信息论与信源编码

刘立柱　平西建　编

（下）

中国人民解放军信息工程学院三系

一九八八年六月

DATA, TEXT & VOICE ENCRYPTION WORLDWIDE MARKETS

International Resource Development Inc.

Report #727

February 1987

由无失真信源编码定理和信道编码定理可知，只要信息传输率R大于信源的熵并且小于信道的容量，总能找到一种编码，使通过信道传输再现的信号中不产生失真。

但是实际上信源的输出常常是连续的消息（即取值是无限、不可数的），信源的熵H为无限大，那么，若要求无失真地传送连续信源的消息，信息传输率R也应为无限大。而在信道中，由于带宽总是有限的，其容量总要受到限制。因此在实际的通信系统中，信息传输率总是大大超过信道容量，不可能实现无失真地传送信源信息。

此外，随着科学技术的发展，由于数字系统在抗干扰、纠错、保密、适应性等方面有明显的优越性，并能利用现代计算技术进行各种信息处理，数字系统得到了愈来愈广泛的应用。但是模拟信号经A／D转换后，数据量大、占用频带宽，给信息的传输或存贮带来了很大困难。为了提高传输和存贮的效率，就必须进行数据压缩，这样也会损失一定的信息，带来失真。

而在实际生活中，人们并不要求无失真地恢复消息，通常总是要求在保证一定质量（一定保真度）的条件下近似地再现原来的消息，即允许一定的失真存在。例如在传递语音信号时，由于人耳接收的带宽和分辨率是有限的，即便语音信号中有一些失真，人耳还是可以分辨或感觉出所传输的语音信号。又如传送图象信号时，由于人们的视觉特性和心理特性的作用，对一定限度内的失真也不敏感。

信息率失真理论告诉我们，若传输的信息允许一定的失真，则信

息率可以进一步减小。于是产生了以容许的失真量换取更大的数码压缩率的限失真信源编码。下面两章所讨论的预测编码和正交变换编码就是限失真编 码中最常用的两类方法。

6.1 概述

语音、图象等模拟信号转换成数字信号后，由于模拟信号的变化相对于采样速率（满足Nyquist率要求）很慢，数据序列中数值的变化多是缓慢的。这就使得信源输出的数据序列中数据之间有较强的相关性。这种信源可近似地看作有记忆的m阶马尔柯夫（Markov）信源。由第一章的讨论我们知道，在这种信源中有较大的信息冗余度。如果设法去除数据之间的相关性，便可减小冗余，实现数据压缩。

预测编码法就是直接利用数据间的相关性进行去相关、去冗余，实现数据压缩的编码方法。在预测编码法中，编码器传输或存贮的并不是信源输出的数据本身，而是此数据的预测值（或称为估计值）与其实际值之间的差值，即所谓预测误差。

此处我们首先分析一下预测编码法的基本原理。

设信源输出的数据X_i之间具有某种程度的相关性，则可根据X_i以前的某些数据对X_i作出预测或估计$\hat{X}_i$，定义两者之间的误差为：

$$e_i = X_i - \hat{X}_i \qquad (6.1)$$

e_i称为预测误差或差值信号。

由于预测值$\hat{X}_i$是利用数据间的相关性做出的，在由它和实际数据

的差值所构成的差值信号序列{ e_i }中，数据间的相关性较小。于是差值信号序列的信息冗余度降低。同时从统计观点讲，利用相关性求出的 $\hat{X}_i$ 多数与 X_i 比较接近，那么差值信号 e_i 的数值分布区间虽然较大，但大部分 e_i 的取值都落在零附近的一个较小的区间内。理论研究表明，差值信号 e_i 的概率分布形式如图 5.1，它可以用拉普拉斯（Laplace）分布作为其数学模型。即

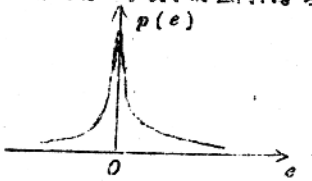

图 5.1 差值信号的概率分布

$$p(e)=\frac{1}{\sqrt{2}\sigma_e}\,exp\left(-\frac{\sqrt{2}}{\sigma_e}|e|\right) \qquad (5.2)$$

其中 e 为 e_i 的取值，σ_e 为差值信号的均方根值。

如果信源的输出是已经数字化的数字序列，对于差值信号序列{ e_i }可以不附加新的量化而直接编码，这样可以构成信息保持型的预测编码。若根据{ e_i }的分布特点（如图 5.1）对 e_i 重新进行量化，则可实现更大的数码压缩率。但是由于量化产生的失真是不可逆的，则使接收端再现的数据产生附加畸变，即产生了失真。若将此时的平均失真控制在某一限度之内，便构成了限失真的预测编码法。这种限失真的编码方法不仅利用了数据的统计特性，而且根据不同问题和不同的要求利用了人的生理特性和心理特性，能够实现更大的数据压缩，进一步提高了系统的有效性。因此，限失真的预测编码方法成为数据压缩处理的主要方法之一。本章的重点是讨论限失真的编码方法。

最典型的限失真预测编码方法是差值脉冲编码调制，通常记作DPCM(*Differential Pulse Code Modulation*)。DPCM系统的原理框图如图5.2所示。图中输入信号为X_i，由前面的某些数据作出的X_i的预测值为$\hat{X}_i$，e_i为差值信号。量化器产生的量化误差为：

$$q_i = e_i' - e_i \qquad (5.3)$$

在接收端，解码输出的再现数据为X_i'；于是，经过系统的传输，输入数据与输出再现数据之间产生的误差为：

$$X_i - X_i' = X_i - (X_i + \hat{e}_i') = X_i - X_i - \hat{e}_i'$$

$$= e_i - \hat{e}_i' = q_i \qquad (5.4)$$

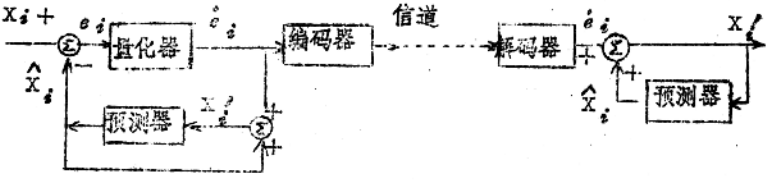

图5.2 DPCM系统框图

可见，再现数据中产生的失真正是发送端量化器中所造成的量化误差，它于解码器无关。

在图5.2的系统框图中，如果舍弃量化器，直接对预测误差

e_i进行编码，再现数据中便不会产生失真，成为信息保持型DPCM。在输入X_i是数字信号时，预测误差的值也是数字的。若进行等长编码，所需比特数比输入信号的编码多一个比特。但是，采用变字长的熵编码法，则可降低平均编码比特数。这时，传输每个预测误差所需的平均比特数，其理论下限为预测误差的熵值。

5．2 最佳线性预测

由预测编码法的基本原理可知，为了提高编码效率，首先应当对X_i具有准确的预测。怎样才能获得最佳的预测值呢？本节讨论应用均方误差（MSE）为极小值的准则来获得最佳线性预测值$\hat{X}_i$的方法。

设X为一个均值为零，方差为σ^2的离散平稳随机过程。X_n的线性预测值$\hat{X}_n$可以由前面$N-1$个数据X_1，X_2，…X_{N-1}得出，即：

$$\hat{X}_n = \sum_{i=1}^{N-1} a_i X_i \qquad (5.5)$$

其中a_i $i=1, 2, \cdots, N-1$为待定的常数，称为预测系数。

若各个预测系数a_i为已知，则可按式（5．5）构成线性预测器。这种线性预测系统的框图如图5．3所示。图中略

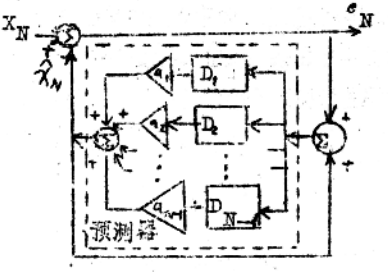

图5．3 线性预测编码器的框图

去了量化器，D_i　$i=1, 2, \cdots N-1$ 代表 i 个节拍延迟的延时元件。

此时，预测编码器输出器输出的差值信号 e_N 为：

$$e_N = X_N - \hat{X}_N \qquad (5.6)$$

下面，我们应用均方误差为极小值的准则，求出一组最佳预测系数，以获得 X_N 的最佳线性预测 $\hat{X}_N$。

X_N 的均方误差为：$E\{(X_N-\hat{X}_N)^2\}$

根据最小均方误差准则，应当求出满足 $E\{(X_N-\hat{X}_N)^2\}$ 为最小值时的一组预测系数

$$a_i \qquad i=1, 2, \cdots, N-1.$$

故将 $E\{(X_N-\hat{X}_N)^2\}$ 对各个 a_i 求偏导，并令其为零，即：

$$\frac{\partial E\{(X_N-\hat{X}_N)^2\}}{\partial a_i} = \frac{\partial}{\partial a_i} E\{[X_N-(a_1X_1+a_2X_2+\cdots a_{N-1}X_{N-1})]^2\}$$

$$= -2E\{[X_N-(a_1X_1+a_2X_2+\cdots a_{N-1}X_{N-1})]X_i\}$$

$$= 0 \qquad i=1, 2, \cdots, N-1$$

此式即：

$$E\{[X_N-(a_1X_1+a_2X_2+\cdots+a_{N-1}X_{N-1})]X_i\} = 0 \qquad (5.7)$$

或

$$E\{(X_N-\hat{X}_N)X_i\} = 0 \qquad (5.8)$$

$$i=1, 2, \cdots, N-1$$

(5.7) 式或 (5.8) 式表示了一个由 $N-1$ 个方程构成的线性方程组。

若假设 X_i 和 X_j 的协方差为：

$$R_{ij} = E\{ X_i X_j \}$$

式(6.7)表示的 $N-1$ 个方程可以改写为：

$$E\{ X_N X_i \} = a_1 E\{ X_1 X_i \} + a_2 E\{ X_2 X_i \} + \cdots + a_{N-1} E\{ X_{N-1} X_i \}$$

即：

$$R_{Ni} = a_1 R_{1i} + a_2 R_{2i} + \cdots + a_{N-1} R_{N-1,i} \qquad (6.9)$$

$$i = 1, 2, \cdots, N-1$$

在这 $N-1$ 个方程构成的线性方程组中，共有 $N-1$ 个未知的待定常数。若所有的协方差（共有 $N(N-1)$ 个）都已知，$N-1$ 个预测系数 a_i 便能够由式(6.9)表示的方程组解出。由于此时均方误差 $E\{ X_N - \hat{X}_N \}^2$ 为最小，$\hat{X}_N$ 是 X_N 的最佳线性预测，相应的 $N-1$ 个预测系数称为最佳线性预测系数。

求出最佳线性预测系数 a_i，$i = 1, 2, \cdots, N-1$ 后，按(6.5)式便可作出 X_N 的最佳线性预测 $\hat{X}_N$。预测误差信号的方差为：

$$\sigma_e^2 = E\{ (X_N - \hat{X}_N)^2 \} = E\{ (X_N - \hat{X}_N) X_N \}$$

$$= R_{NN} - (a_1 R_{N1} + a_2 R_{N2} + \cdots + a_{N-1} R_{N,N-1}) \qquad (6.10)$$

式(6.10)右端第一项 $R_{NN} = E\{ X_N^2 \}$

恰为输入数据序列 X 的方差 σ^2（∵已假定输入数据的均值为零）。于是，上式成为：

$$\sigma_e^2 = \sigma^2 - (a_1 R_{N1} + a_2 R_{N2} + \cdots + a_{N-1} R_{NN-1}) \qquad (6.11)$$

此式表明了输入信号序列{ X_i }和差值信号序列{ e_i }的方差之间的关系。由于式(5.11)是在均方误差最小的准则下得到的，该式表明：

$$\sigma_e^2 < \sigma^2$$

甚至可能有 $\sigma_e^2 \ll \sigma^2$

如果输入信号序列是互不相关的，即 $R_{ij}=0$ $(i \neq j)$，则有 $\sigma_e^2 = \sigma^2$，预测误差序列的方差便不会减少。反之，若输入信号的相关性愈强，方差 σ_e^2 减小的效果就会显著。

可见，由于利用了输入信号序列的相关性，误差序列的方差与信号序列的方差相比总是要小一些，甚至可以小很多。换句话说，误差序列{ e_i }的相关性比起输入信号序列{ X_i }的相关性要弱一些，甚至弱很多。

如果输入的数据序列{ X_i }是一个 m 阶的马尔柯夫过程，即数据 X_N 只与前面 m 个数据有关，与更前面的数据无关，则只需采用前 m 个数据对 X_N 作出预测，并且这样得到的误差序列{ e_i }将是完全去相关的。

5.3 差值脉冲编码调制（DPCM）

最佳线性预测理论可以使预测误差在最小均方误差的准则下达到最小，但是需要计算输入数据 X_i $i=1, 2, \cdots$ 的协方差，再由 $(N-1)$ 个方程构成的线性方程组求解出 $N-1$ 个最佳线性预测系数，需要的计算量很大。在实际中常根据不同的信源与要求将问题简化，以便于实现。这一节我们主要介绍构成 DPCM 系统的基本方法和性能分析。

6.13.1 预测方案

通常，输入信号序列$\{X_i\}$是一个随机过程。为了将问题简化，根据实际情况和具体要求，常常限定相关数据的距离，即假定$\{X_i\}$是一个平稳的m阶马尔柯夫过程。对于语音信号，认为是一维的相关序列，对于图象数据，根据要求和条件可看作一维、二维或三维的相关过程。对于不同的假定，采用的预测方案不同。实际中常用的预测方案大致有下述几种。

1．前值预测

前值预测是一种最简单的预测方案。

假定信源序列是一个一阶的马尔柯夫过程，数据X_N的预测值便可由其最接近的前一个数据X_{N-1}作出，即：

$$\hat{X}_N = aX_{N-1} \qquad (6.12)$$

由于一般的信源数据序列并不满足一阶马尔柯夫的限制，前值预测方案虽然简单，由于利用数据间的相关性很不充分，预测误差较大，数据压缩的效果也较差。

2．一维预测

假定信源数据序列为一个k阶的马尔柯夫过程，数据X_N的预测值便可由前面k个数据作出，即：

$$\hat{X}_N = a_1 X_{N-1} + a_2 X_{N-2} \cdots + a_k X_{N-k} \qquad (6.13)$$

$$= \sum_{i=1}^{k} a_i X_{N-i}$$

此时，最佳线性预测系数可由式(6.9)所表示的k个方程联立解

出。实际中阶数 k 常根据要求和设备能力来决定。为了避免繁杂的运算，亦常由实验确定预测系数 a_i $i=1, 2, \cdots k$ 的值。

一维预测方案可以用于语音数据序列，亦可用于图象数据。对于图象数据，此方案由同一扫描行中前面 k 个数据进行预测。由于它没有利用不同扫描行之间的相关性，去除冗余的能力仍较差。

为了更充分地利用图象数据间的相关性，对于静止图象可以采用二维预测技术，对于活动图象则还应利用帧间的相关性，采用三维的预测技术。

3. 二维预测

设输入图象为 $X(m, n)$，其线性预测值为：

$$\hat{X}(m, n) = \sum_{(k, l)\in Z} a_{k,l} X(m-k, n-l) \qquad (5.14)$$

相应的预测误差为：

$$e(m, n) = X(m, n) - \hat{X}(m, n)$$

$$= X(m, n) - \sum_{(k, l)\in Z} a_{k,l} X(m-k, n-l) \qquad (5.15)$$

其中 Z 为对象素 $X(m, n)$ 进行预测的相关点的集合。原则上说，Z 应取与 $X(m, n)$ 相关性较大的点。实际中有时考虑到象素出现的顺序及因果性，即应由已传送的点来计算未经传送的点。故 Z 不能任意选择。对于一般的数据传输系统，为了实时处理，通常采用一因果预测的方法，如图 (5.4.a)。然而一般图象相关性较大的点常常是周围的许多点，如图 (5.4.b)，它们是非因果型的。在非实时处理时亦可采用非因果的预测方法，此时，必须在输入图象相关部分输入后才能开

始计算其预测值和差值。

无论是因果型的还是非因果型的，都是利用数据间的相关性对某数据进行预测。因此，可在最小均方误差的意义下求出最佳预测系数 $a_{i,j}$，由于计算最佳线性预测系数需要附加许多计算量，实际中也常由实验确定一组固定的预测系数进行线性预测。

图5.4a 符因果型预测区域

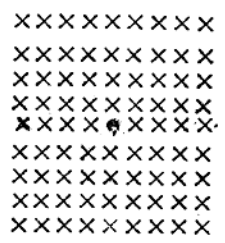

图5.4b 非因果型预测区域

5.3.2 量化编码

5.3.2.1 信息保持型编码

5.1节的分析已经指出，差值信号 $\{e_i\}$ 呈拉普拉斯分布（如图5.1所示）。根据拉普拉斯分布的特点，如果将差值信号进行变长编码，如霍夫曼编码，便可降低平均编码比特数。这时，传输每个预测误差所需的平均比特数，其理论下限由预测差值信号的熵值决定。此时，由于是无失真编码，亦称作信息保持型编码。信息保持型编码可应用于不允许产生失真的数据传输系统。如通过地球资源卫星拍摄的卫星照片，它们首先要转换成PCM的形式，以便于进行计算机的数字处理。这时，除了PCM的线性量化外，不允许其它任何不可逆的编码运算。这时的编码处理既要实现有效的数据压缩，又必须精确地恢复卫星图象中的科学数据。

无失真编码方法很多，此处我们介绍一种结合线性预测技术对预

测误差信号$\{e_i\}$进行无失真编码的一种方法：差分脉码移位编码。

设输入数据和其预测值都是数字式的，则需进行编码的预测差值信号$\{e_i\}$也是数字式的，并且差值信号的取值可能数将比输入数据要增加一个比特表示。

但是，差值信号$\{e_i\}$的分布呈拉普拉斯分布，其基本特点是差值信号e_i的取值在零附近有很大的概率，而距零较远的取值发生的可能性很小。假设大部分差值信号都落在± 8的范围内，便可用一个含有16个码字C_1、C_2，…，C_{16}的循环码来对每一差值信号e_i进行编码。这样就能保证对于绝大多数差值信号都可以用C_1，C_2，…，C_{15}这14个码字中的一个进行编码，而对于那些取值在± 8以外的差值信号e_i，则可利用另两个码字C_1和C_{16}作为标志，重复利用$C_2 \sim C_{15}$的14个码字对其进行编码。

例如，设C_1，C_2，…，C_{16}分别表示0000，0001，…，1111。当e_i取值在-7到$+8$之间时，我们可以用C_2到C_{15}分别对这14种取值情况一一对应编码，如图5.5所示。用C_1和C_{16}分别表示e_i取值小

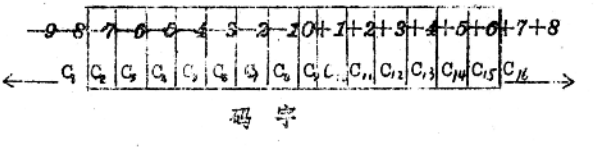

图5.5 e_i取值在$-7 \sim +8$范围内时，码字分配方案图

于－7和大于＋6。当出现这种情况时，这14个码字分别向负方向或正方向移动，用以对－21到－8或＋7到＋20的差值信号编码。图5．6表示了当e_i取值超过＋6时，对e_i取值在＋7到＋20编码的情况。如果e_i的取值还超过＋20，则再次利用C_{16}（C_{16}出现两次），然后把移位码C_2—C_{15}上移到＋21—＋34来进行编码。例如＋24的编码为C_{16} C_{16} C_9。于是只需利用从C_1到C_{16}的16个码，便可对差值信号（e_i）进行一一对应的无失真编码了。

用16进制的等长码进行这种编码时，大部分差值信号都可用4比特码字编码，只有少数差值信

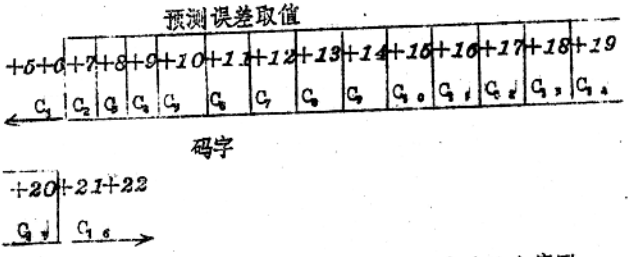

图5．16　e_i取值在＋7—＋20范围时，码字分配方案图

号需用8比特以至于12比特。16比特码字进行描述。从统计的观点可看出，每一数据的平均码长减小。

进一步减小编码率的方法是用变长编码。由于大部分差值信号的取值集中在零附近，在图5．5中码字C_9及其两旁的码字出现的概率将明显大于C_2，C_{15}等靠近两侧的那些码字。这提示我们用交

长码表示 $c_1 - c_4$，将比上面的等长码将会更有效。

5．3．2．2 限失真编码

对于一般的语音、图象等信息的传输系统，一般不要求无失真编码，这时可进行限失真的编码以实现更大的数据压缩率。

在预测编码系统中，由于需量化、编码并传输的预测误差信号 e_N 的方差较小，量化器的动态范围可以缩小，量化器分层的总数可以减小。因而可降低每一数据的编码比特数而不致产生显著失真。

对差值信号重新进行的量化方法，通常有下面三种。

1．均匀量化

由最佳线性预测理论可知，在预测编码系统中，需要量化、编码的差值信号 e_N 的方差较小，故可用较小的量化层次对 e_N 进行量化。最简单的量化方法就是均匀量化，即将 e_N 的取值范围按等距离分割，量化电平一般取在每一量化区间的中间点。

2．非均匀量化

非均匀量化是根据差值信号的概率分布规律（如式（5．2））确定量化间隔和量化值。

由于差值信号呈拉普拉斯分布，采用相应的非均匀量化可以改善输出图象的质量并节省比特率。

3．最佳量化方法

对于要求不高的数据压缩处理，从实际的简单易行方面考虑，可按上述均匀或非均匀量化的方法对 e_N 进行量化编码。但是这两种方法都无法控制总的失真程度，它们都不是最佳的。

最佳量化器的设计有两种方法。它们分别依赖于数据压缩处理质量的两种不同评价方式。

第一种方法是依靠客观评价准则。设计的基本方法是：给定量化器分层总数，根据量化误差的均匀方值为最小值来设计量化器。

第二种方法依赖于主观评价准则。通过给出主观评价的一个限度，使量化分层总数尽量小，并保证量化误差不超出此主观评价的限度。

下面分别介绍这两种最佳量化器的设计原理。

根据量化误差的均方值（MSE）为最小值的准则设计量化器的步骤，是Max（1960）提出的。假设e_1，e_2，…e_k…，e_K为量化输出电平；d_1，d_2，…d_k…d_{K+1}为量化判决电平。当DPCM系统量化器输入端的差值信号e满足下列关系（参见图5）：

$$d_k < e \leq d_{k+1}, \quad k=1, 2, \cdots K \qquad (5.16)$$

时，量化器输出为e_k，且量化量输出的量化误差为$(e-e_k)$。假若量化器输入连续信号e，则量化输出的量化误差均方值为：

$$\sigma_q^2 = \sum_{k=1}^{k} \int_{d_k}^{d_{k+1}} (e-e_k)^2 P(e)de \qquad (5.17)$$

其中$P(e)$是差值信号e的概率密度函数。为了使式（5.17）量化误差的均方值σ_q^2取极小值，必须满足下述两个条件（MaX作了证明）：

$$\int_{d_k}^{d_{k+1}} (e-e_k)P(e)de = 0 \qquad (5.18)$$

其中

$$d_k = \frac{1}{2}(e_{k-1}+e_k) \qquad (5.19)$$