第五章离散信道编码

第三节信道分组编码

陈兴同

中国矿业大学 数学学院

2021 年 8 月

内容提要

内容提要

内容提要

二进信道分组编码有关概念：

信道编码一般都是将来自信源编码器输出的消息序列进行分组，然后对所有不同的分组进行编码，这种编码自然要满足信道编码定理的要求，才能保证信息在信道中能够可靠地、有效地传输。为了方便只讨论工程中常用的二元信道编码问题。

设来自二元信源编码器输出的 k 长消息为 $w^{(k)}$，用 k 维向量 $(w_1, w_2, \cdots, w_k)$ 表示，每个消息字符取自 $\mathcal{W} = \{0, 1\}$，这种 k 长的消息总共有 2^k 个。对每个消息进行 n 长的信道编码，码字记为 $x^{(n)}$，用 n 维向量 $(x_1, x_2, \cdots, x_n)$ 表示，每个 x_i 取自 $\mathcal{X} = \{0, 1\}$，图2.1 5.4 是一个信道编码示意图。这种对固定长度的消息进行固定长度信道编码方法称为分组编码，生成的码称为 (n, k) **分组码**，其中 n 称为**码字长度**，k 长为**信息长度**，$r = n - k$ 称为**冗余位长度或校验位长度**，而 $R = k/n$ 称为**码率**。

分组编码图：

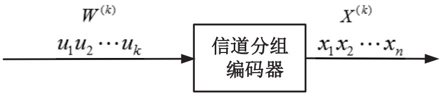

Figure: 图 5-4

检错与纠错：

由于信道有噪声干扰，故输入的 n 长码字 $x^{(n)}$，在传输过程中某些位可能会出错，导致 n 长输出序列 $y^{(n)}$ 与输入序列 $x^{(n)}$ 不同。如何控制这种差错？通常差错控制能力都是在编码与译码中实现的，包括检错与纠错。**检错**是指输出端的译码器能检测出接收到的 n 长输出序列是不是码字，有几位错，但不能确定错在哪些位上；**纠错**则是指译码器能确定 n 长输出序列哪些位发生错误，并且将错误位纠正，译成正确的码字。

例题 5.3.1：

设来自二进信源的消息有两个即 $W=0,1$，进行二元 3 重编码即 n=3,k=1,r=2。
编码规则

$$f: 0\rightarrow 000 \quad , \quad 1\rightarrow 111,$$

它的汉明距离为 3bits，译码规则

$$g: \left.\begin{matrix}000\\001\\010\\100\end{matrix}\right\}\rightarrow 000, \quad \left.\begin{matrix}011\\101\\110\\111\end{matrix}\right\}\rightarrow 111$$

解：

（1） 这种编码方案具备检错能力。因为只要输出不错成另一个码字，编码方案就可以在译码时检测出是错一位，还是错 2 位。比如若输出序列为 001，则译码器根据编码规则会知道它不是码字，肯定是传错了，但是究竟是码字 000 错 1 位还是由码字 111 错 2 位，译码器搞不清楚。它只能检测到传输出错，最多可以检出 2 位错。

（2） 具有一位纠错功能。因为根据译码规则，每个码字的一位错都可以被纠正。比如如果输入码字 000 被传错一位即错成 001 或 100 或 010 按译码规则也照样可以译对；但如果码字 000 被错传二位即错成 011 或 101 或 110，则会译成另一个码字，这就译错了。

（3） 这种编码方案具备 2 位检错与 1 位纠错能力。

例题 5.3.2：

设来自二进信源的消息有两个即 $W=0,1$，考虑进行二元 4 重编码即 n=4,k=1,r=3。
编码规则

$$f:0\rightarrow 0000 \quad , \quad 1\rightarrow 1111,$$

它的汉明距离为 4bits，译码规则

$$g:\quad \left.\begin{matrix}0000\\0001\\0010\\0100\\1000\end{matrix}\right\}\rightarrow 0000, \quad \left.\begin{matrix}0111\\1110\\1011\\1101\\1111\end{matrix}\right\}\rightarrow 1111, \quad \left.\begin{matrix}0011\\0101\\0110\\1010\\1001\\1100\end{matrix}\right\}\rightarrow ?,$$

解：

（1） 这种编码方案具备检错能力，最多可以要检出 3 位错。比如若输出序列为 0001，则译码器根据编码规则会知道它不是码字，肯定是传错了，但是究竟是码字 0000 错 1 位还是由码字 1111 错 3 位，译码器搞不清楚；再如：若输出 0011，则译码器根据编码规则会知道它不是码字，肯定是传错了，但是究竟是码字 0000 错 2 位还是由码字 1111 错 2 位，译码器搞不清楚；

（2） 具有一位纠错功能。比如输入码字 0000 被传错一位即错成 0001 或 0010 或 0100 或 1000 也照样可以译对；但如果码字被错传二位即错成 0011 或 1001 或 1100 或 1010，则无法译码。同理如果输入码字为 1111 也会发生这些情况。

（3） 这种编码方案具备 3 位检错与 1 位纠错能力。

定义 5.3.1：码距

分组码的最小汉明距离与检错、纠错能力紧密相关。

(n,k) *分组码* $\mathcal{C}$ *的最小汉明距离（简称**码距**）定义为*

$$d = \min_{c_i, c_j \in \mathcal{C}, i \neq j} d(c_i, c_j),$$

即它的任何两个不同码字之间汉明距离的最小值。

比如例题 5.1.5 中的编码的最小汉明距离为 3bits。

例题 5.3.3：

设码字集
$\mathcal{C} = \{c_1 = 11100, c_2 = 01001, c_3 = 10010, c_4 = 00111\}$。(1) 求编码 $\mathcal{C}$ 的最小汉明距离 d。
(2) 根据最小汉明距离译码规则对输出
$b_1 = 10000, b_2 = 01100, b_3 = 00100$ 分别进行译码。
(3) 计算码 $\mathcal{C}$ 的码率。

解：

（1）根据 $d(c_i,c_j)=c_i\oplus c_j$ 中非 0 元个数可求得两个码之间的汉明距离为

$$d(c_1,c_2)=3, d(c_1,c_3)=3, d(c_1,c_4)=4,$$

$$d(c_2,c_3)=4, d(c_2,c_4)=3, d(c_3,c_4)=3,$$

所以这个码的最小汉明距离为 3bits。

（2）检查 b_1 与每个码字的汉明距离：$d(b_1,c_1)=2, d(b_1,c_2)=3, d(b_1,c_3)=1, d(b_1,c_4)=4$，故与输出 b_1 汉明距离最小的码字为 c_3，从而将输出 b_1 译成输入码字 c_3；检查 b_2 与每个码字的汉明距离：$d(b_2,c_1)=1, d(b_2,c_2)=2, d(b_2,c_3)=4, d(b_1,c_4)=3$，故与输出 b_2 汉明距离最小的码字为 c_1，从而将输出 b_2 译成输入码字 c_1；检查 b_3 与每个码字的汉明距离：$d(b_3,c_1)=2, d(b_3,c_2)=3, d(b_3,c_3)=3, d(b_3,c_4)=2$，故与输出 b_3 汉明距离最小的码字为 c_1,c_4，从而将输出 b_3 译成输入码字 c_1 或 c_4。

（3）$R=\frac{1}{[illegible]}\log 4=0.4bits$

定理 5.3.1:

最小汉明距离是衡量编码抗干扰能力的重要指标，与检错纠错有非常大的关系。

设一个二进 (n, k) 分组码 $\mathcal{C}$ 的最小汉明距离为 d，采用最小汉明距离译码方法。

（1） *这个码 $\mathcal{C}$ 能检测出 s 位错误（指传错的位数不超过 s 位时译码器都能发现）当且仅当 $d \geq s + 1$。*

（2） *这个码能纠正 t 位错（指传错的位数不超过 t 位时译码器都能发现并纠正成原发码字）当且仅当 $d \geq 2t + 1$。*

（3） *这个码能纠正 t 位错又能检测 s 位错误当且仅当 $d \geq s + t + 1$。*

证明：

考虑带有汉明距离的度量空间 $\mathcal{X}^n$，它包含了信道所有可能的 n 长输出。
（1）如果任意一个码字 c_i 的输出 $y^{(n)}$ 有 s 位错误，则这个输出就在以 c_i 为中心半径为 s 的 n 维球内部。如果要想检出 s 位错误，则任一个码字的半径为 s 的球内不能包含另一个码字（图图5-5），从而必有 $s \leq d-1$ 即 $d \geq s+1$。反之如果 $d \geq s+1$ 则任何码字的半径为 s 的 n 维球都不包含另一个码字，从而可以检出 s 位错误。

图 5-5:

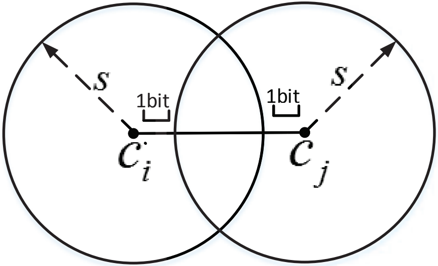

Figure: 图 5-5

图 5-6:

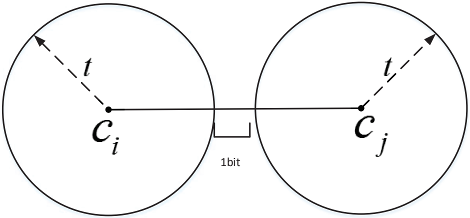

Figure: 图 5-6

续证明：

（2）如果能纠正 t 位输出错误，则任意两个不同码字 c_i, c_j 传错 t 位后的输出 $x^{(n)}, y^{(n)}$ 应当在它们各自互不相交的半径为 t 的 n 维球内部（图 5-6 和 5-7），很明显 $d \geq 2t+1$。反之，当 $d \geq 2t+1$ 时，两个不同码字为中心半径为 t 的 n 维球必定不相交；当输出 $y^{(n)}$ 有不超过 t 位错误时，它必定位于某个码字 c_i 的球内，该输出与这个码字距离最小；事实上：

$$d(y^{(n)}, c_j) > d - d(y^{(n)}, c_i) \geq d - t \geq t + 1 > t.$$

根据最小汉明距离译码规则，将这个输出译成它所在球的码字，这就达到了纠错的目的。

图 5-7:

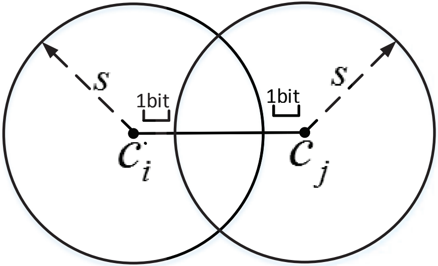

Figure: 图 5-7

续证明:

（3）可以区分出检错球：半径为 s 的球；纠错球：半径为 t 的球。纠错球的半径要小。因此要检错与纠错同时具有，必须使码字 c_i 的检错球与码字 c_j 的纠错球不相交；同时也必须使码字 c_j 的检错球与码字 c_i 的纠错球不相交（图 5-8）！故
$d \geq s + t + 1$。

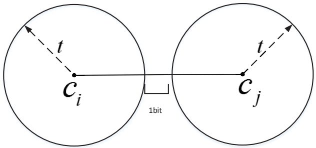

Figure: 图 5-8

练习：

从网络上搜索学习常用检错码，比如：奇偶校验码、水平校验码、群计数码等，并写成报告。