C(bit/秒)，如果R<=C,则存在一种编码方法，使信源的输出能以任意小的错误概率在信道上传输，如果R>C，就不可能有象R<=C时的编码方法
  - ✓定理本身没有给出编码方法，而是存在性！
- 信源和信道给定后，编码器的设计分为两大类
  - ✓变换信源符号间的概率分布使之达到最佳，从而使信息传输速率任意逼近信道容量，这类编码称为有效性编码，主要针对信源特性实现信源与信道之间的匹配，也称为信源编码
  - ✓变换信源符号间的规律性或相关性，使之在一定传信条件下错误概率任意小，这类编码主要为了提高抗干扰性，称为抗干扰编码，针对信道特性而采取的措施，也称为信道编码
  - ✓有噪信道的编码定理在原则上给出了有效性和抗干扰性两方面的最佳统一，**但目前没找到！**

2.6 信源与信道匹配的编码-信源最佳化

- **通信系统的传输效率就是指给定信道的信道容量利用率，表示信道的实际传信率与信道容量的比值**

$$\eta = \frac{R}{C}$$

- **无噪声信道中** $\eta = \dfrac{H'(x)}{C} = \dfrac{H'(x)}{\max R} = \dfrac{H'(x)}{\max H(x)}$

- **提高传输效率，基本任务就是要改造信源，使其熵最大化，此时** $\eta \to 1$**，达到这个目的的过程就是信源最佳化的过程。而信源熵**$H(x)$最大化的实际上就是找到一种最佳的概率分布

- 根据熵函数的性质，在离散信源情况下，当各个符号彼此独立并且概率相等时，信源熵达到最大值

2.6 信源与信道匹配的编码-信源最佳化(续1)

◆信源熵的最大化

- ➢首先：进行符号独立化，解除符号间的相关性
- ➢各个符号概率均匀化

◆实际操作

- ➢如果符号已经相互独立了，那就只需要概率均匀化
- ➢一般先解除符号间的相关性，再进行概率均匀化

◆问题：Huffman编码属于那一种？为什么？

2.6 信源与信道匹配的编码-符号独立化

◆符号独立化

- ➢解除信源各个符号间的相关性
- ➢弱记忆信源和强记忆信源

◆弱记忆信源(弱相关信源)

- ➢信源符号序列中，只有相邻的少数几个符号之间有统计相关性，而距离较远的符号之间的统计相关性可以忽略不计
- ➢怎么去相关?
  - ✓把相关性强的几个符号看成一个符号，这样这些大符号之间的相关性就小得多(延长或合并法)
  - ✓等价于把基本信源空间变成多重空间，重数越多越好吗?
    - ●复杂性

2.6 信源与信道匹配的编码-符号独立化(例)

◆二元序列:000101101011…,变成二重空间的四元序列00，01，01，10，10，11…,新序列的符号熵?

◆原始信源序列的熵为条件熵

$$H=H(x_m/x_n)=-\sum_{n=1}^{2}\sum_{m=1}^{2}P(x_n,x_m)\log P(x_m/x_n)$$

$$=-\sum_{n=1}^{2}\sum_{m=1}^{2}P(x_n,x_m)\log P(x_m,x_n)$$

$$+\sum_{n=1}^{2}\sum_{m=1}^{2}P(x_n,x_m)\log P(x_n)$$

$$=-\sum_{n=1}^{2}\sum_{m=1}^{2}P(x_n,x_m)\log P(x_m,x_n)+$$

$$+\sum_{n=1}^{2}P(x_n)\log P(x_n)$$

$$=H_2-H_1(bit/sign)$$

2.6 信源与信道匹配的编码-符号独立化(续1)

- **强记忆信源(相关信源)**
  - 信源符号序列具有很强的相关性，根据一部分可以推测出其余部分的符号
  - 可以根据统计特性做预测，能预测的不需要传送，例如只传送预测差值，从而节约信道容量--------预测法
  - 例如增量调制，差分编码调制(DPCM)
- **别的方法也可以解除相关，压缩信源和提高传输效率**
  - 模式识别技术，变换编码技术，相关编码技术等

2.6 信源与信道匹配的编码-概率均匀化

- **解除相关性后，使用概率均匀化，可以进一步改造有剩余信源的输出，去掉多余度，提高熵速率，使传信率接近信道容量**
  - 称为信源的有效性编码
  - **Huffman**编码
  - **Shannon-Fano**编码
  - 通过什么手段达到？
    - ✓出现概率大的符号编成短码，出现概率较小的符号编成长码
- **编码器** $X=\{x_1,x_2,\ldots,x_n\}$通过编码器中的符号 $A=\{a_1,a_2,\ldots,a_D\}$转换成$W=\{w_1,w_2,\ldots,w_n\}$
  - 二元代码，同价代码，等长代码，单义代码，非续长代码

2.7 信息率失真函数

- **从失真角度看，编码可以分为两类**
  - 没有失真，无损的
  - 有失真
- 数据压缩是信息传输和处理的重要研究内容，率失真理论研究的就是在允许一定失真的前提下，对信源的压缩编码。率失真信源编码定理（香农第三定理）指出：率失真函数$R(D)$就是在给定失真测度条件下，对信源熵可压缩的最低程度。

2.7 信息率失真函数(续1)

◆失真测度与平均失真

- 在允许一定失真的前提下，从提高传输效率的角度出发，可以对信源信息量事先进行压缩再予传输，这章要讨论的问题就是给定一个失真度，求出在平均失真小于给定值的条件下，信源所能压缩的最低程度，即率失真函数$R(D)$
- **失真测度$d(x, y)$:**给定离散信源，信道输出符号y_j引起的失真用 $d(x_i, y_j)$（$i=1, 2, \ldots, I \quad j=1, 2, \ldots, J$）表示，简记为$d_{ij}$，将所有的$d_{ij}$列出来，可以得到下面的失真测度矩阵

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

$$[d] = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1J} \\ d_{21} & d_{22} & \cdots & d_{2J} \\ \vdots & \vdots & \cdots & \vdots \\ d_{I1} & d_{I2} & \cdots & d_{IJ} \end{bmatrix}$$

2.7 信息率失真函数(续2)-失真测度举例

◆几个例子

- 汉明（Hamming）失真测度
- 信源输出符号$X = \{x_1, x_2, \ldots, x_K\}$，信道输出符号$Y = \{y_1, y_2, \ldots, y_K\}$，约定失真测度 $\begin{cases} y_i = x_i & \text{无误码} & d_{ii} = 0 \\ y_j = x_i (i \neq j) & \text{误码} & d_{ij} = 1 \end{cases} \quad i, j = 1,2,\cdots,K$
- 上述约定可以用矩阵表示如下，式中$d_{ij} \geq 0 \quad i, j = 1, 2, \ldots, K$为信源方发送符号$x_i$而信宿方判为$y_j$引起的失真度

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

- 对于矢量传输情况，若信道的输入、输出均为N 长序列$\boldsymbol{X} = X_1 X_2 \ldots X_N$，$\boldsymbol{Y} = Y_1 Y_2 \ldots Y_N$，定义失真测度为

$$d^{(N)}(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{N}\sum_{k=1}^{N} d(X_k, Y_k)$$

2.7 信息率失真函数(续3)-失真测度举例

【例】 **平方误差失真测度**

信源输出符号 $X=\{0, 1, 2\}$，信道输出符号 $Y=\{0, 1, 2\}$，给出失真测度 $d_{ij}=(x_i - y_j)^2 \quad i, j = 0, 1, 2$

则失真测度矩阵为

$$[d]=\begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ 4 & 1 & 0 \end{bmatrix}$$

【例】 **绝对值误差失真测度**

信源输出符号 $X=\{0, 1, 2\}$，信道输出符号 $Y=\{0, 1, 2\}$，给出失真测度 $d_{ij}=|x_i - y_j| \quad i, j = 0, 1, 2$

则失真测度矩阵为

$$[d]=\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

2.7 信息率失真函数(续4)-平均失真

◆ 离散信源 $\begin{bmatrix} X \\ p(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_I \\ p(x_1) & p(x_2) & \cdots & p(x_I) \end{bmatrix}$，经有扰信道传输，信道输出符号为$Y=\{y_1, y_2, \ldots, y_J\}$，平均失真即对$d_{ij}$（$i=1, 2, \ldots, I$; $j=1, 2, \ldots, J$）求统计平均值，记为

$$\overline{D} = \sum_{i=1}^{I}\sum_{j=1}^{J} p(x_i y_j) d_{ij} = \sum_{i=1}^{I}\sum_{j=1}^{J} p(x_i) p(y_j | x_i) d_{ij}$$

平均失真 $\overline{D}$ 是对在给定信源分布$p(x)$条件下，通过有扰信道传输而引起失真的统计平均度量。

2.7 信息率失真函数(续5)

◆定义

给定信源，即信源概率分布$p(x)$一定，给定失真测度矩阵$[d]=[d_{ij}]$，寻找信道，记它的转移概率矩阵为$P(Y/X)$，要求满足

$$\boxed{\overline{D}=\sum_i\sum_j p(x_i)p(y_j|x_i)d_{ij}\le D}$$

式中D是预先给定的失真度，上式称为**保真度准则**。

2.7 信息率失真函数(续6)

当信源$p(x)$一定时，平均互信息量$I(X;Y)$是信道转移概率函数$p(y|x)$的∪型凸函数，这意味着可以关于$p(y|x)$对平均互信息量$I(X;Y)$求得极小值，定义这个极小值为**率失真函数*R*(*D*)**，即：

$$R(D) \triangleq \min_{p(y|x)} \left\{ I(X;Y):\ \overline{D} \le D \right\}$$

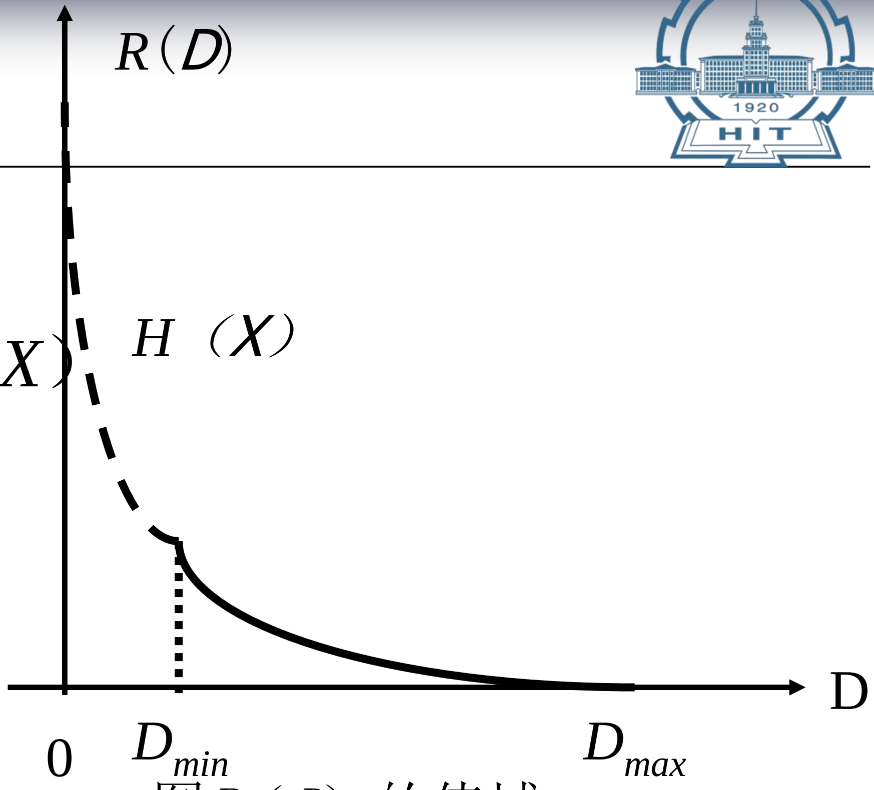
意义在于，选择$p(y|x)$即选择某种编码方法在满足$\overline{D} \le D$的前提下，使$I(X;Y)$达到最小值$R(D)$，这就是满足平均失真条件下的信源信息量可压缩的最低程度。

2.7 信息率失真函数(续7)

◆率失真函数的定义域和值域

➢**R(D)的值域**　$0 \le R(D) \le H(X)$

➢**R(D)的定义域**



图$R(D)$的值域

（1）D的最小值$D_{\min}$

在给定的失真测度矩阵中，对每一个x_i，找一个最小的 d_{ij}，然后对所有的$i=1, 2, \ldots, I$求统计平均值，就是D的最小值，即

$$D_{\min} \triangleq \sum_i p(x_i) \min_{y_j} d_{ij}$$

2.7 信息率失真函数(续8)

（2）D的最大值D_{max}

当$R(D)$ 达到其最小值R_{min}（D）= 0时，对应的失真最大，这种情况下D对应着$R(D)$ 函数定义域的上界值D_{max}。

$$D_{max} \triangleq \min\{D : R(D) = 0\} = \min\{D: I(X; Y) = 0\}$$

求出计算D_{max}的显式:

思考:率失真函数的计算方法

$$D_{max} = \min_j \sum_{i=1}^{I} p(x_i) d_{ij} \qquad j = 1, 2, \ldots, J$$

◆ 纵上所述，$R(D)$的定义域为:

$D_{min} \le D \le D_{max}$，式中$D_{min}$和$D_{max}$可分别由上两个公式求出

2.8 基于熵的几种基本度量

第二章 信息论基础小结

◆2.5 基本信息论

➢ 2.5.1 通信基本模型
➢ 2.5.2 信息度量
➢ 2.5.3 离散信源熵
➢ 2.5.4 连续信源熵
➢ 2.5.5 熵速率与信道容量
➢ 2.5.6 离散有噪信道中的信道容量
➢ 2.5.7 连续有噪信道中的信道容量

◆2.6 信源与信道匹配的编码

➢ 2.6.1 编码定理
➢ 2.6.2 信源最佳化
➢ 2.6.3 符号独立化
➢ 2.6.4 概率均匀化

◆2.7 信息率失真函数

第二章 参考文献

- 郑君里，杨为理，应启珩等，信号与系统(第二版)，上下册，高等教育出版社，2000.5
- 丁玉美，高西全.数字信号处理(第二版),西安电子科技大学出版社，2001.1
- 贾世楼，信息论理论基础(第二版)，哈尔滨工业大学出版社,2001.2
- Thomas M.over， JoyA.Thomas著,阮吉寿，张华译.信息论基础(第二版)，机械工业出版社，2009.11

课后阅读掌握

- 率失真函数的计算方法！
- 如何用信息论的观点来解释数据压缩中用到的各种技术：例如预测，变换，Huffman编码等

END

第2章 信号处理与信息论基础

服务

- 内容处理(Content Processing)
- 文档(Documents)
- 安全(Security)
- 语义(Semantics)
- 同步(Synchronization)
- 成组通信(Group Communications)

系统

- 多媒体数据库(MM-Database)
- 程序设计(Programming)
- 媒体服务器(Media-Server)
- 多媒体操作系统(MM Operating System)
- 通信(Communication)
- 光存储器(Optical Memories)
- 服务质量(Quality of Service)
- 网络(Networks)

基础

- 计算机结构(Compute Architectures)
- 压缩(Compression)
  - 图像和图形(Image & Graphics)
  - 视像(Video)
  - 声音(Audio)
  - 文本(Text)
  - 动画(Animation)

底层

- 信息论
- 信号处理
- 数字化
  - HVS-HAS
  - D/A-A/D
  - 量化、编码
  - 多媒体表示