纠错编码入门

〔美〕S·林 著

陈 太 一 译

人 民 邮 电 出 版 社

纠 错 编 码 入 门

(美)S. 林 著
陈 太 一 译

人民邮电出版社

Shu Lin

An Introduction to

Error Correcting Codes

1970

内 容 提 要

本书专讲数据传输中纠正错误所需的纠错编码技术，共十 三章，介绍了各种近代通信用的信道编码和与之对应的译码方法及 其纠错 能力；不仅讨论了纠个别错误的码，还讨论了纠成群的突发错 误以及同时纠突发与随机错误的码。书中比较着重物理意义的阐述，有 一定的实用参考意义和一定的系统性，可供电信工程技术人员参考阅 读，为深入研究编码问题打下基础。

纠 错 编 码 入 门

（美）S·林 著

陈 太 一 译

人民邮电出版社出版

北京东长安街27号

河北省邮电印刷厂印刷

新 华 书 店 发 行

限 国 内 发 行

开本：787×1092 1/32　1976年4月 第 一 版

印张：10 4/32 页数 162　1976年4月 河北第一次印刷

字数： 280千字　印数：1— 12,000 册

统一书号：15045·总2044－无612

定价：1.00 元

译　者　序

数字通信之所以受到人们的重视和得到迅速的发展，主要是由于：电子计算机的广泛应用，须有数字传输与之相配合；快速电报的发展；数字电话信号经密码加密后，具有高度的保密性能；数字通信最终可能将电话、电报、传真和数据传输等通信网纳入一个统一的数字通信系统。

数字通信的发展，对通信技术提出了新的要求，其中很重要的一个问题是：某些数据，要求传输过程中产生的差错概率很低。解决的途径是：采用新的可靠的传输系统（如卫星通信系统），更新原有的传输系统以减少差错，及采用纠错编码技术。后者就是本书所讨论的问题。

纠错编码的研究已有二十五年的历史，特别是近年来，它得到迅速的进展，其主要原因是：数字通信的发展促进了纠错编码的研究；近代代数的理论为代数编码提供了理论基础；大规模集成电路的进展，为它提供了物质基础。

纠错编码技术在通信方面的应用已日益广泛，凡研究通信系统总体及从事设备研制的工程技术人员，均须要了解与掌握这一新的技术。为此，我们根据“**洋为中用**”的方针，将“纠错编码入门”一书译出，以供参考。

本书比较着重物理意义的阐述，较为联系实际，有一定的系统性，可供工程技术人员参阅。

本书译稿经过西北电讯工程学院一系编码讨论班的同志们

认真审校，改正了不少原书的错误，并加了许多必要的注释，谨在此表示谢意。

对于译文中的缺点和不妥之处，尚请读者批评指正。

译　者

目　　录

第一章　通信与编码

1.1　引言

近年来，由于自动数据处理机的应用日益增多和对远程通信的需要不断增长，更加需要有效而又可靠的数字数据传输系统。任何高速数据传输系统的严重问题之一是发生错误。如何控制这些错误是一个基本的重要问题。

目前的多数数字计算机和数字数据通信系统，是按二进制形式处理信息，更具体地讲，就是把信息编码成二进制数字“0”或“1”。图1.1是典型数字数据通信系统（或信息存储系统）的方框图。

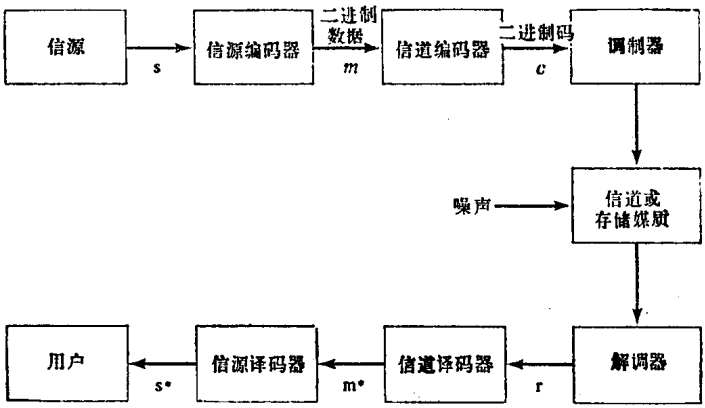

图 1.1　典型数据通信系统或存储系统的方框图

此系统的第一单元是信息源，它可以是人或机器（例如数

字计算机）。信源的输出可以是一个连续波形，或者是离散的符号（或字母）序列。信道是传输含有有用信息的信号的媒质。电话线、高频无线电线路、宇宙通信线路以及包括存储系统用的写头和读头的磁带装置，都是信道的典型例子。信道常受各类自然的或人为的噪声干扰。例如，在电话线上，干扰可来自热噪声、闪电、脉冲噪声或其它线路来的串话。在磁带上，磁带的缺陷可看作一种干扰。信源编码器将信源的输出s变换成二进制符号序列m，即0和1组成的序列，这称作信息序列。这个变换应满足下列两项要求：（1）代表信源输出s所需的每单位时间内的二进制数字的个数要少；（2）可以从信息序列m重新构成信源输出。信道编码器根据某些规则，将输入的信息序列m变换成称为m的码字的较长的二进制序列c，此变换叫做信道编码。二进制数字不宜在实际信道上传输。调制器的功能是把信道编码器的每位输出数字编码成持续时间为T秒的两种实际波形中的一种。例如，“1”可编码成持续时间为T的正脉冲，“0”编码成持续时间为T的负脉冲（或空号）。调制器的输出信号进入信道并被噪声干扰。解调器对每个持续时间为T的接收信号进行判决，以确定发送的是1还是0。于是，解调器的输出是二进制数字序列r。此序列r称为接收序列。由于信道噪声的干扰，接收序列r可能与码字c不一致。它们不同的位称为传输错误（简称错误）。例如，若发送的是c=(110011000111011)，收到的是r=(110001000101011)，那么在第5及11位发生了错误。信道编码器应设计得使它的输出码字具有抵抗传输错误的能力。信道译码器根据接收序列r信道编码规则及信道特性，完成以下两项任务：（1）它试图纠正r中的传输错误，并产生真正发送的码字c的估值c^*。（2）它变换c^*为信息序列m^*，m^*是发送信息序列m的估值。

根据信源编码规则，信源译码器将m^*变换成真正信源输出s的估值s^*，并送至用户。若信道平静（无噪），则c^*、m^*、s^*分别为c、m、s的重现。若信道噪声很大，s^*可能与真正信源输出s差别十分大。

一个主要的通信工程问题是设计这样的信道编码器—译码器对，使得：（1）二进制数据可尽快地在噪声信道上传输；（2）在信道译码器的输出端能够可靠地重现信息序列m。信道编码器—译码器对的设计主要根据信道特性。

1.2 分组码和最大似然译码

若我们仅讨论信道编码和译码，图1.1的系统可简化成图1.2的形式。图中信源由原来信源和信源编码器组成；信道由调制器、原来信道和解调器组成。在这种情况，信道接收二进制数据，并把受损的二进制数据传送至译码器。

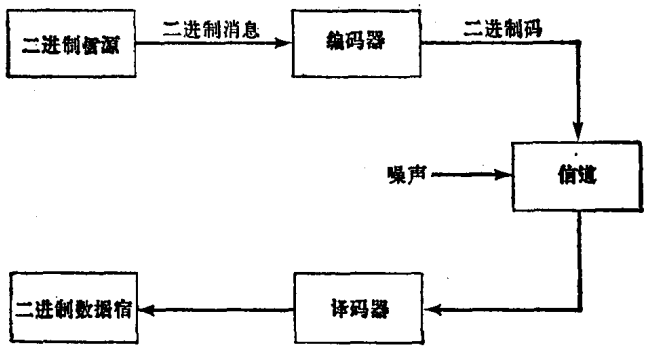

图 1.2 信道模型

下面讨论编码方案的一个例子。信源的输出信息序列首先分成消息组；每个消息组m包含k位信息数字，总共有2^k种不

同的消息。编码器将每个输入消息组m变换成一个有 n 位数字的较长的二进制序列c，称之为m的*码字*。故对应于 2^k 个可能的输入消息，在编码器的输出端有2^k个可能的码字。此 2^k个码字的集合称为*分组码*。编码器使每个消息组附加上的 $n-k$ 位数字称为*多余数字*。这些多余数字不带新的信息，它们的功能是使码具有纠正传输中造成的错误的能力。如何构成这些多余数字以使码具有良好的纠错能力是设计编码器主要考虑的问题。比率$R=k/n$称为*编码速率*，即每发送一位信道数字，进入编码器中的信息数字的数目。

设$\mathbf{c}=(c_0,c_1,c_2,\cdots\cdots,c_{n-1})$为发送码字，$\mathbf{r}=(r_0,r_1,r_2,\cdots\cdots,r_{n-1})$为信道输出端的接收序列。由于信道噪声的干扰，接收序列可能与发送码字不同。根据r及信道特性，在译码器中对发送码字作出判决。此判决过程称为*译码*。若译码器未能识别出（重现）真正的发送码字，就会作出错误的译码。错误译码的概率取决于所用的码、信道特性及译码器所用的译码方法。若所有码字都是以等可能发送的，则最佳的译码方案如下：在收到序列r后，译码器对所有2^k个码字计算它们的条件概率$P(\mathbf{r}|\mathbf{c}_i)$。若条件概率$P(\mathbf{r}|\mathbf{c}_i)$最大，则认为码字 $\mathbf{c}_i$ 就是发送码字。这种译码方案称为*最大似然译码*。

关于在噪声信道中传输信息的重要理论是香农的“编码定理”[21,22]。此定理说：每个信道具有确定的容量C，对于任何小于C的速率R，存在速率为R的码，若用最大似然译码时，其错误译码概率$P(\varepsilon)$为任意小。更具体地讲，对于任何给定速率$R<C$及长度n，都存在一种分组码，可使错误译码概率为

$$P(\varepsilon) \leqslant e^{-nE(R)}, \qquad (1.1)$$

式中对于$R<C$，$E(R)$是R的正函数，并由信道转移概率

确定。所以增加码长n并保持速率R小于信道容量C，就可使错译码概率小到我们所期望的那样小。香农定理仅说明存在给定错误译码概率任意小的码，但并没有指出如何构成这些码。所以我们面临的问题是如何构成香农定理所预示的这些大 n 的好码。

如（1.1）式所示，要一个码有效的话，码长 n 必须长。若我们用存储 $2^{nR}(=2^k)$ 个码字的编码器和实行最大似然译码的译码器来实现这种码，则编码器和译码器都将非常复杂。译码器必须计算 2^{nR} 个条件概率 $P(\mathbf{r}|\mathbf{c}_l)$。所以，人们面临下列三个问题：（1）寻求长的好码；（2）寻求实用的编码方法；（3）寻求实用的译码方法。

一个码的错误译码概率取决于传输错误的统计特性。在理论研究中，几乎一致地假定：所收到的符号与所发送的符号相同的概率为 $q_0>\frac{1}{2}$，收到的符号与发送的符号相反的概率 $p_0=1-q_0$，

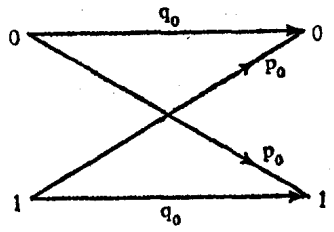

图 1.3　二进制对称信道

如图1.3所示。还假设信道对这些发送符号发生的影响是独立的。满足这些条件的信道称为二进制对称信道(BSC)。概率 p_0 称为转移概率。对于二进制对称信道，条件概率$P(\mathbf{r}|\mathbf{c}_l)$可表达如下：

$$P(\mathbf{r}|\mathbf{c}_l) = \prod_{i=0}^{n-1} P(r_i|\mathbf{c}_{li}), \tag{1.2}$$

式中，当$r_i=\mathbf{c}_{li}$，$P(r_i|\mathbf{c}_{li})=q_0$，当$r_i\neq\mathbf{c}_{li}$，$P(\mathbf{r}_i|\mathbf{c}_{li})=p_0$。令 d_l 为码字 $\mathbf{c}_l$ 与接收序列 r 不同的位的数目。则(1.2)式变成

$$P(\mathbf{r}|\mathbf{c}_l) = q_0^{n-d_l} q_0^{d_l} \tag{1.3}$$

由于$q_0>p_0$，$P(\mathbf{r}|\mathbf{c}_l)$随 d_l 增大而单调地减小。所以，按$P(\mathbf{r}|\mathbf{c}_l)$最大来求码字$\mathbf{c}_l$，等效于求与接收序列$\mathbf{r}$不同的位最小的码字$\mathbf{c}_l$。

二进制对称信道中引入的传输错误，称为随机错误。能抵抗这种错误的码称为纠随机错误码。不幸的是，很少实际信道与二进制对称信道相似。通常在连续发送的符号中，错误有着密切的依赖性。噪声干扰一例如一个闪电的冲击或人为的电干扰一常常影响几个相邻的符号。磁记录装置中的缺陷亦常影响不止一个符号。于是错误突发出现。对这种现象的注意，导致研究各种纠突发错误码。遗憾的是，它们也没有解决问题，因为有时突发错误是一阵阵突然到来的。一个信道可能长时间内非常好，然后一会儿很坏，这“一会儿”也许很短，但也可能长得足以使错误纠正，即便不是不可能的话，也是非常困难的。对于这种信道，单采用错误纠正只能取得非常有限的改进，而需采用纠错与检错和请求重发相结合（该请求信号经反向信道发回到发送端）。在本书中，我们仅研究采用单向信道的数字通信系统中使用的各种纠错编码技术。

1.3 例子

下面用一个纠错码的简单例子来说明我们陈述的方式，并介绍一些术语。

假设我们有由一12位二进制信息数字组成的消息要发送。这消息的12位数字可排列成一个3×4的矩形阵列：

$$\begin{matrix} X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \\ X_{31} & X_{32} & X_{33} & X_{34} \end{matrix}$$

现给每行和每列再加一个符号，而选定使每行和每列中1的个数为偶数。最后，在右下角加上一个符号，并选择这一符号使最后一行中1的个数也是偶数。这样最后一列中1的数目也会是偶数：

$$\begin{matrix} X_{11} & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & X_{22} & X_{23} & X_{24} & X_{25} \\ X_{31} & X_{32} & X_{33} & X_{34} & X_{35} \\ X_{41} & X_{42} & X_{43} & X_{44} & X_{45} \end{matrix}$$

对于所示信息，应如图选择校验符号。

1 1 1 0	1
1 1 0 0	0
0 1 0 1	0
0 1 1 1	1

若将整个阵列逐行(或逐列)发送，则如沿该方向若有一个符号被改变——即有一个错误——该错误能被纠正。纠正是通过校验每行和每列是否符合偶数校验(1的个数为偶数)来进行的。若有单个错误，则必有一列校验和一行校验不符合，错误必定在其交点上。于是便可纠正错误。这20个发送符号构成一个码字。这码共有$n=20$个符号，其中$k=12$是信息符号，这叫做(20，12)码。此码有$n-k=20-12=8$个校验符号。这些多余的数字，加到消息上去使码字具有纠错的能力。

本书叙述的所有码，都建立在与这(20，12)码所用的相类似的数学概念上。这些概念一般地在数学上要深些，并导出更好的码。例如，纠正单个错误的汉明码[15]所需的校验符号的数目最少，要比上述码要求的数目少得多。

两个n位符号的码字之间的距离是它们之间不同的位的个数。码的最小距离是两码字之间的最小距离。上面例举的码其最小距离为4——改变矩形四个角上的符号，就把一个码字变成另一个码字，因为这样仍保持了所有行和列的一致校验关系。为要纠正t个或少于t个错误的所有组合，最小距离至少为$2t+1$。这保证若发生t个错误，接收序列仍更接近于真正发送的

码字，而不会接近其它可能的码字。在此情况，真正发送的码字 c_i 将有最大条件概率 $P(r|c_i)$。所以为了纠正单个错误，最小距离至少应为 3 。

在所例举的码中，最小距离为 4 保证能纠正所有的单个错误。至于所有两个错误，将有两行校验和（或）两列校验不符合偶数校验，尽管可以检查出有错误，但不能确定位置。在各种不同位置上的三个错误的组合，可能表现为共有二、四或六行及列的校验不符。一般可检查出三个错误，并与单个错误不同，但若三个错误恰巧位于一个矩形的三个角上，将会有一列和一行校验不符，译码器可能错误地判断为这是单个错误。同样，四个错误的大多数组合可被检查出来，但位于矩形四角上的四个错误，会产生一个没有一致校验不符的接收序列。

一般说来，既能纠正 t 个错误的各种组合又能同时检测 $l \geqslant t$ 个错误的各种组合，要求最小距离为 $t+l+1$ 。此外，本书中介绍的任何码，可以把纠错由其最大可能数 t 限制到某较小数，而保证检测出较大一类错误。

主 要 参 考 资 料

1. *Abramson, N., Information Theory and Coding*, 1963.
2. *Ash, R. B., Information Theory*, 1965.
3. *Berlekamp, E. R., Algebraic Coding Theory*, 1968.
4. *Elias, P., "Error Free Coding" IRE Trans. on Information Theory, IT-4, pp. 29-37, September* 1954 .
5. *Elias, P., "Coding for Noisy Channels," IRE Convention Record, Part 4, pp.* 37-46, 1955.
6. *Fano, R. M. Transmission of Information*, 1961.
7. *Feinstein, A., Foundations of Information Theory.* 1958.
8. *Forney, G. D., Concatenated Codes*, 1967.
9. *Gallager, R. G., "A Simple Derivation of The Coding Theorem and Some Applications," IEEE Trans-*