**第 10 次课 学时 2**

|  |  |
| --- | --- |
| **授课章节内容** | 第七章无失真信源编码 |
|  | 7.1信源编码器概述 |
| 7.2无失真信源编码定理 |
| **教学目标** | 教学目标5 |
| **支撑毕业要求** | 毕业要求2.1 |

**教学要求：**

1. 知识目标

信源编码器，

无失真信源编码定理

1. 能力目标

通过学习，能够理解信源编码器的本质及无失真信源编码所能达到的极限

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定理的学习，具备理论与实践协调统一的思维

**教学重点与难点**：

掌握信源编码器概念及属性分类；理解无失真信源编码定理

**教学过程设计：**

通过雨课堂提前发布：课前观看视频7.1-7.4预习，课上通过案例导入本次课程

| **讲授与指导内容** | | **讲课、互动内容设计** |
| --- | --- | --- |
| 第七章 无失真信源编码  信源编码包括无失真信源编码和限失真信源编码。两种编码目的都是为了提高通信系统的有效性，即用较少的码率传送同样多的信息。  无失真信源编码，其理论基础是香农第一定理。编码方法亦称为概率匹配编码，即用较短码长的码字代替概率较大的原始信源符号或序列，达到压缩码率的目的，以至于解除或部分消除原始信源符号概率分布的不均匀性或记忆性。  本章在介绍香农第一定理的基础上，给出一些实用的无失真信源编码方法。  7.1 信源编码器概述  信源编码的实质是对原始信源符号按照一定的规则进行变换，以码字代替原始信源符号，使变换后得到的码符号接近等概率分布，从而提高信息传输的有效性。  需要指明的是，在研究信源编码时，通常将信道编码和信道译码看作是信道的一部分，而且不考虑信道干扰问题。  7.1.1信源编码的基本概念  信源编码的数学模型分以下两种。  1.单符号的无失真信源编码器  如图7.1，编码器将信源符号变换成码字，表示为    其中，，*X*为构成码字的码符号集。    图7.1 单符号的无失真信源编码器  编码器的输入为原始信源*S*，样本空间为；编码器输出的码字集合为*C*，共有等*q*个码字，它与*S*中的*q*个信源符号一一对应。  2.N次扩展信源的无失真信源编码器  如图7.2，编码器将长度为N的信源符号序列变换成码字：      图7.2 N次扩展信源的无失真信源编码器  7.1.2信源编码的分类  【**例7.1**】设一个离散无记忆信源的概率空间为    采用两种信源编码方案编出的码字如下表所示。   |  |  |  |  | | --- | --- | --- | --- | | 信源符号 | 概率 | 方案一的码字 | 方案二的码字 | | S1 | 1/8 | 00 | 000 | | S2 | 1/8 | 01 | 001 | | S3 | 1/4 | 10 | 01 | | S4 | 1/2 | 11 | 1 |   试问方案一和方案二的码字哪个有效性较好？   1. 定长码和变长码   如果码字集合*C*中所有码字的码长都相同，称为定长码（或等长码）。反之，如果码字长度不同，则称为变长码。   1. 二元码   如果码元集*X*={0,1}，对应的码字称为二元码。   1. 奇异码和非奇异码   一般说来，无论是定长码还是变长码，如果码字集合*C*中含有相同码字，则称为奇异码。否则称为非奇异码。非奇异性是正确译码的必要条件。   1. 唯一可译码和非唯一可译码   如果一种码的任何一串有限长的码元序列，只能被唯一地译成对应的信源符号序列，则称该码为唯一可译码。否则称为非唯一可译码。为了实现无失真信源编码，必须采用唯一可译码。  例如，{0,10,11}是一种唯一可译码，因为任意一串有限长的码序列，比如10001100，只能被分割为{10,0,0,11,0,0,}，任何其它分割方法都会产生一些非定义的码字。   1. 即时码和非即时码   在唯一可译码中有一类码，它在译码时无需参考后续的码元就能立即做出判断，译成对应的信源符号序列，则这类码称为即时码。否则称为非即时码。即时码不能在一个码字后面添上一些码元构成另一个码字，即任一码字都不是其它码字的前缀，所以它也称为异前缀码。即时码一定是唯一可译码，但唯一可译码不一定是即时码。  非奇异码，唯一可译码和即时码三者之间的关系如图7.3。    图7.3 非奇异码，唯一可译码和即时码三者之间的关系  7.1.3唯一可译码和即时码  即时码可以采用码树法来构造。例如即时码，可以按以下步骤得到即时码的码树。   1. 最上端为树根，从根开始，画出两条分支（树枝），每条分支代表一个码元。因为的码长，任选一条分支的终点（称为节点）来表示。 2. 从没有选用的分支终点再画出两条分支，因为，选用其中的一个分支终点来表示。 3. 继续下去，直到所有的都有分支终点来表示为止。 4. 从顶点（树根）到，要经过条分支，把这些分支所代表的码元依照先后次序就可以写出该码字。   码树还可以用来对即时码进行译码。例如收到一串码字100110010，从码树的树根出发，第一个符号为1，向右走一节；第二个码符号为0，向左走一节，遇到了码字。然后再回到树根，从头开始，遇到了码字后又回到树根。这样就可完成对即时码的即时译码。码字100110010译码得到的码字分别为，，，，如右图7.4所示。  图7.4 对即时码100110010进行译码  对较简单的信源，可以很方便地用码树法直观地构造出即时码。但是当信源较复杂时，直接画码树就比较复杂。1949年L.G.Kraft提出一个在数学上与码树等效的、表达即时码存在充要条件的不等式。  **定理7.1** 对于码长分别为的*r*元码，若此码为即时码，则必定满足  反之，若码长满足上式，则一定存在具有这种码长的*r*元即时码。  克拉夫特（Kraft）不等式是即时码存在的充要条件。其中，*r*为码元的进制数，*q*为信源的符号数，为信源符号对应的码字长度。注意的是，上述不等式只是即时码存在的充要条件，而不能作为判断依据。后来麦克米伦（B.McMillan）证明唯一可译码也满足克拉夫特不等式。这说明在码长选择的条件上，即时码与唯一可译码是一致的。  【**例7.2**】对于二元码，即*r*=2，如果*q*=4，，，，，是否存在这样的唯一可译码和即时码？  解：因为  所以满足克拉夫特不等式，则一定可以构成至少一种具有这样码长的唯一可译码和即时码。  7.1.4编码效率  衡量信源编码的效果是通过以下三种方式。   1. 平均码长   平均码长表示编码后每个信源符号平均所需的码元个数。单位为“码元/信源符号”。   1. 对单个信源符号进行编码   对单个信源符号编码，码字分别为，各码字对应的码长分别为。  则该码的平均码长为  （码元/信源符号）   1. 对N次扩展信源符号进行编码   对长度为N的信源符号序列编码，码字分别为，各码字对应的码长分别为。则对N长的信源符号序列编出的码字平均码长为  （码元/信源符号序列）  所以，信源各符号编码的平均码长为  （码元/信源符号）   1. 编码后信道的信息传输率*R*   编码后信息传输率*R*又称为码率，是指编码后平均每个码元载荷的信息量。单位为“比特/码元”或“比特/码符号”。  当原始信源*S*给定时，信源熵*H(S)*就给定了，而编码后每个信源符号平均用个码元来表示，故编码后信息传输率  （比特/码元）   1. 编码效率   编码效率表示编码后实际信息量和能载荷最大信息量的比值。   1. 定义:每个码元载荷的平均信息量与它所能载荷的最大信息量的比值。      1. 编码效率也可以表示为     其中，称为编码后信源信息率，它表示编码后平均每个码字能载荷的最大信息量。编码效率表征了信源熵和编码后平均每个信源符号能载荷的最大信息量的比值。  7.2 无失真信源编码定理  对N次扩展信源进行等长编码，如果要求编得的等长码是唯一可译码，由于扩展信源符号共有个，则相应的输出码字应不少于个，即必须满足    其中，是等长码的码长，码符号有*r*种可能值，表示长度为的等长码数目。两边取对数，得到    可见对于等长唯一可译码，平均每个信源符号所需的码元个数至少为个。当采用二元码时，即*r=*2时，上式变为。例如英文电报中有32个符号（26个字母加6个字符），即*q*=32，如采用等长码，为了实现无失真信源编码，则每个信源符号至少需要5位二元码符号编码才行。  实际英文电报符号信源，在考虑了符号出现概率以及符号之间的依赖性后，平均每个英文电报符号所能提供的信息量约等于1.4比特，即编码后的5个二元码符号最大能载荷的信息量为5比特。可见，这种编码方式的传输效率极低。  我们提出问题：   1. 能否使每个信源符号所需要的编码符号数减少，以提高传输效率呢？ 2. 最小平均码长为多少时，才能得到无失真的译码？ 3. 若小于这个平均码长是否还能无失真地译码？   这就是无失真信源编码定理要研究的内容。  7.2.1无失真定长信源编码定理  设离散无记忆信源的熵为，其N次扩展信源用个码符号进行定长编码，对于任意，只要满足    则当N足够大时，可使译码错误概率为任意小。  反之，当时，则不可能实现无失真编码，而当N足够大时，译码几乎必定出错。  定理蕴含了如下思想：  （1）定长无失真信源编码的错误概率可以任意小，但并非为零。  （2）定长无失真信源编码通常是对非常长的消息序列进行的，特别是信源符号序列长度Ｎ趋于无穷时，才能实现Shannon意义上的有效信源编码。  为什么不等概信源的每个符号平均所需的码元数可以减少呢？  对不等概信源*S*进行若干次扩展，可以推想，当扩展次数*N*足够长时，则扩展信源中一部分序列出现概率将比其他符号序列出现的概率大得多，整个扩展信源可划分为高概率集合和低概率集合，在一定的允许误差条件下，如果舍弃扩展信源中的低概率集合，而只对高概率集合进行等长编码，这样所需的码元数就可以减少。  定长无失真信源编码定理给出了对信源进行等长编码所需的理论极限值。  由定理可知，当时，可以实现几乎无失真编码。这个不等式的左边表示长为的码元序列所能载荷的最大信息量，而右边代表长度为Ｎ的信源符号序列平均携带的信息量。  定长编码定理表明：只要码字所能载荷的信息量大于信源序列携带的信息量，总可以实现几乎无失真编码。  由可得：编码信息率  可见，信源平均符号熵为一个临界值，只要，这种编码器就可以做到几乎无失真，条件是Ｎ足够大。  编码效率：。  编码定理从理论上阐明了编码效率接近1的理想编码器的存在。  当二元编码时（*r*=2）  编码器容许的输出信息率    编码效率    **译码错误概率**  由于等长编码时，舍弃了扩展信源中的低概率集合，而只对高概率集合进行等长编码，所以会产生译码错误。通过推导可得，译码错误概率为：。式中，  为信源符号的自信息方差，为一个正数。  当和均为定值时，只要信源序列长度Ｎ足够大，可以小于任一正数。即或时，能达到差错率要求。如果取足够小的，就可几乎无差错地译码，而所需的编码信息率不会超过。  【**例7.3**】设一个离散无记忆信源的概率空间为。若采取等长二元编码，要求编码效率，允许译码错误概率，试计算信源序列长度Ｎ？  解：信源熵：  自信息量的方差：  因为编码效率，可得  因此：  所以，信源序列长度达到以上，才能实现给定的要求，这在实际中是很难实现的。因此等长编码没有实际意义，实际中一般都采用变长编码。  7.2.2无失真变长信源编码定理  在变长编码中，码长是变化的，根据信源符号的统计特性，对概率大的符号用短码，而对概率小的符号用较长的码，这样平均码长就可以降低，从而提高编码效率。当平均码长时，编码效率达到1。是否此时的平均码长就是无失真信源编码的最小值？无失真变长信源编码定理将回答这个问题。  **定理7.2（单个符号的无失真变长编码定理）** 一个符号熵为的离散无记忆信源，每个信源符号用*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，其码字平均长度满足    定理说明：码字的平均码长不能小于极限值，否则唯一可译码不存在。可以看出变长编码的极限值和定长编码定理的极限值是一样的。  **定理7.3（离散平稳无记忆序列的无失真变长编码定理）** 一个平均符号熵为  的离散平稳无记忆信源，若对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  定理7.3又称为香农第一定理，前面的定理7.2可以看作是它的特例。香农第一定理的结论同样适用于平稳遍历的有记忆信源。  香农第一定理是香农信息论的主要定理之一。定理指出：要实现无失真的信源编码，每个信源符号的平均码长的极限值就是原始信源的熵值。当编码的平均码长小于信源的熵值，则唯一可译码不存在，在译码时必然带来失真或差错。同时它还表明：通过对扩展信源进行变长编码，当时，平均码长可达到这个极限值。  香农第一定理也可陈述为：若，就存在唯一可译码变长编码；若，唯一可译变长编码不存在，不能实现无失真的信源编码。  编码效率：  二元编码时，编码效率：  由香农第一定理可以看出：当平均码长达到极限值时，编码效率为1。这时编码后的信道信息传输率，即*r*个码符号独立等概分布，达到最大熵。  无失真信源编码的实质就是对离散信源进行适当的变换，使变换后新的码元概率尽可能等概，以使每个码元平均所含的信息量达到最大。  请思考以下问题：   1. 为什么一般不采用定长码，而采用变长码？ 2. 无失真定长信源编码定理（或无失真变长信源编码定理）指出：平均码长的极限值为多少？ 3. 无失真变长信源编码定理指出：当编码效率达到1时，这时编码后的信道信息传输率*R*为多少？此时变换后新的码元概率分布如何？ | | 编码器的本质就是用更短的符号序列代替原始符号序列  码字的属性分析  唯一可译码是难点  克拉夫特（Kraft）不等式是唯一可译码的存在定理  这几个参数是无失真信源编码的核心指标  无失真信源编码定理是香农第一定理，首先从定长信源编码定理讲授  要求学生现场回答这三个问题 |
| **小结** | 无失真信源编码器；香农第一定理 | |
| **复习要点** | 掌握无失真信源编码原理；理解香农第一定理 | |
| **思考题** | 无失真信源编码最短码长上限是多少？ | |
| **作业题** | 7.8 7.11 | |

作者签名：

****