试论信源编码与信道编码的相似性

许 织 新

提 要 本文从乏晰集合（Fuzzy Sets）理论的角度，论述了信源编码和信道编码的原理，从而论证了它们间的相似性，得出了四个相应的概念。最后说明了伴随式法用作数据压缩的方法和计算机模拟结果。

一、引 言

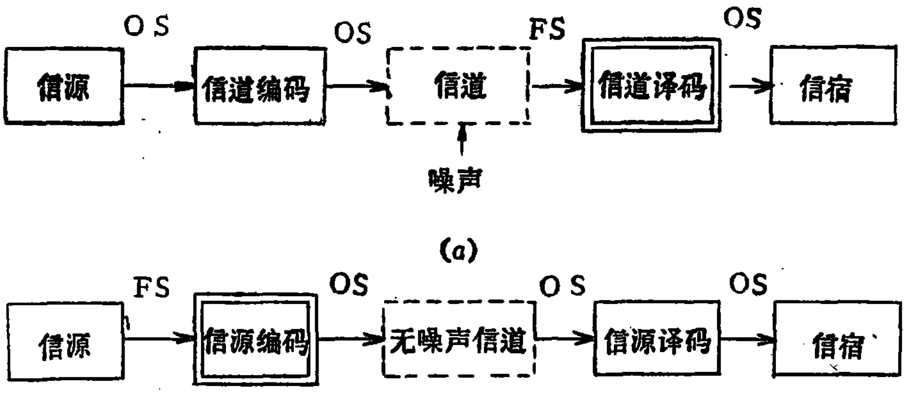
大家知道，经典的香农(Shannon)信息论，一直是沿着信源编码和信道编码这两大分支而发展起来的。时至今日，这两支理论已经都有了很大的进展，并已相继应用到实际的信息科学领域里去了。从性质上来看，信源编码是一种针对信源而言的信息处理，而信道编码则是针对信道而论的又一种信息处理方法。显然它们是有重大差异的。但是从一开始，人们似乎就看到了它们之间的联系和类同。例如，信道容量与信源熵之间的确定关系式，信道编码的可纠正错误的限界和码率失真函数之间的关系[1]，以及应用等长码（如分组码、树码、栅格码）进行信源编码的存在性定理[2]等等，都是信源编码和信道编码统一性的若干重大体现。

近十几年来，数学界提出了一种新的理论，叫"乏晰集合论"（FS, Fuzzy Sets, 又叫模糊集合论），它很快在信息科学范畴内得到应用。G.Longo 教授认为：应该把信源和信道的输出看成是一种乏晰集合，而信源编码器和信道译码器的任务，都是将此乏晰集合变换成普通集合[3]。这样，就在新的基础上把信源编码和信道编码统一起来了。依照这种论点，我们不妨将信源和信道编码的模型建立起来，见图1-1。

在图1-1(a)中，信源并不经过有效编码，只经过信道编码后进入信道。在此以前，我们都把它们看成是普通集合(OS，Ordinary Sets)，只是经信道噪声迭加后，其输出可看成是乏晰集合。而信道译码器的任务，就是将其处理成普通集合（或至少说处理成一种乏晰性更小的 集合）。同样，在图(b)中，因为专门研究 信源 特性，所以一开始就将信源输出看成是乏晰集合。通过信源编码器，就处理成为普通集合。尔后经无噪声信道，信源译码器一直到信宿。这一段，我们可以认为始终是普通集合。

下面，将深入地阐明这种由 FS 变成 OS 的具体理解，并论述实际的将信道编码技术用于信源编码的可能方法，且示出了计算机模拟结果以证明其正确性。

本文 1982 年 7 月26日收到



(b)

OS：普通集合，　FS：乏晰集合

图 1-1　信源和信道编码模型

二、用乏晰集合论解释信道译码原理

我们先以(3, 1)码来说明问题的本质，这种码的生成多项式为

$$G_1=[1\ \ 1\ \ 1] \tag{2-1}$$

即信息位重复 3 次的意思。这里命名为 C_1 码。于是按照图 1-1 的模型，就可以得出下述那样的译码过程，现示于图 2-1 中。

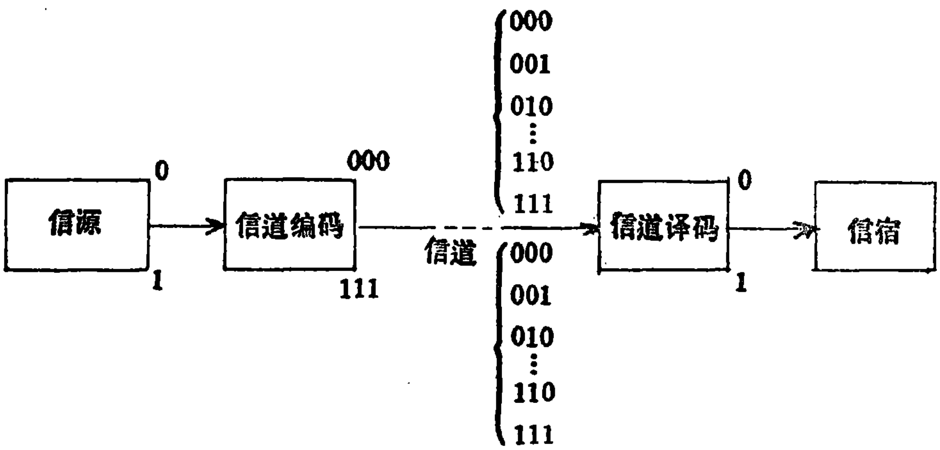


图 2-1　(3,1) 码编译码过程

由图可见，信源是个二进制对称信源，只送出 1 和 0 二个字符。信道编码的输出就成为 111 和 000 三个码字。在经过信道噪声迭加后，不管上述任一种码字输入，它的输出都可能是[000, 001, 010, 100, 011, 101, 110, 111]这八个码字中的一个。于是这八个码字中哪个更可能是输入码字？它们的可能性要用什么来描述？应用乏晰集合理论是最适宜不过了。

根据乏晰集合理论[4]，我们定义一个参考集 Q

$$Q=\begin{Bmatrix} x_0, & x_1, & x_2, & x_3, & x_4, & x_5, & x_6, & x_7 \\ (000) & (001) & (010) & (100) & (011) & (101) & (110) & (111) \end{Bmatrix} \tag{2-2}$$

其中 x_i $(i=0,1,\cdots,7)$ 是参考集中的八个元素。于是，例如针对信道输入为000而言，可定义在Q中的一个乏晰子集 $\underset{\sim}{A_0}$ 为

$$\mu_{\underset{\sim}{A_0}}(x)=\left\{\frac{x_0}{1},\ \frac{x_1}{0.66},\ \frac{x_2}{0.66},\ \frac{x_3}{0.66},\ \frac{x_4}{0.33},\ \frac{x_5}{0.33},\ \frac{x_6}{0.33},\ \frac{x_7}{0}\right\} \tag{2-3}$$

其中，每一项的"分母"表示该项的隶属度函数 $\mu_{\underset{\sim}{A_0}}(x_i)$。注意，这仅仅是一种表示方法而不代表一般的分式意义。这就是说，x_0 比其它元素要更象 000，而 x_2 则又要比x_4，x_5，x_6 更象一些。而对 x_7 来说，我们则认为面目全非，故其隶属度为 0 .

同样，针对信道输入为 111 而言，又可定义在Q中的另一个乏晰子集 $\underset{\sim}{A_1}$ 为

$$\mu_{\underset{\sim}{A_1}}(x)=\left\{\frac{x_0}{0},\ \frac{x_1}{0.33},\ \frac{x_2}{0.33},\ \frac{x_3}{0.33},\ \frac{x_4}{0.66},\ \frac{x_5}{0.66},\ \frac{x_6}{0.66},\ \frac{x_7}{1}\right\} \tag{2-4}$$

其概念可同样得到解释。

于是不难写出它们的隶属度函数的一般式为

$$\mu_{\underset{\sim}{A_j}}(x_i)=\left(1-\frac{d(x_i)}{n}\right) \tag{2-5}$$

其中，n 为码字长度，$d(x_i)$是第 i 个元素与第 j 个对应物之间的汉明距离（Hamming 距离）。显然，信道误码率愈高，隶属度愈小，也就是乏晰性愈强。

这样，用乏晰集合理论就很容易来解释信道译码的原理了。根据对两个以上乏晰子集的隶属原则[5]，对任一元素x_i来说，首先要判定其更可能隶属于哪个乏晰子集？这就是，若有

$$\mu_{\underset{\sim}{A_j}}(x_i)=\mathrm{Max}\left[\mu_{\underset{\sim}{A_0}}(x_i),\ \mu_{\underset{\sim}{A_1}}(x_i)\right] \qquad i=0,1,\cdots,7 \tag{2-6}$$

则就说，x_i 更可能隶属于$\underset{\sim}{A_j}$. 例如将$\underset{\sim}{A_0}$，$\underset{\sim}{A_1}$同画于图 2-2 中。显见，若信道译码器收到x_3，于是有

$$\begin{aligned}\mathrm{Max}\left[\mu_{\underset{\sim}{A_0}}(x_3),\ \mu_{\underset{\sim}{A_1}}(x_3)\right]&=\mathrm{Max}[0.66,0.33]\\&=0.66=\mu_{\underset{\sim}{A_0}}(x_3)\end{aligned}$$

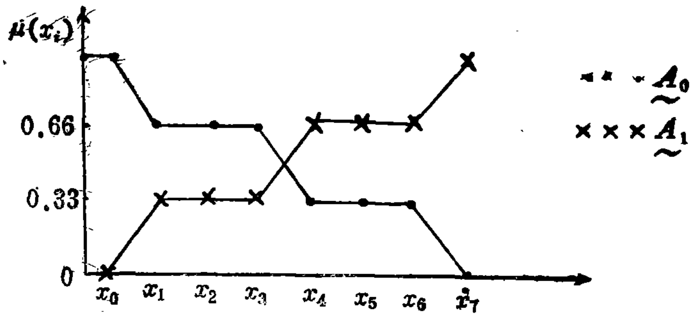


图 2-2 $\underset{\sim}{A_0}$和$\underset{\sim}{A_1}$的隶属度函数

因此说，若收到 100，则可判定它更可能象 000。余此类推。

信道译码的第二步是要将其变成普通集合。也就是说，经过隶属原则确定了x_0, x_1，x_2, x_3 是相对隶属于 $\underset{\sim}{A}_0$ 以后，但它毕竟是乏晰概念，$x_4 \sim x_6$ 仅仅是相对隶属性差一些而已，所以可根据设定一个置信水平 λ[5]，将乏晰集合变为普通集合，λ 实际上是个门坎，介于 0，1 之间的实数，我们规定，当 $\mu_{\underset{\sim}{A}}(x_i) \geqslant \lambda$ 时，便算作$x_i \in \underset{\sim}{A}$，当 $\mu_{\underset{\sim}{A}}(x_i) < \lambda$ 时，便算作 $x_i \overline{\in} A$. 于是得到一个普通子集，记为 A_λ，即

$$A_\lambda = \{x_i: \ x_i \in X, \ \mu_{\underset{\sim}{A}}(x_i) \geqslant \lambda\} \tag{2-7}$$

例如，若设定门坎为$\lambda = 0.5$，则普通子集

$$\mu_{A_0}|_{\lambda=0.5} = \left\{\frac{x_0}{1}, \frac{x_1}{1}, \frac{x_2}{1}, \frac{x_3}{1}, \frac{x_4}{0}, \frac{x_5}{0}, \frac{x_6}{0}, \frac{x_7}{0}\right\}$$

$$= \left\{\frac{x_0}{1}, \frac{x_1}{1}, \frac{x_2}{1}, \frac{x_3}{1}\right\} = \{x_0, x_1, x_2, x_3\} \tag{2-8}$$

此时，x_0, x_1, x_2, x_3的隶属度均为 1，其余均为 0。同理可得

$$\mu_{A_1}|_{\lambda=0.5} = \{x_4, x_5, x_6, x_7\} \tag{2-9}$$

于是可用记号 μ_0，μ_1 来记该普通集合，即

$$\left.\begin{array}{l} \mu_0 = 0 \quad \Rightarrow A_0|_{\lambda=0.5} \\ \mu_1 = 1 \quad \Rightarrow A_1|_{\lambda=0.5} \end{array}\right\} \tag{2-10}$$

这样就完成了信道译码过程。

大家知道，从普通集合的观点去研究信道译码的性能的话，如码字之间的汉明距离愈大，则纠错能力愈强，通常可纠正错误个数 t 为

$$t \leqslant \frac{d_{\min}(c) - 1}{2} \tag{2-11}$$

$d_{\min}(c)$ 为码字间的最小汉明距离。实际上就是说以二个码字为中心，以 t 为半径所形成的二个集合之间的交集要为 0，这样才能正确译码。而乏晰集合理论中的交集却定义为

$$(\underset{\sim}{A}_0 \cap \underset{\sim}{A}_1)(x_i) = \min(\mu_{\underset{\sim}{A}_0}(x_i), \ \mu_{\underset{\sim}{A}_1}(x_i)) \tag{2-12}$$

由于乏晰集合缺乏明确的集合边界，所以上述交集为 0 的原理就无法应用了。但是可以看到，当(2-12)式所示的乏晰交集愈是小的话，那么也可认为纠错能力就愈强。同时也自然希望，对任一元素在不同乏晰子集中的隶属度不能相等，至少要有一定的差别，才能正确运用隶属原则。所以，从乏晰集合观点看，可提出如下概念作为码字纠错能力的评定标准。

概念 1

为保证正确译码，必须满足下式

$$|\mu_{\underset{\sim}{A}_0}(x_i) - \mu_{\underset{\sim}{A}_1}(x_i)| \geqslant \delta \tag{2-13}$$

对任一x_i均成立，其中 δ 为大于 0 的某个实数。

概念 2

当有$j\geqslant 2$个乏晰子集时，对于任一x_i来说，则必有 j 个 $\mu_{\underset{\sim}{A}}(x_i)$ 存在。若设 $\mu_{\underset{\sim}{A_k}}(x_i)$ 与 $\mu_{\underset{\sim}{A_l}}(x_i)$为其中二个最大的隶属度，为保证正确译码起见，就一定要使

$$|\mu_{\underset{\sim}{A_k}}(x_i)-\mu_{\underset{\sim}{A_l}}(x_i)|\geqslant\delta \tag{2-14}$$

成立，其中 δ 为大于 0 的某个实数。

这二个概念实质上就是要使隶属原则有唯一解。上述码 C_1 即足以说明概念 1 。下面我们举码 C_2，即(4,2)码[6]来说明概念 2 .

(4,2)码的信息组是 00，10，01，11。它们分别所对应的许用码字是 0000，1001，0111，1110。但实际上可能存在的码字共有$2^4=16$个。我们将其编为有16个元素的参考集，并构成以 4 个许用码字为对象的 4 个乏晰子集，如下：

元素号：	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
码　字：	0000	1001	0111	1110	1000	0001	1111	0110	0100	1101	0011	1010	0010	1011	0101	1100
$\mu_{\underset{\sim}{A_0}}(x_i)=$	1	0.5	0.25	0.25	0.75	0.75	0	0.5	0.75	0.25	0.5	0.5	0.75	0.25	0.5	0.5
$\mu_{\underset{\sim}{A_1}}(x_i)=$	0.5	1	0.25	0.25	0.75	0.75	0.5	0	0.5	0.75	0.5	0.5	0.5	0.75	0.5	0.5
$\mu_{\underset{\sim}{A_2}}(x_i)=$	0.25	0.25	1	0.5	0	0.5	0.75	0.75	0.5	0.5	0.75	0.5	0.5	0.5	0.75	0.25
$\mu_{\underset{\sim}{A_3}}(x_i)=$	0.25	0.25	0.25	1	0.5	0	0.75	0.75	0.5	0.5	0.25	0.75	0.5	0.5	0.25	0.75
	许用码字				禁用码字											

可以看到，信道输出的 16 个可能码字中，除了 x_4，x_5，x_6，x_7 以外，其余均能正确译码。因为都将符合上述概念。但对$x_4(x_5)$来说，$\mu_{\underset{\sim}{A_0}}=\mu_{\underset{\sim}{A_1}}$，对$x_6(x_7)$来说，$\mu_{\underset{\sim}{A_2}}=\mu_{\underset{\sim}{A_3}}$，所以均不能正确译码。

以上所述，已经阐明了用乏晰集合理论如何来解释信道译码的原理，以及用乏晰集合理论来初步分析译码性能的若干概念。笔者推测，对于用大数判决作为译码原理的信道译码，上述解释都将是成立的。

三、用乏晰集合论解释信源编码原理

正如前一个问题那样，我们假设信源送出如下八种消息，即000，001，010，100，011，101，110，111。每一消息由 0 或 1 所构成的三个字符所组成。那么应该怎样来分

析这样一种信源呢？按照普通集合论来讲，主要考虑这八个事件x_i（$i=0,\ 1,\ \cdots 7$）的随机性，便可确定该信源所具有的信息量，但是，今天用乏晰集合论来观察该信源时，不但仍有其随机性的一面，同时还有另一面，即乏晰性，是必须考虑的。

这里仍按(2—2)式定义由八个元素组成为参考集。它们各自的概率可写成

$$Q=\begin{Bmatrix} x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ p_0 & p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 \end{Bmatrix} \tag{3-1}$$

同时可以定义以 000 及 111 为针对物的两个乏晰子集为$\underset{\sim}{A_0}$和$\underset{\sim}{A_1}$（与 2-3，2-4 式同）。这就是说，当信源编码器只“认识”二个状态，即只能“认识”000 或 111 二个状态时，从它来观察信源，就形成了$\underset{\sim}{A_0}$和$\underset{\sim}{A_1}$二个乏晰子集。不管信源送出为何，信源编码器总是输出000 和 111 二个状态，所以就构成如图3-1 所示那样的模型。

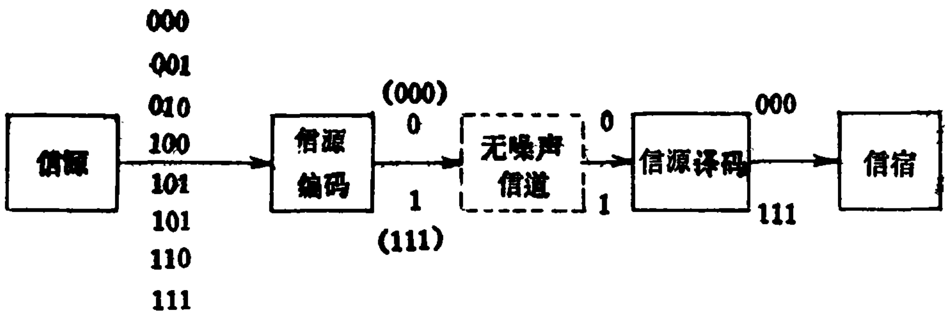


图 3-1　信源编码过程

当我们用这样的观点去观察信源时，应该看到信源有二种不肯定性(uncertainty)。一是由于概率不同而反映出来的，所谓经典的不肯定性。另一个是由于信源有某种乏晰性，或者叫不确实性（incertitude）而引起的不肯定性。因此，这二种不肯定性的解除，都将得到信息量。所以，从乏晰观点看信源的熵，应该考虑到这二方面。

应该指出，学术界对于乏晰集合中的熵的定义方法，目前至少看到有三种[7][8][9]。笔者认为，意大利学者 A.De Luca 和 S.Termini 在 1972 年的观点可能较为直观和贴切。在此，我们试照这种观点加以阐述。

如果观察者（信源编码器）通过某种判决策略，比如令置信水平λ为某值，从而将乏晰集合变为普通集合。这就解除了它的不确实性，于是得到一定的乏晰熵，它可仿照香农熵的定义方法，定义乏晰熵$H(\mu_{\underset{\sim}{A}})$为

$$\begin{aligned} H(\mu_{\underset{\sim}{A}}) &= -K\Big[\sum_{i=0}^{n-1}\mu_{\underset{\sim}{A}}(x_i)\log\mu_{\underset{\sim}{A}}(x_i)+\sum_{i=0}^{n-1}(1-\mu_{\underset{\sim}{A}}(x_i))\log(1-\mu_{\underset{\sim}{A}}(x_i))\Big] \\ &= K\sum_{i=0}^{n-1}S(\mu_{\underset{\sim}{A}}(x_i)) \end{aligned} \tag{3-2}$$

其中，$S(\mu_{\underset{\sim}{A}}(x_i))$称为香农函数，即

$$S(\mu_{\underset{\sim}{A}}(x_i))=-\mu_{\underset{\sim}{A}}(x_i)\log\mu_{\underset{\sim}{A}}(x_i)-(1-\mu_{\underset{\sim}{A}}(x_i))\log(1-\mu_{\underset{\sim}{A}}(x_i)) \tag{3-3}$$

而K为正实数。如$K=\frac{1}{n}$，则称为归一化熵。对数则以 2 为底。

如果按照这种定义方法，乏晰熵满足如下三条性质：

(a) 当且仅当$\underset{\sim}{A}$变成普通集合时，即 $\mu_{\underset{\sim}{A}}(x_i)$非 1 即 0 时，$H(\mu_{\underset{\sim}{A}})=0$。这就是说，既无不确实性可言，自然也无乏晰熵可得。

(b) 当且仅当 $\mu_{\underset{\sim}{A}}(x_i)=\frac{1}{2}$，对所有$x_i$时，则$H(\mu_{\underset{\sim}{A}})$为最大。显然，因为$\mu_{\underset{\sim}{A}}(x_i)=\frac{1}{2}$就是说明最为似是而非，解除这种不确实性，应该得到最大信息量。

(c) 当 $\mu_{\underset{\sim}{A}}(x_i)\geqslant\frac{1}{2}$时，如 $\mu^*_{\underset{\sim}{A}}(x_i)\geqslant\mu_{\underset{\sim}{A}}(x_i)$，则$H\geqslant H^*$.

当 $\mu_{\underset{\sim}{A}}(x_i)<\frac{1}{2}$时，如 $\mu^*_{\underset{\sim}{A}}(x_i)<\mu_{\underset{\sim}{A}}(x_i)$，则$H<H^*$

这个性质的意义，可用图 3-2 来表示。显然，当由此一乏晰集合变换成另一乏晰集合时，也可能部分消除这不确实性，因而也能得到

$$[H(\mu_{\underset{\sim}{A}})-H(\mu_{\underset{\sim}{B}})]$$

信息量，[假定 $H(\mu_{\underset{\sim}{A}})>H(\mu_{\underset{\sim}{B}})$]. 当$\underset{\sim}{B}$是普通集合时，则 $H(\mu_{\underset{\sim}{B}})=0$，故消除的不确实性即为$H(\mu_{\underset{\sim}{A}})$.

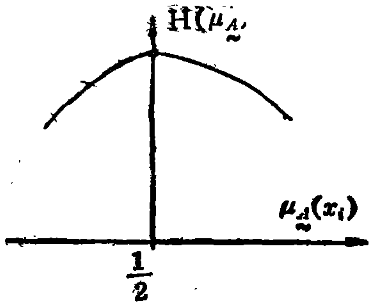


图 3-2 乏晰熵的性质

上面所述，都是在假设集合中各个事件是非随机事件而得出的结论。我们知道，乏晰集合可以是非统计性的，又可以是统计性的。下面来进一步说明统计性的乏晰熵。

首先考虑由随机性而引起的熵，这显然是

$$H(p_0,p_1,\cdots,p_{n-1})=-\sum_{i=0}^{n-1}p_i\log p_i \tag{3-4}$$

而由乏晰性所引起的熵，则应该改写(3-2)式为

$$H(\mu,p_0,\cdots,p_{n-1})=\sum_{i=0}^{n-1}p_i S(\mu_{\underset{\sim}{A}}(x_i)) \tag{3-5}$$

于是，此乏晰集合总的熵为

$$H_{\mathrm{T}}=H(p_0,\cdots,p_{n-1})+H(\mu,p_0,\cdots,p_{n-1}) \tag{3-6}$$

可见，此式综合了随机性和乏晰性的两种因素。当无随机因素发生时，即固定某一消息x_i，那么就认为$p_i=1$，于是随机熵为 0 ，总熵等于乏晰熵为$H_T=S(\mu_{\underset{\sim}{A}}(x_i))$. 如果有随机性时，且 p_i 为等概率分布，即$p_i=\frac{1}{n}$，则(3-5)式就与 (3-2) 式完全一致。当仅有随

机性而无乏晰性时，则(3-5)式为0，总熵即等于香农熵。

于是，我们再回过来看图3-1那样的信源编码原理。在信源编码器前后各自是乏晰集合和普通集合。因此，在信源各消息是等概率分布的前提下，由信源编码器所引起的乏晰熵要减少，其减少量恰为

$$I(\underset{\sim}{\mu_A};\underset{\sim}{B})=H(\underset{\sim}{\mu_A})-H(\underset{\sim}{\mu_B})=\frac{1}{n}\sum_{i=0}^{n-1}S[\underset{\sim}{\mu_A}(x_i)]-\frac{1}{m}\sum_{i=0}^{m-1}S[\underset{\sim}{\mu_B}(x_i)]. \tag{3-7}$$

这就相当于经典信息论中的相互信息量的概念。众所周知，码率失真函数$R(D)$就等于试验信道的相互信息量的最小值[2]. 因此，不妨形成如下概念。

概念 3

由信源编码所造成的乏晰熵的减少量，就是用来确定码率与失真之间关系的基本依据。在码率确定了的前提下，乏晰熵减小的多，则造成的再现失真就愈大。同样，当要求再现失真不超过某个容界时，那么如果乏晰熵减少的愈多，就必须用较大的码率才能实现。

例如，图3-1所示的减少量可计算得为

$$H(\underset{\sim}{\mu_A})=0.69\text{比特/消息}$$

而如果引进另一种码C_3，叫做(5,1)码，它的生成多项式为

$$G_3=[1\ 1\ 1\ 1\ 1]$$

即对信息位重复五次。此时，算出的减少量为

$$H(\underset{\sim}{\mu_A})=0.83\text{比特/消息}$$

这二种码率是一样的。显然后者所造成的再现失真为之较大。

综上所述，用乏晰集合论来观察信源编码和信道译码，可以发现它们有许多相通性。因此，将纠错码用之于数据压缩，自然是行得通的了。应该指出，上述用法有人称之为“码字法”。也就是说将码字作为信源消息，将对应的信息位作为信源的压缩输出。但是，这种码字法虽然原理上是成立的，但应用上却受到许多限制。在文献[3]中已有较详细论述。归之为一点，就是码字的特性不能很好地和信源特性相匹配。因之，还必须寻求其它有效方法来实现之。

四、用伴随式法进行数据压缩

根据信道编码的基本原理，如一个(n，k)线性分组码的码字为$\underline{v}$，信道所产生的误差图样为$\underline{e}$. 那么接收端所收到的码字为

$$\underline{r}=\underline{v}\oplus\underline{e} \tag{4-1}$$

其中，$\oplus$为模二加。这时信道译码的原理是将此码字向量与监督矩阵H相乘，所得之结果称为伴随式$\underline{s}$，即

$$\begin{aligned}\underline{s}&=H\underline{r}=H(\underline{v}\oplus\underline{e})\\&=H\underline{v}\oplus H\underline{e}\end{aligned}$$

$$=H\underset{\sim}{e} \tag{4-2}$$

然后，根据$\underline{s}$再计算出误差图样$\hat{\underline{e}}$来，将$\hat{\underline{e}}$与$\underline{r}$再模二加，就得到再现的码字$\hat{\underline{v}}$。只要在纠错能力范围之内，$\underline{v}$就会与$\hat{\underline{v}}$相等，图 4-1 则示出了一个伴随式译码器原理图。

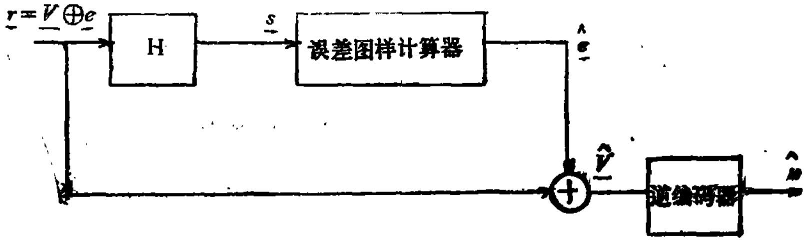


图 4-1 伴随式译码原理图

如果我们仍以(3, 1)码为例来说明，其译码过程可说明如下：

$\underline{v}$	$\underline{e}$	$\underline{r}$	$\underline{s}$	$\hat{\underline{e}}$	$\hat{\underline{v}}$	$\hat{\underline{u}}$
×××	0 0 0	×××	0 0	0 0 0	×××	
	0 0 1	××△	1 0	0 0 1	×××	×
	0 1 0	×△×	0 1	0 1 0	×××	
	1 0 0	△××	1 1	1 0 0	×××	
	0 1 1	×△△	1 1	1 0 0	△△△	
	1 0 1	△×△	0 1	0 1 0	△△△	△
	1 1 0	△△×	1 0	0 0 1	△△△	
	1 1 1	△△△	0 0	0 0 0	△△△	

这里，“×”可以指“0”或“1”，则“△”相应指“1”或“0”。可见相对于$\underline{v}$来说，$\underline{r}$或$\underline{e}$都形成了乏晰集合。同样$\hat{\underline{e}}$也是一个乏晰子集合。只是到了$\hat{\underline{v}}$及$\hat{\underline{u}}$那里，才变为普通集合。于是按以上讲过的原理，可定义$\underline{e}$为乏晰子集

$$\underset{\sim}{E}=\left\{\frac{e_0}{1},\ \frac{e_1}{0.66}\ \frac{e_2}{0.66}\ \frac{e_3}{0.66}\ \frac{e_4}{0.33}\ \frac{e_5}{0.33}\ \frac{e_6}{0.33}\ \frac{e_7}{0}\right\} \tag{4-1}$$

其中e_i的下标i则与前面顺序一致。但对于$\hat{\underline{e}}$来说，它与$\underline{e}$有着一定的关系。如果用乏晰关系来看，这二个乏晰子集间存在着一个普通关系R，可写为

$$R=\begin{array}{c|c} & \hat{e}_i \\ e_i & \left\{\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}\right\} \end{array} \tag{4-2}$$

于是，$\hat{\underset{\sim}{E}}$可通过乏晰变换而得，即

$$\hat{\underset{\sim}{E}}=\underset{\sim}{E}\circ R \tag{4-3}$$

其中，"∘"表示乏晰集合的内积运算。于是根据(4-1)及(4-2)式可求出$\hat{\underset{\sim}{E}}$为

$$\hat{\underset{\sim}{E}}=\left\{\frac{\hat{e}_0}{1},\ \frac{\hat{e}_1}{0.66},\ \frac{\hat{e}_2}{0.66},\ \frac{\hat{e}_3}{0.66},\ \frac{\hat{e}_4}{0},\ \frac{\hat{e}_5}{0},\ \frac{\hat{e}_6}{0},\ \frac{\hat{e}_7}{0}\right\} \tag{4-4}$$

它比起$\underset{\sim}{E}$来讲，已经比较接近普通集合了，当然，再经过向$\hat{\underline{v}}$子集的变换，便完全成为普通集合了。

于是，就可以用这种伴随式译码原理来完成数据压缩，其原理示于图 4-2 中。根据

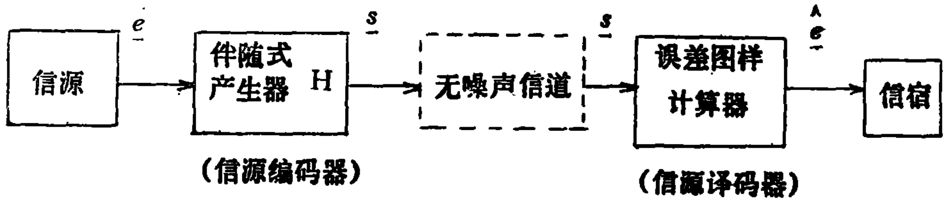


图 4-2 伴随式译码原理图

对乏晰熵的定义，在等概率分布的前提下，它们的乏晰熵分别为

$$H(\mu_{\underset{\sim}{E}})=0.69\text{比特/消息}$$

$$H(\mu_{\hat{\underset{\sim}{E}}})=0.35\text{比特/消息}$$

乏晰熵的减少量为

$$H(\mu_{\underset{\sim}{E}})-H(\mu_{\hat{\underset{\sim}{E}}})\doteq 0.34\text{比特/消息}$$

这就表明有一定的信息传输至收端。

在这里显然可以看到，对$e_0\sim e_3$来讲，并不造成再现失真，只是对$e_4\sim e_7$会造成较大的失真。如果$\underset{\sim}{E}$乏晰子集中各元素的概率分布得极不对称，比如前四个元素的概率远大于后四个元素的概率，那么所造成的平均失真就大大降低。也就是说，此时乏晰熵的减少量就会降低，那么在码率固定情况下自然会减低平均失真。

还应指出，对无失真的四个元素来讲，每个码字的 1 和 0 的分布是不对称的。这一点正好与实际的许多信源有着 0 多 1 少的分布相吻合。而且信源的本身的记忆性愈强，連 0 的机会就更会增多，所以说更适宜用伴随式来压缩数据。

这样，我们对于信道编码和信源编码之间的相似性的分析的结论，似可阐述为如下概念：

概念 4

应该把信源和信道的输出看成是乏晰集合。而信道译码的任务和信源编码（或者信源编/译码)的任务，就是将此乏晰集合变换成普通集合（或者变换成另一个乏晰性更小的乏晰集合）。这就是信源编码理论和信道编码理论之间相通性的一种重大体现。

五、计算机模拟

以上所述的伴随式法之用于数据压缩，已经为许多学者所证实[14][10]. 我们也同样进行了必要的计算机模拟。

比如，选择码 C_4, 即(15,11)汉明码，以及码 C_5, 即 (15,7) BCH码作为我们所用的纠错码。在130机上用BASIC语言进行模拟计算，其结果是令人鼓舞的。只要在纠错能力范围之内，数据得到压缩，再现时并不产生失真。其数据压缩比，显然是 $n/(n-k)$。

六、结　语

信源编码和信道编码之间的相似性，二重性，或者叫统一性，已经成为近代信息论的发展方向之一[13]，被提到当前的议事日程上来了。本文所论及的内容，仅仅从某个角度，即是从乏晰集合理论的角度，去解释它们之间的直接关系。显然这些内容是带有探讨性的，也是不全面的。尽管如此，我们把它还是归结为上述的四个概念，以作为今后工作的起点。

这里还要指出，香农的有关信道编码理论的非结构性，已经被当前迅速发展的代数编码理论所解决了。同样，信源编码理论也是缺乏这种结构性的。但是，直到现在为止，信源编码的结构性理论还未系统地、统一地形成。只能说，有一些散在的不系统的结果已经得到，比如著名的Huffman编码原理等等，因此，一些代数编码理论能否和如何应用到信源编码中来，正好是当前学术界的希望所在。

当然，信源编码之所以没有形成自己的结构性理论是有其内在原因的。这大致是：

(a) 由于信源特性远比信道特性复杂得多，所以，信源编码的难度显然要比信道编码的难度大得多。比如说，信道输出的概率分布，通常集中在码字附近。但是对信源来讲，却不总是会这样分布的。这就给信源编码带来了特殊的困难。

(b) 实际上可允许的信源编码所造成的失真，通常是很小的，因而使得许多信源编码技术都不能得到好的压缩效果。

(c) 在实际中存在的大多数（或许多）信源，是十分难于用某个数学模型来表述清楚的。同样，对于失真的测量究竟用什么数学模型为好，也缺乏明确的定论。

应该看到，学术界提出这些问题所在，并不是一种悲观的宿命论。恰恰相反，正好给人们指出了问题的症结和努力的方向。当前许多在信道编码领域内造诣极深的学者，正转过来从事于两者相似性的研究，正是说明了当前的趋势。在这一点上，期望我国的科学工作者迎头赶上，为发展我国近代信息论作出积极的贡献。这也是笔者的微小心愿。谨请信息论的老前辈和同志们多加赐教和指正。

参 考 文 献

[1] J.K.Omura: Source and Channel Coding with Block and Convolutional Codes, «CISM» 1978, P.2～80。

[2] 有本 卓: Rate-Distortion理论, «电子通信学会誌», 1978, Vol.5～9.

[3] G.Longo: Fuzzy Sets, Graphs and Source Coding, «New Direc. in Sing.Proc.in Commun.and Cont.» NATO, 1975, P.27～34.

[4] A.考夫芒著:模糊集理论习题集(汪浩译),国防科技大学数学教研室,1979.

[5] 汪培庄: 模糊数学簡介(Ⅰ), «数学的实践和认识», 1980,2, P.45～59.

[6] 王新梅: 纠错码浅说, 邮电出版社, 1976.

[7] A.De Luca; S.Termini: A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory, «Information & Control», May 1972, P30～312.

[8] T.Okuda等: A Formulation of Fuzzy Decision with Fuzzy Information using Probability Measures of Fuzzy Events «Information & Control», Aug. 1978, P. 135～147.

[9] 山本雅晴: Fuzzy 代数とその应用 (連載), «数理科学», 1970, 9, P.58～62.

[10] 铃木辉晓等: 单纯マルコフ情报源に対するシントローム情报源符号化について «电子通信学会论文誌» 79/10 P.736～743.

[11] F.J.Macwilliams 等: Theory of Error-Correcting Codes, North-Holland Publishing Comp., 1977.

[12] T.C.Ancheta: Syndrome Source Coding and Its Universal Generalization, «NASA-CR-143102», 1975.

[13] R.J.Mc Eliece: Information Theory and Coding, Addison-Wesley Publishing Company, 1977.

[14] J. L. Massey: The Codeword and Syndrome Methods for Data Compression with Error-Correcting Codes, «数字信息传输专题译丛», 1980, 国防科技大学。

On the Similarity between Source Coding and Channel Coding

Xu Zhi-xin

Abstract

On the basis of Fuzzy Sets theory, the similarity between source coding and channel coding was discussed in this paper. And then four concepts can be suggested. Finally, the data compression principal by syndrome method and the results of computer simulation are presented.