通 信 安 全

L2—信息论基础

- 教师：崔爱娇

- 编号：ELEC3019
- 学时：32学时

概 要

信息论基础

- 信息的特征
- 信息量
- 保密通信系统
- 小结

概 要

信息论基础

- 信息的特征
- 信息量
- 保密通信系统
- 小结

信息的特征

信息奠基人香农认为： 信息是用来消除随机不确定性的东西。[1]

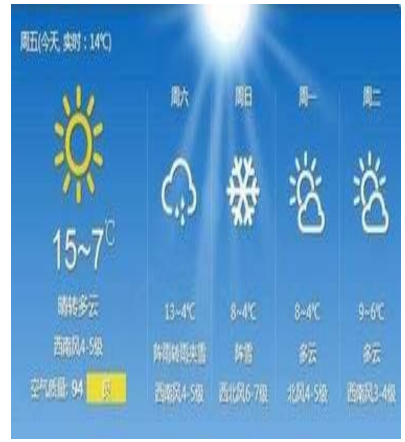
依附性　　可再生　　可传递性　　可贮存性

[1] Claud Shannon, “A mathematical theory of communication,” in *Bell System Technical Journal*, 1948.

信息的特征

可压缩性

可共享性

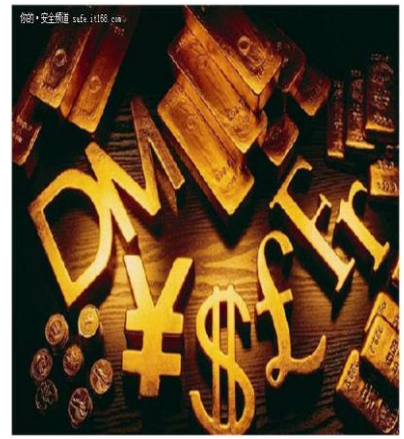


可预测性

有效性和无效性

信息的特征

可处理性



有价值

概 要

信息论基础

- 信息的特征
- 信息量
- 保密通信系统
- 小结

信息量

信息：指音讯、消息、通讯系统传输和处理的对象，泛指人类社会传播的一切内容[1]

信息可以被量化吗？

信息量： 信息多少的量度。[2]

信息量 $\xleftrightarrow{\text{直接}}$ 不确定性

信息量的度量就等于不确定性的多少

$I(x) \to p(x)$ **单调函数；** $I(x,y) = I(x) + I(y)$, $p(x, y) = p(x)p(y)$

$\boxed{I(x) = -\log p(x)}$ **负号是用来保证信息量是正数或者零，基数为2，信息单位bits**

$I(x)$:也被称为自信息量：随机变量的**某个事件发生所带来的信息量。**

单调关系图?

信息熵

平均信息量： $H(X)=-\sum_{x} p(x)\log p(x)=-\sum_{i=1}^{n} p(x_i)\log p(x_i)$

$H(x)$：随机变量x的熵，它表示随机变量不确定性；
对所有可能发生的事件产生的信息量的期望。

- 随机变量的取值个数越多，状态数也就越多，信息熵就越大，混乱程度就越大；
- 当随机分布为均匀分布时，熵最大；
- $0 \le H(x) \le \log n$
- 熵只依赖于随机变量的分布, 与随机变量取值无关，所以也可以将 X 的熵记作 $H(p)$
- 令0log0=0(因为某个取值概率可能为0)。

一维——多维

联合熵：$H(X,Y)=-\sum_{x,y} p(x,y)\log p(x,y)=-\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_i)\log p(x_i,y_i)$

性质

Eg1. 考虑一个随机变量 x。这个随机变量有4种可能的状态，每个状态都是等可能的。为了把 x 的值传给接收者，我们需要传输几比特的消息？

$$H(X) = -4 \times \frac{1}{4}\log_2\frac{1}{4} = 2bits$$

Eg2.现在考虑一个具有4种可能的状态 $\{a, b, c, d\}$ 的随机变量，每个状态各自的概率为 (1/2,1/4,1/8,1/8)

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{8}\log_2\frac{1}{8} = 1.75bits$$

非均匀分布比均匀分布的熵要小

编码

非均匀分布比均匀分布的熵要小　　？？？

使用更短的编码来描述更可能的事件，使用更长的编码来描述不太可能的事件。

Eg2.现在考虑一个具有4种可能的状态 $\{a, b, c, d\}$ 的随机变量，每个状态各自的概率为 (1/2,1/4,1/8,1/8)

0、10、110、111来表示状态 $\{a, b, c, d\}$。传输的编码的平均长度就是

平均码长 $=\frac{1}{2}\times 1+\frac{1}{4}\times 2+\frac{1}{8}\times 3\times 2=1.75bits$　　熵和最短编码长度的关系？？

香农编码定理表明：

熵是传输一个随机变量状态值所需的比特位下界（最短平均编码长度）

信息熵可应用在信息压缩领域。

条件熵

条件熵$H(Y|X)$：表示在已知随机变量X的条件下随机变量Y的不确定性。

$H(Y \mid X)$ 定义为X 给定条件下Y 的条件概率分布的熵对X 的数学期望。

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

- 条件熵相当于联合熵$H(X,Y)$减去单独的熵$H(X)$，即$H(Y \mid X) = H(X,Y) - H(X)$
- 描述X和Y所需的信息是描述X自己所需的信息，加上给定X的条件下具体化Y所需的额外信息；

熵的性质

1. 非负性

$H(x) \geq 0$。事件概率为1时，熵为0。

2. 可加性

统计独立的信源，联合信源的熵等于它们各自的熵之和。

3. 对称性

熵与具体自变量的出现位置的差异无关。熵只与随机变量的总体结构有关，与信源总体的统计特性有关。

4. 最大熵特性（极值性）

信源各状态为等概率分布时，熵值最大。$H(x) \leq \log n$

互信息量

互信息：一个随机变量中包含的关于另一个随机变量的信息量，或者说是一个随机变量由于已知另一个随机变量而减少的不肯定性。

$$I(x;y) = H(x) - H(x \mid y)$$

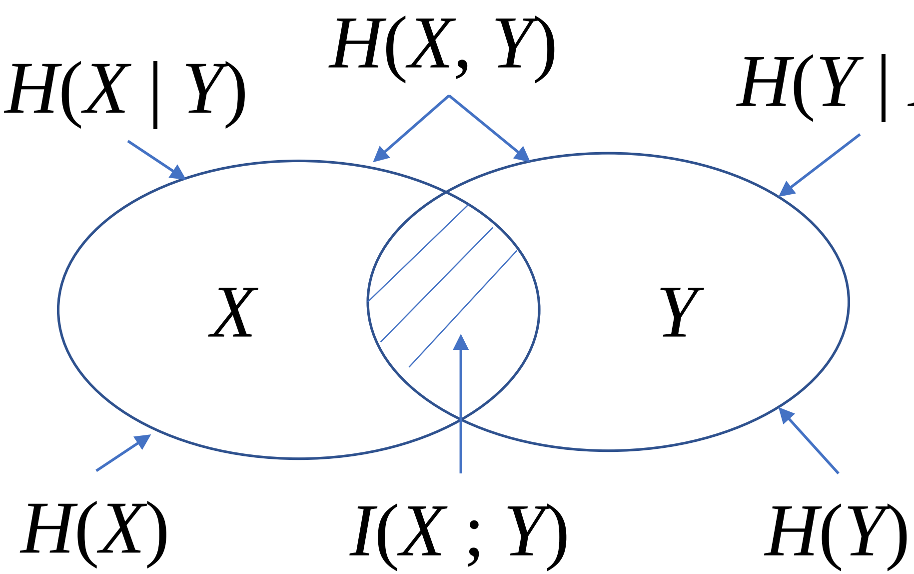
它反映了处理器处理信息的能力

先验概率：信源发出x的概率$p(x)$

后验概率：信宿收到y后推测信源发出x的概率

y 对 x 的互信息：x 的后验概率与先验概率比值的对数

互信息量



- 信息接受处理后，必能减少事件的不确定性（增加负熵，至多不变）；
- 处理次数足够多的，不确定量一定趋于零，根本上消除不确定性。

$I(x;y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X \mid Y) \\
&= H(X) + H(Y) - H(X,Y) \\
&= \sum_x p(x)\log\frac{1}{p(x)} + \sum_y p(y)\log\frac{1}{p(y)} - \sum_{x,y} p(x,y)\log\frac{1}{p(x,y)} \\
&= \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}
\end{aligned}
$$

互信息的性质

1. 对称性 $I(x; y) = I(y; x)$

2. x 与 y 独立时， $I(x; y) = 0$

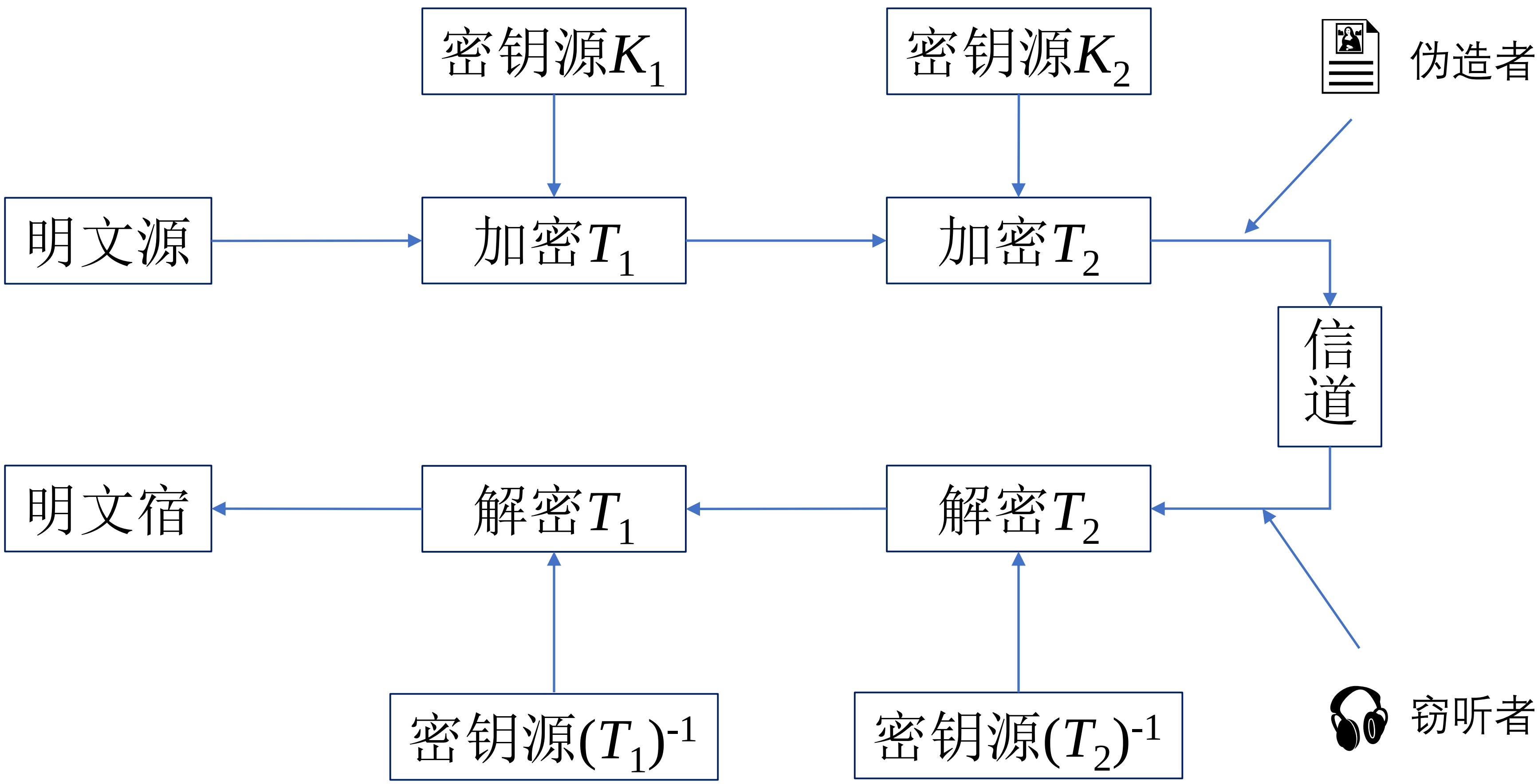
3. 非负性，$I(x; y) \geq 0$

概 要

信息论基础

- 信息的特征
- 信息量
- 保密通信系统
- 小结

保密通信结构

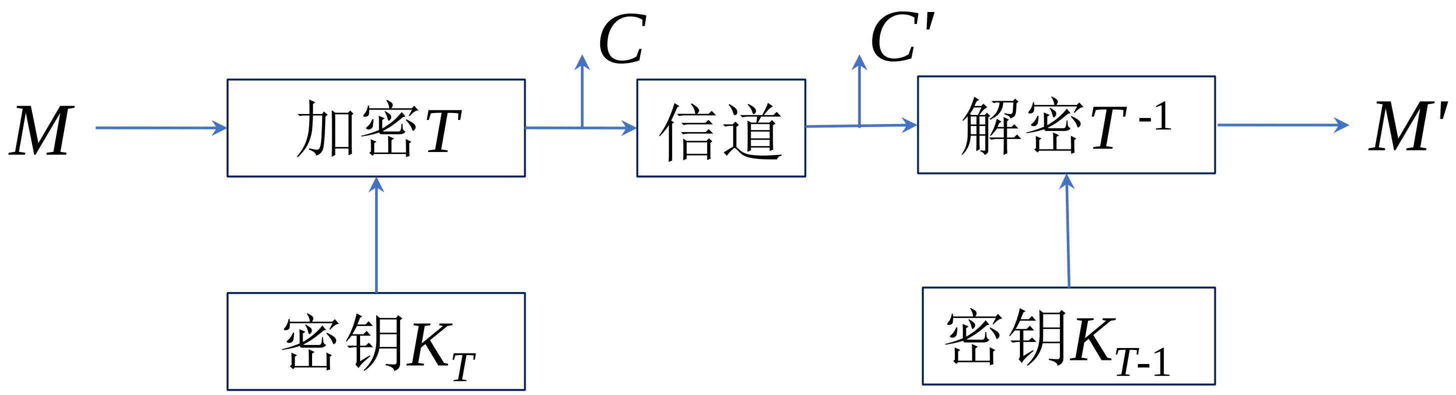

密钥源K_1
密钥源K_2
伪造者
明文源
加密T_1
加密T_2
信道
明文宿
解密T_1
解密T_2
密钥源$(T_1)^{-1}$
密钥源$(T_2)^{-1}$
窃听者

香农保密定理

保密系统需满足的两个要求：

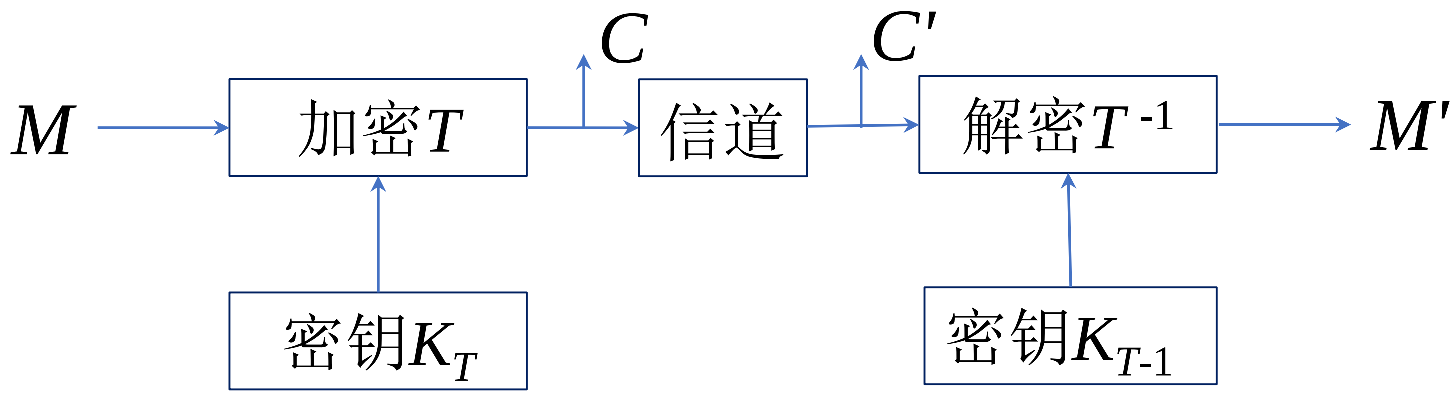
1）非目的用户不能通过密文直接获得明文；

2）目的用户通过密钥解密密文获得明文。



定理： 不考虑信道干扰等条件下，具有加、解密对（T，T_1）的保密通信系统，当明文和密文互相独立时，系统达到完全保密状态；当加密与解密的关系满足T_1=T^{-1}时，保密系统对目的用户能完全获得发送信息。

香农保密定理



一个完全保密系统的充要条件：

$I(M; C') = 0$ (非目的用户)

$H(M) = H(M')$ (目的用户)

$C = C'$

$$I(M; C') = H(M) - H(M \mid C') = H(M) - H(M \mid C) = 0$$

$H(M) = H(M \mid C) \rightarrow p(m) = p(m|c)$ 明文与密文互相独立！

$$H(M') = H(T_1(C')) = H(T_1(C)) = H(T_1(T(M))) = H(M)$$

$$T_1(T(M)) = T_1T(M) = M \rightarrow T_1 = T^{-1}$$

概 要

信息论基础

信息的特征

信息量

保密通信系统

小结

作业

1. 信源消息X={0,1,2}的概率模型如下：$p(x_i=0)=1/3$, $p(x_i=1)=1/6$, $p(x_i=2)=1/2$, 计算该信源各消息的自信息量。

2. 根据下图的信源概率信息，计算信源的熵。

x_i	000	001	010	011	100	101	110	111
$q(x_i)$	1/4	1/4	1/8	1/8	1/16	1/16	1/16	1/16

3. 当信源的n个消息均匀分布时，熵H(X)最大，据此，请计算H(X) 的最大值。

4. 从互信息的角度，阐述处理器与输入输出信号之间的关系。