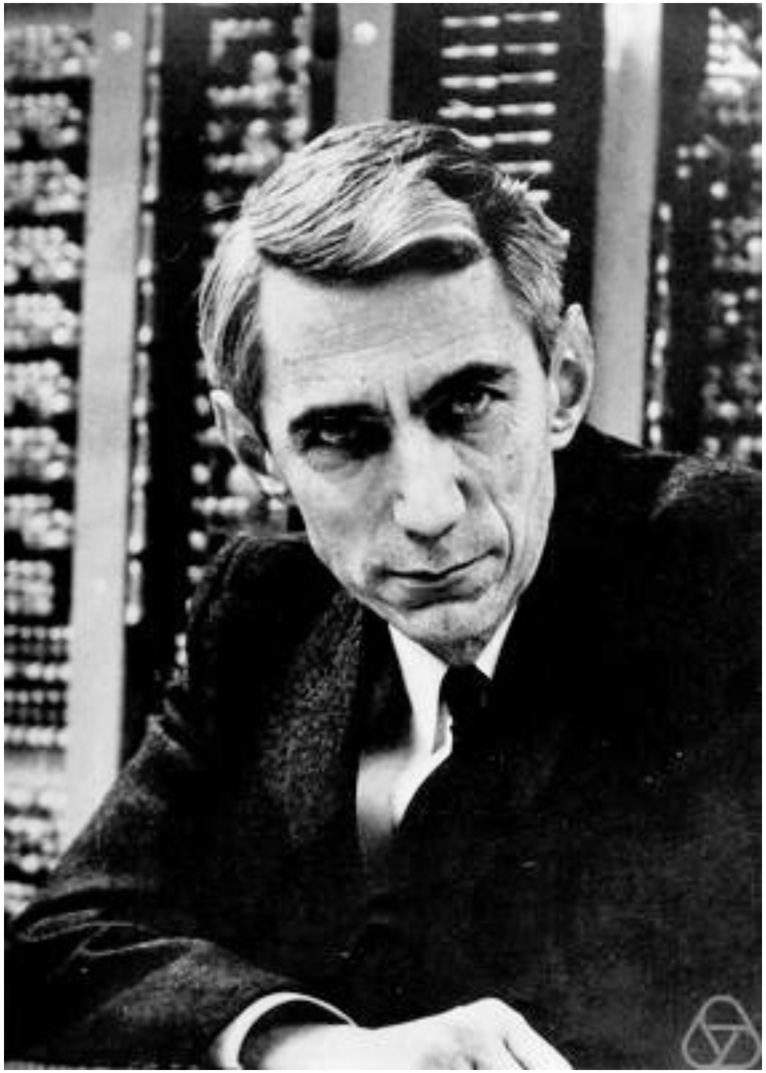
信息论与编码

信息论创始人

C. E. Shannon

美国科学家



信息论是由Shannon奠基的一门数学学科，他在《贝尔系统电话杂志》发表了题为《通信的数学理论》的长篇论文，他创立了信息论，但是却没有给出信息的确切定义， 他认为“信息就是一种消息”。

4.1 信息

4.2 信源编码

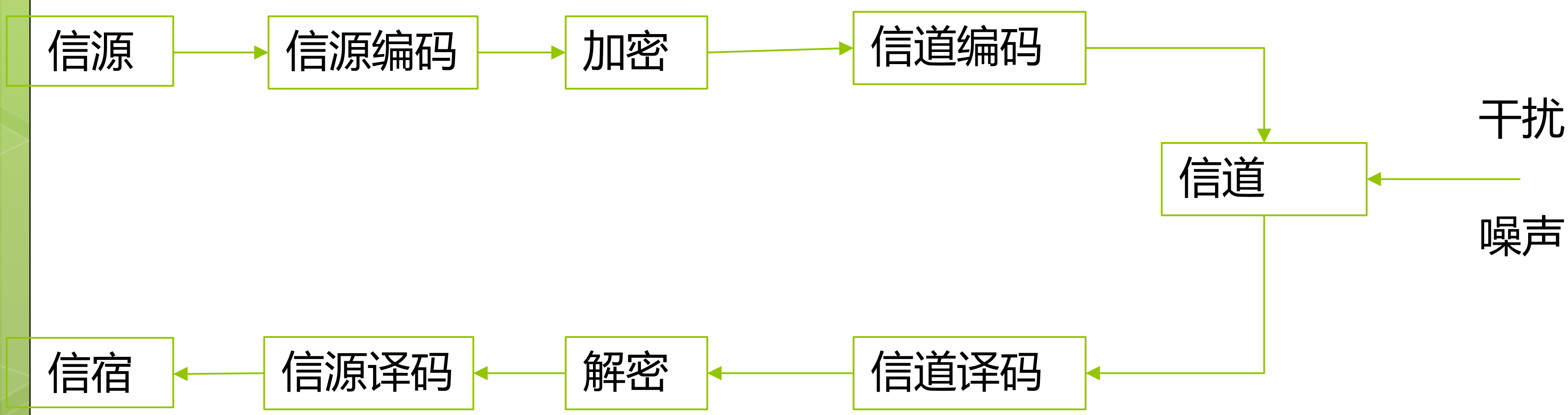
4.3 信道容量

信息的定义

- 古时的通信： 烽火台
- 信息传播五阶段：
- 手势和语言－－文字－－印刷术－－电磁波－－计算机和通信
- 微电子技术、通信技术和计算机技术促进了信息技术发展
- 信息产业的发展促进了社会产业结构的变化与发展

通信系统模型

信源：消息的来源
信道：消息的传送媒介
信宿：消息的目的地



- 通信系统的性能指标
  - 有效性 - - 信源编码，去除冗余
  - 可靠性 - - 信道编码，添加冗余
  - 安全性 - - 加密编码
  - 经济性

  ？ 如何优化， 使得通信系统的这些指标最佳

Shannon信息论的中心问题

- 信息论，又称为“通信的数学理论”，是研究信息的传输、存储、处理的科学。
- 信息论的中心问题：为设计有效而可靠的通信系统 提供理论依据。

- 问题一：信源消息常常不能够完全发送。（否则发送量巨大，比如：无尽的天空，因此优先拣“有用的发送”）
- 问题二：信道因干扰而出现差错，如何进行检错和纠错。

What's the information?

- Although there is no accurate definition of information, but it has two obvious characteristic:
- Widespread and abstractive

- Widespread

  The objective world is filled with information

  The humanity cannot leave information

  The knowledge, the books are the useful information accumulation.

- Signal is physical expression level of the information. It is the most concrete level.
- Message: (also named symbol) is the mathematical expression level of the information.
- Information: It is in higher level of philosophy abstract, i.e., it is the higher expression level of signal and message. In these three levels, signal is the most concrete and information is most abstract.

- 香农信息论主要讨论的是语法信息中的概率信息，我们也以概率信息为主要研究对象。

信息的直观认识1

信道上传送的是随机变量的值

* 这就是说， 我们收到消息之前，并不知道消息的内容，否则消息就没必要发送了。
* 消息随机变量有一个概率分布。
* 消息随机变量的一个可能取值就称为一个事件。

信息的直观认识2

- 事件发生的概率越小，此事件含有的信息量就越大。（不太可能发生的事件发生了，另人震惊）
- 事件： 中国足球队5:0力克韩国足球队此事件含有的信息量大（小概率事件发生了，事件的信息量大）

  事件：中国足球队0:1负于韩国足球队此事件含有的信息量小，（大概率事件发生了，事件的信息量小）

信息的直观认识3

- 消息随机变量的随机性越大，此消息随机变量含有的信息量越大。
- 例　消息随机变量 X = 中国足球队与巴西足球队比赛的结果

  则消息随机变量 X含有的信息量小。

- 例　消息随机变量 Y=意大利足球队与德国足球队比赛的结果。

  则消息随机变量 Y含有的信息量大。

信息的直观认识4

- 两个消息随机变量的相互依赖性越大，它们的互信息量就越大。

- 例 X=呼和浩特明日平均气温， Y=包头明日平均气温，Z=北京明日平均气温， W=纽约明日平均气温，则

  X与Y互信息量大。

  X与Z互信息量小得多。

  X与W互信息量几乎为0。

信息的度量

- 把信源和信宿看作两个随机变量
- 设X为一个离散型随机变量，概率分布如下：

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} \sim \begin{bmatrix} x_1, & \ldots, & x_n \\ P(x_1), & \ldots & P(x_n) \end{bmatrix}$$

自信息

- 定义: 假定一个随机变量X 的所有取值为 $x_i, i=1,2,\ldots,n,$

而事件 $X=x_i$ 的概率为 $P(x_i)$. 则事件 $X=x_i$

的自信息(self-information)定义为

$$I(x_i)=\log_2\left(\frac{1}{P(x_i)}\right)=-\log_2 P(x_i)$$

- 单位：比特、奈特、哈特
- 三个信息单位之间的转换关系如下：

$$1nat = \log_2 e \approx 1.433bit$$

$$1Hart = \log_2 10 \approx 3.322bit$$

$$1bit \approx 0.693nat$$

$$1bit \approx 0.301Hart$$

注： 1 事件的概率越低，事件发生时所带来的信息越多。小概率事件带来更多信息。

2 计算自信息的公式中底数可以取任何底数。通常取2为底数的对数，此时得出的单位为比特。

- 例1 投掷均匀硬币可能出现两种结果，该信源的每一个输出所含的信息量为

$$I(x_i) = -\log_2 P(x_i) = -\log_2(0.5) = 1(bit), i = 0,1.$$

- 例2 信源为连续投掷硬币n次的输出结果。用0表示正面朝上，1表示正面朝下，x_i 表示第i次的投掷结果。则将输出结果看作随机事件X时，X的状态可以为 $x_1, x_2, \ldots, x_n$ 的任意值， 而且等可能地发生。因此每一个输出结果所携带的信息为

$$I(x_1, \ldots, x_n) = -\log_2 P(x_1, \ldots, x_n) = -\log_2 \prod_{i=1}^{n} P(x_i) = n(bit)$$

互信息

假定X和Y是两个随机变量，它们的状态分别取自有限集合

$\{x_1, x_2, \ldots, x_m\}$ 和 $\{y_1, y_2, \ldots, y_n\}$

定义　事件 x_i和 y_j 之间的互信息为

$$I(x_i; y_j) = \log_2(P(x_i \mid y_j) / P(x_i))$$

注： 1 $I(x_i; y_j) = I(y_j; x_i).$

2. 如果事件 x_i 和y_j 独立，则有

$$I(x_i; y_j) = \log_2\left(\frac{P(x_i \mid y_j)}{P(x_i)}\right) = \log_2 1 = 0$$

条件自信息

- 定义　在事件 y_j 发生的前提下， 关于事件 x_i 的自信息称为条件自信息(conditional self-information), 表示为

$$I(x_i \mid y_j) = \log_2\left(\frac{1}{P(x_i \mid y_j)}\right) = -\log_2(P(x_i \mid y_j))$$

因此互信息可以写为

$$I(x_i; y_j) = \log_2\left(\frac{P(x_i \mid y_j)}{P(x_i)}\right) = \log_2 P(x_i \mid y_j) - \log_2 P(x_i) = I(x_i) - I(x_i \mid y_j).$$

上式给出了互信息、自信息和条件信息之间的关系。

随机变量的信息度量

- 定义 假设随机变量X和Y所有可能的状态分别来自有限集合 $\{x_1,x_2,\dots,x_n\}$和 $\{y_1,y_2,\dots,y_m\}$。则X 和Y之间的平均互信息定义为

$$I(X;Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_i,y_j)I(x_i;y_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_i,y_j)\log_2(\frac{P(x_i,y_j)}{P(x_i)P(y_j)})$$

上述随机变量的互信息的意义可以理解为所有可能的事件对 (x_i,y_j) 的互信息 的加权平均，而权重就是这两个事件

共同发生的概率。

性质：

- 性质1 ：$I(X;Y)=I(Y;X)\geq 0.$
- 性质2： 当且仅当X和 Y统计独立时,有

$$I(X;Y)=I(Y;X)=0.$$

随机变量的熵(entropy)　信源熵

定义：信源熵指信源随机变量X的各离散自信息量的数学期望，即信源的平均信息量

$$H(X)=E(I(x_i))=\sum_{i=1}^{n}P(x_i)I(x_i)=-\sum_{i=1}^{n}P(x_i)\log_2 P(x_i)$$

注： 上述对随机变量平均自信息的表示符号为H(X), 它表示一个系统的不确定性程度，而在统计动力学中表示系统不确定性程度的量称为系统的熵。因此这里采用同样的符号， 而且也称H(X) 为随机变量的熵。

信源熵，香农熵，无条件熵； 单位：比特 / 字符

条件熵

- 定义： 随机变量X 对于Y的条件平均自信息或条件熵定义为

$$I(X|Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_i,y_j)\log_2\frac{1}{P(x_i|y_j)}$$

根据关于平均互信息、平均自信息和条件平均自信息的定义不难得到下述等式：

$$I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X).$$

平均互信息的物理意义

1. $$I(X;Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_i y_j)(\log P(x_i \mid y_j)-\log P(x_i))$$
$$=H(X)-H(X \mid Y)$$

$$H(X \mid Y)$$

---损失熵

表示收到 Y后，对X仍存在不确定度，代表信道中损失的信息。

平均互信息的物理意义

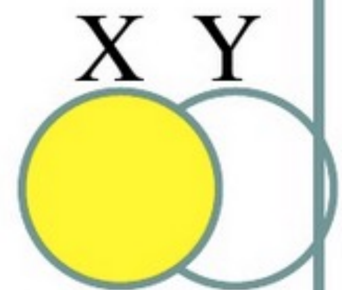
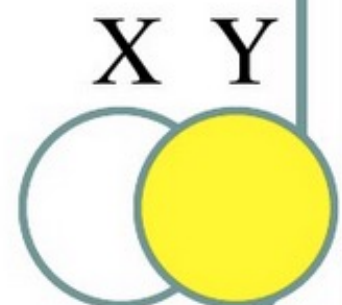
2 $I(Y;X)=H(Y)-H(Y|X)$

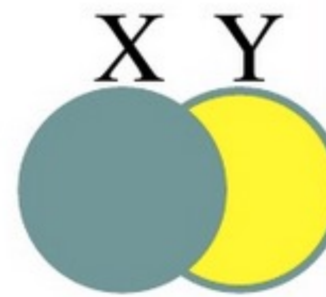
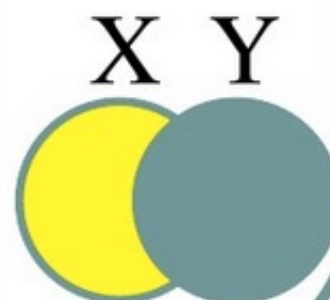
$H(Y|X)$ ---噪声熵
表示发出X后，对Y仍存在不确定性，由于信道中的噪声引起的

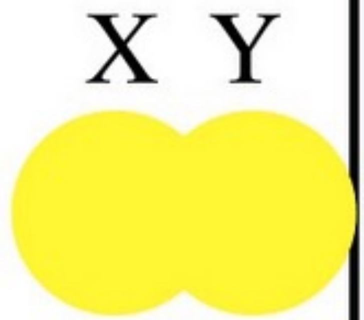
联合熵

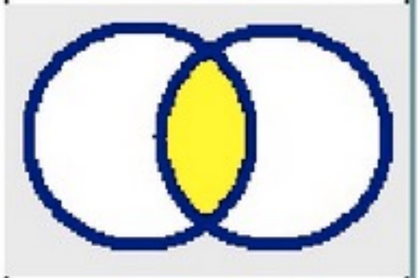
$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y) log\, p(x,y)$$

H(X, Y) 或 H（XY）表示一个多维随机变量的随机系统获得的 信息量。

名称	符号	关系式	图示
无条件熵	$H(X)$	$H(X)=H(X/Y)+I(X;Y)$ $\geq H(X/Y)$ $H(X)=H(XY)-H(Y/X)$	
	$H(Y)$	$H(Y)=H(Y/X)+I(X;Y)$ $\geq H(Y/X)$ $H(Y)=H(XY)-H(X/Y)$	

名称	符号	关系式	图示
条件熵	$H(Y/X)$	$H(Y/X)=H(XY)-H(X)$ $=H(Y)-I(X;Y)$	
	$H(X/Y)$	$H(X/Y)=H(XY)-H(Y)$ $=H(X)-I(X;Y)$	

名称	符号	关系式	图示
联合熵	$H(XY)$	$H(XY)=H(X)+H(Y/X)$ $=H(Y)+H(X/Y)$ $=H(X)+H(Y)-I(X;Y)$ $=H(X/Y)+H(Y/X)+I(X;Y)$	

名称	符号	关系式	图示
交互熵	$I(X;Y) = I(Y;X)$	$I(X;Y) = H(X) - H(X/Y)$ $= H(Y) - H(Y/X)$ $= H(X) + H(Y) - H(XY)$ $= H(XY) - H(X/Y) - H(Y/X)$	

信息熵的基本性质

- 定理 (最大离散熵定理)

  信源中包含n 个不同的离散消息时，信源熵H(X)有

$$H(X) \le \log_2 n$$

当且仅当X中各个消息出现的概率全相等时，上式取等号。

$$H(X)-\log_2 n$$

$$=\sum_{i=1}^{n}P(x_i)\log_2\frac{1}{P(x_i)}-\sum_{i=1}^{n}P(x_i)\log_2 n$$

$$=\sum_{i=1}^{n}P(x_i)\log_2\frac{1}{nP(x_i)}$$

令 $z=\frac{1}{nP(x_i)}$，引用 $\ln z\le z-1, z>0$，并注意 $\log z=\ln z\log e$，

$$H(X)-\log n$$

$$\le\sum_{i=1}^{n}[P(x_i)(\frac{1}{nP(x_i)}-1)]\log e=[\sum_{i=1}^{n}\frac{1}{n}-P(x_i)]=0.$$

当且仅当 $z=\frac{1}{nP(x_i)}=1$，即 $P(x_i)=\frac{1}{n}$，等式成立。

对于单符号离散信源，当信源呈等概率分布时具有最大熵。