哈爾濱工業大學(深圳)
HARBIN INSTITUTE OF TECHNOLOGY, SHENZHEN

通 信 安 全

L3—离散信源编码理论

- 教师：崔爱娇
- 编号：ELEC3019
- 学时：32学时

概 要

离散信源编码理论

- 信源编码基本概念
- 离散信源编码理论-定长编码
- 变长编码
- 小结

概 要

离散信源编码理论

- 信源编码基本概念
- 离散信源编码理论-定长编码
- 变长编码
- 小结

为什么要进行信源编码

信源的两个重要问题？？

信源输出的信息量计算

更有效的表示信源输出

信源信息变换 信源编码 → 提高通信效率

使信源减少冗余，更加有效、经济地传输，最常见的应用形式就是压缩。

为什么要进行信源编码？？

无失真传送 → 信源无失真编码

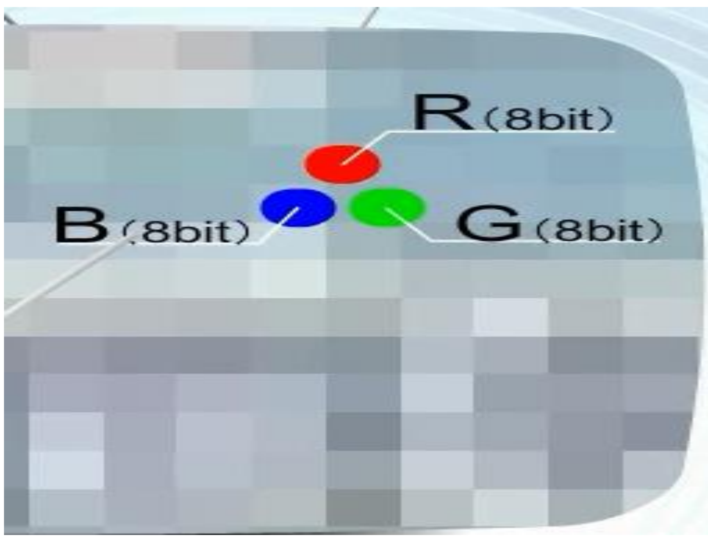
模拟信号转换数字信号 ← 数字技术应用广泛

高效传输 → 数据压缩

为什么要进行信源编码


625 lines
ZTE中兴

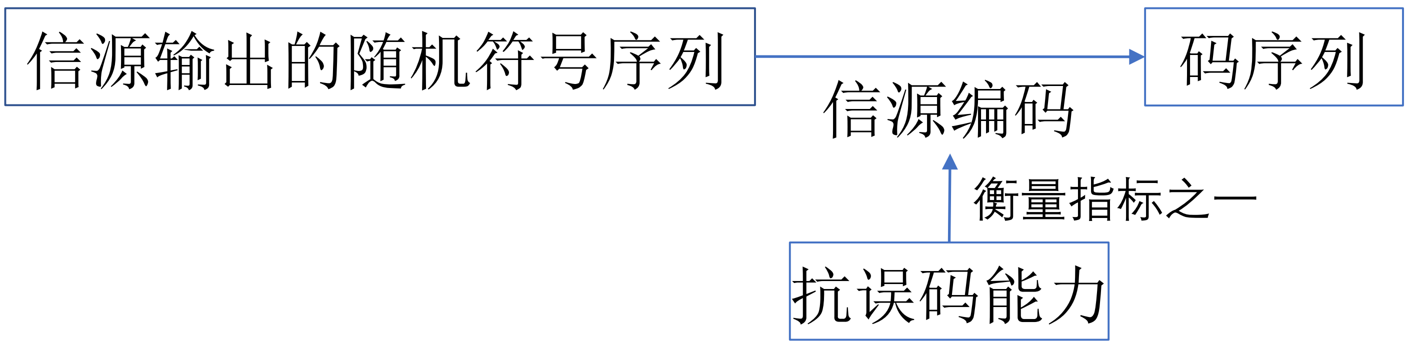

4a
3a
ZTE中兴
目标
愿景
使命

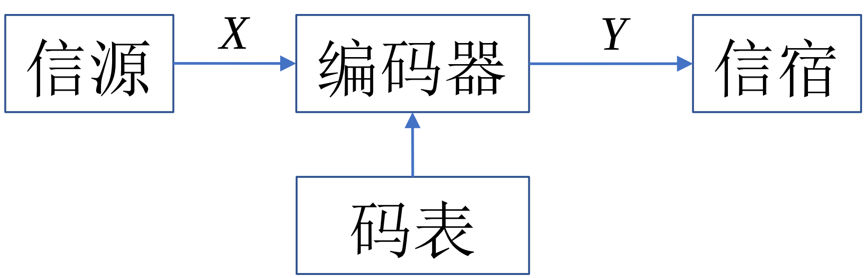

R(8bit)
B(8bit)
G(8bit)


每帧象素数：
N=625×(625×4/3)=520833 pixel/frame
每帧比特数：
M=8×3×N=12.5 Mbit
若每秒发出25帧，所需信息的传输速率为：
R=25f/s×12.5Mbit/f=312.5 Mbit/s
若信道的利用率为2bit/(s*Hz), 信道带宽为：
B=R/2=156.25 MHz
信源压缩倍数为：
n=B/(6×10⁶)=156.25/6=26
6MHz
Cable TV

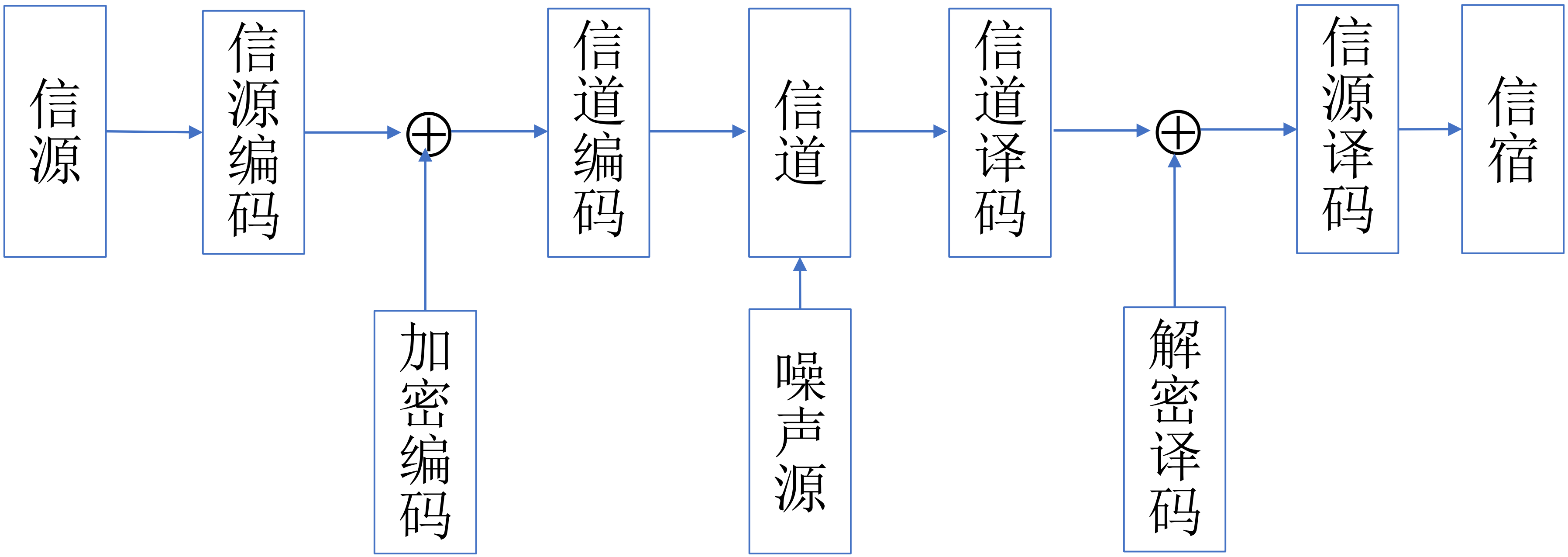
信源编码的概念

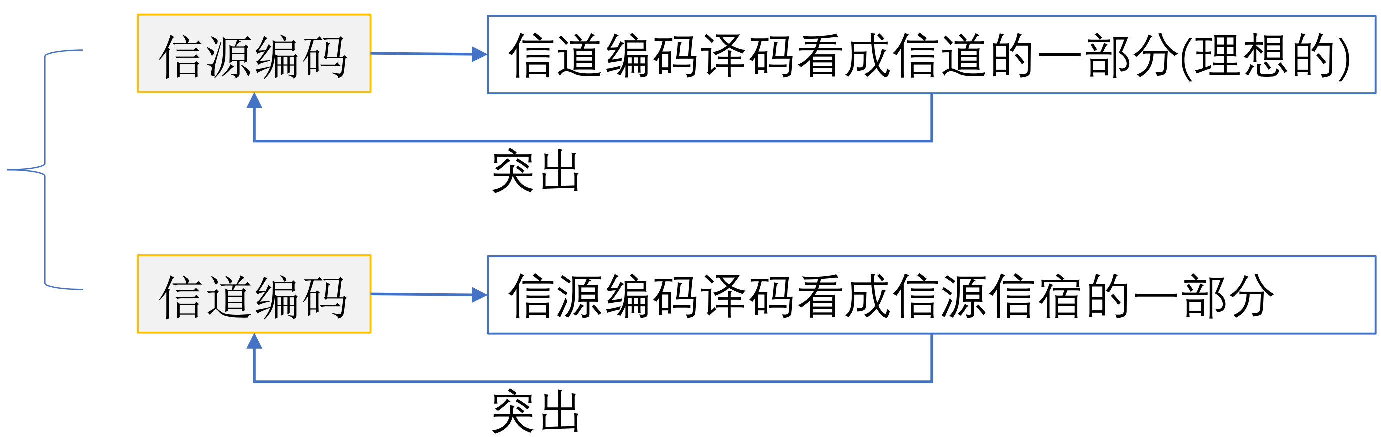
信源编码：指定能够满足信道特性（适合于信道传输）的符号序列–码序列，来代表信源输出的消息。





信源与信道编码


信源
信源编码
加密编码
信道编码
信道
噪声源
信道译码
解密译码
信源译码
信宿


信源编码
信道编码译码看成信道的一部分(理想的)
突出
信道编码
信源编码译码看成信源信宿的一部分
突出

信源编码

$\vec{x_i} = (x_{i1}, x_{i2}, \cdots, x_{iN}) \longrightarrow \vec{y_i}$ 分组码（块码）

信源输出的随机符号序列 $\xrightarrow{\text{信源编码}}$ 码序列（码字）

d元码

码符号集合：$D = \{0, 1, \ldots, d\text{-}1\}$

码字C_i的符号数量称为码字长度L_i

码字集合，码集：$C = \{C_1, C_2, \ldots, C_q\}$

信源编码码表

信源符号	符号概率	码表	
		码1	码2
a_1	$p(a_1)$	00	0
a_2	$p(a_2)$	01	10
a_3	$p(a_3)$	10	110
a_4	$p(a_4)$	11	111

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ p(a_1) & p(b_1) & p(c_1) & p(d_1) \end{bmatrix}$$

码1：4个码字，码字长度均为2；

码2：码字不等长，变长码。

信源编码分类

◆ 非奇异编码：$\overrightarrow{x_i} \neq \overrightarrow{x_j} \Rightarrow C_i \neq C_j$

◆ 奇异编码： $\overrightarrow{x_i} \neq \overrightarrow{x_j} \Rightarrow C_i = C_j$

◆ 唯一可译码：编码的码字只有一种可能信源序列与之对应。

$a_1 \to 0, a_2 \to 01, a_3 \to 001$ 给定码字001，译码结果？ a_1a_2 or a_3

◆ 即时码：可直接根据当前的码符号序列正确译出相应。

➢ 即时码的充要条件：任一码字非其他码字的前缀，唯一可译码
➢ 唯一可译码不一定是即时码。

$a_1 \to 0, a_2 \to 10, a_3 \to 110, a_4 \to 111$ 0110111100110？ $a_1a_3a_4a_2a_1a_3$

信源编码分类

- 码
  - 非分组码
  - 分组码
    - 奇异码
    - 非奇异码
      - 非唯一可译码
      - 唯一可译码
        - 非即时码
        - 即时码

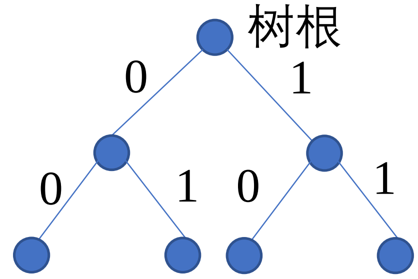
信源编码举例

信源消息	出现概率	码1	码2	码3	码4
x_1	1/2	0	0	1	1
x_2	1/4	11	10	10	01
x_3	1/8	00	00	100	001
x_4	1/8	11	01	1000	0001

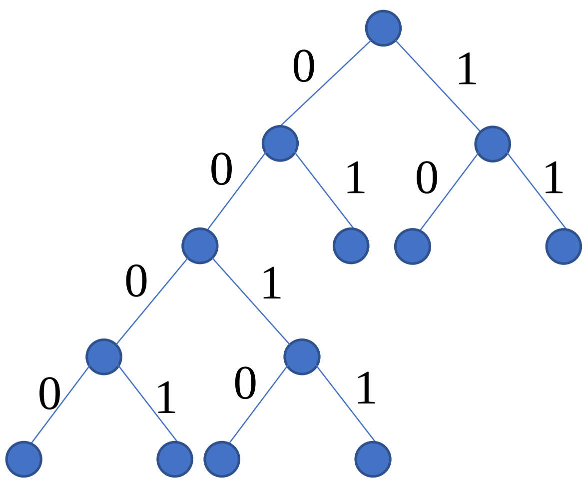
- 码1：非唯一可译码；奇异码　　$x_4x_3x_1$, $x_4x_1x_3$, $x_1x_2x_3$ or $x_1x_2x_1x_1$
- 码2：非奇异码，非唯一可译码。收到01000？
- 码3：唯一可译码，译码有延时；
- 码4：唯一可译码，无译码延时；码字中的1起到了逗点的作用，称逗点码。

信源编码分类

◆ m元树码（码树）：



满 树



非 满 树

- 树根：顶部起点
- 树枝：根部引出的线段
- 一级节点：从根部起，通过一条树枝到达的节点。一级节点最多m个
- n级节点：通过n条树枝到达的节点。最多有m^n个。
- 终端节点：下面不再有树枝的节点。
- 中间节点：除了根与终节点以外的节点。
- 联枝：串联的树枝。
- 满树：每个码字的串联枝数都相同时，定长码，此时的码树称满树。

树码

- 树码一定是即时码，反过来，任何即时码都可以用树码表示。
- 树码可以推导出唯一可译码的充要条件。

定理（克拉夫特不等式）

采用d进制符号集进行信源编码，产生的码字所对应的C_i所对应的长度为L_i，即时码存在的充要条件为

$$\sum_{i=1}^{q} d^{-L_i} \le 1$$

树码

- 设二进制码树中X $\in(a_1, a_2, a_3, a_4)$，各码字长度为K_1=1， K_2=2，K_3=2， K_4=3，判断是否存在满足这种码字长度的即时码。

$$\sum_{i=1}^{4} 2^{-L_i} = 2^{-1} + 2^{-2} + 2^{-2} + 2^{-3} = 9/8$$

香农信息论三大定理

- 第一定理：无失真信源编码定理

- 第二定理：信道编码定理
（包括离散和连续信道）

- 第三定理：限失真信源编码定理

概 要

离散信源编码理论

- 信源编码基本概念
- 离散信源编码理论-定长编码
- 变长编码
- 小结

无失真信源编码

若信源符号序列的长度$L\geq 1$，即$X=$（X_1，$X_2,\ldots,X_l,\ldots,X_L$），$X_l\in\{a_1, a_2,\ldots,a_n\}$,变换成由$K_L$个符号组成的码符号序列，$Y=$（$Y_1$，$Y_2,\ldots,Y_k,\ldots,Y_{KL}$）,$Y_k \in\{b_1, b_2,\ldots,b_m\}$

变换要求：能够无失真的从Y恢复X，即能正确的译码，同时希望传送Y所需的信息率最小。

无失真信源编码

$Y_k \in$有m种可能取值，平均每个码符号能输出最大信息量为$\log m$，K_L长码符号序列能输出的最大信息量为$K_L \log m$。用该码符号序列表示L长的信源序列，则每个信源符号所需的信息率平均为

$$\bar{K} = \frac{K_L}{L}\log_2 m$$

希望找到一种编码方式能够使上述信息率的值达到最小。

定长编码定理

定长编码定理： 由L个符号组成的，平均符号熵为$H(x)$的平稳无记忆符号序列X， 可用由K个码符号(每个有m种取值)组成的码序列作定长编码。对任意$\varepsilon > 0,\ \delta > 0$,有

（1）正定理：只要当L足够大时，译码差错必小于δ

$$\frac{K}{L}\log_2 m \geq H\left(X\right)+\varepsilon$$

（2）逆定理：当 $\frac{K}{L}\log_2 m \leq H\left(X\right)-2\varepsilon$

时译码差错必为有限值，且当L足够大时，译码几乎必定出错。

说明

（1）信息率（编码速率）：$R=(K/L)\log_2 m$

m——码序列中，每个符号的可能取值数。

K——定长编码，每个码的长度。

m^K——码字的总数。

$\log_2 m$ ——m个取值等概率，每个码符号的最大熵。

$K\log_2 m$ ——无记忆，码序列总信息量等于各符号信息量之和。

$(K/L)\log_2 m$ ——编码后平均每个符号的信息量，即平均符号熵；传递一个符号所需的信息量。

说 明

（2）正定理指出：当信息率R略大于单符号熵$H(X)$时可做到几乎无失真译码，条件是L足够大。

即编码后发送一个消息符号所需的平均信息量大于信源平均每消息符号的信息量时，可以使传输几乎无失真。可以证明，只要 $L \geq \frac{\sigma^2(x)}{\varepsilon^2\delta}$，译码差错率必小于$\delta$。

$\sigma^2(x) = E\left\{\left[I(x_i) - H(x)\right]^2\right\}$ ——信源序列自信息方差。

结　论

- （单符号）信源熵H（X）实为一个界限；
- $R > H(X)$时——无失真译码；
- $R < H(X)$时——有失真译码。

定长编码举例

设离散无记忆信源概率空间为：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ 0.4 & 0.18 & 0.1 & 0.1 & 0.07 & 0.06 & 0.05 & 0.04 \end{bmatrix}$$

信源熵为：$H(X) = -\sum_{i=1}^{8} p_i \log_2 p_i = 2.55\,\text{bit/符号}$

自信息方差为：$\sigma^2\left[I(x_i)\right] = E\left[I^2(x_i)\right] - H^2(X) = \sum_{i=1}^{8} p_i\left[-\log_2 p_i\right]^2 - H^2(X) = 1.3082$

对信源符号采用定长二元编码，要求编码效率$\eta = 90\%$，

无记忆信源有$H_L(X)=H(X)$，因此：$\eta = \dfrac{H(X)}{H(X)+\varepsilon} = 90\%$

可以得到 $\varepsilon = 0.2836$：

要求译码错误概率$\delta \le 10^{-6}$，则 $L \ge \dfrac{\sigma^2\left[I(x_i)\right]}{\varepsilon^2\delta} = 1.6256\times10^7$

分析

在对编码效率和译码错误概率的要求不是十分苛刻的情况下，就需要1600多万个信源符号一起进行编码

对存储和处理技术的要求太高，不现实。

对上述信源中8种可能的取值编定长码，每种取值3比特时，可实现译码无差错，单编码效率只有2.55/3=85%。

当L有限时，高传输效率的定长码，往往引入一定的失真和译码错误

变长编码

概 要

离散信源编码理论

- 信源编码基本概念
- 离散信源编码理论-定长编码
- 变长编码
- 小结

变长编码定理

对离散无记忆信源，消息长度为L，符号熵为$H(X)$，对信源进行m元变长编码，一定存在无失真信源编码方法

码字平均长度 $\overline{K}$ 满足 $1+\dfrac{H(X)}{\log_2 m} > \overline{K} \geq \dfrac{H(X)}{\log_2 m}$

码字平均信息率R满足$H(X) \leq R < H(X)+\varepsilon$

变长编码定理

- 对信源进行变长编码，一般要求的信源符号长度L比定定长编码小得多

- 当L足够大时，可使 $\dfrac{\log_2 m}{L} < \varepsilon \longrightarrow H(X) \le R < H(X) + \dfrac{\log_2 m}{L}$

编码效率 $\eta = \dfrac{H(X)}{R} > \dfrac{H(X)}{H(X) + \dfrac{\log_2 m}{L}}$

例 子

- 在定长编码例子中，$H(X)$=2.5525(bit/signal)，现在对其进行2元变长编码，即m=2，$\log_2 m$=1。

- 若要求编码效率大于90%，即

$$\eta = \frac{H(X)}{R} > \frac{H(X)}{H(X)+\frac{\log_2 m}{L}} = \frac{2.5525}{2.5525+\frac{1}{L}} = 0.9 \quad L=3.5261$$

只要4个符号一起编码，即可满足对编码效率的要求

概 要

离散信源编码理论

- 信源编码基本概念
- 离散信源编码理论-定长编码
- 变长编码
- 小结

作 业

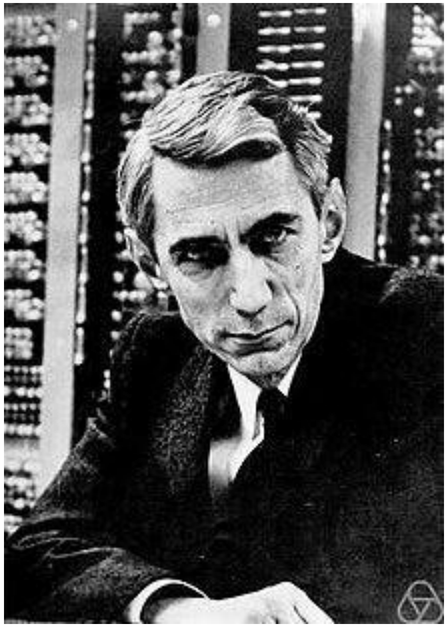
1. 设二进制码树中X $\in(a_1, a_2, a_3, a_4)$，各码字长度为K_1=1， K_2=2，K_3=3， K_4=3，判断是否存在满足这种码字长度的即时码。如果有，请举例给出一个可行的编码方案。

2. 给出6种编码例子

信源消息	出现概率	奇异码	非奇异码	唯一可译码	非唯一可译码	即时码	非即时码
x_1	¼						
x_2	¼						
x_3	1/8						
x_4	1/8						
x_5	1/8						
x_6	1/16						
x_7	1/16						

作 业

3. 克劳德.香农生平读后感



他不只是提出问题然后解决问题而已。他会依照自己的方法，发展出一个过程，来透过现象看本质

寻找答案当然很重要，但是很多人都忽略问题本身。**提问本身是有方法的，提出能解答、有答案的问题，也很重要。**

香农的推理方式是，我们必须排除所有和问题不相关的细节，才能看清问题的核心，之后就能自然地找到答案。

如果你没找到问题核心，却拿着一堆错误的细节来解决问题，你只会不断地找到更多错误的细节，最后深陷细节迷雾，更是不复得路了。

有些人，你告诉他一个点子，他能回你半个点子；还有些人，你告诉他一个点子，他能回你两个点子。”

香农实际上想说的是，要想有好点子，就要善于增加每次输入的精华。如果你没拿捏好点子的本质，想法就会平庸。你越能触及问题的本质，得出真知灼见的效率就越高。