**第 11 次课 学时 2**

|  |  |
| --- | --- |
| **授课章节内容** | 第七章 无失真信息编码 |
|  | 7.3 常见的无失真信源编码方法 |
| 7.4 实用编码方法 |
| **教学目标** | 教学目标5 |
| **支撑毕业要求** | 毕业要求1-4 |

**教学要求：**

1. 知识目标

* 掌握常见的无失真信源编码方法；了解实用编码方法。

1. 能力目标

* 能运用无失真信源编码方法完成信源压缩。

1. 素质目标

* 激发学生对电子信息相关专业课程的学习热情，通过对无失真信源编码方法学习，提升解决工程问题的素质

**教学重点与难点**：

香农编码，费诺编码，哈夫曼编码，代数编码

**教学过程设计：**

课前观看视频7.5,7.6，导入新课，课程小结

| **讲授与指导内容** | | **讲课、互动内容设计** |
| --- | --- | --- |
| 7.3 常见的无失真信源编码方法  原始信源普遍存在剩余度，香农信息论认为信源的剩余度主要来自两个方面：一是信源符号间的相关性，二是信源符号概率分布的不均匀性。为了去除信源剩余度，提高信源的信息传输率，必须对信源进行压缩编码。  目前去除信源符号间相关性的主要方法是预测编码和变换编码（两种编码方法不是本书讨论重点），而去除信源符号概率分布不均匀性的主要方法是统计编码。  本节主要介绍对统计特性已知的无记忆离散信源进行的统计编码，统计编码又称为匹配编码，因为编码过程需要匹配信源的统计特性。对于无记忆信源，其剩余度主要体现在各个信源符号概率分布的不均匀性上。统计编码能实现压缩的关键是编码器在编码过程中尽可能等概率地使用各个码符号，从而使原始信源各符号出现概率的不均匀性在编码后得以消除。  7.3.1香农（Shannon）码  香农第一定理的证明过程给出了一种编码方法，称为香农编码。其编码方法是选择每个码字长度满足  (7.1)  可以证明，这样的码长一定满足Kraft不等式，所以一定存在这样码长的唯一可译码和即时码。  香农编码的基本思想是概率匹配原则，即概率大的信源概率符号用短码，概率小的信源符号用长码，以使减小平均码长，提高编码效率。  香农编码的步骤如下：   1. 将信源发出的*q*个消息，按出现概率递减顺序进行排列； 2. 计算各消息的； 3. 确定满足式（7.1）的整数码长； 4. 计算第i个消息的累积分布函数； 5. 将累积分布函数变换成二进制数； 6. 取二进制数的小数点后位作为第*i*个符号的二进制码字。   【**例7.4**】已知信源共6个符号，其概率空间为    试进行香农编码。  解：以消息*s5*为例来介绍。计算，取整数作为*s5*的码长。计算*s1,s2,s3,s4*的累积分布函数    将0.74变换成二进制小数，取小数点后面三位101作为*s5*的代码。  香农编码是依据香农第一编码定理而来的，有着重要的理论意义。但香农编码的冗余度稍大，实用性不强。比如信源有3个符号，概率分布为（0.5,0.4,0.1），根据香农编码方法求出各个符号的码长对应为1,2,4，码字为（0,10,1110）。下面将看到如果采用霍夫曼编码，可以构造出平均码长更短的即时码（0,10,11）。  7.3.2霍夫曼（Huffman）码  对于某一信源和某一码元集来说，若有一个唯一可译码，其平均长度不大于其他唯一可译码的平均长度，则称此码为最佳码（紧致码）。具有以下性质：   1. 若，则。即概率大的信源符号所对应的码长不大于概率小的信源符号所对应的码长。 2. 对于二元最佳码，两个最小概率的信源符号所对应的码字具有相同的码长。而且这两个码字，除了最后一位码元不同之外，前面各位码元都相同。   1952年霍夫曼提出了一种构造最佳码的方法。所得的码字是即时码，在所有的唯一可译码中，它的平均码长最短，是一种最佳变长码。  二元霍夫曼码的编码步骤如下：   1. 将*q*个信源符号*si*按出现概率递减次序排列起来； 2. 取两个概率最小的符号，其中一个符号编为0，另一个符号编为1；并将这两个概率相加作为一个新符号的概率，从而得到包含（*q*-1）个符号的新信源，称为缩减信源。 3. 把缩减信源中的（*q*-1）个符号重新以概率递减的次序排列，重复步骤（2）。 4. 依次继续下去，直至所有概率相加得到1为止； 5. 从最后一级开始，向前返回，得到各个信源符号所对应的码元序列，即相应的码字。   【例7.5】某离散无记忆信源共有8个符号消息，其概率空间为     1. 计算信源熵*H(S)*和信源的冗余度。 2. 进行霍夫曼编码，并计算编码后的信息传输率、编码效率和码冗余度。   解：（1）信源熵：  信源的冗余度：  （2）     |  |  |  | | --- | --- | --- | | 信源符号 | 码字 | 码长 | |  | 1 | 1 | |  | 001 | 3 | |  | 011 | 3 | |  | 0000 | 4 | |  | 0100 | 4 | |  | 0101 | 4 | |  | 00010 | 5 | |  | 00011 | 5 |   平均码长：  信息传输率：  编码效率：  码冗余度：  **说明：**  1.霍夫曼编码方法得到的码并非唯一。   1. 每次对信源缩减时，赋予信源最后的两个概率最小的符号，用0和1是可以任意的，所以可以得到不同的霍夫曼码，但是不会影响码字的长度。 2. 对信源进行缩减时，两个概率最小的符号合并后的概率与其他信源符号的概率相同时，这两者在缩减信源中进行概率排序，其位置放置次序是可以任意的，故会得到不同的霍夫曼码。这时将影响各码字的长度，但是平均码长相同。一般将合并后的概率放在上面，这样可以获得较小的码方差。   2.需要大量的存储设备来缓冲码字长度的差异，这是码方差小的码质量好的原因。  【**例7.6**】某离散无记忆信源共有5个符号消息，其概率空间为    两种霍夫曼编码如图7.4所示。    图7.4 两种霍夫曼编码  两种码有相同的平均码长和编码效率，但第一种霍夫曼编码的码长方差比第二种霍夫曼编码的码长方差小许多，所以第一种霍夫曼编码的质量较好。  进行霍夫曼编码时，应把合并后的概率总是放在其他相同概率的信源符号之上，以得到码长方差最小的码。  【**例7.7**】设离散无记忆信源的概率空间为，对信源进行N次扩展，采用二元霍夫曼编码。当时的平均码长和编码效率为多少？  解：（1）N=1时，将*s*1编成0，将*s*2编成1，则    又因为信源熵比特/符号  所以编码效率  （2）如果对长度N=2的信源序列进行霍夫曼编码，编码结果如下表所示。   |  |  |  | | --- | --- | --- | | 信源序列 |  | 霍夫曼码 | |  | 0.81 | 1 | |  | 0.09 | 01 | |  | 0.09 | 001 | |  | 0.01 | 000 |   此时，信源序列的平均码长  二元码符号/信源符号序列  则单个符号的平均码长  二元码符号/信源符号  所以对长度为2的信源序列进行变长编码，编码后的编码效率     1. 用同样的方法进一步将信源序列的长度增加，对N=3，N=4的序列进行最佳编码，可得编码效率为   ，  显然，随着信源序列的长度增加，编码效率越来越接近1。  （4）时，由香农第一定理，必然存在唯一可译码，使    而霍夫曼编码为最佳码，即平均码长最短的码，故    即  ***r*元霍夫曼码**  二元霍夫曼编码方法可以推广到r元编码中来。不同的是每次将*r*个概率最小的符号合并成一个新的信源符号，并分别用等码元表示。  为了使短码得到充分利用，平均码长最短，必须使最后一步的缩减信源有*r*个信源符号，因此对于*r*元编码，信源的符号个数*q*必须满足  （7.2）  其中表示缩减的次数，（*r*-1）为每次缩减所减少的信源符号个数。  对于*r*元码，*q*为任意整数时不一定能找到一个满足式（7.2）。若*q*不满足时，不妨人为地增加一些概率为零的符号。设增加*t*个信号源符号  并使它们对应的概率为零，即  设*n=q+t*，此时*n*满足  然后取概率最小的*r*个符号合并成一个新符号，并把这些符号的概率相加作为该节点的概率，重新按概率由大到小顺序排队，再取概率最小的*r*个符号并列；如此下去直至树根。  【例7.8】已知离散无记忆信源    试编出三元霍夫曼码和四元霍夫曼码，并计算编码效率。  解：三元霍夫曼码和四元霍夫曼码如下所示。    三元霍夫曼编码 四元霍夫曼编码  三元霍夫曼编码结果   |  |  |  | | --- | --- | --- | | 信源符号 | 码字 | 码长 | | *S*1 | 0 | 1 | | *S*2 | 1 | 1 | | *S*3 | 20 | 2 | | *S*4 | 21 | 2 | | *S*5 | 22 | 2 |   四元霍夫曼编码结果   |  |  |  | | --- | --- | --- | | 信源符号 | 码字 | 码长 | | *S*1 | 0 | 1 | | *S*2 | 1 | 1 | | *S*3 | 2 | 1 | | *S*4 | 30 | 2 | | *S*5 | 31 | 2 |   **说明**：要发挥霍夫曼编码的优势，一般情况下，信源符号数应远大于码元数。本例中，若编五元码，只能对每个信源符号赋予一个码元，相当于没有编码，当然无压缩可言。  霍夫曼码为最佳码，其平均码长满足    7.3.3费诺（Fano）码  费诺编码属于统计匹配编码。它不是最佳码，但有时也能得到与霍夫曼编码相同的性能。二元费诺编码的步骤如下：   1. 将信源符号按其出现的概率由大到小依次排列； 2. 将依次排列的信源符号按概率值分为两大组，使两个组的概率之近于相同，并对各组分别赋予一个二进制码元“0”和“1”。 3. 将每一大组的信源符号进一步再分成两组，使划分后的两个组的概率之和近于相同，并又分别赋予一个二进制码元“0”和“1”。 4. 如此重复，直至每组只剩下一个信源符号为止。 5. 信源符号所对应的码字即为费诺码。   【**例7.9**】某离散无记忆信源共有8个符号消息，其概率空间为    解：费诺编码如下表7.1所示。  表7.1 费诺编码过程   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 信源  符号 | 概率 | 第一次  分组 | 第二次  分组 | 第三次  分组 | 第四次  分组 | 所得  码字 | 码长 | | *S*1 | 0.4 | 0 | 0 |  |  | 00 | 2 | | *S*2 | 0.18 | 0 | 1 |  |  | 01 | 2 | | *S*3 | 0.10 | 1 | 0 | 0 |  | 100 | 3 | | *S*4 | 0.10 | 1 | 0 | 1 |  | 101 | 3 | | *S*5 | 0.07 | 1 | 1 | 0 | 0 | 1100 | 4 | | *S*6 | 0.06 | 1 | 1 | 0 | 1 | 1101 | 4 | | *S*7 | 0.05 | 1 | 1 | 1 | 0 | 1110 | 4 | | *S*8 | 0.04 | 1 | 1 | 1 | 1 | 1111 | 4 |   7.4 实用编码方法  7.4.1游程编码  游程编码是适用于相关信源的有效编码方法，尤其适用于二元相关信源。游程编码已在图文传真、图像传输等实际工程中得到应用。工程中，游程编码常与其他编码方法混合使用。  游程定义：指字符序列中各种字符连续地重复出现而形成的字符串的长度，又称游程长度或游长。  游程编码：将字符序列映射成串的字符、串的长度和串的位置的标志序列。游程编码适用于一维字符序列，也适用于二维字符序列。  二元相关信源，仅输出0游程和1游程。  *l*(0) —— 0游程长度 *l*(1) —— 1游程长度  二元相关信源的0游程和1游程总是交替出现。游程长度的取值：1,2,3,…，直至无穷值。可规定二元序列总是从0游程开始。游程长度常用自然数标记，所以，游程序列是自然数序列。且这种游程映射是可逆的，是无失真的。  【**例7.10**】某二元序列为：00001100011111100000001…，可以映射成游程序列：42367…；  规定从0 游程开始，由上面的游程序列极易恢复二元序列，也包括序列最后一个1符号。因为*L*(0) = 7的游程后面必定是符号1。  一般为二元数字信道，还需将游程序列变换成二元码字序列，即对游程长度进行二次编码。二次编码可以采用：等长编码或变长编码。本例用等长游程编码，因最大游程长度为 7，三位码元编码。本例码字序列为：100,010,011,110,111,…。也可用变长游程编码，即对游程长度再进行其他变长编码，如Huffman编码，以进一步提高通信效率。需要测定0游程 、1游程长度的概率分布，再设计变长码。一般，0游程和1游程长度分别编码，建立两种码表，且两码表中一般码字是不同的。0 游程与 1 游程的两码表之间的码字不一定要满足非延长码的前缀条件。理论上，游程长度可以从1至无限大。但可采用一定技术，用有限码表实现编码。  给出一种编码方法（设*l*为游程长度）：  （1）选取适当*n*值，将*l=*1,2,…, 2*n*长的2*n*个游程进行Huffman编码，其中2*n*长游程码为*C*。  （2）*l*>2*n*的游程长度用2*n*长游程码 *C*来处理。  方法：将2*n*<*l*<2*n+*1的2*n*个游程用*C*加*n*位的自然码*A*表示。*A*代表余数，用以区分2*n~*2*n+*1之间的不同长度。  如：游程长度  ↔  游程长度  ↔  ……  游程长度  ↔  （3）将 2*n+*1<*l*<2*n+*2的2*n*个游程，就用两个CA为码字。  如：游程长度  ↔  游程长度  ↔  ……  游程长度  ↔  依次类推，可得到所有游程长度的一一对应的码字。0游程和1游程的码表中，为 2n的码字*C*必须不同，且必须与码表中的其他码字，都满足非延长码的前缀条件，以保证译码。  ——0游程、1游程长度的Huffman码的编码效率。由编码效率定义，可得0和1游程的平均码长为  ，  可推导出对应的二元序列的编码效率    假设可以证明  7.4.2数字传真编码  文件传真是指一般文件、图纸、手写稿、表格、报纸等文件的传真，是黑白二值的（二元信源，*r* = 2）。为测定信源的概率分布，CCITT精选8种标准文件样本，如图示。  ccitt1_打字的商业信函（英文）  图7.5 打字的商业信函（英文）  ccitt2_电路图（手绘）  图7.6 电路图（手绘）  ccitt3_印刷和打字的发票（法文）  图7.7 印刷和打字的发票（法文）  ccitt4_密集打字报告（法文）  图7.8 密集打字报告（法文）  ccitt5_包括插图与公式的科技论文（法文）  图7.9 包括插图与公式  ccitt6_带有印刷解说词的图（法文）  图7.10 带有印刷解说词  ccitt7_密集文件（日文假名）  图7.11 密集文件（日文假名）  ccitt8_具有很大黑白字母的手写备忘录（英文）  图7.12 具有很大黑白字母  数字文件传真：将一页文件分成为n×m个像素。像素只有白、黑两种灰度值，用 0,1 编码表示。一页文件的码元数就是该页的二值图像的像素数，这种编码称为直接编码。  分辨率：定义为单位长度（1mm）所包含的像素数。分辨率愈高，细节愈清晰，文件质量愈高，但数据量就愈多。例如，一页A4幅面文件（210mm×297mm），分辨率5样点/ mm。直接编码需传210×297×52≈1.56Mbit，用2.4 kbit/s码率传送约需时间11分钟。  CCITT对选用的8种标准文件样本建议使用以下两种分辨率：  （1）1728像素/行（8样点/mm），3.85 行/ mm；  （2）1728像素/行（8样点/ mm），7.7 行/ mm。  直接表达数字传真文件的数据量非常大，必须进行数据压缩。  MH编码：一维编码方案，即对一行一行的数据进行编码。  特点：将游程编码和Huffman码相结合，是一种标准的改进Huffman码。CCITT的T.4 推荐MH编码为文件传真三类机（G3）一维压缩编码的国际标准。MH编码过程是查表，可实时，易扩展且基本适合中文传真。因每行标准像素：1728 个，样张统计，黑、白游程长度在 0 ~ 63 的情况居多，MH码设计为：终端码（结尾码）和组合码（形成码）两种。  MH码的编码规则如下:  （1）游程长度在 0 ~ 63 时，码字直接用相应的终端码（结尾码）表示。  例：一行中，连续 15 个白，接着连续 30 个黑，即白游程长度为 15，接着黑游程长度为 30。查表得码字为： 110101, 000001101000  表7.2 MH码表（1），终端码（结尾码）   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **RL** | **白游程码** | **黑游程码** | **RL** | **白游程码** | **黑游程码** | **RL** | **白游程码** | **黑游程码字** | | 0 | 00110101 | 0000110111 | 22 | 0000011 | 00000110111 | 44 | 00101101 | 000001010100 | | 1 | 000111 | 010 | 23 | 0000100 | 00000101000 | 45 | 00000100 | 000001010101 | | 2 | 0111 | 11 | 24 | 0101000 | 00000010111 | 46 | 00000101 | 000001010110 | | 3 | 1000 | 10 | 25 | 0101011 | 00000011000 | 47 | 00001010 | 000001010111 | | 4 | 1011 | 011 | 26 | 0010011 | 000011001010 | 48 | 00001011 | 000001100100 | | 5 | 1100 | 0011 | 27 | 0100100 | 000011001011 | 49 | 01010010 | 000001100101 | | 6 | 1110 | 0010 | 28 | 0011000 | 000011001100 | 50 | 01010011 | 000001010010 | | 7 | 1111 | 00011 | 29 | 00000010 | 000011001101 | 51 | 01010100 | 000001010011 | | 8 | 10011 | 000101 | 30 | 00000011 | 000001101000 | 52 | 01010101 | 000000100100 | | 9 | 10100 | 000100 | 31 | 00011010 | 000001101001 | 53 | 00100100 | 000000110111 | | 10 | 00111 | 0000100 | 32 | 00011011 | 000001101010 | 54 | 00100101 | 000000111000 | | 11 | 01000 | 0000101 | 33 | 00010010 | 000001101011 | 55 | 01011000 | 000000100111 | | 12 | 001000 | 0000111 | 34 | 00010011 | 000011010010 | 56 | 01011001 | 000000101000 | | 13 | 000011 | 00000100 | 35 | 00010100 | 000011010011 | 57 | 01011010 | 000001011000 | | 14 | 110100 | 00000111 | 36 | 00010101 | 000011010100 | 58 | 01011011 | 000001011001 | | 15 | 110101 | 000011000 | 37 | 00010110 | 000011010101 | 59 | 01001010 | 000000101011 | | … | … | … | … | … | … | … | … | … |   表7.3 MH 码表（2），组合基干码   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | RL | 白游程码 | 黑游程码 | RL | 白游程码 | 黑游程码 | | 64 | 11011 | 0000001111 | 960 | 011010100 | 0000001110011 | | 128 | 10010 | 000011001000 | 1024 | 011010101 | 0000001110100 | | 192 | 010111 | 000011001001 | 1088 | 011010110 | 0000001110101 | | 256 | 0110111 | 000001011011 | 1152 | 011010111 | 0000001110110 | | 320 | 00110110 | 000000110011 | 1216 | 011011000 | 0000001110111 | | 384 | 00110111 | 000000110100 | 1280 | 011011001 | 0000001010010 | | 448 | 01100100 | 000000110101 | 1344 | 011011010 | 0000001010011 | | 512 | 01100101 | 0000001101100 | 1408 | 011011011 | 0000001010100 | | 576 | 01101000 | 0000001101101 | 1472 | 010011000 | 0000001010101 | | 640 | 01100111 | 0000001001010 | 1536 | 010011001 | 0000001011010 | | 704 | 011001100 | 0000001001011 | 1600 | 010011010 | 0000001011011 | | 768 | 011001101 | 0000001001100 | 1664 | 011000 | 0000001100100 | | 832 | 011010010 | 0000001001101 | 1728 | 010011011 | 0000001100101 | | 896 | 011010011 | 0000001110010 | EOL | 000000000001 | 000000000001 |   表7.4 MH 码表（3），供加大纸宽用的组合基干码 （1792 ~ 2560，黑、白相同）   |  |  |  |  | | --- | --- | --- | --- | | **游程长度** | **组合基干码** | **游程长度** | **组合基干码** | | 1792 | 00000001000 | 2240 | 000000010110 | | 1856 | 00000001100 | 2304 | 000000010111 | | 1920 | 00000001101 | 2368 | 000000011100 | | 1984 | 000000010010 | 2432 | 000000011101 | | 2048 | 000000010011 | 2496 | 000000011110 | | 2112 | 000000010100 | 2560 | 000000011111 | | 2176 | 000000010101 |  |  |   （2）游程长度在 64 ~ 1728，用“组合码 + 终端码”表示。  如：a. 白游程长度 65（= 64+1），表示为白游程长度组合码 64 + 白游程长度终端码1，查表得码字为11011, 000111  b. 黑游程长度 855 = 832 + 23 = 64×13+23，故查表得码字为0000001001101, 00000101000。  （3）规定每行从白游程开始。若实际为黑游程开始，则行首加上零长度白游程码字。每行结束用一个结束码（EOL）。  （4）每页文件开始第一个数据前加一个结束码。每页尾连续使用6个结束码表示结尾。  （5）每行恢复成1728个像素，否则有错。  （6）为实现同步操作，规定T 为每编码行的最小传输时间。一般规定20m <T< 5s。若行传输小于T，则在结束码之前填充0码元（填充码）。    图7.13 传真信息传输格式  【**例7.11**】某页传真文件中某一扫描行的像素点为：16 白，6 黑，57 白，8 黑，1641 白。  该扫描行MH码为：  原一行为 1728 像素，需 1728 位二元码元。现 MH 码则只需用 53 位二元码元。数据压缩比为1728:53=32.6，压缩效率很高。  7.4.3 算术编码  1.积累概率的计算  设信源为：  各符号的积累概率为：  可得：  以 *r* + 1 个点：    图7.14 区间分割  可以完整地分割区间 [0,1)，如图所示。可以发现：  （1）与信源符号建立一一对应关系；  （2）用区间内任一点可作为该对应信源符号的代码；  （3）只要代码长度与概率相匹配，就得到高效编码。  信源符号序列的积累概率递推计算：  信源***X***输出的N长序列为：  输出的可能序列的总数为：  **说明**：算术编码时，一般N比较大，甚至是一个文件的长度。  实际中，很难得到对应信源序列的概率和积累概率，一般从已知的信源符号概率P= (*p*0,*p*1,…,*p*r-1)递推得到。为了简单，先讨论独立二元序列，再推广到一般。  设二元序列：=011，3长二元序列共有8个，按自然二进数排列为：  000、001、010、011、100、101、110、111，相当于。  比如=011对应，其积累概率为    设想扩展序列长为 4，则总序列数为 16 个。其中由=011扩展的两个符号为0110或 0111，按自然二进数排序，在0110前有6个序列，在0111前有7个序列，计算积累概率如下    因二元信源的积累概率为：，，且有 ：可进一步得统一公式：  可以证明对于多元序列，有一般的递推关系    其中，为多元序列，为扩展字符，是字符的发生概率。  2．码区间分割与代码  （1）码区间分割：计算各个值，完成在半开区间 [0,1)上的码区间分割；每个值是分割线，每个小区间的长度是信源符号概率。  （2）代码：小区间内任一点的坐标值，可以代表该信源符号；特别地，用二进数表示，就得到二元编码。类似地，对于字符串，码区间分割用下式进行：    （3）具体的编码过程：  代码长度*l*的计算：，其中，—— 字符串出现的概率，—— 大于等于*x*的最小整数。取积累概率的二进小数前*l*位作为码字，若有尾数，则进位到第*l*位。  【**例7.12**】计算得 = 0.10110001（二进数），= 1/7，则，该字符序列的算术码长为：110（有进位）。  3.码字的唯一性  可证明，上述规则编成的码字能唯一译出所代表的字符序列 。  4.算术码的编码  计算步骤：（待编码信源序列为）  （1）初始化:  （2）输入信源符号，计算:  （3）序列输入是否结束，若未结束，则，转（2）;否则，执行（4）。  （4）计算:，取二进数的形式的前*l*位作为码字，若后面有尾数就进位到第*l*位。  【**例7.13**】设二元独立信源 ：    求信源序列10111101的算术编码。  解：信源符号的概率为：，  累积概率为：，  信源序列算术编码的相关数据如下表所示。得序列10111101的编码为0110001。  表7.5 信源序列 10111101 的算术编码过程   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **序列** |  |  |  | **序列的码字** | |  | **0** | **1** | 0 |  | | 1 | **0.01** | **0.11** | 1 | 1 | | 10 | **0.01** | **0.0011** | 3 | 010 | | 101 | **0.010011** | **0.001001** | 3 | 011 | | 1011 | **0.01010101** | **0.00011011** | 4 | 0110 | | 10111 | **0.0101101111** | **0.0001010001** | 4 | 0110 | | 101111 | **0.011000001101** | **0.000011110011** | 5 | 01101 | | 1011110 | **0.011000001101** | **0.00000011110011** | 7 | 0110001 | | 10111101 | **0.0110000111000011** | **0.0000001011011001** | 7 | 0110001 |   设，表中数据的计算过程如下（黑体数字表示二进小数）：  输入1：  输入0：  输入1：  输入1：  输入1：  输入0：  输入1：  输入1：    图7.15 算术编码过程区间宽度减小图解  【**例7.14**】设四元无记忆信源 ：  求信源序列的算术编码。  解：信源符号的概率为：，，  累积概率为：，  ，  信源序列算术编码的相关数据如下表所示，得序列的编码为 0101110110。  表7.6 信源序列算术编码   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **序列** | *Q*(***α***) | *P*(***α***) | *l* | **序列的码字***w* | |  | **0** | **1** | 0 |  | |  | **0** | **0.1** | 1 | 0 | |  | **0.01** | **0.001** | 3 | 010 | |  | **0.010111** | **0.000001** | 6 | 010111 | |  | **0.010111** | **0.0000001** | 7 | 0101110 | |  | **0.010111011** | **0.0000000001** | 10 | 0101110110 |   设，表中数据的计算过程如下（黑体数字表示二进小数）：  输入：  输入：  输入：  输入：  输入：    图7.16 算术编码过程区间宽度减小图解  5．算术码的译码  算术码的译码过程就是一系列的比较过程。接收到的编码*w*是信源序列累积概率的小数部分，则累积概率数值*C*一定满足：    以二元码为例，给出的译码规则为：若，则译码输出符号为0；若，则译码输出符号为1。  **本章要点**  1.定义 信源编码器  编码器将长度为N的信源符号序列变换成码字    2.性质 即时码存在的充要条件  满足克拉夫特（Kraft）不等式    3.定义 编码后信道的信息传输率R  （比特/码元）  4.定义 编码效率    5.离散平稳无记忆序列的无失真变长编码定理  对N次扩展信源符号序列用*r*进制码元进行变长编码。一定存在一种无失真编码方法，构成唯一可译码，使得平均码长满足    当时，有  6.无失真信源编码方法  香农码，霍夫曼码，费诺码，算术编码，LZW码等。 | | 香农编码最接近香农定理思路  霍夫曼（Huffman）编码是紧致码，应用广泛  提醒同学：信源统计特性的获得方法  代数编码需要自学掌握  总结本章核心知识点 |
| **小结** | 常用及实用的无失真信源编码方法 | |
| **复习要点** | 掌握香农编码，费诺编码，哈夫曼编码，代数编码等 | |
| **思考题** | 给出各种实用编码方法的优缺点及应用场合？ | |
| **作业题** | 7.20 | |

作者签名：

****