信 息 论 基 础 教 程

李亦农　李梅　编著

北 京 邮 电 大 学 出 版 社
·北 京·

内 容 简 介

信息论是研究信息传输和信息处理过程中的一般规律的一门学科,也是现代信息通信领域的一门基础理论。本书以香农的3个编码定理为中心,重点讲述了相关的基本概念、基本原理和基本方法。本书是作者根据多年教学经验编著而成的。鉴于目前学生可选择的课程越来越多,每门课的学时数不会很大,因此本书只讲述经典香农信息论的内容,而没有涉及过多分支。

图书在版编目(CIP)数据

信息论基础教程/ 李亦农,李梅编著 .—北京:北京邮电大学出版社,2004

ISBN 7-5635-0910-0

Ⅰ.信 ... Ⅱ.①李 ...②李 ... Ⅲ.信息论—高等学校—教材 Ⅳ.G201

中国版本图书馆 CIP 数据核字(2004)第 060781 号

书　　名:信息论基础教程

编　　著:李亦农　李梅

责任编辑:李欣一

出 版 者:北京邮电大学出版社(北京市海淀区西土城路 10 号)

邮编:100876　电话:62282185　62283578

电子信箱:publish @ bupt .edu .cn

经　　销:各地新华书店

印　　刷:北京市忠信诚胶印厂

印　　数:1—5 000 册

开　　本:787 mm×1 092 mm　1/ 16　印张:14 .75　字数:321 千字

版　　次:2005 年 1 月第 1 版　2005 年 1 月第 1 次印刷　2006 年 2 月第 2 次印刷

ISBN 7-5635-0910-0/ TN·331　定　价:22.00 元

前　言

信息论是研究信息传输和信息处理过程中一般规律的一门学科,也是现代信息科学和通信科学领域的一门基础理论,因此目前各高等院校相关专业的本科生、研究生都开有信息论这门专业基础课。

本书是作者参考借鉴了众多国内外优秀的信息论教材及参考书之后根据多年教学实践的经验编著而成的。鉴于目前学生可选择的课程越来越多,而每门课的学时数较少,因此本书只讲述了经典香农信息论的内容,而没有涉及过多的分支。本书可作为通信、电子、计算机专业本科生教材使用,也可作为相关专业科研人员的参考书。

在本书以香农的 3 个编码定理为中心,重点讲述了相关的基本概念、基本原理和基本方法,力图以读者最易接受的方式介绍信息论的基本内容及应用。

本书一共分为 7 章,外加两个附录。第 1 章绪论主要介绍香农信息论的研究对象、目的和内容等;第 2 章介绍关于信息度量的几个重要概念:自信息、互信息、信息熵及平均互信息以及数据处理定理;第 3 章研究定量度量信源产生信息的能力和度量信源冗余度的问题;第 4 章研究定量描述信道传递信息能力的问题,并介绍了几种特殊信道的信道容量的计算方法;第 5 章的核心内容是香农的无失真信源编码定理,围绕这个定理介绍了无失真信源编码的基本概念,讲解了定理的基本内容及其证明,另外还讲述了几种实用的无失真信源编码算法;第 6 章讲述香农的有噪信道编码定理以及纠错编码,涉及信道编码的基本概念、基本技巧和基本理论知识,同时较为详细地讲述了纠错编码的有关知识;第 7 章介绍香农的第三编码定理——限失真信源编码定理,引入了

信息率失真函数的概念并介绍了信息率失真函数的性质以及计算方法,最后介绍了几种常用的熵压缩编码算法;附录 A 为上机作业;附录 B 是学习本课程需要用到的一些数学预备知识。

第 1～4 章,6.3、6.4 节以及附录 B.1～B.4 由李梅编写,其余各章节由李亦农编写,全书由李梅统稿。

在本书的编写过程中,参阅了国内外一些经典著作(列于本书参考文献中),同时参考了很多国外知名大学信息论基础课程的课后习题及解答,在此向有关作者表示感谢!在本书编写的过程中,同时得到了陈剑、李洪、徐益民、唐晓晟等同志的大力支持,在此也表示感谢!

由于作者能力所限,难免有疏漏之处,敬请读者赐教。

作 者

2004 年 5 月

/

目　录

第 1 章　绪　论

1.1　信息的概念 …… 1
1.2　信息论的研究对象、目的和内容 …… 3

第 2 章　信息的度量

2.1　自信息和互信息 …… 7
2.1.1　自信息 …… 7
2.1.2　互信息 …… 9
2.2　平均自信息 …… 10
2.2.1　平均自信息的概念 …… 10
2.2.2　熵函数的性质 …… 11
2.2.3　联合熵与条件熵 …… 15
2.3　平均互信息 …… 19
2.3.1　平均互信息的概念 …… 19
2.3.2　平均互信息的性质 …… 20
2.3.3　数据处理定理 …… 24
习　题　2 …… 25

第 3 章　信源及信源熵

3.1　信源的分类及其数学模型 …… 29
3.2　离散单符号信源 …… 30
3.3　离散多符号信源 …… 31
3.3.1　离散平稳无记忆信源 …… 32
3.3.2　离散平稳有记忆信源 …… 33

3.3.3 马尔可夫信源 …… 36
3.3.4 信源的相关性和剩余度 …… 40
*3.4 连续信源 …… 44
3.4.1 连续信源的微分熵 …… 44
3.4.2 连续信源的最大熵 …… 48
3.4.3 连续信源的熵功率 …… 49
习 题 3 …… 50

第4章 信道及信道容量

4.1 信道的分类 …… 55
4.2 离散单符号信道及其信道容量 …… 56
4.2.1 离散单符号信道的数学模型 …… 56
4.2.2 信道容量的概念 …… 58
4.2.3 几种特殊信道的信道容量 …… 59
4.2.4 离散对称信道的信道容量 …… 61
4.2.5 一般离散信道的信道容量 …… 65
4.2.6 信道容量定理 …… 68
*4.2.7 信道容量的迭代算法 …… 72
4.3 离散多符号信道及其信道容量 …… 75
4.4 组合信道及其信道容量 …… 78
4.4.1 独立并联信道 …… 78
4.4.2 级联信道 …… 79
*4.5 连续信道及其信道容量 …… 80
4.5.1 连续随机变量的互信息 …… 80
4.5.2 高斯加性信道的信道容量 …… 81
4.5.3 多维高斯加性信道的信道容量 …… 83
*4.6 波形信道及其信道容量 …… 84
习 题 4 …… 85

第5章 无失真信源编码

5.1 信源编码的相关概念 …… 89
5.1.1 编 码 器 …… 89
5.1.2 码的分类 …… 91
5.2 定长码及定长编码定理 …… 94

5.3　变长码及变长编码定理 …… 99
5.3.1　Kraft 不等式和 McMillan 不等式 …… 99
5.3.2　唯一可译码的判别准则 …… 102
5.3.3　无失真变长编码定理 …… 103
5.3.4　香农第一编码定理 …… 106
5.4　变长码的编码方法 …… 109
5.4.1　香农编码 …… 110
5.4.2　香农-费诺-埃利斯编码 …… 111
5.4.3　霍夫曼编码 …… 111
5.4.4　r 元霍夫曼编码 …… 115
5.4.5　费诺编码 …… 115
5.5　实用的无失真信源编码方法 …… 117
5.5.1　游程编码 …… 117
5.5.2　算术编码 …… 119
5.5.3　LZW 码 …… 120
习　题　5 …… 123

第 6 章　有噪信道编码

6.1　信道编码的相关概念 …… 127
6.1.1　错误概率和译码规则 …… 128
6.1.2　错误概率与编码方法 …… 134
6.2　有噪信道编码定理 …… 141
6.2.1　有噪信道编码定理 …… 141
6.2.2　有噪信道编码逆定理 …… 144
6.2.3　错误概率的上界 …… 145
6.3　纠错编码 …… 147
6.3.1　纠错码分类 …… 147
6.3.2　纠错码的基本概念 …… 149
*6.4　几种重要的纠错码 …… 150
6.4.1　线性分组码 …… 150
6.4.2　汉明码 …… 163
6.4.3　循环码 …… 165
6.4.4　卷积码 …… 170
习　题　6 …… 173

第 7 章 限失真信源编码

7.1 失真测度 …… 177
7.1.1 失真函数 …… 177
7.1.2 平均失真 …… 179
7.2 信息率失真函数 …… 181
7.2.1 D 失真许可信道 …… 181
7.2.2 信息率失真函数的定义 …… 181
7.2.3 信息率失真函数 $R(D)$的性质 …… 182
7.3 信息率失真函数的计算 …… 188
7.3.1 应用参量表示式计算 $R(D)$ …… 188
7.3.2 二元信源和离散等概信源的 $R(D)$函数 …… 195
7.4 限失真信源编码定理和逆定理 …… 198
7.4.1 限失真信源编码定理 …… 198
7.4.2 限失真信源编码逆定理 …… 202
7.5 熵压缩编码具体方法 …… 204
7.5.1 标量量化 …… 204
7.5.2 矢量量化 …… 205
7.5.3 变换编码 …… 206
*7.5.4 预测编码 …… 206
习 题 7 …… 207

附录 A 上机作业

A.1 信道容量的迭代计算 …… 209
A.2 唯一可译码判决准则 …… 210
A.3 Huffman 编码 …… 211
A.4 LZW 编码 …… 211

附录 B 数学预备知识

B.1 概率论简单回顾 …… 212
B.2 Jensen 不等式 …… 213
B.3 马尔可夫链 …… 214
B.4 信道容量定理的引理 …… 218
B.5 契比雪夫不等式 …… 219

B.6 大数定理 …… 219
B.7 渐进等同分割性和 ε 典型序列 …… 220
B.8 n 维欧式空间 …… 224

参考文献 …… 226

第 1 章

绪　　论

克劳德·艾尔伍德·香农(Claude Elwood Shannon ,1916～2001 年)美国数学家,信息论的创始人 .1948 年香农在《贝尔系统技术杂志》(Bell System Technical Journal)上连载发表了著名的论文《通讯的数学原理》.1949 年,香农又在该杂志上发表了另一影响深远的论文《噪声下的通信》.在这两篇论文中,香农阐明了通信的基本问题,给出了通信系统的模型,提出了信息量的数学表达式,并解决了信道容量、信源统计特性、信源编码、信道编码等一系列基本问题.这两篇论文成为信息论的奠基性著作 .

1.1　信息的概念

信息论是通信的数学基础,它是随着通信技术的发展而形成和发展起来的一门新兴的横断学科 .

信息论创立的标志是 1948 年香农发表的论文“A Mathematical Theory of Communi-

cation”.为了解决在噪声信道中有效传输信息的问题,香农在这篇文章中创造性地采用概率论的方法来研究通信中的问题,并且对信息给予了科学的定量描述,第一次提出了信息熵的概念.

在日常生活中,人们往往对消息和信息不加区别,消息被认为就是信息.例如,当人们收到一封电报,或者听了天气预报,人们就说得到了信息.

人们收到消息后,如果消息告诉了我们很多原来不知道的新内容,我们会感到获得了很多的信息,而如果消息是我们基本已经知道的内容,我们得到的信息就不多,所以信息应该是可以度量的.那么怎样度量信息呢?人们需要有一个可以用数学模型来表示的信息概念.

1928 年,哈特莱(Hartley)首先提出了对数度量信息的概念,即一个消息所含有的信息量用它的所有可能的取值的个数的对数来表示.比如,抛掷一枚硬币可能有两种结果:正面和反面,所以当我们得知抛掷结果后获得的信息量是 $\log_2 2 = 1$ bit.而一个十进制数字可以表示 0～9 中的任意一个符号,所以一个十进制数字含有 $\log_2 10 = 3.321\ 9$ bit 的信息量.这里对数取以 2 为底,信息量的单位为 bit.

哈特莱的工作给了香农很大的启示,他进一步注意到消息的信息量不仅与它的可能值的个数有关,还与消息本身的不确定性有关.例如,抛掷一枚偏畸硬币,如果正面向上的可能性是 90%,那么当我们得知抛掷结果是反面时得到的信息量会比得知抛掷结果是正面时得到的信息量大.

一个消息之所以会含有信息,正是因为它具有不确定性,一个不具有不确定性的消息是不会含有任何信息的,而通信的目的就是为了消除或部分消除这种不确定性.比如,在得知硬币的抛掷结果前,我们对于结果会出现正面还是反面是不确定的,通过通信,我们得知了硬币的抛掷结果,消除了不确定性,从而获得了信息.因此,信息是对事物运动状态或存在方式的不确定性的描述.这就是香农信息的定义.

用数学的语言来讲,不确定性就是随机性,具有不确定性的事件就是随机事件.因此,可运用研究随机事件的数学工具——概率——来测度不确定性的大小.在信息论中,我们把消息用随机事件表示,而发出这些消息的信源则用随机变量来表示.比如,抛掷一枚硬币的试验可以用一个随机变量来表示,而抛掷结果可以是正面或反面,这个具体的消息则用随机事件表示.

我们把某个消息 x_i 出现的不确定性的大小定义为自信息,用这个消息出现的概率的对数的负值来表示:

$$I(x_i) = -\log_2 p(x_i) \tag{1.1}$$

自信息同时表示这个消息所包含的信息量,也就是最大能够给予收信者的信息量.如果消息能够正确传送,收信者就能够获得这么大小的信息量.

信源所含有的信息量定义为信源发出的所有可能消息的平均不确定性,香农把信源所含有的信息量称为信息熵.信息熵定义为自信息的统计平均,即

$$H(X) = -\sum_{i=1}^{q} p(x_i)\log_2 p(x_i) \tag{1.2}$$

这里的 q 表示信源消息的个数.信息熵表示信源的平均不确定性的大小,同时表示信源输出的消息平均所含的信息量.因此,虽然信源产生的消息可能会含有不同的信息量,比如抛掷一枚偏畸硬币的结果是正面和是反面这两个消息所含的信息量不同,但是可以用它们的平均值来表示这个信源(抛掷一枚偏畸硬币的试验)的平均不确定性.

在收信端,信源的不确定性得到了部分或全部的消除,收信者就得到了信息.信息在数量上等于通信前后"不确定性"的消除量(减少量).

这种建立在概率模型上的信息概念排除了日常生活中"信息"一词主观上的含义和作用,而只是对消息的统计特性的定量描述,所以信息可以度量,而且与日常生活中信息的概念并不矛盾,因此是一个科学的定义.根据这样的信息定义,同样一个消息对于任何一个收信者来说,所含有的信息量都是一样的.而事实上信息有很强的主观性和实用性,同样一个消息对不同的人常常有不同的主观价值或主观意义.例如,同一则气象预报对在室外工作的人和在室内工作的人可能会有不同的意义和价值,因此所提供的信息量也应该不同.所以香农信息的定义在某些情况下也具有一定的局限性.

1.2　信息论的研究对象、目的和内容

信息论从诞生到现在,虽然只有短短的 50 多年,但它的发展对学术界及人类社会的影响是相当广泛和深刻的.如今,信息论的研究内容不仅仅包括通信,而且包括所有与信息有关的自然和社会领域,如模式识别、机器翻译、心理学、遗传学、神经生理学、语言学、语义学甚至包括社会学中有关信息的问题.香农信息论迅速发展成为涉及范围极广的广义信息论——信息科学.

信息论的研究对象是广义的通信系统,它把所有的信息流通系统都抽象成一个统一的模型,如图 1.1 所示.

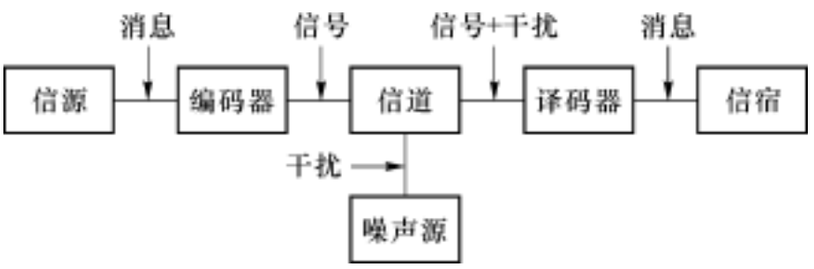


图 1.1　通信系统模型

这个模型不仅包括电话、电报、传真、电视、雷达等狭义的通信系统,还包括生物有机体的遗传系统、神经系统、视觉系统,甚至人类社会的管理系统.信息以消息的形式在这个

通信系统中传递,人们通过研究通信系统中消息的传输和处理来得到信息传输和处理的规律,目的是提高通信的可靠性和有效性 .

在任何一个信息流通系统中,都有一个发出信息的发送端(信源),有一个接收信息的接收端(信宿),以及信息流通的通道(信道) .在信息传递的过程中不可避免的会有噪声,所以有一个噪声源 为了把信源发出的消息变成适合在信道中传输的信号,还需要加入编码器,在送到信宿之前要进行反变换,所以要加入译码器 .

这个通信系统主要分成 5 个部分:

(1) 信源

顾名思义,信源是产生消息和消息序列的源 .信源可以是人、生物、机器或其他事物 .比如,各种气象状态是信源,能够产生独特的气味吸引蜜蜂来采花蜜的花朵是信源,人脑的思维活动也是一种信源 .信源的输出是消息(或消息序列) .

消息有着各种不同的形式,例如:文字、符号、语言、图片、图像、气味等等 .消息以能被通信双方所理解的形式,通过通信进行传递和交换 .消息携带着信息,是信息的载体 .信源输出的消息是随机的、不确定的,但又有一定的规律性,因此用随机变量或随机矢量等数学模型来表示信源 .

(2) 编码器

编码就是把消息变成适合在信道传输的物理量,这种物理量称为信号(如电信号、光信号、声信号、生物信号等等) .信号携带着消息,它是消息的载体 .

编码器可分为信源编码器和信道编码器 .信源编码的目的是压缩信源的冗余度(即多余度),提高信息传输的效率,这是为了提高通信系统的有效性 .信源编码又可分为无失真信源编码和限失真信源编码 .信道编码是为了提高信息传输的可靠性而有目的地对信源编码器输出的代码组添加一些监督码元,使之具有纠、检错能力 比如,老师讲课需要把知识进行加工和提炼,以提高信息传输的有效性,而为了让学生听得明白,有时又需要适当地重复,这是为了提高信息传输的可靠性 .

在实际的通信系统中,可靠性和有效性常常是相互矛盾的,提高有效性必须去掉信源符号的冗余部分,但是这会导致可靠性的下降,而提高可靠性就需要增加监督码元,这又降低了有效性 .有时为了兼顾有效性,就不一定要求绝对准确地在接收端再现原来的消息,而是可以允许一定的误差或失真,也就是说允许近似地再现原来的消息 .

(3) 信道

信道是指通信系统把载荷消息的信号从发送端送到接收端的媒介或通道,是包括收发设备在内的物理设施 .信道除了传播信号以外,还有存储信号的作用 .在狭义的通信系统中,实际信道有明线、电缆、光缆、无线电波传播空间、磁盘、光盘等,这些都属于传输电磁波能量的信道 对于广义的通信系统来说,信道还可以是其他的传输媒介 .

在信道中引入噪声和干扰,这是一种简化的表达方式 .为了分析方便起见,把在系统其他部分产生的干扰和噪声都等效地折合成信道干扰,看成是由一个噪声源产生的,它将

作用于所传输的信号上.这样,信道输出的已是叠加了干扰的信号.噪声源的统计特性是划分信道的依据,并且是信道传输能力的决定因素.由于干扰或噪声往往具有随机性,所以信道用输入和输出之间的条件概率分布来描述.

(4) 译码器

译码就是把信道输出的已叠加了干扰的编码信号进行反变换,变成信宿能够理解的消息.译码器也可分成信源译码器和信道译码器.译码器需要尽可能准确地再现信源输出的消息.

(5) 信宿

信宿是消息传送的对象,即接受消息的人、机器或其他事物.

以上我们考虑的是收发两端单向通信的情况,它只有一个信源和一个信宿,信息传输也是单向的.在组网通信的情况下(比如电话网、计算机网等等),可能有很多分开的信源、信道和信宿同时进行信息交换.例如,广播信道是一个输入、多个输出的单向信道,而卫星通信则是多个输入、多个输出的多向传输的通信,这就需要把两端单向通信的模型做适当的修正,得出多用户通信系统的模型,把两端单向通信的信息理论发展成为多用户通信信息理论.

信息论研究的是关于这个通信系统的最根本、最本质的问题.例如:

① 什么是信息?如何度量信息?

② 怎样确定信源的输出中含有多少信息量?

③ 对于一个信道,它传输信息量的最高极限(信道容量)是多少?

④ 为了能够无失真地传输信源信息,对信源编码时所需的最少的码符号数是多少?这是无失真信源编码,即香农第一定理的内容.

⑤ 在有噪信道中有没有可能以接近信道容量的信息传输率传输信息而错误概率几乎为零?这是有噪信道编码,即香农第二定理的内容.

⑥ 如果对信源编码时允许一定量的失真,所需的最少的码符号数又是多少?这是限失真信源编码,即香农第三定理的内容.

毫无疑问,如果我们对这些问题都有了确定的答案,那么在设计通信系统时就有了目标和指导方向,同时也有了评价通信系统优劣的标准.

在这里,我们举几个成功地应用信息论的概念和方法指导通信系统设计的例子.

(1) 无失真信源编码的应用:计算机文件的压缩

由于数据库的广泛应用,存储计算机文件所需的存储量问题日益突出.在过去的 20 多年中,至少已有 20 种不同的对计算机文件的压缩算法问世,其中较好的算法能使文件压缩后所需的存储量只为原文件的 30%左右.

(2) 有噪信道编码的应用:模拟话路中数据传输速率的提高

最早的调制解调器速率只有 300 bit/ s,此后,调制解调器的速率从 4 800 bit/ s 到 33.6 kbit/ s,已经非常接近于理论极限(56 K 调制解调器由于下行只经过一次模/ 数转换,

所以下行速率更快一些) .

(3) 限失真信源编码的应用:语音信号压缩

按照信息理论的分析,语音信号(也就是话音信号)所需的编码速率可以远远低于按 Nyquist 采样定理和量化噪声理论所确定的编码速率 .几十年来,人们在这方面的工作取得了巨大的进展 .CCITT 关于长途电话网的语音编码速率标准已从 1972 年的64 kbit/ s降低到 1992 年的 16 kbit/ s .在移动通信中,1988 年欧洲 GSM 标准中的语音编码速率为 13 2 kbit/ s,而 1989 年美国 CTIA 标准中的语音编码速率仅为 7 .95 kbit/ s .目前,声码器的速率可低于 100 bit/ s,已接近信息论指出的极限 .

目前,对信息论的研究范围一般有 3 种理解:

(1) 狭义信息论:又称香农信息论 .主要通过数学描述与定量分析,研究通信系统从信源到信宿的全过程,包括信息的测度、信道容量以及信源和信道编码理论等问题,强调通过编码和译码使收、发两端联合最优化,并且以定理的形式证明极限的存在 这部分内容是信息论的基础理论 .

(2) 一般信息论:也称工程信息论 .主要也是研究信息传输和处理问题,除香农信息论的内容外,还包括噪声理论、信号滤波和预测、统计检测和估计、调制理论、信息处理理论以及保密理论等 .

(3) 广义信息论:也称信息科学,不仅包括上述两方面内容,而且包括所有与信息有关的自然和社会科学领域,如模式识别、机器翻译、心理学、遗传学、神经生理学、语言学、语义学甚至包括社会学中有关信息的问题 .

本课程主要研究香农信息论的内容 .

第 2 章

信息的度量

关于信息的度量有几个重要的概念:

(1) 自信息(量):一个事件(消息)本身所包含的信息量,它是由事件的不确定性决定的,比如抛掷一枚硬币的结果是正面这个消息所包含的信息量 .

(2) 互信息(量):一个事件所给出关于另一个事件的信息量,比如今天下雨所给出关于明天下雨的信息量 .

(3) 平均自信息(量),或称信息熵:事件集(用随机变量表示)所包含的平均信息量,它表示信源的平均不确定性,比如抛掷一枚硬币的试验所包含的平均信息量 .

(4) 平均互信息(量):一个事件集所给出关于另一个事件集的平均信息量,比如今天的天气所给出关于明天的天气的信息量 .

我们在最简单的离散随机变量的情况下引入这些概念 .

2 .1　自信息和互信息

2 .1 .1　自 信 息

在绪论中我们讲过,信源发出的消息(事件)具有不确定性,而事件发生的不确定性与

事件发生的概率大小有关,概率越小,不确定性越大,事件发生以后所含有的信息量就越大.小概率事件,不确定性大,一旦出现必然使人感到意外,因此产生的信息量就大,特别是几乎不可能出现的事件一旦出现,必然产生极大的信息量;大概率事件,是预料之中的事件,不确定性小,即使发生,也没什么信息量,特别是概率为 1 的确定事件发生以后,不会给人以任何信息量.因此随机事件的自信息量 $I(x_i)$ 是该事件发生概率 $p(x_i)$ 的函数,并且 $I(x_i)$ 应该满足以下公理化条件:

(1) $I(x_i)$ 是 $p(x_i)$ 的严格递减函数.当 $p(x_1) < p(x_2)$ 时,$I(x_1) > I(x_2)$,概率越小,事件发生的不确定性越大,事件发生以后所包含的自信息量越大.

(2) 极限情况下,当 $p(x_i) = 0$ 时,$I(x_i) \to \infty$;当 $p(x_i) = 1$ 时,$I(x_i) = 0$.

(3) 从直观概念上讲,由两个相对独立的不同的消息所提供的信息量应等于它们分别提供的信息量之和,即自信息量满足可加性.

可以证明,满足以上公理化条件的函数形式是对数形式.

定义 2.1 随机事件的自信息量定义为该事件发生概率的对数的负值.设事件 x_i 的概率为 $p(x_i)$,则它的自信息量定义为

$$I(x_i) = -\log_2 p(x_i) = \log_2 \frac{1}{p(x_i)} \tag{2.1}$$

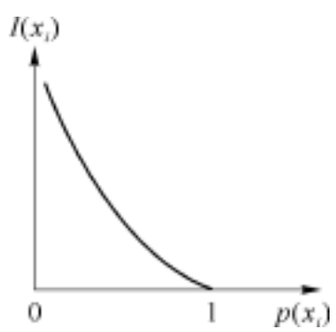


图 2.1 自信息量

从图 2.1 中可以看到上述自信息量的定义正是满足公理性条件的函数形式.在它的定义域[0,1]内,自信息量是非负的.

$I(x_i)$ 代表两种含义:在事件 x_i 发生以前,等于事件 x_i 发生的不确定性的大小;在事件 x_i 发生以后,表示事件 x_i 所含有或所能提供的信息量.在无噪信道中,事件 x_i 发生以后,能正确无误地传输到收信者,所以 $I(x_i)$ 就等于收信者接受到 x_i 后所获得的信息量.这是因为消除了 $I(x_i)$ 大小的不确定性,才获得这么大小的信息量.

自信息量的单位与所用对数的底有关.

(1) 通常取对数的底为 2,信息量的单位为比特(bit, binary unit).当 $p(x_i) = 1/2$ 时,$I(x_i) = 1$ bit,即概率等于 1/2 的事件具有 1 bit 的自信息量.例如,一枚均匀硬币的任何一种抛掷结果均含有 1 bit 的信息量.比特是信息论中最常用的信息量单位,当取对数的底为 2 时,2 常省略.注意:计算机术语中 bit 是位的单位(bit, binary digit),与信息量单位不同,但有联系,1 位的二进制数字最大能提供 1 bit 的信息量.

(2) 若取自然对数(以 e 为底),自信息量的单位为奈特(nat, natural unit).理论推导中或用于连续信源时用以 e 为底的对数比较方便.

$$1 \text{ nat} = \log_2 e \text{ bit} = 1.443 \text{ bit}$$

(3) 工程上用以 10 为底较方便.若以 10 为对数底,则自信息量的单位为哈特莱

(Hartley),用来纪念哈特莱首先提出用对数来度量信息.

$$1\ \text{Hartley} = \log_2 10\ \text{bit} = 3.322\ \text{bit}$$

(4) 如果取以 r 为底的对数($r>1$),则 $I(x_i) = -\log_r p(x_i)$　r 进制单位

$$1\ r\ \text{进制单位} = \log_2 r\ \text{bit}$$

【例 2.1】

(1) 英文字母中“a”出现的概率为 0.064,“c”出现的概率为 0.022,分别计算它们的自信息量.

(2) 假定前后字母出现是互相独立的,计算“ac”的自信息量.

(3) 假定前后字母出现不是互相独立的,当“a”出现以后,“c”出现的概率为 0.04,计算“a”出现以后,“c”出现的自信息量.

解

(1) $I(\text{a}) = -\log_2 0.064 = 3.96\ \text{bit}$

$I(\text{c}) = -\log_2 0.022 = 5.51\ \text{bit}$

(2) 由于前后字母出现是互相独立的,“ac”出现的概率为 0.064×0.022,所以

$I(\text{ac}) = -\log_2(0.064\times0.022) = -(\log_2 0.064 + \log_2 0.022) = I(\text{a}) + I(\text{c}) = 9.47\ \text{bit}$

即两个相对独立的事件的自信息量满足可加性,也就是由两个相对独立的事件的积事件所提供的信息量应等于它们分别提供的信息量之和.

(3)“a”出现的条件下,“c”出现的概率变大,它的不确定性变小.

$$I(\text{c}|\text{a}) = -\log_2 0.04 = 4.64\ \text{bit}$$

2.1.2　互信息

定义 2.2　一个事件 y_j 所给出关于另一个事件 x_i 的信息定义为互信息,用 $I(x_i;y_j)$ 表示.

$$I(x_i;y_j) = I(x_i) - I(x_i|y_j) = \log_2 \frac{p(x_i|y_j)}{p(x_i)} \tag{2.2}$$

互信息 $I(x_i;y_j)$ 是已知事件 y_j 后所消除的关于事件 x_i 的不确定性,它等于事件 x_i 本身的不确定性 $I(x_i)$ 减去已知事件 y_j 后对 x_i 仍然存在的不确定性 $I(x_i|y_j)$.互信息的引出,使信息的传递得到了定量的表示.

【例 2.2】

某地二月份天气出现的概率分别为晴 1/2,阴 1/4,雨 1/8,雪 1/8.某一天有人告诉你:“今天不是晴天”,把这句话作为收到的消息 y_1,求收到 y_1 后,y_1 与各种天气的互信息量.

解

把各种天气记作 x_1(晴),x_2(阴),x_3(雨),x_4(雪).收到消息 y_1 后,各种天气发生的

概率变成了后验概率:

$$p(x_1 \mid y_1) = \frac{p(x_1 y_1)}{p(y_1)} = 0$$

$$p(x_2 \mid y_1) = \frac{p(x_2 y_1)}{p(y_1)} = \frac{1/4}{1/4 + 1/8 + 1/8} = \frac{1}{2}$$

$$p(x_3 \mid y_1) = \frac{p(x_3 y_1)}{p(y_1)} = \frac{1/8}{1/4 + 1/8 + 1/8} = \frac{1}{4}$$

同理

$$p(x_4 \mid y_1) = \frac{1}{4}$$

根据互信息量的定义,可计算出 y_1 与各种天气之间的互信息:

$$I(x_1; y_1) = \log_2 \frac{p(x_1 \mid y_1)}{p(x_1)} = \infty$$

$$I(x_2; y_1) = \log_2 \frac{p(x_2 \mid y_1)}{p(x_2)} = \log_2 \frac{1/2}{1/4} = 1 \text{ bit}$$

$$I(x_3; y_1) = \log_2 \frac{p(x_3 \mid y_1)}{p(x_3)} = \log_2 \frac{1/4}{1/8} = 1 \text{ bit}$$

$$I(x_4; y_1) = \log_2 \frac{p(x_4 \mid y_1)}{p(x_4)} = \log_2 \frac{1/4}{1/8} = 1 \text{ bit}$$

2.2 平均自信息

2.2.1 平均自信息的概念

自信息量是信源发出某一具体消息所含有的信息量,发出的消息不同它的自信息量就不同,所以有信息量本身为随机变量,不能用来表征整个信源的不确定度.我们用平均自信息量来表征整个信源的不确定度.平均自信息量又称为信息熵、信源熵,简称熵.

因为信源具有不确定性,所以把信源用随机变量来表示,用随机变量的概率分布来描述信源的不确定性.通常把一个随机变量的所有可能的取值和这些取值对应的概率[X, $P(X)$]称为它的概率空间.

假设随机变量 X 有 q 个可能的取值 x_i, $i = 1, 2, \cdots, q$,各种取值出现的概率为 $p(x_i)$, $i = 1, 2, \cdots, q$,它的概率空间表示为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} X = x_1 & \cdots & X = x_i & \cdots & X = x_q \\ p(x_1) & \cdots & p(x_i) & \cdots & p(x_q) \end{bmatrix}$$

这里要注意，$p(x_i)$满足概率空间的基本特性：非负性 $0 \leq p(x_i) \leq 1$ 和完备性 $\sum_{i=1}^{q} p(x_i) = 1$.

定义 2.3　随机变量 X 的每一个可能取值的自信息 $I(x_i)$ 的统计平均值定义为随机变量 X 的平均自信息量 .

$$H(X) = E[I(x_i)] = -\sum_{i=1}^{q} p(x_i)\log_2 p(x_i) \tag{2.3}$$

这里 q 为 X 的所有可能取值的个数 .

熵的单位也是与所取的对数底有关，根据所取的对数底不同，可以是比特/ 符号、奈特/ 符号、哈特莱/ 符号或者是 r 进制单位/ 符号，通常用比特/ 符号为单位 .

熵这个名词是香农从物理学中热熵的概念借用过来的，热熵是表示分子混乱程度的一个物理量，因此，香农用熵来描述信源的平均不确定性 .但是在热力学中任何孤立系统的演化，热熵只能增加不能减少，而在信息论中，信息熵正相反，只会减少，不会增加，所以有人称信息熵为负热熵 .

信息熵 $H(X)$ 是对信源的平均不确定性的描述 .在第 5 章无失真信源编码定理和它的逆定理会进一步证明，要对信源输出的消息进行无失真的编码，平均每个信源符号至少需要用 $H(X)$ 个码符号 .

一般情况下，信息熵并不等于收信者平均获得的信息量 .只有在无噪情况下，收信者才能正确无误地接收到信源所发出的消息，全部消除了 $H(X)$ 大小的平均不确定性，所以获得的平均信息量就等于 $H(X)$，而一般情况下，因为干扰和噪声的存在，收信者不能全部消除信源的平均不确定性，获得的信息量将小于信息熵 .

【例 2.3】

假设随机变量 X 的概率分布为 $p(x_i) = 2^{-i}$，$i = 1, 2, 3, \cdots$，求 $H(X)$.

解

$$H(X) = \sum_{i=1}^{\infty} 2^{-i}\log_2 \frac{1}{2^{-i}} = \sum_{i=1}^{\infty} i2^{-i} = 2 \text{ 比特/ 符号}$$

2.2.2　熵函数的性质

信息熵 $H(X)$ 是随机变量 X 的概率分布的函数，所以又称为熵函数 .如果把概率分布 $p(x_i)$，$i = 1, 2, \cdots, q$，记为 $p_1, p_2, \cdots, p_q$，则熵函数又可以写成概率矢量 $p = [p_1, p_2, \cdots, p_q]$ 的函数形式，记为 $H(p)$.

$$H(X) = -\sum_{i=1}^{q} p_i \log_2 p_i = H[p_1, p_2, \cdots, p_q] = H(p) \tag{2.4}$$

因为概率空间的完备性，即 $\sum_{i=1}^{q} p_i = 1$，所以 $H(p)$ 是 $(q-1)$ 元函数 .当 $q = 2$ 时，因

为 $p_1 + p_2 = 1$,若令其中一个概率为 p,则另一个概率为$(1 - p)$,熵函数可以写成 $H(p)$.

熵函数 $H(p)$具有以下性质:

1.对称性

$$H(p_1, p_2, \cdots, p_q) = H(p_2, p_1, \cdots, p_q) = \cdots = H(p_q, p_1, \cdots, p_{q-1}) \tag{2.5}$$

也就是说概率矢量 $p = (p_1, p_2, \cdots, p_q)$ 各分量的次序可以任意变更,熵值不变.对称性说明熵函数仅与信源的总体统计特性有关.

例如,3 个信源 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1(\text{红}) & x_2(\text{黄}) & x_3(\text{蓝}) \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$, $\begin{bmatrix} Y \\ P(Y) \end{bmatrix} = \begin{bmatrix} y_1(\text{红}) & y_2(\text{黄}) & y_3(\text{蓝}) \\ 1/6 & 1/2 & 1/3 \end{bmatrix}$, $\begin{bmatrix} Z \\ P(Z) \end{bmatrix} = \begin{bmatrix} z_1(\text{晴}) & z_2(\text{雾}) & z_3(\text{雨}) \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$ 的信息熵都相等,因为 3 个信源的总体统计特性都相同,香农熵只抽取了信源信息输出的统计特征,而没有考虑信息的具体含义和效用.

2.确定性

$$H(1,0) = H(1,0,0) = H(1,0,0,0) = \cdots = H(1,0,\cdots,0) = 0 \tag{2.6}$$

在概率矢量 $p = (p_1, p_2, \cdots, p_q)$ 中,只要有一个分量为 1,其他分量必为 0,它们对熵的贡献均为 0,因此熵等于 0,也就是说确定信源的平均不确定度为 0.

3.非负性

$$H(p) = H(p_1, p_2, \cdots, p_q) \geq 0 \tag{2.7}$$

对确定信源,等号成立.

信源熵是自信息的数学期望,自信息是非负值,所以信源熵必定是非负的.离散信源熵才有这种非负性,以后会讲到连续信源的相对熵则可能出现负值.

4.扩展性

$$\lim_{\varepsilon \to 0} H_{q+1}(p_1, p_2, \cdots, p_q - \varepsilon, \varepsilon) = H_q(p_1, p_2, \cdots, p_q) \tag{2.8}$$

这是因为$\lim_{\varepsilon \to 0} \varepsilon \log_2 \varepsilon = 0$.

这个性质的含义是:增加一个基本不会出现的小概率事件,信源的熵保持不变.虽然小概率事件出现给予收信者的信息量很大,但在熵的计算中,它占的比重很小,可以忽略不计,这也是熵的总体平均性的体现.

5.连续性

$$\lim_{\varepsilon \to 0} H(p_1, p_2, \cdots, p_{q-1} - \varepsilon, p_q + \varepsilon) = H(p_1, p_2, \cdots, p_q) \tag{2.9}$$

即信源概率空间中概率分量的微小波动,不会引起熵的变化.

6．递增性

$$H(p_1, p_2, \cdots, p_{n-1}, q_1, q_2, \cdots, q_m) = H(p_1, p_2, \cdots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \cdots, \frac{q_m}{p_n}\right) \tag{2.10}$$

这个性质表明，假如有一信源的 n 个元素的概率分布为 $(p_1, p_2, \cdots, p_n)$，其中某个元素 x_n 又被划分成 m 个元素，这 m 个元素的概率之和等于元素 x_n 的概率，这样得到的新信源的熵增加了一项，增加的一项是由于划分产生的不确定性．

【例 2.4】

利用递推性计算 $H(1/2, 1/8, 1/8, 1/8, 1/8)$．

解

$$\begin{aligned} & H(1/2, 1/8, 1/8, 1/8, 1/8) \\ &= H(1/2, 1/2) + \frac{1}{2} \times H(1/4, 1/4, 1/4, 1/4) \\ &= 1 + \frac{1}{2} \times 2 \\ &= 2 \text{ 比特/符号} \end{aligned}$$

7．极值性

$$H(p_1, p_2, \cdots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) = \log_2 n \tag{2.11}$$

式中 n 是随机变量 X 的可能取值的个数．

极值性表明离散信源中各消息等概率出现时熵最大，这就是最大离散熵定理．连续信源的最大熵则还与约束条件有关．

极值性可看成

$$H(p_1, p_2, \cdots, p_n) \leq -\sum_{i=1}^{n} p_i \log_2 q_i \tag{2.12}$$

的特例情况．下面先证明式(2.12)．

证明

利用 Jensen 不等式(参见附录 B.2)，有

$$\begin{aligned} & H(p_1, p_2, \cdots, p_n) + \sum_{i=1}^{n} p_i \log_2 q_i \\ &= -\sum_{i=1}^{n} p_i \log_2 p_i + \sum_{i=1}^{n} p_i \log_2 q_i = \sum_{i=1}^{n} p_i \log_2 \frac{q_i}{p_i} \leq \log_2 \sum_{i=1}^{n} \left(p_i \cdot \frac{q_i}{p_i}\right) = 0 \end{aligned}$$

当 $\frac{q_i}{p_i} = 1$，$i = 1, 2, \cdots, n$ 时，等号成立．

证毕

式(2.12)表明任一随机变量的概率分布 p_i，对其他概率分布 q_i 定义的自信息

$-\log_2 q_i$的数学期望,必不小于概率分布 p_i 本身定义的熵 $H(p_1, p_2, \cdots, p_n)$.

如果取 $q_i = \frac{1}{n}$, $i = 1, 2, \cdots, n$ 时,由式(2 .12)就得到

$$H(p_1, p_2, \cdots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) = \log_2 n \tag{2 .13}$$

当 $p_i = \frac{1}{n}$, $i = 1, 2, \cdots, n$ 时,等号成立 .

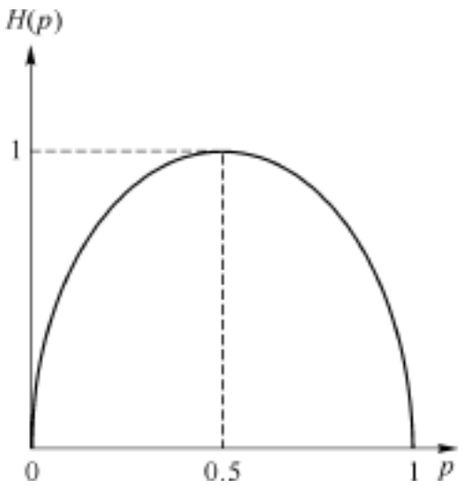


图 2 .2 二元熵函数

一个二元信源的熵函数如图 2 .2 所示,当信源输出的消息等概分布时,信源熵达到最大值——1 比特/符号 .因此当二元数字是由等概的二元信源输出时,每个二元数字提供 1 bit 的信息量,否则,每个二元数字提供的信息量小于1 bit .这就是信息量的单位比特和计算机术语中位的单位比特的关系 .

8 .上凸性

$H(p)$是严格的上凸函数,设 $p = (p_1, p_2, \cdots, p_q)$, $p' = (p_1', p_2', \cdots, p_q')$, $\sum_{i=1}^{q} p_i = 1$, $\sum_{i=1}^{q} p_i' = 1$,则对于任意小于 1 的正数 α, $0 < \alpha < 1$,以下不等式成立:

$$H[\alpha p + (1 - \alpha) p'] > \alpha H(p) + (1 - \alpha) H(p') \tag{2 .14}$$

证明

因为 $0 \leq p_i \leq 1$, $0 \leq p_i' \leq 1$, 且 $0 < \alpha < 1$, 所以, $0 \leq \alpha p_i + (1 - \alpha) p_i' \leq 1$, 并且 $\sum_{i=1}^{q} [\alpha p_i + (1 - \alpha) p_i'] = 1$, 所以, $\alpha p + (1 - \alpha) p'$可以看作是一种新的概率分布 .

$$\begin{aligned} H[\alpha p + (1 - \alpha) p'] &= -\sum_{i=1}^{q} [\alpha p_i + (1 - \alpha) p_i'] \log_2 [\alpha p_i + (1 - \alpha) p_i'] \\ &= -\alpha \sum_{i=1}^{q} p_i \log_2 [\alpha p_i + (1 - \alpha) p_i'] - (1 - \alpha) \sum_{i=1}^{q} p_i' \log_2 [\alpha p_i + (1 - \alpha) p_i'] \\ &\geq -\alpha \sum_{i=1}^{q} p_i \log_2 p_i - (1 - \alpha) \sum_{i=1}^{q} p_i' \log_2 p_i' \\ &\geq \alpha H(p) + (1 - \alpha) H(p') \end{aligned}$$

当 $p \neq p'$ 时,有 $\frac{\alpha p_i + (1 - \alpha) p_i'}{p_i} \neq 1$,式(2 .12)中等号不成立,所以

$$H[\alpha p + (1 - \alpha) p'] > \alpha H(p) + (1 - \alpha) H(p') \tag{2 .15}$$

成立 .

证毕

上凸函数在定义域内的极值必为极大值,可以利用熵函数的这个性质证明熵函数的极值性.请读者自行证明.

直观来看,随机变量的不确定程度并不都是一样的.例如,抛掷一枚均匀硬币结果所得到的信息量会比抛掷一枚偏畸硬币所得到的信息量大;投掷一颗均匀骰子的试验比抛掷一枚均匀硬币的试验所得到的信息量大.怎么度量这种不确定性呢?香农指出,存在这样的不确定性的度量,它是随机变量的概率分布的函数,而且必须满足 3 个公理性条件:

(1) 连续性条件: $f(p_1, p_2, \cdots, p_n)$ 应是 p_i, $i = 1, 2, \cdots, n$ 的连续函数;

(2) 等概时为单调函数: $f(1/n, 1/n, \cdots, 1/n)$ 应是 n 的增函数;

(3) 递增性条件:当随机变量的取值不是通过一次试验而是若干次试验才最后得到时,随机变量在各次试验中的不确定性应该可加,且其和始终与通过一次试验取得的不确定程度相同,即

$$f(p_1, p_2, \cdots, p_n) = f[(p_1 + p_2 + \cdots + p_k), p_{k+1}, \cdots, p_n] + (p_1 + p_2 + \cdots + p_k) f(p_1', p_2', \cdots, p_k')$$

其中, $p_k' = p_k/(p_1 + p_2 + \cdots + p_k)$.

香农根据这 3 个公理性条件于 1948 年先提出了熵的概念,他当时并没有像我们现在这样把熵看成自信息的均值.后来,Feinstein(范恩斯坦)等人从数学上严格地证明了当满足上述条件时,信息熵的表达形式是唯一的.

2.2.3　联合熵与条件熵

一个随机变量的不确定性可以用熵来表示,这一概念可以方便地推广到多个随机变量.

定义 2.4　二维随机变量 XY 的概率空间表示为

$$\begin{bmatrix} XY \\ P(XY) \end{bmatrix} = \begin{bmatrix} x_1 y_1 & \cdots & x_i y_j & \cdots & x_n y_m \\ p(x_1 y_1) & \cdots & p(x_i y_j) & \cdots & p(x_n y_m) \end{bmatrix}$$

其中, $p(x_i y_j)$ 满足概率空间的非负性和完备性: $0 \leq p(x_i y_j) \leq 1$, $\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i y_j) = 1$.

二维随机变量 XY 的联合熵定义为联合自信息的数学期望,它是二维随机变量 XY 的不确定性的度量.

$$H(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i y_j) I(x_i y_j) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i y_j) \log_2 p(x_i y_j) \tag{2.16}$$

考虑在给定 $X = x_i$ 的条件下,随机变量 Y 的不确定性为

$$H(Y \mid x_i) = -\sum_{j} p(y_j \mid x_i) \log_2 p(y_j \mid x_i) \tag{2.17}$$

由于不同的 x_i, $H(Y \mid x_i)$ 是变化的,对 $H(Y \mid x_i)$ 的所有可能值进行统计平均,就得出给

定 X 时，Y 的条件熵 $H(Y|X)$.

定义 2.5

$$\begin{aligned} H(Y|X) &= \sum_i p(x_i)\, H(Y|x_i) \\ &= -\sum_i \sum_j p(x_i)\, p(y_j|x_i) \log_2 p(y_j|x_i) \\ &= -\sum_i \sum_j p(x_i y_j) \log_2 p(y_j|x_i) \end{aligned} \tag{2.18}$$

其中，$H(Y|X)$ 表示已知 X 时，Y 的平均不确定性 .

同理

$$H(X|Y) = -\sum_i \sum_j p(x_i y_j) \log_2 p(x_i|y_j) \tag{2.19}$$

各类熵之间的关系：

(1) 联合熵与信息熵、条件熵的关系

$$H(XY) = H(X) + H(Y|X) \tag{2.20}$$

证明

$$\begin{aligned} H(XY) &= -\sum_{i=1}^{n} \sum_{y=j}^{m} p(x_i y_j) \log_2 p(x_i y_j) \\ &= -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i y_j) \log_2 [\, p(x_i)\, p(y_j|x_i)\,] \\ &= -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i y_j) \log_2 p(x_i) - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i y_j) \log_2 p(y_j|x_i) \\ &= -\sum_{i=1}^{n} \Big[\sum_{j=1}^{m} p(x_i y_j) \Big] \log_2 p(x_i) - \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i)\, p(y_j|x_i) \log_2 p(y_j|x_i) \\ &= -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) - \sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p(y_j|x_i) \log_2 p(y_j|x_i) \\ &= H(X) + \sum_{i=1}^{n} p(x_i)\, H(Y|x_i) \\ &= H(X) + H(Y|X) \end{aligned}$$

上述证明还可以更简洁地表示成：

$$\begin{aligned} H(XY) &= E\Big[\log_2 \frac{1}{p(x\ y)} \Big] \\ &= E\Big[\log_2 \frac{1}{p(x)\, p(y|x)} \Big] \\ &= E\Big[\log_2 \frac{1}{p(x)} + \log_2 \frac{1}{p(y|x)} \Big] \end{aligned}$$

$$= E\left[\log_2 \frac{1}{p(x)}\right] + E\left[\log_2 \frac{1}{p(y|x)}\right]$$
$$= H(X) + H(Y|X)$$

即两个随机变量 X 和 Y 的联合熵等于 X 的熵加上在 X 已知条件下 Y 的条件熵，这个关系可以方便地推广到 N 个随机变量的情况，即

$$H(X_1 X_2 \cdots X_N) = H(X_1) + H(X_2 | X_1) + \cdots + H(X_N | X_1 X_2 \cdots X_{N-1}) \quad (2.21)$$

称为熵函数的链规则．

证毕

推论　当二维随机变量 X, Y 相互独立时，联合熵等于 X 和 Y 各自熵之和．

$$H(XY) = H(X) + H(Y) \quad (2.22)$$

证明

因为随机变量 X, Y 相互独立，所以有

$$p(x_i y_j) = p(x_i)\, p(y_j)$$
$$H(XY) = E[-\log_2 p(xy)]$$
$$= E[-\log_2 (p(x)\, p(y))]$$
$$= E[-(\log_2 p(x) + \log_2 p(y))]$$
$$= E[-\log_2 p(x)] + E[-\log_2 p(y)]$$
$$= H(X) + H(Y)$$

证毕

如果 N 个随机变量 X_1, X_2, …, X_N 相互独立，则有

$$H(X_1 X_2 \cdots X_N) = \sum_{i=1}^{N} H(X_i) \quad (2.23)$$

(2) 条件熵与信息熵的关系

$$H(X|Y) \leq H(X) \quad (2.24)$$
$$H(Y|X) \leq H(Y) \quad (2.25)$$

证明

利用式(2.12)先证明式(2.24)．

$$-\sum_i \sum_j p(x_i y_j) \log_2 p(x_i | y_j)$$
$$= -\sum_i \sum_j p(y_j)\, p(x_i | y_j) \log_2 p(x_i | y_j)$$
$$= -\sum_j p(y_j) \sum_i p(x_i | y_j) \log_2 p(x_i | y_j)$$

$$\leq -\sum_j p(y_j)\sum_i p(x_i|y_j)\log_2 p(x_i)$$

$$= -\sum_i\sum_j p(x_i y_j)\log_2 p(x_i)$$

$$= -\sum_i p(x_i)\log_2 p(x_i) = H(X)$$

当 $p(x_i|y_j) = p(x_i)$时等号成立。

类似地,可以证明 $H(Y|X) \leq H(Y)$.

证毕

(3) 联合熵和信息熵的关系

$$H(XY) \leq H(X) + H(Y) \tag{2.26}$$

当 X、Y 相互独立时等号成立 .

证明

$$H(XY) = H(X) + H(Y|X) \leq H(X) + H(Y)$$

当 X、Y 相互独立时等号成立 .

推广到 N 个随机变量的情况:

$$H(X_1 X_2 \cdots X_N) \leq H(X_1) + H(X_2) + \cdots + H(X_N) \tag{2.27}$$

当 $X_1, X_2, \cdots, X_N$ 相互独立时,等号成立 .

证毕

【例 2.5】

随机变量 X, Y 的联合概率分布如表 2.1 所示,求联合熵 $H(XY)$和条件熵 $H(Y|X)$.

解

$$H(XY) = \frac{1}{4}\log_2\frac{1}{1/4} + \frac{1}{4}\log_2\frac{1}{1/4} + \frac{1}{2}\log_2\frac{1}{1/2}$$

$$= \frac{2}{4}\log_2 4 + \frac{1}{2}\log_2 2$$

$$= \frac{2}{4}\times 2 + \frac{1}{2}\times 1$$

$$= \frac{3}{2}$$

由联合概率分布得到 X 的边沿概率分布: $P_r\{X=0\} = \frac{1}{2}$, $P_r\{X=1\} = \frac{1}{2}$和条件概率分布 $P(y_j|x_i)$(如表 2.2 所示),得到 $H(Y|X=0) = 1$, $H(Y|X=1) = 0$ 和 $H(Y|X) = \frac{1}{2}\times 1 + \frac{1}{2}\times 0 = \frac{1}{2}$.

注意到 $H(Y) = H(1/4) = 0.811\,3 > 1/2 = H(Y|X)$.

表 2.1　X, Y 的联合概率分布 $P(XY)$

X \ Y	0	1	$p(x_i)$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{2}$	0	$\frac{1}{2}$
$p(y_j)$	$\frac{3}{4}$	$\frac{1}{4}$	1

表 2.2　条件概率分布 $P(Y|X)$

X \ Y	0	1
0	$\frac{1}{2}$	$\frac{1}{2}$
1	1	0

2.3　平均互信息

2.3.1　平均互信息的概念

互信息 $I(x_i;y_j)$ 表示某一事件 y_j 所给出的关于另一个事件 x_i 的信息，它随 x_i 和 y_j 的变化而变化，为了从整体上表示从一个随机变量 Y 所给出关于另一个随机变量 X 的信息量，定义互信息 $I(x_i;y_j)$ 在 XY 的联合概率空间中的统计平均值为随机变量 X 和 Y 间的平均互信息.

定义 2.6

$$\begin{aligned} I(X;Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_j)\, I(x_i;y_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_j)\log_2\frac{p(x_i|y_j)}{p(x_i)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_j)\log_2\frac{1}{p(x_i)} - \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_j)\log_2\frac{1}{p(x_i|y_j)} \\ &= H(X) - H(X|Y) \end{aligned} \tag{2.28}$$

条件熵 $H(X|Y)$ 表示给定随机变量 Y 后，对随机变量 X 仍然存在的不确定度.所以 Y 关于 X 的平均互信息是收到 Y 前后关于 X 的不确定度减少的量，也就是从 Y 所获得的关于 X 的平均信息量.

【例 2.6】

掷骰子，若结果是 1，2，3 或 4，则抛一次硬币；如果结果是 5 或者 6，则抛两次硬币，试计算从抛硬币的结果可以得到多少掷骰子的信息量.

解

本题的题意是根据抛硬币出现正面的次数 Y 来获得关于掷骰子结果 X 的信息(两种结果).

设掷骰子结果是 1，2，3，4 的事件为 $X=0$，结果是 5，6 的事件为 $X=1$，随机变量

$Y=0$表示抛币出现 0 次正面，$Y=1$ 表示出现 1 次正面，$Y=2$ 表示出现 2 次正面 .

随机变量 X 的概率空间为 $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$

条件概率矩阵 $P_{Y|X} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$

Y 的概率分布为

$$P_Y = P_X P_{Y|X} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{5}{12} & \frac{1}{2} & \frac{1}{12} \end{pmatrix}$$

所以 Y 的信息熵

$$\begin{aligned} H(Y) &= p_Y(0)\log_2\frac{1}{p_Y(0)} + p_Y(1)\log_2\frac{1}{p_Y(1)} + p_Y(2)\log_2\frac{1}{p_Y(2)} \\ &= \frac{5}{12}\log_2\frac{12}{5} + \frac{1}{2}\log_2 2 + \frac{1}{12}\log_2 12 \\ &= 1.325 \end{aligned}$$

又可以根据 X 的概率分布和条件概率分布 $P_{Y|X}$ 求出

$$H(Y|X) = \frac{2}{3}\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right) + \frac{1}{3}\left(\frac{1}{4}\log_2 4 + \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4\right) = 1.166$$

所以

$$I(X;Y) = H(Y) - H(Y|X) = 1.325 - 1.166 = 0.159$$

即从抛硬币出现正面次数平均得到关于掷骰子结果的信息量为 0.159 比特/ 符号 .

2.3.2 平均互信息的性质

1. 非负性

$$I(X;Y) \geq 0 \tag{2.29}$$

证明

$$\begin{aligned} -I(X;Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i y_j)\log_2\frac{p(x_i)\,p(y_j)}{p(x_i y_j)} \\ &\leq \log_2\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i y_j)\frac{p(x_i)\,p(y_j)}{p(x_i y_j)} \\ &= \log_2\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)\,p(y_j) \\ &= 0 \end{aligned}$$

所以　　$I(X;Y)\geq 0$

证毕

平均互信息是非负的,说明给定随机变量 Y 后,一般来说总能消除一部分关于 X 的不确定性 .

2 . 互易性(对称性)

$$I(X;Y)=I(Y;X) \tag{2.30}$$

证明

$$\begin{aligned} I(X;Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_j)\log_2\frac{p(x_i\mid y_j)}{p(x_i)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_j)\log_2\frac{p(x_iy_j)}{p(x_i)\,p(y_j)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_j)\log_2\frac{p(y_j\mid x_i)}{p(y_j)} \\ &= I(Y;X) \end{aligned}$$

证毕

对称性表示从 Y 中获得关于 X 的信息量等于从 X 中获得关于 Y 的信息量 .

3 . 平均互信息和各类熵的关系

平均互信息和各类熵的关系如图 2 .3 所示,即

$$\begin{aligned} I(X;Y) &= H(X)-H(X\mid Y) \\ &= H(Y)-H(Y\mid X) \\ &= H(X)+H(Y)-H(XY) \end{aligned} \tag{2.31}$$

当 X, Y 统计独立时, $I(X;Y)=0$.

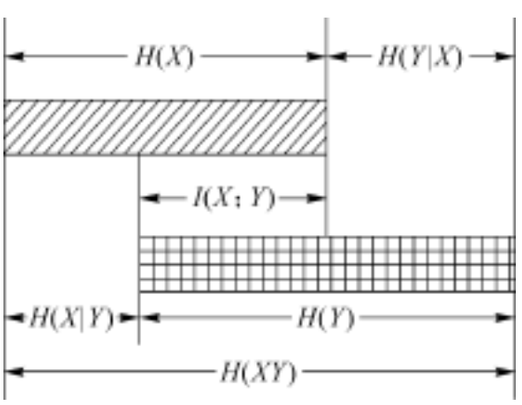


图 2 .3　平均互信息和各类熵之间的关系

4．极值性

$$I(X;Y)\leq H(X),\ I(X;Y)\leq H(Y) \tag{2.32}$$

由于 $I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)$，而条件熵 $H(X|Y)$、$H(Y|X)$ 是非负的(请读者自己证明)，所以可得到 $I(X;Y)\leq H(X)$，$I(X;Y)\leq H(Y)$．极值性说明从一个事件获得的关于另一个事件的信息量至多只能是另一个事件的平均自信息量，不会超过另一事件本身所含的信息量．最好的情况是通信后 $I(X;Y)=H(X)=H(Y)$，最坏的情况是当 X，Y 相互独立时，从一个事件不能得到另一个事件的任何信息，即 $I(X;Y)=0$，等效于通信中断．

5．凸函数性

定理 2.1 当条件概率分布$\{p(y_j|x_i)\}$给定时，平均互信息 $I(X;Y)$ 是输入分布 $\{p(x_i)\}$ 的上凸函数．

证明

设给定条件概率分布$\{p(y_j|x_i)\}$，$p_1(x_i)$ 和 $p_2(x_i)$ 为信源的两种不同的概率分布，相应的平均互信息记为 $I[p_1(x_i)]$ 和 $I[p_2(x_i)]$，再选择信源符号集的另一种概率分布 $\{p(x_i)\}$，且令

$$p(x_i)=\theta p_1(x_i)+(1-\theta)p_2(x_i) \tag{2.33}$$

其中，$0<\theta<1$，相应的平均互信息记为 $I[p(x_i)]$，根据上凸函数的定义，需要证明：

$$I[p(x_i)]\geq\theta I[p_1(x_i)]+(1-\theta)I[p_2(x_i)] \tag{2.34}$$

根据平均互信息的定义，有

$$\theta I[p_1(x_i)]+(1-\theta)I[p_2(x_i)]-I[p(x_i)]$$

$$=\theta\sum_{i,j}p_1(x_i)p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p_1(y_j)}+(1-\theta)\sum_{i,j}p_2(x_i)p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p_2(y_j)}-\sum_{i,j}p(x_i)p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p(y_j)}$$

$$=\theta\sum_{i,j}p_1(x_i)p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p_1(y_j)}+(1-\theta)\sum_{i,j}p_2(x_i)p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p_2(y_j)}-\theta\sum_{i,j}p_1(x_i)p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p(y_j)}-(1-\theta)\sum_{i,j}p_2(x_i)p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p(y_j)}$$

(将式(2.33)代入)

$$=\theta\sum_{i,j}p_1(x_i)p(y_j|x_i)\log_2\frac{p(y_j)}{p_1(y_j)}+(1-\theta)\sum_{i,j}p_2(x_i)p(y_j|x_i)\log_2\frac{p(y_j)}{p_2(y_j)}$$

(合并同类项)

$$\leq\theta\log_2\sum_{i,j}p_1(x_i)p(y_j|x_i)\frac{p(y_j)}{p_1(y_j)}+(1-\theta)\log_2\sum_{i,j}p_2(x_i)p(y_j|x_i)\frac{p(y_j)}{p_2(y_j)}$$

(Jensen 不等式)

$$= \theta\log_2 \sum_j \frac{p(y_j)}{p_1(y_j)} \sum_i p_1(x_i y_j) + (1-\theta)\log_2 \sum_j \frac{p(y_j)}{p_2(y_j)} \sum_i p_2(x_i y_j)$$

$$= \theta\log_2 1 + (1-\theta)\log_2 1$$

$$= 0$$

所以,证得式(2 34)成立 以上证明中应用了 $p_1(x_i)\,p(y_j \mid x_i) = p_1(x_i y_j)$, $p_2(x_i)\,p(y_j \mid x_i) = p_2(x_i y_j)$, $\sum_i p_1(x_i y_j) = p_1(y_j)$, $\sum_j \frac{p(y_j)}{p_1(y_j)} p_1(y_j) = 1$, $\sum_i p_2(x_i y_j) = p_2(y_j)$, $\sum_j \frac{p(y_j)}{p_2(y_j)} p_2(y_j) = 1$ 的概率关系 .

证毕

由上凸函数的定义可知,当条件概率分布$\{p(y_j \mid x_i)\}$给定时,平均互信息 $I(X;Y)$是输入分布$\{p(x_i)\}$的上凸函数 .如果把条件概率分布$\{p(y_j \mid x_i)\}$看成信道的转移概率分布,那么存在一个最佳信道输入分布$\{p(x_i)\}$使 $I(X;Y)$的值最大 .

定理 2 .2　对于固定的输入分布$\{p(x_i)\}$,平均互信息量 $I(X;Y)$是条件概率分布$\{p(y_j \mid x_i)\}$的下凸函数 .

证明

设固定的信源分布为$\{p(x_i)\}$,$\{p_1(y_j \mid x_i)\}$和$\{p_2(y_j \mid x_i)\}$为信道的两种不同的转移概率分布,相应的平均互信息记为 $I[p_1(y_j \mid x_i)]$和 $I[p_2(y_j \mid x_i)]$,再选择信道的另一种转移概率分布$\{p(y_j \mid x_i)\}$,且令

$$p(y_j \mid x_i) = \theta p_1(y_j \mid x_i) + (1-\theta)\,p_2(y_j \mid x_i) \tag{2 35}$$

其中,$0 < \theta < 1$,相应的平均互信息记为 $I[p(y_j \mid x_i)]$,根据下凸函数的定义,需要证明

$$\theta I[p_1(y_j \mid x_i)] + (1-\theta)\,I[p_2(y_j \mid x_i)] \geq I[p(y_j \mid x_i)] \tag{2 36}$$

根据平均互信息的定义,有

$$I[p(y_j \mid x_i)] - \theta I[p_1(y_j \mid x_i)] - (1-\theta)\,I[p_2(y_j \mid x_i)]$$

$$= \sum_{i,j} p(x_i)\,p(y_j \mid x_i)\log_2 \frac{p(x_i \mid y_j)}{p(x_i)} - \theta \sum_{i,j} p(x_i)\,p_1(y_j \mid x_i)\log_2 \frac{p_1(x_i \mid y_j)}{p(x_i)} -$$

$$(1-\theta) \sum_{i,j} p(x_i)\,p_2(y_j \mid x_i)\log_2 \frac{p_2(x_i \mid y_j)}{p(x_i)}$$

$$= \theta \sum_{i,j} p(x_i)\,p_1(y_j \mid x_i)\log_2 \frac{p(x_i \mid y_j)}{p(x_i)} + (1-\theta) \sum_{i,j} p(x_i)\,p_2(y_j \mid x_i)\log_2 \frac{p(x_i \mid y_j)}{p(x_i)} -$$

$$\theta \sum_{i,j} p(x_i)\,p_1(y_j \mid x_i)\log_2 \frac{p_1(x_i \mid y_j)}{p(x_i)} - (1-\theta) \sum_{i,j} p(x_i)\,p_2(y_j \mid x_i)\log_2 \frac{p_2(x_i \mid y_j)}{p(x_i)}$$

(将式(2 .35)代入)

$$= \theta\sum_{i,j} p(x_i)\,p_1(y_j|x_i)\log_2\frac{p(x_i|y_j)}{p_1(x_i|y_j)} + (1-\theta)\sum_{i,j} p(x_i)\,p_2(y_j|x_i)\log_2\frac{p(x_i|y_j)}{p_2(x_i|y_j)}$$

(合并同类项)

$$\leq \theta\log_2\sum_{i,j} p(x_i)\,p_1(y_j|x_i)\frac{p(x_i|y_j)}{p_1(x_i|y_j)} + (1-\theta)\log_2\sum_{i,j} p(x_i)\,p_2(y_j|x_i)\frac{p(x_i|y_j)}{p_2(x_i|y_j)}$$

(Jensen 不等式)

$$= \theta\log_2\sum_j p(y_j)\sum_i p(x_i|y_j) + (1-\theta)\log_2\sum_j p(y_j)\sum_i p(x_i|y_j)$$

$$= \theta\log_2 1 + (1-\theta)\log_2 1$$

$$= 0$$

所以，证得式(2.36)成立．以上证明中应用了 $p(x_i)\,p_1(y_j|x_i) = p_1(x_iy_j) = p_1(x_i|y_j)\,p(y_j)$，$p(x_i)\,p_2(y_j|x_i) = p_2(x_iy_j) = p_2(x_i|y_j)\,p(y_j)$，$\sum_i p(x_i|y_j) = 1$，$\sum_j p(y_j) = 1$ 的概率关系．

证毕

因此，由下凸函数的定义可知，在给定输入分布的情况下，平均互信息量 $I(X;Y)$ 是条件概率分布 $\{p(y_j|x_i)\}$ 的下凸函数．如果把条件概率分布 $\{p(y_j|x_i)\}$ 看成信道的转移概率分布，那么对于给定的输入分布，必存在一种最差的信道，此信道的干扰(噪声)最大，收信者获得的信息量最小．在第 7 章讨论信息率失真函数时会用到这个定理．

2.3.3 数据处理定理

为了表述数据处理定理，需要引入三元随机变量 X，Y，Z 的平均条件互信息和平均联合互信息的概念．

定义 2.7 平均条件互信息

$$I(X;Y|Z) = E[I(x;y|z)] = \sum_x\sum_y\sum_z p(xyz)\log_2\frac{p(x|yz)}{p(x|z)} \tag{2.37}$$

它表示随机变量 Z 给定后，从随机变量 Y 所得到的关于随机变量 X 的信息量．

定义 2.8 平均联合互信息

$$I(X;YZ) = E[I(x;yz)] = \sum_x\sum_y\sum_z p(xyz)\log_2\frac{p(x|yz)}{p(x)} \tag{2.38}$$

它表示从二维随机变量 YZ 所得到的关于随机变量 X 的信息量．

可以证明

$$I(X;YZ) = \sum_x\sum_y\sum_z p(xyz)\log_2\frac{p(x|z)\,p(x|yz)}{p(x)\,p(x|z)}$$

$$= I(X;Z) + I(X;Y|Z) \tag{2.39}$$

同理

$$I(X;YZ)=I(X;Y)+I(X;Z|Y) \tag{2.40}$$

定理 2.3　(数据处理定理)

如果随机变量 X, Y, Z 构成一个马尔可夫链,则有以下关系成立:

$$I(X;Z)\leq I(X;Y),\ I(X;Z)\leq I(Y;Z) \tag{2.41}$$

等号成立的条件是对于任意的 x, y, z,有 $p(x|yz)=p(x|z)$和 $p(z|xy)=p(z|x)$.

证明

当 X, Y, Z 构成一个马尔可夫链时, Y 值给定后, X, Z 可以认为是互相独立的,所以,

$$I(X;Z|Y)=0$$

又因为 $I(X;YZ)=I(X;Y)+I(X;Z|Y)=I(X;Z)+I(X;Y|Z)$, 并且 $I(X;Y|Z)\geq 0$,所以 $I(X;Z)\leq I(X;Y)$.

当 $p(x|yz)=p(x|z)$时, Z 值给定后, X 和 Y 相互独立,所以

$$I(X;Y|Z)=0$$

因此

$$I(X;Z)=I(X;Y).$$

这时 $p(x|yz)=p(x|z)=p(x|y)$. Y, Z 为确定关系时显然满足该条件.

同理可以证明 $I(X;Z)\leq I(Y;Z)$,并且当 $p(z|xy)=p(z|x)$时, 等号成立.

证毕

$I(X;Z)\leq I(X;Y)$表明从 Z 所得到的关于 X 的信息量小于等于从 Y 所得到的关于 X 的信息量.如果把 $Y\rightarrow Z$ 看作数据处理系统,那么通过数据处理后,虽然可以满足我们的某种具体要求,但是从信息量来看,处理后会损失一部分信息,最多保持原有的信息,也就是说,对接收到的数据 Y 进行处理后,决不会减少关于 X 的不确定性.这个定理称为数据处理定理.数据处理定理与日常生活中的经验是一致的.比如:通过别人转述一段话或多或少会有一些失真,通过书本得到的间接经验总不如直接经验来得详实.

数据处理定理再一次说明,在任何信息传输系统中,最后获得的信息至多是信源所提供的信息,一旦在某一过程中丢失一些信息,以后的系统不管如何处理,如不触及丢失信息的输入端,就不能再恢复已丢失的信息,这就是信息不增性原理,它与热熵不减原理正好对应,反映了信息的物理意义.

习　题　2

2.1　同时掷 2 颗骰子,事件 A、B、C 分别表示:(A)仅有一个骰子是 3;(B)至少有一个骰子是 4;(C)骰子上点数的总和为偶数.试计算事件 A、B、C 发生后所提供的信息量.

2.2 设有 n 个球,每个球都以同样的概率 $1/N$ 落入 N 个格子($N \geq n$)的每一个格子中.假定:(A)某指定的 n 个格子中各落入一个球;(B)任何 n 个格子中各落入一球.试计算事件 A、B 发生后所提供的信息量.

2.3 一信源有 4 种输出符号 x_i, $i = 0, 1, 2, 3$,且 $p(x_i) = 1/4$。设信源向信宿发出 x_3,但由于传输中的干扰,接收者收到 x_3 后,认为其可信度为 0.9.于是信源再次向信宿发送该符号(x_3),信宿无误收到.问信源在两次发送中发出的信息量各是多少?信宿在两次接收中得到的信息量又各是多少?

2.4 用递推性计算熵函数 $H(1/3, 1/3, 1/6, 1/6)$的值.

2.5 一信源有 6 种输出状态,概率分别为

$$p(A) = 0.5,\quad p(B) = 0.25,\quad p(C) = 0.125,\quad p(D) = p(E) = 0.05,\quad p(F) = 0.025$$

试计算 $H(X)$.然后求消息 $ABABBA$ 和 $FDDFDF$ 的信息量(设信源先后发出的符号相互独立),并将之与长度为 6 的消息序列的信息量期望值相比较.

2.6 中国国家标准局所规定的二级汉字共 6763 个.设每字使用的频度相等,求一个汉字所含的信息量.设每个汉字用一个 16×16 的二元点阵显示,试计算显示方阵所能表示的最大信息.显示方阵的利用率是多少?

2.7 已知信源发出 a_1 和 a_2 两种消息,且 $p(a_1) = p(a_2) = \frac{1}{2}$.此消息在二进制对称信道上传输,信道传输特性为 $p(b_1 \mid a_1) = p(b_2 \mid a_2) = 1 - \varepsilon$, $p(b_1 \mid a_2) = p(b_2 \mid a_1) = \varepsilon$.求互信息量 $I(a_1; b_1)$和 $I(a_2; b_2)$.

2.8 已知二维随机变量 XY 的联合概率分布 $p(x_i y_j)$为:$p(0,0) = p(1,1) = 1/8$, $p(0,1) = p(1,0) = 3/8$,求 $H(X \mid Y)$.

2.9 X 和 Y 是$\{0, 1, 2, 3\}$上的独立、均匀分布的随机变量,求:

(1) $H(X + Y)$, $H(X - Y)$, $H(X \cdot Y)$

(2) $H(X + Y, X - Y)$, $H(X + Y, X \cdot Y)$

2.10 棒球比赛中大卫和麦克在前面的比赛中打平,最后 3 场与其他选手的比赛结果将最终决定他们的胜、负或平.

(1) 假定最后 3 场他们与其他选手的比赛结果胜负的可能性均为 0.5,把麦克的最终比赛结果{胜、负、平}作为随机变量,计算它的熵;

(2) 假定大卫最后三场比赛全部获胜,计算麦克的最终比赛结果的条件熵.

2.11 X, Y, Z 为 3 个随机变量,证明以下不等式成立并指出等号成立的条件:

(1) $H(XY \mid Z) \geq H(X \mid Z)$

(2) $I(XY; Z) \geq I(X; Z)$

(3) $H(XYZ) - H(XY) \leq H(XZ) - H(X)$

(4) $I(X; Z \mid Y) \geq I(Z; Y \mid X) - I(Z; Y) + I(X; Z)$

2.12　找出一个概率分布$\{p_1, p_2, \cdots, p_5\}$,并且 $p_i > 0$,使得 $H(p_1, p_2, \cdots, p_5) = 2$.

2.13　有两个二元随机变量 X 和 Y,它们的联合概率分布如题表 2.1,同时定义另一随机变量 $Z = X \cdot Y$(一般乘积).试计算:

(1) 熵 $H(X)$, $H(Y)$, $H(Z)$, $H(XZ)$, $H(YZ)$ 和 $H(XYZ)$;

(2) 条件熵 $H(X|Y)$, $H(Y|X)$, $H(X|Z)$, $H(Z|X)$, $H(Y|Z)$, $H(Z|Y)$, $H(X|YZ)$, $H(Y|XZ)$ 和 $H(Z|XY)$;

(3) 互信息 $I(X;Y)$, $I(X;Z)$, $I(Y;Z)$, $I(X;Y|Z)$, $I(Y;Z|X)$ 和 $I(X;Z|Y)$.

2.14　假定 $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ 形成一个马尔可夫链,那么 $p(x_1 x_2 \cdots x_n) = p(x_1)\,p(x_2 \mid x_1) \cdots p(x_n | x_{n-1})$,请化简 $I(X_1; X_2 \cdots X_n)$.

2.15　给定 X, Y 的联合概率分布如题表 2.2:

题表 2.1

X \ Y	0	1
0	$\frac{1}{8}$	$\frac{3}{8}$
1	$\frac{3}{8}$	$\frac{1}{8}$

题表 2.2

X \ Y	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

求:(1) $H(X)$, $H(Y)$

(2) $H(X|Y)$, $H(Y|X)$

(3) $H(XY)$

(4) $H(Y) - H(Y|X)$

(5) $I(X;Y)$

2.16　(1) 假定 X 是一个离散随机变量,$g(X)$ 是 X 的函数,证明: $H[g(X)] \leq H(X)$.

(2) 假定 X 是一个定义在$\{0,1,2,3,4\}$上的等概分布的离散随机变量,$g(X) = \cos\frac{\pi X}{2}$, $f(X) = x^2$,比较它们的熵的大小。

2.17　考虑两个发射机和一个接收机之间的平均联合互信息 $I(X_1 X_2; Y)$,证明:

(1) $I(X_1 X_2; Y) \geq I(X_1; Y)$,也就是用两台发射机比用一台发射机的效果好;

(2) 如果 X_1 和 X_2 相互独立,那么 $I(X_2; Y|X_1) \geq I(X_2; Y)$;

(3) 如果 X_1 和 X_2 相互独立,那么 $I(X_1 X_2; Y) \geq I(X_1; Y) + I(X_2; Y)$,也就是同时用两台发射机比单独用两台发射机的效果好.

2.18　在一个布袋中有三枚硬币,分别用 H、T、F 表示,H 的两面都是正面,T 的两面都是反面,而 F 是一个一正一反的均匀硬币。随机选择一枚硬币并投掷两次,用 X 表示所选择的硬币, Y_1, Y_2 表示两次投掷的结果, Z 表示两次投掷中出现正面的次数。

求:(1) $I(X;Y_1)$

(2) $I(X;Z)$

(3) $I(Y_1;Y_2)$

2.19 猜宝游戏 三扇门中有一扇门后藏有一袋金子,并且三扇门后面藏有金子的可能性相同 如果有人随机打开一扇门并告诉你门后是否藏有金子,他给了你多少关于金子位置的信息量?

题表 2.3

实际 / 预报	下雨	晴天
下雨	$\frac{1}{8}$	$\frac{3}{16}$
晴天	$\frac{1}{16}$	$\frac{10}{16}$

2.20 一个年轻人研究了当地的天气纪录和气象台的预报纪录后,得到实际天气和预报天气的联合概率分布如题表 2.3 所示.他发现预报只有 12/16 的准确率,而不管三七二十一都预报明天不下雨的准确率却是 13/16.他把这个想法跟气象台台长说了后,台长却说他错了.请问这是为什么?

第 3 章

信源及信源熵

信源(Information Source)是信息的来源,是产生消息(符号)、时间离散的消息序列(符号序列)以及时间连续的消息的来源 .

信源输出的消息都是随机的,因此可用概率来描述其统计特性 .在信息论中,用随机变量 X、随机矢量 X、随机过程$\{X(e, t)\}$分别表示产生消息、消息序列以及时间连续消息的信源 .

信源的主要问题:

如何描述信源(信源的数学建模问题);

怎样定量描述信源输出信息的能力;

怎样有效地表示信源输出的消息,也就是信源编码问题 .

本章介绍第 1、2 个问题,在第 5、7 章介绍第 3 个问题 .我们将分类介绍信源的数学模型及其信源熵的计算 .

3 .1 信源的分类及其数学模型

第 2 章已经介绍了离散随机变量及信息熵,离散随机变量表示信源输出的是一个符号的消息,比如掷一颗骰子的试验;而通常实际信源输出的消息是时间(或空间)的函数,

比如掷多颗骰子的试验;消息的取值还可能是连续的,比如跳远比赛的结果 .

信源的分类有多种方法,我们常根据信源输出的消息在时间和取值上是离散或连续进行分类,如表 3 .1 所示:

表 3 .1 信源的分类

时间(空间)	取值	信源种类	举例	数学描述
离散	离散	离散信源 (数字信源)	文字、数据、 离散化图像	离散随机变量序列 $P(X)=P[X_1X_2\cdots X_N]$
离散	连续	连续信源	多人跳远比赛的结果、 语音信号抽样以后	连续随机变量序列 $P(X)=P[X_1X_2\cdots X_N]$
连续	连续	波形信源 (模拟信源)	语音、音乐、热噪声、 图形、图像	随机过程$\{X(e,t)\}$
连续	离散		不常见	

实际信源输出的消息,比如平时说话的语声和图像,在时间(或空间)和取值上都是连续的,这样的信源称为波形信源,用随机过程$\{X(e,t)\}$来描述 .对于频带受限的随机过程,根据抽样定理,我们通常把它转化成时间离散的随机序列来处理,这样的信源称为连续信源 .抽样后的值通常还是连续的,因此还可以进一步经过分层量化,将连续随机变量转化成离散随机变量、连续信源变成离散信源来处理 .

此外,还可以根据各维随机变量的概率分布是否随时间的推移而变化将信源分为平稳信源和非平稳信源,根据随机变量间是否统计独立将信源分为有记忆信源和无记忆信源 .

一个实际信源的统计特性往往是相当复杂的,要想找到精确的数学模型很困难 .实际应用时常常用一些可以处理的数学模型来近似 .比如语音信号就是非平稳随机过程,但常常用平稳随机过程来近似 .平稳随机过程抽样后的结果就是平稳随机序列 .在数学上,随机序列是随机过程的一种,是时间参数离散的随机过程,在这里我们把它单列出来 .随机序列,特别是离散平稳随机序列是我们研究的主要内容 .实际信源分类如下:

- 随机过程:波形信源
  - 平稳信源
    - 离散平稳信源
      - 离散无记忆信源
      - 离散有记忆信源
        - 记忆长度无限
        - 记忆长度有限(马尔可夫信源)
    - 连续平稳信源
  - 非平稳信源

3 .2 离散单符号信源

输出离散取值的单个符号的信源称为离散单符号信源 .它是最简单也是最基本的信

源,是组成实际信源的基本单元,用一个离散随机变量表示.信源所有可能输出的消息和消息对应的概率共同组成的二元序对[X, $P(X)$] 称为信源的概率空间:

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} X = x_1 & \cdots & X = x_i & \cdots & X = x_q \\ p(x_1) & \cdots & p(x_i) & \cdots & p(x_q) \end{bmatrix}$$

其中, X 表示信源输出的消息的整体, x_i 表示某个消息, $p(x_i)$ 表示消息 x_i 出现的概率. q 是信源可能输出的消息数(信源可能输出的消息数可以是有限个,也可以是可数无限个,通常是有限个),这些消息两两不相容,信源每次输出其中的一个消息.这里要注意, $p(x_i)$ 满足概率空间的非负性和完备性: $0 \leq p(x_i) \leq 1$, $\sum_{i=1}^{q} p(x_i) = 1$.

信源输出的所有消息的自信息的统计平均值定义为信源的平均自信息量(信息熵),它表示离散单符号信源的平均不确定性:

$$H(X) = E[-\log_2 p(x_i)] = -\sum_{i=1}^{q} p(x_i)\log_2 p(x_i) \tag{3.1}$$

【例 3.1】

二元信源 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & q \end{bmatrix}$, $p + q = 1$, 求 $H(X)$.

解

$$\begin{aligned} H(X) &= -\sum_{i=1}^{q} p_i \log_2 p_i \\ &= -p\log_2 p - (1-p)\log_2(1-p) \\ &= H(p) \end{aligned}$$

如图 2.2 所示, $H(X)$ 是概率 p 的函数,通常用 $H(p)$ 表示, p 取值于[0,1]区间.

如果输出符号是确定的,即 $p = 1$ 或 $p = 0$,则 $H(p) = 0$,信源不提供任何信息.而当 $p = 0.5$,即符号 0、1 以等概率输出时,信源熵达到极大值,平均每符号等于 1 bit 信息量.

3.3　离散多符号信源

前面介绍的单符号信源是最简单的信源模型,用一个离散随机变量表示.实际信源输出的往往是符号序列,称为离散多符号信源,通常用离散随机变量序列(随机矢量)来表示: $X = X_1 X_2 \cdots$. 例如,电报系统发出的是一串有无脉冲的信号(用有脉冲表示 1,无脉冲表示 0),因此电报系统是输出一串 0、1 序列的二元信源.

为了简单起见,这里只研究离散平稳信源,也就是统计特性不随时间改变的信源.这里先给出离散平稳信源的严格数学定义.

定义 3.1　对于离散随机变量序列 $X_1, X_2, \cdots$,在任意两个不同时刻 i 和 j(i 和 j 为

大于 1 的任意整数),信源发出的消息序列的概率分布完全相同,即对于任意的 $N=0,1,2,\cdots$, $X_iX_{i+1}\cdots X_{i+N}$ 和 $X_jX_{j+1}\cdots X_{j+N}$ 具有相同的概率分布,也就是

$$P(X_i)=P(X_j) \tag{3.2}$$

$$P(X_iX_{i+1})=P(X_jX_{j+1}) \tag{3.3}$$

$$\cdots$$

$$P(X_iX_{i+1}\cdots X_{i+N})=P(X_jX_{j+1}\cdots X_{j+N}) \tag{3.4}$$

即各维联合概率分布均与时间起点无关的信源称为离散平稳信源.

根据式(3.2)～(3.4)以及联合概率与条件概率的关系可得:

$$P(X_{i+1}\mid X_i)=P(X_{j+1}\mid X_j) \tag{3.5}$$

$$\cdots$$

$$P(X_{i+N}\mid X_iX_{i+1}\cdots X_{i+N-1})=P(X_{j+N}\mid X_jX_{j+1}\cdots X_{j+N-1}) \tag{3.6}$$

即离散平稳信源的条件概率分布均与时间起点无关,只与关联长度 N 有关.

这样,很容易推出:

$$H(X_1)=H(X_2)=\cdots=H(X_N) \tag{3.7}$$

$$H(X_2\mid X_1)=H(X_3\mid X_2)=\cdots=H(X_N\mid X_{N-1}) \tag{3.8}$$

$$H(X_3\mid X_1X_2)=H(X_4\mid X_2X_3)=\cdots=H(X_N\mid X_{N-2}X_{N-1}) \tag{3.9}$$

$$\cdots$$

对于离散单符号信源,用信息熵来表示信源的平均不确定性.对于离散多符号信源,怎样表示信源的平均不确定性呢?我们引入熵率的概念,它表示信源输出的符号序列中,平均每个符号所携带的信息量.

定义 3.2 随机变量序列中,对前 N 个随机变量的联合熵求平均:

$$H_N(X)=\frac{1}{N}H(X_1X_2\cdots X_N) \tag{3.10}$$

称为平均符号熵.如果当 $N\to\infty$ 时上式极限存在,则 $\lim\limits_{N\to\infty}H_N(X)$ 称为熵率,或称为极限熵,记为

$$H_\infty=\lim_{N\to\infty}H_N(X) \tag{3.11}$$

3.3.1 离散平稳无记忆信源

一般情况下,信源输出序列中每一位出现什么符号是随机的,但是前后符号的出现有一定的统计关系.简单起见,先假定消息符号序列中前后符号的出现是无关的,即首先讨论无记忆信源.

离散平稳无记忆信源输出的符号序列是平稳随机序列,并且符号之间是无关的,即统计独立的.为了研究离散平稳无记忆信源的熵率,假定信源每次输出的是 N 长符号序列,这可以看作是一个新信源,称为离散平稳无记忆信源的 N 次扩展信源,它的数学模型是

N 维离散随机变量序列(随机矢量): $X = X_1 X_2 \cdots X_N$,其中每个随机变量之间统计独立. 同时,由于是平稳信源,每个随机变量的统计特性都相同,因此还可以把一个输出 N 长符号序列的信源记为: $X = X_1 X_2 \cdots X_N = X^N$.

根据统计独立的多维随机变量的联合熵与信息熵之间的关系,可以推出:

$$H(X) = H(X^N) = N H(X) \tag{3.12}$$

即 N 次扩展信源的熵等于单符号离散信源熵的 N 倍,信源输出的 N 长符号序列平均提供的信息量是单符号离散信源平均每个符号所提供信息量的 N 倍.这似乎很好理解,比如,抛掷一枚均匀硬币的试验每次可以得到 1 bit 的信息量,抛掷 N 枚均匀硬币的试验则可以得到 N bit 的信息量.

离散平稳无记忆信源的熵率:

$$H_\infty = \lim_{N\to\infty} H_N(X) = \lim_{N\to\infty} \frac{1}{N} \cdot NH(X) = H(X) \tag{3.13}$$

【例 3.2】

设有一离散无记忆信源 X,其概率空间为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

求该信源的熵率及其二次扩展信源(信源每次输出两个符号)的熵.

解

单符号离散信源熵

$$H(X) = -\sum_{i=1}^{q} p_i \log_2 p_i = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 = 3/2 \text{ 比特/符号}$$

二次扩展信源的熵　　$H(X) = 2H(X) = 3$ 比特/二个符号

注意, $H(X)$ 的单位在这里是“比特/二个符号”,其中每个符号提供的信息量仍然是 1.5 bit.

熵率　　$H_\infty = \lim_{N\to\infty} H_N(X) = \lim_{N\to\infty} \frac{1}{N} \times NH(X) = 3/2$ 比特/符号

3.3.2　离散平稳有记忆信源

前面讲了离散平稳信源中最简单的离散平稳无记忆信源,而实际信源往往是有记忆信源.假定信源输出 N 长的符号序列,则它的数学模型是 N 维随机变量序列(随机矢量): $X = X_1, X_2, \cdots, X_N$,其中每个随机变量之间存在统计依赖关系.

对于相互间有依赖关系的 N 维随机变量的联合熵可以用式(3.14)表示,称为熵函数的链规则:

$$H(X) = H(X_1 X_2 \cdots X_N)$$
$$= H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1 X_2) + \cdots + H(X_N \mid X_1 X_2 \cdots X_{N-1}) \tag{3.14}$$

即 N 维随机变量的联合熵等于起始时刻随机变量 X_1 的熵与各阶条件熵之和.

定理 3.1 对于离散平稳信源,有以下几个结论:

(1) 条件熵 $H(X_N \mid X_1 X_2 \cdots X_{N-1})$ 随 N 的增加是递减的;

(2) N 给定时平均符号熵大于等于条件熵,即

$$H_N(X) \geq H(X_N \mid X_1 X_2 \cdots X_{N-1}) \tag{3.15}$$

(3) 平均符号熵 $H_N(X)$ 随 N 的增加是递减的;

(4) 如果 $H(X_1) < \infty$,则 $H_\infty = \lim_{N\to\infty} H_N(X)$ 存在,并且

$$H_\infty = \lim_{N\to\infty} H_N(X) = \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1}) \tag{3.16}$$

证明

(1) $H(X_N \mid X_1 X_2 \cdots X_{N-1}) \leq H(X_N \mid X_2 \cdots X_{N-1})$ (条件熵小于等于无条件熵)

$= H(X_{N-1} \mid X_1 X_2 \cdots X_{N-2})$ (序列的平稳性)

所以,条件熵 $H(X_N \mid X_1 X_2 \cdots X_{N-1})$ 随着 N 的增加是递减的.

这表明记忆长度越长,条件熵越小,也就是序列的统计约束关系增加时,不确定性减少.

(2) $N\ H_N(X)$

$= H(X_1 X_2 \cdots X_N)$

$= H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1 X_2) + \cdots + H(X_N \mid X_1 X_2 \cdots X_{N-1})$

$= H(X_N) + H(X_N \mid X_{N-1}) + \cdots + H(X_N \mid X_1 X_2 \cdots X_{N-1})$ (序列的平稳性)

$\geq N\ H(X_N \mid X_1 X_2 \cdots X_{N-1})$ (条件熵小于等于无条件熵)

所以,$H_N(X) \geq H(X_N \mid X_1 X_2 \cdots X_{N-1})$,即 N 给定时平均符号熵大于等于条件熵.

(3) $N\ H_N(X) = H(X_1 X_2 \cdots X_N)$

$= H(X_N \mid X_1 X_2 \cdots X_{N-1}) + H(X_1 X_2 \cdots X_{N-1})$

$= H(X_N \mid X_1 X_2 \cdots X_{N-1}) + (N-1) H_{N-1}(X)$

$\leq H_N(X) + (N-1) H_{N-1}(X)$ (利用式(3.15)的结果)

所以,$H_N(X) \leq H_{N-1}(X)$,即序列的统计约束关系增加时,由于符号间的相关性,平均每个符号所携带的信息量减少.

(4) 只要 X_1 的样本空间是有限的,则必然 $H(X_1) < \infty$.因此,$0 \leq H(X_N \mid X_1 X_2 \cdots X_{N-1}) \leq H(X_{N-1} \mid X_1 X_2 \cdots X_{N-2}) \leq \cdots \leq H(X_1) < \infty$,所以,$H(X_N \mid X_1 X_2 \cdots X_{N-1})$,$N = 1, 2, \cdots$ 是单调有界数列,极限 $\lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1})$ 必然存在,且极限为 0 和 $H(X_1)$ 之间的某一值.

对于收敛的实数列,有以下结论成立:

如果 $a_1, a_2, a_3, \cdots$ 是一个收敛的实数列,那么

$$\lim_{N\to\infty}\frac{1}{N}(a_1 + a_2 + \cdots + a_N) = \lim_{N\to\infty} a_N \tag{3.17}$$

利用式(3.17)可以推出:

$$\lim_{N\to\infty} H_N(X) = \lim_{N\to\infty}\frac{1}{N}[H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1 X_2) + \cdots + H(X_N \mid X_1 X_2 \cdots X_{N-1})]$$
$$= \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1})$$

证毕

该定理表明,由于信源输出序列前后符号之间的统计依赖关系,随着序列长度 N 的增加,也就是随着统计约束条件不断增加,平均符号熵 $H_N(X)$ 及条件熵 $H(X_N \mid X_1 X_2 \cdots X_{N-1})$ 均随之减小.当 $N\to\infty$ 时,$H_N(X) = H(X_N \mid X_1 X_2 \cdots X_{N-1})$,即为熵率,它表示信源输出的符号序列中,平均每个符号所携带的信息量.所以在求熵率时可以有两种途径:可以求它的极限平均符号熵,也可以求它的极限条件熵,即

$$H_\infty = \lim_{N\to\infty}\frac{1}{N}H(X_1 X_2 \cdots X_N) = \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1})$$

一般情况下,平稳信源输出的符号序列中,符号之间的相关性可以追溯到最初的一个符号,比如,一篇文章的最后一句话可以一直追溯到和开篇第一句话相关.要准确地计算出这个熵率,必须测定信源的无穷维联合概率和条件概率分布,这是相当困难的.为了简化分析,往往用 N 不太大时的平均符号熵或条件熵作为熵率的近似值.比如,英语的熵率通常用记忆长度为 5 个字母的条件熵近似.

而有一类信源,它在某时刻发出的符号仅与在此之前发出的有限个符号有关,而与更早些时候发出的符号无关,这称为马尔可夫性,这类信源称为马尔可夫信源.马尔可夫信源可以在 N 不很大时得到 H_∞.如果信源在某时刻发出的符号仅与在此之前发出的 m 个符号有关,则称为 m 阶马尔可夫信源,它的熵率:

$$H_\infty = \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1})$$
$$= \lim_{N\to\infty} H(X_N \mid X_{N-m} X_{N-m+1} \cdots X_{N-1}) \quad (\text{马尔可夫性})$$
$$= H(X_{m+1} \mid X_1 X_2 \cdots X_m) \quad (\text{平稳性}) \tag{3.18}$$

$H(X_{m+1} \mid X_1 X_2 \cdots X_m)$ 通常记作 H_{m+1}.

【例 3.3】

信源 X 的信源模型为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{4} & \frac{4}{9} & \frac{11}{36} \end{bmatrix}$$

输出符号序列中,只有前后两个符号有记忆,条件概率 $P(X_2 \mid X_1)$ 列于表 3.2.

表 3.2 条件概率 $P(X_2 \mid X_1)$

X_1 \ X_2	x_1	x_2	x_3
x_1	$\frac{7}{9}$	$\frac{2}{9}$	0
x_2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$
x_3	0	$\frac{2}{11}$	$\frac{9}{11}$

求熵率,并比较 $H(X_2 \mid X_1)$、$\frac{1}{2}H(X_1X_2)$ 和 $H(X)$的大小.

解

熵率: $H_\infty = H_2 = H(X_2 \mid X_1) = 0.870$ 比特/符号

如果不考虑符号间的相关性,则信源熵为

$$H(X) = \frac{1}{4}\log_2 4 + \frac{4}{9}\log_2 \frac{9}{4} + \frac{11}{36}\log_2 \frac{36}{11} = 1.542 \text{ 比特/符号}$$

可见, $H(X_2 \mid X_1) < H(X) = H(X_2)$,这是由于 X_1 和 X_2 之间存在统计依赖关系,在 X_1 已知的情况下, X_2 的不确定度减少,即条件熵 $H(X_2 \mid X_1)$小于无条件熵 $H(X_2)$. 因此,在考虑序列符号之间的相关性之后,序列的熵减小.

如果把信源输出的符号序列看成是分组发出的,每两个符号作为一组,这样,可以把符号序列看成是由一个新信源发出的,新信源每次发出的是由两个符号构成的消息.新信源的数学模型是一个二维随机变量,新信源的熵为

$$H(X_1X_2) = H(X_1) + H(X_2 \mid X_1) = 1.542 + 0.870 = 2.412 \text{ 比特/两个符号}$$

平均符号熵为

$$\frac{1}{2}H(X_1X_2) = 1.206 \text{ 比特/符号}$$

可见, $H(X_2 \mid X_1) < \frac{1}{2}H(X_1X_2) < H(X)$,这是因为 $H(X_1X_2)$考虑了同一组的两个符号之间的相关性,因此 $H(X_1X_2)$小于不考虑符号间相关性时的信源熵 $H(X)$,但是 $H(X_1X_2)$没有考虑前一组的后一符号与后一组的前一符号之间的关联,因此 $H(X_1 \mid X_2) < \frac{1}{2}H(X_1X_2)$.

3.3.3 马尔可夫信源

前面讨论了离散平稳信源的熵率,由于符号间的相关性可以追溯到很远,使得熵率的计算比较复杂.

马尔可夫信源是一类相对简单的有记忆信源,信源在某一时刻发出某一符号的概率除与该符号有关外,只与此前发出的有限个符号有关.例如 m 阶马尔可夫信源只与前面发出的 m 个符号有关,而 1 阶马尔可夫信源只与前面 1 个符号有关.因此把前面若干个符号看作一个状态(若信源有 q 个可能的输出符号,则一共有 q^m 个可能的状态),可以认为,信源在某一时刻发出某一符号的概率除了与该符号有关外,只与该时刻信源所处的状态有关,而与过去的状态无关.信源发出一个符号后,信源所处的状态即发生改变,这些状态的变化组成了马氏链.因此可以把对马尔可夫信源的研究转化为对马氏链的研究.

如图 3.1 所示,信源在某时刻处于某一状态 s_i,当它发出一个符号 $x_{i_{m+1}}$ 后,所处的状态就变了,转移到状态 s_j,因此,信源输出的符号序列 $X_1 X_2 \cdots X_m X_{m+1} \cdots$ 变换成信源状态序列 $S_1 S_2 \cdots S_L S_{L+1} \cdots$,于是一个讨论信源输出符号不确定性的问题变成讨论信源状态转换的问题.

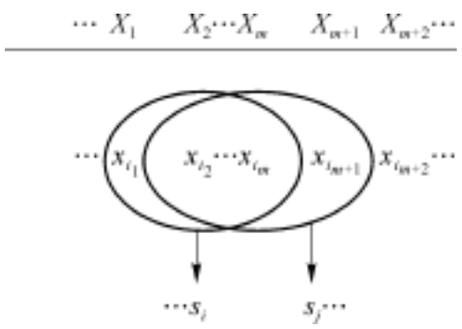


图 3.1　马尔可夫信源

状态之间的一步转移概率 $p(s_j \mid s_i) = P_r\{S_{L+1} = s_j \mid S_L = s_i\}$ 表示前一时刻(即 L 时刻)信源处于 s_i 状态下,在下一时刻(即 $(L+1)$ 时刻)信源处于 s_j 状态的概率.可以用马尔可夫链的状态转移图来描述离散马尔可夫信源的状态转移概率.

【例 3.4】

设一个二元一阶马尔可夫信源,信源符号集为 $X = \{0, 1\}$,信源输出符号的条件概率为

$$p(0|0) = 0.25,\ p(0|1) = 0.5,\ p(1|0) = 0.75,\ p(1|1) = 0.5$$

求状态转移概率.

解

这里 $q = 2$, $m = 1$, $q^m = 2$,共有两种状态: $s_1 = 0$, $s_2 = 1$.由信源输出符号的条件概率可求得马尔可夫链的状态转移概率:

$$p(s_1 \mid s_1) = 0.25,\ p(s_1 \mid s_2) = 0.5,\ p(s_2 \mid s_1) = 0.75,\ p(s_2 \mid s_2) = 0.5$$

信源的状态转移概率还可以用如图 3.2 所示的状态转移图表示.

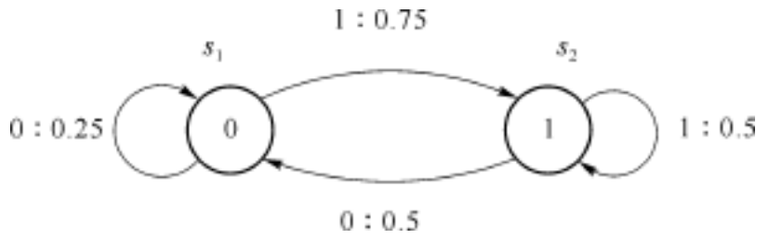


图 3.2　一阶马尔可夫信源状态转移图

对于一阶马尔可夫信源,它的状态转移概率和信源输出符号的条件概率(即符号转移

概率)相同 .

【例 3 .5】

设有一个二元二阶马尔可夫信源,其信源符号集为 $X=\{0,1\}$,输出符号的条件概率为 $p(0|00)=p(1|11)=0.8$, $p(0|01)=p(0|10)=p(1|01)=p(1|10)=0.5$, $p(1|00)=p(0|11)=0.2$.

求状态转移概率矩阵 .

解

这里 $q=2$, $m=2$,故共有 $q^m=4$ 个可能的状态: $s_1=00$, $s_2=01$, $s_3=10$, $s_4=11$.但由于信源只可能发出 0 或 1,所以信源下一时刻只可能转移到其中的两种状态之一 .比如,如果信源原来所处状态为 $s_1=00$,则下一时刻信源只可能转移到 00 或 01 状态,而不会转移到 10 或 11 状态 .

由输出符号的条件概率容易求得状态转移概率:

$p(s_1|s_1)=p(s_4|s_4)=0.8$, $p(s_3|s_2)=p(s_1|s_3)=p(s_4|s_2)=p(s_2|s_3)=0.5$, $p(s_2|s_1)=p(s_3|s_4)=0.2$

其余状态转移概率为 0,该信源的状态转移图如图 3 3 所示 .

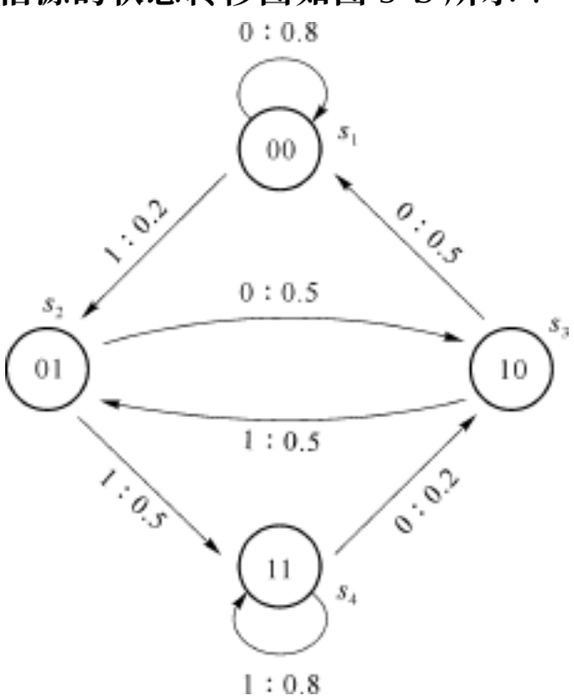


图 3 3 二阶马尔可夫信源状态转移图

信源的状态转移概率矩阵为

$$P=\begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix}$$

对于一个 m 阶马尔可夫信源,它的概率空间可以用它的所有可能的输出符号及输出符号的条件概率表示:

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_q \\ p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m}) \end{bmatrix}$$

令 $s_i = x_{i_1} x_{i_2} \cdots x_{i_m}$, $i_1, i_2, \cdots, i_m \in \{1,2,\cdots,q\}$, 则由信源输出符号的条件概率 $p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m})$可以确定状态转移概率 $p(s_j \mid s_i)$, $i, j \in \{1,2,\cdots,q^m\}$, 从而得到马尔可夫信源的状态空间:

$$\begin{bmatrix} s_1 & \cdots & s_i & \cdots & s_j & \cdots & s_{q^m} \\ & & p(s_j \mid s_i) & & & & \end{bmatrix}$$

状态空间由所有状态及状态间的状态转移概率组成.因此通过引入状态转移概率,可以把对马尔可夫信源的研究转化为对马尔可夫链的研究.

下面主要研究遍历的 m 阶马尔可夫信源的熵率.

当时间足够长后,遍历的马尔可夫信源可以视作平稳信源来处理,又因为 m 阶马尔可夫信源发出的符号只与最近的 m 个符号有关,所以

$$\begin{aligned} H_\infty &= \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1}) \\ &= \lim_{N\to\infty} H(X_N \mid X_{N-m} X_{N-m+1} \cdots X_{N-1}) \text{ (马尔可夫性)} \\ &= H(X_{m+1} \mid X_1 X_2 \cdots X_m) \text{ (序列的平稳性)} \\ &= H_{m+1} \end{aligned} \tag{3.19}$$

即 m 阶马尔可夫信源的极限熵 H_∞ 等于条件熵 H_{m+1}. H_{m+1} 表示已知前面 m 个符号的条件下,输出下一个符号的平均不确定性.

对于齐次遍历的马尔可夫链,其状态 s_i 由 $x_{i_1} x_{i_2} \cdots x_{i_m}$ 唯一确定,因此有

$$p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m}) = p(x_{i_{m+1}} \mid s_i) = p(s_j \mid s_i) \tag{3.20}$$

所以

$$\begin{aligned} H_{m+1} &= H(X_{m+1} \mid X_1 X_2 \cdots X_m) \\ &= E[p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m})] \\ &= E[p(x_{i_{m+1}} \mid s_i)] \\ &= -\sum_{i=1}^{q^m} \sum_{i_{m+1}=1}^{q} p(s_i)\, p(x_{i_{m+1}} \mid s_i) \log_2 p(x_{i_{m+1}} \mid s_i) \\ &= \sum_i p(s_i) H(X \mid s_i) \qquad (3.21) \\ &= -\sum_i \sum_j p(s_i)\, p(s_j \mid s_i) \log_2 p(s_j \mid s_i) \qquad (3.22) \end{aligned}$$

其中，$p(s_i)$是马尔可夫链的平稳分布或称状态极限概率；$H(X|s_i)$表示信源处于某一状态s_i时发出下一个符号的平均不确定性；$p(s_j|s_i)$表示下一步状态转移概率．

【例 3.6】

求图 3.3 中的二阶马尔可夫信源的极限熵．

解

由图 3.3 可判断，图中的 4 个状态是不可约的非周期常返态，因此是遍历的（参见附录 B.3）．

设状态的平稳分布为 $W=(W_1 \quad W_2 \quad W_3 \quad W_4)$，其中 $W_1=p(s_1)$，$W_2=p(s_2)$，$W_3=p(s_3)$，$W_4=p(s_4)$，根据马尔可夫链遍历的充分条件：$WP=W$，得

$$\begin{cases} 0.8W_1+0.5W_3=W_1 \\ 0.2W_1+0.5W_3=W_2 \\ 0.5W_2+0.2W_4=W_3 \\ 0.5W_2+0.8W_4=W_4 \end{cases}$$

并且满足 $W_1+W_2+W_3+W_4=1$，因此可解得

$$W_1=p(s_1)=5/14,\ W_2=p(s_2)=1/7,\ W_3=p(s_3)=1/7,\ W_4=p(s_4)=5/14$$

所以，

$$\begin{aligned} H_\infty &= H_{m+1} \\ &= H_3 \\ &= \sum_i p(s_i)H(X|s_i) \\ &= \frac{5}{14}H(0.8,0.2)+\frac{1}{7}H(0.5,0.5)+\frac{1}{7}H(0.5,0.5)+\frac{5}{14}H(0.8,0.2) \\ &= 0.80 \text{ 比特/符号} \end{aligned}$$

注意，这时符号的平稳概率分布为

$$p(0)=0.8p(s_1)+0.5p(s_2)+0.5p(s_3)+0.2p(s_4)=0.5$$

$$p(1)=0.2p(s_1)+0.5p(s_2)+0.5p(s_3)+0.8p(s_4)=0.5$$

它与状态的平稳分布是有区别的．

如果不考虑符号间的相关性，则由符号的平稳概率分布可得信源熵 $H(X)=1$ 比特/符号，考虑符号间的相关性后，该信源的熵率为

$$H_\infty = H_{m+1} = H_3 = 0.80 \text{ 比特/符号}$$

3.3.4 信源的相关性和剩余度

在前面几节，讨论了离散平稳信源及其熵率．实际的离散信源可能是非平稳的，对于

非平稳信源来说,其 H_∞ 不一定存在,但为了方便通常假定它是平稳的,用平稳信源的 H_∞ 来近似.然而即使对于一般的离散平稳信源,求 H_∞ 值也是很困难的,那么进一步假定它是 m 阶马尔可夫信源,用 m 阶马尔可夫信源的条件熵 H_{m+1} 来近似(大多数平稳信源都可以用马尔可夫信源来近似,即认为输出符号只与前 m 个符号有关).当 $m=1$ 时是最简单的离散平稳有记忆信源,这时 $H_{m+1}=H_2=H(X_2|X_1)$.若再进一步简化信源模型,则可以假设信源为离散平稳无记忆信源,这时可用单符号离散信源的平均自信息量来近似,$H_1=H(X)$.最后,还可以假定信源输出的符号是等概率分布的,因此可以用最大熵来近似,$H_0=\log_2 q$.所以,对于一般的离散信源,根据研究的目的不同,可以用不同的信源模型来近似.

根据定理 3.1 可得:　$\log_2 q=H_0\geq H_1\geq H_2\geq\cdots\geq H_{m+1}\geq\cdots\geq H_\infty$.

对于一个离散平稳信源,其输出的每个符号实际所携带的平均信息量用熵率 H_∞ 表示.由于信源输出符号间的依赖关系也就是信源的相关性,使信源的 H_∞ 减小.信源输出符号间统计约束关系越长,信源的 H_∞ 越小.当信源输出符号间彼此不存在依赖关系且为等概率分布时,信源的 H_∞ 等于最大熵 H_0.例如,信源符号集有 4 个符号,最大熵为 2 比特/符号,输出一个由 10 个符号构成的符号序列,最多可以包含 $10\times2=20$ bit 的信息量.假如由于符号间的相关性或不等概分布,使信源的极限熵减小到 1.2 比特/符号,输出的符号序列平均所含有的信息量为 $10\times1.2=12$ bit,而如果信源输出符号间没有相关性并且符号等概分布,则输出 12 bit 的信息量只需输出 6 个符号就可以了,说明信源存在剩余.因此引入信源剩余度(冗余度)概念.

定义 3.3　一个信源的熵率(极限熵)与具有相同符号集的最大熵的比值称为熵的相对率.其表达式为

$$\eta=\frac{H_\infty}{H_0} \tag{3.23}$$

信源剩余度为

$$\gamma=1-\eta=1-\frac{H_\infty}{H_0}=1-\frac{H_\infty}{\log_2 q} \tag{3.24}$$

H_0-H_∞ 越大,信源的剩余度越大.

信源的剩余度来自两个方面:一是信源符号间的相关性,相关程度越大,符号间的依赖关系越长,信源的 H_∞ 越小;另一方面是信源输出消息的不等概分布使信源的 H_∞ 减小.当信源输出符号间不存在相关性并且输出消息为等概分布时信源的 H_∞ 最大,等于 H_0.对于一般平稳信源来说,其极限熵 H_∞ 远小于 H_0.传送一个信源的信息实际只需要传送的信息量为 H_∞,如果用二元符号来表示,只需用 H_∞ 个二元符号.为了最有效地传递信源的信息,需要掌握信源全部的概率统计特性,即任意维的概率分布,这显然是不现实的.实际上,往往只能掌握有限 N 维的概率分布,这时需要传送 H_N 个二元符号,与理论值 H_∞ 相比,相当于多传送了 H_N-H_∞ 个二元符号.

为了更经济有效地传送信息,需要尽量压缩信源的剩余度,压缩剩余度的方法就是尽

量减小符号间的相关性,并且尽可能地使信源输出消息等概率分布,在第 5 章无失真信源编码中,会看到具体的信源剩余度压缩方法 .

下面以英文字母为例来说明,信源模型的近似程度不同,计算的信源熵不同 .

① 英文字母共 26 个,加上空格 27 个符号,则最大熵为

$$H_0 = \log_2 27 = 4.76 \text{ 比特/ 符号}$$

② 对在英文书中各字母出现的概率加以统计,可以得到各个字母的概率分布,如表 3 3 所示 .

表 3 3 英文字母概率表

字母	P_i	字母	P_i	字母	P_i
空格	0 .2	S	0 .050 2	Y、W	0 .012
E	0 .105	H	0 .047	G	0 .011
T	0 .072	D	0 .035	B	0 .010 5
O	0 .065 4	L	0 .029	V	0 .008
A	0 .063	C	0 .023	K	0 .003
N	0 .059	F、U	0 .022 5	X	0 .002
I	0 .055	M	0 .021	J、Q	0 .001
R	0 .054	P	0 .017 5	Z	0 .001

因此,如果认为英语字母间是离散无记忆的,根据表中的概率可求得

$$H_1 = -\sum_{i=1}^{27} p(x_i)\log_2 p(x_i) = 4.03 \text{ 比特/ 符号}$$

③ 若考虑前后两个、三个、若干个字母之间存在相关性,则可根据字母出现的条件概率求得:

$$H_2 = 3.32 \text{ 比特/ 符号}$$

$$H_3 = 3.1 \text{ 比特/ 符号}$$

$$\cdots$$

$$H_5 = 1.65 \text{ 比特/ 符号}$$

$$H_\infty = 1.4 \text{ 比特/ 符号} \quad (\text{利用统计推断方法})$$

当考虑 5 个字母间的相关性(也就是约等于英文单词的平均长度 4 .5)时,所计算的信源熵已非常接近英文符号的实际信源熵 H_∞ .

$$\eta = \frac{H_\infty}{H_0} = \frac{1.4}{4.76} = 0.29, \quad \gamma = 1 - \eta = 0.71$$

这说明,写英语文章时,71%是由语言结构定好的,是多余成分,只有 29%是写文章的人可以自由选择的 .直观地说,100 页英文书,理论上看仅有 29 页是有效的,其余 71 页是多余的 .正是由于这一多余量的存在,才有可能对英文信源进行压缩编码 .如果对英文

信源进行恰当地编码,传递或存储这些符号时,可大量压缩篇幅,100 页的英语,大约只要 29 页就行了 .

下面是 5 种语言文字在不同近似程度下的熵,如表 3 .4 所示 .

表 3 .4　5 种文字在不同近似程度下的熵

文字	H_0	H_1	H_2	H_3	…	H_∞	η	γ
英文	4 .7	4 .03	3 .32	3 .1		1 .4	0 .29	0 .71
法文	4 .7					3	0 .63	0 .37
德文	4 .7					1 .08	0 .23	0 .77
西班牙文	4 .7					1 .97	0 .42	0 .58
中文 (按 8 000 汉字计算)	≈13	9 .41	8 .1	7 .7		4 .1	0 .315	0 .685

【例 3 .7】

计算汉字的剩余度 .假设常用汉字约为 10 000 个,其中 140 个汉字出现的概率占 50%, 625 个汉字(含 140 个)出现的概率占 85%,2 400 个汉字(含 625 个)出现的概率占 99 .7%,其余 7 600 个汉字出现的概率占 0 .3%,不考虑符号间的相关性,只考虑它的概率分布,在这一级近似下计算汉字的剩余度 .

解

为了计算方便,假设每类中汉字出现是等概的,即可得表 3 .5 .

表 3 .5　汉字的近似概率表

类别	汉字个数	所占概率	每个汉字的概率
1	140	0 .5	0 .5/ 140
2	625 - 140 = 485	0 .85 - 0 .5 = 0 .35	0 .35/ 485
3	2 400 - 625 = 1 775	0 .997 - 0 .85 = 0 .147	0 .147/ 1 775
4	7 600	0 .003	0 .003/ 7 600

不考虑符号间的相关性,只考虑它的概率分布,因此信源的实际熵近似为 $H(X)$ = 9 .773 比特/ 汉字,而 H_0 = 13 .288 比特/ 汉字,所以

$$\gamma = 1 - \frac{H(X)}{H_0} = 0.264$$

从提高信息传输效率的观点出发,人们总是希望尽量去掉剩余度 .比如发电报,我们都知道尽可能把电文写得简洁些以去除相关性,如“母病愈”三个字的中文电报就可以表达母亲身体情况好转的消息 .

但是从提高抗干扰能力角度来看,却希望增加或保留信源的剩余度,因为剩余度大的消息抗干扰能力强 .比如,收到电文“母亲病 X,身体健康”,很容易把电文纠正为“母亲病

愈,身体健康”,而收到电文“母病 X”我们就不知道对方发的是“母病愈”还是“母病危”.

从第 5 章开始,将讨论信源编码和信道编码.通过讨论,可以进一步理解:信源编码是减少或消除信源的剩余度以提高信息的传输效率,而信道编码则通过增加冗余度来提高信息传输的抗干扰能力.

*3.4 连续信源

3.4.1 连续信源的微分熵

连续随机变量的取值是连续的,一般用概率密度函数来描述其统计特征.

单变量连续信源的数学模型为 $X:\begin{bmatrix} \mathrm{R} \\ p(x) \end{bmatrix}$,并满足 $\int_{\mathrm{R}} p(x)\mathrm{d}x = 1$,R 是实数域,表示 X 的取值范围.

对于取值范围有限的连续信源还可以表示成 $X:\begin{bmatrix} (a,b) \\ p(x) \end{bmatrix}$,并满足 $\int_a^b p(x)\mathrm{d}x = 1$,$(a,b)$ 是 X 的取值范围.

通过对连续变量的取值进行量化分层,可以将连续随机变量用离散随机变量来逼近.量化间隔越小,离散随机变量与连续随机变量越接近.当量化间隔趋于 0 时,离散随机变量就变成了连续随机变量.通过对离散随机变量的熵取极限,可以推导出连续随机变量熵的计算公式.

假定概率密度函数 $p(x)$如图 3.4 所示,我们把连续随机变量 X 的取值分割成 n 个

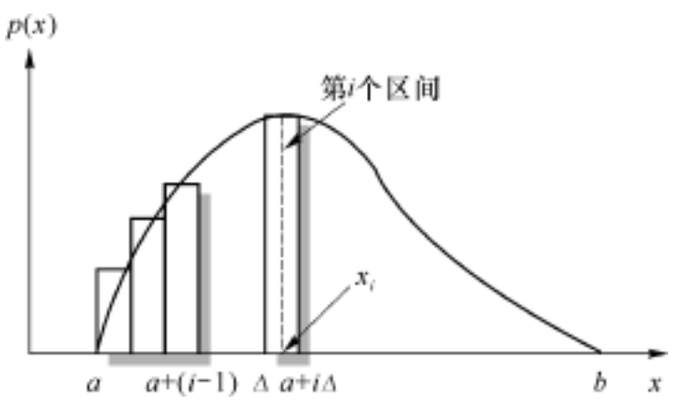


图 3.4 连续随机变量的概率密度函数

小区间,各小区间等宽,区间宽度 $\Delta = \dfrac{b-a}{n}$,则变量落在第 i 个小区间的概率为

$$P_r\{a+(i-1)\Delta \leq x \leq a+i\Delta\} = \int_{a+(i-1)\Delta}^{a+i\Delta} p(x)\mathrm{d}x = p(x_i)\Delta \tag{3.25}$$

其中，x_i 是 $a+(i-1)\Delta$ 到 $a+i\Delta$ 之间的某一值．当 $p(x)$ 是 X 的连续函数时，由中值定理可知，必存在一个 x_i 值使式(3.25)成立，这样，连续变量 X 就可用取值为 x_i，$i=1,2,\cdots,n$ 的离散变量来近似，连续信源就被量化成离散信源，这 n 个取值对应的概率分布为 $p_i = p(x_i)\Delta$，这时的离散信源熵是

$$H(X) = -\sum_{i=1}^{n} p(x_i)\Delta\log_2[p(x_i)\Delta] = -\sum_{i=1}^{n} p(x_i)\Delta\log_2 p(x_i) - \sum_{i=1}^{n} p(x_i)\Delta\log_2\Delta \tag{3.26}$$

当 $n\to\infty$ 时，$\Delta\to 0$，如果式(3.26)极限存在，离散信源熵就变成了连续信源的熵：

$$\lim_{\substack{n\to\infty\\ \Delta\to 0}} H(X) = \lim_{\substack{n\to\infty\\ \Delta\to 0}} -\sum_{i=1}^{n} p(x_i)\Delta\log_2 p(x_i) - \lim_{\substack{n\to\infty\\ \Delta\to 0}} \sum_{i=1}^{n} p(x_i)\Delta\log_2\Delta \tag{3.27}$$

$$= -\int_a^b p(x)\log_2 p(x)\mathrm{d}x - \lim_{\substack{n\to\infty\\ \Delta\to 0}} \log_2\Delta\int_a^b p(x)\mathrm{d}x \tag{3.28}$$

$$= -\int_a^b p(x)\log_2 p(x)\mathrm{d}x - \lim_{\substack{n\to\infty\\ \Delta\to 0}} \log_2\Delta \tag{3.29}$$

式(3.29)第一项一般是定值，第二项为无穷大量，因此连续信源的熵实际是无穷大量．这一点是可以理解的，因为连续信源的可能取值是无限多的，所以它的不确定性是无限大的，当确知输出为某值后，所获得的信息量也是无限大．在丢掉第二项后，定义第一项为连续信源的微分熵：

$$h(X) = -\int_R p(x)\log_2 p(x)\mathrm{d}x \tag{3.30}$$

微分熵又称为差熵．虽然 $h(X)$ 已不能代表连续信源的平均不确定性，也不能代表连续信源输出的信息量，但是它具有和离散熵相同的形式，也具有离散熵的主要特性，比如可加性，但是不具有非负性．另外，我们在实际问题中常常考虑的是熵差，比如平均互信息，在讨论熵差时，只要两者离散逼近时所取的间隔 Δ 一致，这两个无限大量就将互相抵消，所以熵差具有信息的特性，如非负性．由此可见，连续信源的熵 $h(X)$ 具有相对性．

同样，可以定义两个连续随机变量的联合熵：

$$h(XY) = -\iint_{R^2} p(xy)\log_2 p(xy)\mathrm{d}x\mathrm{d}y \tag{3.31}$$

及条件熵

$$h(Y|X) = -\iint_{R^2} p(xy)\log_2 p(y|x)\mathrm{d}x\mathrm{d}y \tag{3.32}$$

$$h(X|Y) = -\iint_{R^2} p(xy)\log_2 p(x|y)\mathrm{d}x\mathrm{d}y \tag{3.33}$$

并且它们之间也有与离散随机变量一样的相互关系：

$$h(XY) = h(X) + h(Y|X) = h(Y) + h(X|Y) \tag{3.34}$$

$$h(X|Y) \leq h(X) \tag{3.35}$$

$$h(Y|X) \leq h(Y) \tag{3.36}$$

【例 3.8】

X 是在区间 (a, b) 内服从均匀分布的连续随机变量，求微分熵。

$$p(x) = \begin{cases} \dfrac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b) \end{cases}$$

解

$$h(X) = -\int_a^b p(x)\log_2 p(x)\,\mathrm{d}x = -\int_a^b \frac{1}{b-a}\log_2 \frac{1}{b-a}\,\mathrm{d}x = \log_2 (b-a)$$

当 $(b-a) > 1$ 时，$h(X) > 0$

当 $(b-a) = 1$ 时，$h(X) = 0$

当 $(b-a) < 1$ 时，$h(X) < 0$

这说明连续熵不具有非负性，失去了信息的部分含义和性质（但是熵差具有信息的特性）。

【例 3.9】

求均值为 m，方差为 σ^2 的高斯分布的随机变量的微分熵。

解

高斯随机变量的概率密度为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}$$

微分熵为

$$\begin{aligned} h(X) &= -\int_{-\infty}^{+\infty} p(x)\log_2 p(x)\,\mathrm{d}x \\ &= -\int_{-\infty}^{+\infty} p(x)\log_2\left[\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}\right]\mathrm{d}x \\ &= -\int_{-\infty}^{+\infty} p(x)\log_2 \frac{1}{\sqrt{2\pi}\sigma} - \log_2 \mathrm{e}\int_{-\infty}^{+\infty} p(x)\left[-\frac{(x-m)^2}{2\sigma^2}\right]\mathrm{d}x \\ &= \log_2 \sqrt{2\pi}\sigma + \log_2 \mathrm{e}\int_{-\infty}^{+\infty} p(x)\frac{(x-m)^2}{2\sigma^2}\mathrm{d}x \\ &= \log_2 \sqrt{2\pi}\sigma + \frac{1}{2}\log_2 \mathrm{e} \\ &= \log_2 \sqrt{2\pi\mathrm{e}}\sigma \end{aligned}$$

这里对数以 2 为底，所得微分熵的单位为比特/样值，如果对数取以 e 为底，则得到

$$h(X) = \log_2 \sqrt{2\pi\mathrm{e}}\sigma \quad \text{奈特/样值}$$

我们看到，正态分布的连续信源的微分熵与数学期望 m 无关，只与方差 σ^2 有关。

【例 3.10】

求指数分布的随机变量的微分熵.

$$p(x)=\begin{cases}\dfrac{1}{a}\mathrm{e}^{-\frac{x}{a}} & x>0\\ 0 & x\leqslant 0\end{cases}$$

解

$$\begin{aligned}h(X)&=-\int_{-\infty}^{+\infty}p(x)\ln p(x)\,\mathrm{d}x\\&=-\int_{0}^{+\infty}p(x)\ln\left[\frac{1}{a}\mathrm{e}^{-\frac{x}{a}}\right]\mathrm{d}x\\&=-\int_{0}^{+\infty}p(x)\ln\frac{1}{a}\,\mathrm{d}x-\int_{0}^{+\infty}p(x)\ln\mathrm{e}^{-\frac{x}{a}}\,\mathrm{d}x\\&=\ln a\int_{0}^{+\infty}p(x)\,\mathrm{d}x+\frac{1}{a}\ln\mathrm{e}\int_{0}^{+\infty}xp(x)\,\mathrm{d}x\quad\left[\int_{0}^{+\infty}xp(x)\,\mathrm{d}x=a,\int_{0}^{+\infty}p(x)\,\mathrm{d}x=1\right]\\&=\ln a+\ln\mathrm{e}\\&=\ln a\mathrm{e}\end{aligned}$$

所以指数分布的相对熵只取决于信源的均值 a.

【例 3.11】

求 N 维高斯信源的熵.

解

把 N 维高斯信源输出的 N 维连续随机矢量记为列向量,则其转置为行向量: $X=(X_1, X_2, \cdots, X_N)^{\mathrm{T}}$,其均值矢量 $M=(m_1, m_2, \cdots, m_N)^{\mathrm{T}}$,协方差矩阵为

$$R=\begin{pmatrix}r_{11} & r_{12} & \cdots & r_{1N}\\ r_{21} & r_{22} & \cdots & r_{2N}\\ \cdots & \cdots & & \cdots\\ r_{N1} & r_{N2} & \cdots & r_{NN}\end{pmatrix}$$

其中, $r_{ij}=E[(X_i-m_i)(X_j-m_j)]$, $i, j=1, 2, \cdots, N$. N 维联合概率密度为 $p(x_1\,x_2\cdots x_N)=\dfrac{1}{(2\pi)^{\frac{N}{2}}|R|^{\frac{1}{2}}}\exp\left[-\dfrac{1}{2}(X-M)^{\mathrm{T}}R^{-1}(X-M)\right]$.

N 维联合熵为

$$\begin{aligned}&h(X_1\,X_2\cdots X_N)\\&=-\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}p(x_1\,x_2\cdots x_N)\ln p(x_1\,x_2\cdots x_N)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_N\\&=-\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}p(x_1\,x_2\cdots x_N)\left[-\ln\sqrt{(2\pi)^N|R|}-\frac{1}{2}(X-M)^{\mathrm{T}}R^{-1}(X-M)\right]\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_N\\&=\frac{1}{2}\ln[(2\pi)^N|R|]+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{1}{2}(X-M)^{\mathrm{T}}R^{-1}(X-M)\,p(x_1\,x_2\cdots x_N)\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_N\\&=\frac{1}{2}\ln[(2\pi)^N|R|]+\frac{N}{2}\end{aligned}$$

当 $X_1, X_2, \cdots, X_N$ 统计独立时，$|R| = \prod_{i=1}^{N} \sigma_i^2$，这时

$$h(X_1 X_2 \cdots X_N) = \frac{1}{2} \sum_{i=1}^{N} \ln \sigma_i^2 + \frac{N}{2} \ln 2\pi + \frac{N}{2}$$

3.4.2 连续信源的最大熵

离散信源当信源符号为等概分布时有最大熵。连续信源微分熵也有极大值，但是与约束条件有关，当约束条件不同时，信源的最大熵不同。我们一般关心的是下面两种约束下的最大熵。

定理 3.1 在输出幅度受限的情况下，服从均匀分布的随机变量 X 具有最大熵。

即

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{其他} \end{cases}$$

$$h(X) = -\int_a^b p(x) \log_2 p(x) \mathrm{d}x = -\int_a^b \frac{1}{b-a} \log_2 \frac{1}{b-a} \mathrm{d}x = \log_2 (b-a)$$

因此对于输出信号幅度受限的连续信源，当满足均匀分布时达到最大熵。这个结论与离散信源在等概分布时达到最大熵的结论类似。

定理 3.2 对于平均功率受限的连续随机变量，当服从高斯分布时具有最大熵。

对于均值为 m，方差为 σ^2 的连续随机变量，平均功率 = 直流功率 + 交流功率，即 $P = m^2 + \sigma^2$。该定理的证明相当于在如下约束条件下求 $h(X)$ 的极值：

$$\int_{-\infty}^{+\infty} p(x) \mathrm{d}x = 1$$

$$\int_{-\infty}^{+\infty} x p(x) \mathrm{d}x = m$$

$$\int_{-\infty}^{+\infty} (x-m)^2 p(x) \mathrm{d}x = \sigma^2$$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2} \right]$$

可以证明当 $h(x)$ 有极大值时，有（参见例 3.9）

$$h(X) = \log_2 \sqrt{2\pi \mathrm{e}} \sigma$$

这说明，当平均功率受限时，高斯分布的连续信源的熵最大，也就是说，高斯信源输出的每个样值（也称自由度）提供的平均信息量最大，其大小随交流功率 σ^2 的变化而变化。

以上定理的证明需要用到数学上的变分法，请有兴趣的同学参考相关书籍。

3.4.3　连续信源的熵功率

与离散信源一样,在讨论了连续信源的最大熵问题之后,也要考虑没有达到最大熵的信源的冗余度问题.从这个角度出发,引出熵功率的概念.我们知道,在不同的约束条件下,连续信源有不同的最大熵,因为均值为零、平均功率受限的连续信源是实际中最常见的一种信源,我们重点讨论这种信源的冗余问题.

均值为零,平均功率限定为 P 的连续信源当服从高斯分布时达到最大熵:

$$h_0(X)=\log_2\sqrt{2\pi e\sigma^2}=\log_2\sqrt{2\pi eP} \tag{3.37}$$

也就是说高斯信源的熵值与 P 有确定的对应关系:

$$P=\frac{1}{2\pi e}e^{2h_0(X)} \tag{3.38}$$

如果另一信源的平均功率也为 P,但不是高斯分布,那么它的熵值 $h(X)$一定比高斯信源的熵 $h_0(X)$小.反过来说,如果有一个信源与这个高斯信源有相同的熵 $h(X)$,则它的平均功率 $P\geq\overline{P}$,$\overline{P}$ 为高斯信源的平均功率,因为对于非高斯信源,$h(X)\leq\log_2\sqrt{2\pi e\overline{P}}$,而对于高斯信源,$h(X)\leq\log_2\sqrt{2\pi e\overline{P}}$.

现在假定某连续信源的熵为 $h(X)$,平均功率为 P,则与它具有相同熵的高斯信源的平均功率 P 定义为熵功率,即

$$P=\frac{1}{2\pi e}e^{2h(X)} \tag{3.39}$$

所以,$P\leq P$,当该连续信源为高斯信源时等号成立.

P 的大小可以表示连续信源剩余度的大小.如果熵功率等于信源平均功率,表示信源没有剩余;熵功率和信源的平均功率相差越大,说明信源的剩余度越大,所以把信源平均功率和熵功率之差$(P-P)$称为连续信源的剩余度.

习　题　3

3.1　证明 $\lim\limits_{n\to\infty}\frac{1}{2}H(X_nX_{n-1}\mid X_1\cdots X_{n-2})=H_\infty$.

3.2　有一无记忆信源的符号集为$\{0,1\}$,已知信源的概率空间为$\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}0&1\\\frac{1}{4}&\frac{3}{4}\end{bmatrix}$.

(1) 求信源熵;

(2) 求由 m 个“0”和$(100-m)$个“1”构成的某一特定序列的自信息量的表达式;

(3) 计算由 100 个符号构成的符号序列的熵.

3.3 有一离散无记忆信源，其输出为 $X\in\{0,1,2\}$，相应的概率为 $p_0=1/4$，$p_1=1/4$，$p_2=1/2$，设计两个独立实验去观察它，其结果分别为 $Y_1\in\{0,1\}$，$Y_2\in\{0,1\}$，已知条件概率如题表 3.1 所列。

题表 3.1

$P(Y_1\mid X)$	0	1
0	1	0
1	0	1
2	$\frac{1}{2}$	$\frac{1}{2}$

$P(Y_2\mid X)$	0	1
0	1	0
1	1	0
2	0	1

(1) 求 $I(X;Y_1)$和 $I(X;Y_2)$，并判断哪一个实验好些；

(2) 求 $I(X;Y_1Y_2)$，并计算做 Y_1 和 Y_2 两个实验比做 Y_1 或 Y_2 中的一个实验各可多得多少关于 X 的信息；

(3) 求 $I(X;Y_1\mid Y_2)$和 $I(X;Y_2\mid Y_1)$，并解释它们的含义。

3.4 某信源的消息符号集的概率分布和二进制代码如题表 3.2 所示。

题表 3.2

信源符号	u_0	u_1	u_2	u_3
概率	1/2	1/4	1/8	1/8
代码	0	10	110	111

(1) 求信源的符号熵；

(2) 求平均每个消息符号所需要的二进制码元的个数或平均代码长度，进而用这一结果求码序列中的二进制码元的熵；

(3) 当消息是由符号序列组成时，各符号之间若相互独立，求其对应的二进码序列中出现“0”和“1”的无条件概率 $p(0)$和 $p(1)$，求相邻码元间的条件概率 $p(0|1)$、$p(1|0)$、$p(1|1)$和 $p(0|0)$。

3.5 二次扩展信源的熵为 $H(X^2)$，而一阶马尔可夫信源的熵为 $H(X_2\mid X_1)$，试比较两者的大小，并说明原因。

3.6 一个马尔可夫过程的基本符号为 0,1,2，这 3 个符号等概率出现，并且具有相同的转移概率。

(1) 画出一阶马尔可夫过程的状态图，并求稳定状态下的一阶马尔可夫信源熵 H_1；

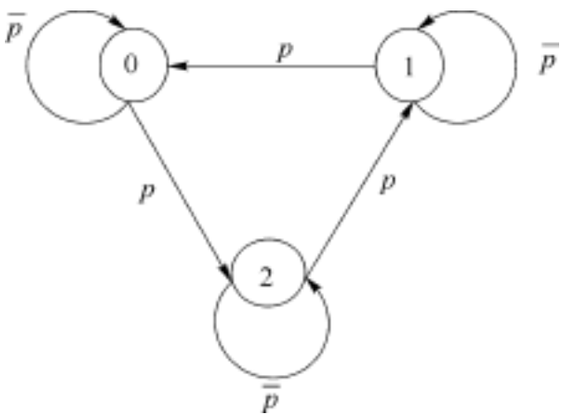
(2) 画出二阶马尔可夫过程的状态图，并求稳定状态下二阶马尔可夫信源熵 H_2 和信源剩余度。

3.7 一阶马尔可夫信源的状态转移图如题图 3.1 所示，信源 X 的符号集为{0,1,2}。

(1) 求平稳后的信源的概率分布;

(2) 求信源熵 H_∞;

(3) 求当 $p=0$ 或 $p=1$ 时信源的熵,并说明其理由.



题图 3.1

3.8　有一个二元无记忆信源,其发 0 的概率为 p,而 $p\approx 1$,所以在发出的二元序列中经常出现的是那些一串为 0 的序列(称高概率序列).对于这样的信源我们可以用另一新信源来代替,新信源中只包含这些高概率序列.这时新信源 $S_n=\{s_1, s_2, s_3, \cdots, s_n, s_{n+1}\}$,共有 $n+1$ 个符号,它与高概率的二元序列的对应关系如下:

二元序列:1,01,001,…, 00…01(共 $n-1$ 个 0),00…000(共 n 个 0);

新信源符号: $s_1, s_2, s_3, \cdots, s_n, s_{n+1}$.

(1) 求 $H(S_n)$;

(2) 当 $n\to\infty$ 时,求信源的熵 $H(S)=\lim\limits_{n\to\infty} H(S_n)$.

3.9　给定状态转移概率矩阵, $P=\begin{pmatrix} 1-p & p \\ 1 & 0 \end{pmatrix}$,求:

(1) 此二状态马尔可夫链的熵率 H_∞;

(2) 此熵率的极大值及相应的 p;

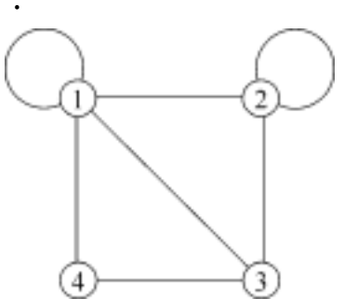
(3) 在达到最大熵率的情况下,求出每一个 n 长序列的概率.

3.10　在一个 3×3 的国际象棋棋盘上,分别计算“王”、“车”、“左象”、“右象”和“后”随机行走的概率.

3.11　题图 3.2 是一张有 4 个节点的随机行走图,从任何一个节点走到下一个节点的概率都相等.

(1) 求随机行走的稳态分布;

(2) 求随机行走的熵率.



题图 3.2

3.12　求具有如下概率密度函数的随机变量的熵.

(1) 指数分布 $f(x)=\lambda e^{-\lambda x}$, $x\geq 0$;

(2) $f(x)=\frac{1}{2}\lambda e^{-\lambda|x|}$;

(3) 单边高斯密度 $f(x)=\frac{2}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}$, $x\geq 0$.

3.13 连续随机变量 X 和 Y 的联合概率密度为

$$p(xy)=\frac{1}{2\pi\sqrt{SN}}\exp\left\{-\frac{1}{2N}\left[x^2\left(1+\frac{N}{S}\right)-2xy+y^2\right]\right\}$$

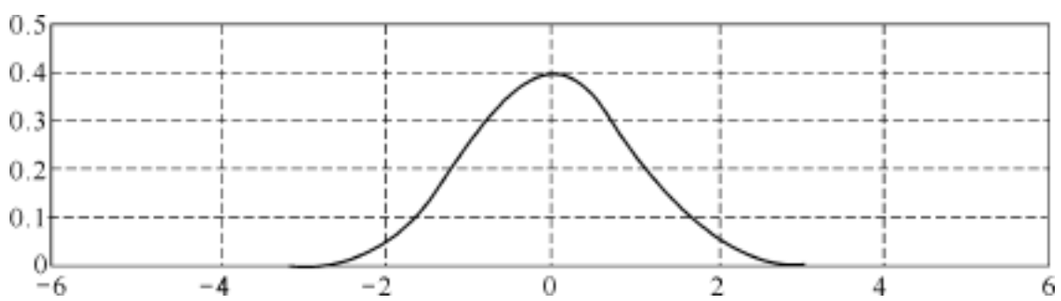
试求 $h(X)$, $h(Y)$, $h(Y|X)$ 和 $I(X;Y)$.

3.14 一信源产生的时不变波形信号(即信号统计特性不随时间而变)的带宽 $W=4$ kHz,幅度分布为

$$p(x)=e^{-x},\ x\geq 0$$

设在信号幅度 0～2 区间按量化单位 $\Delta=0.5$ 做量化,试求该信源的信息输出率 .

3.15 随机变量 X 和 Y 的联合概率密度函数在曲线 $y=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ 和 X 轴所组成的区域内均匀分布,如题图 3.3 所示 .



题图 3.3

(1) 求 $h(X,Y)$;

(2) 求 $h(X)$;

(3) Y 的概率密度函数为 $f(y)=2\sqrt{-2\ln(y\sqrt{2\pi})}$, $0<y\leq 1/\sqrt{2\pi}$.证明:

$$-\frac{1}{2}\ln 2\pi-\frac{1}{2}<h(Y)<-\frac{1}{2}\ln 2\pi .$$

3.16 给定状态转移概率矩阵, $P=\begin{pmatrix}1-\alpha & \alpha\\ \beta & 1-\beta\end{pmatrix}$, 求:此二状态马尔可夫信源的熵率 H_∞ .

3.17 布袋中有手感完全相同的 3 个红球和 3 个蓝球,每次从中随机取出一个球,取出后不放回布袋。用 X_i 表示第 i 次取出的球的颜色, $i=1,2,\cdots,6$, 求:

(1) $H(X_1)$;

(2) $H(X_2)$;

(3) $H(X_2 \mid X_1)$;

(4) 随着 k 的增加, $H(X_k \mid X_1 \cdots X_{k-1})$是增加还是减少？请解释 .

(所有的答案用 $H(p)$ 的形式表示)

3 .18　已知一个二元一阶马尔可夫信源的状态转移概率矩阵为 $P=\begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$.

(1) 求此马尔可夫信源的熵率;

(2) 求符号序列 1000011 的概率(根据平稳分布确定第一个符号的概率);

(3) 计算分布函数 $F(x) = P[(X_1 X_2 \cdots) < x]$当 $x = 1000011$ 时的值 .

第 4 章

信道及信道容量

信道是指信息传输的通道,包括空间传输和时间传输 .在实际通信中所利用的各种物理通道是空间传输信道的最典型的例子,如电缆、光纤、电波的传输空间、载波线路等等;时间传输是指将信息保存,以便以后读取,如磁带、光盘等在时间上将信息进行传输的信道 .有时把为了某种目的而使信息不得不经过的通道也看作信道,例如一个分类器的输入到输出就可以看作是一个信道 这里最关键的是信道有一个输入以及一个与输入有关的输出 .至于信道本身的物理结构,则可能是千差万别的,最简单的如一个放大器的输入到输出,而复杂的如一条国际通信线路,其中可能包括终端设备、电缆、微波等等 .信息论研究的信道其输入点和输出点在一个实际物理通道中所处位置的选择完全取决于研究的目的 .例如,通信中可以把发送天线到接收天线之间的通道看成信道,也可以把从话机到话机之间的通道看作信道 .

关于信道的主要问题有:

信道的建模(信道的统计特性的描述);

信道传输信息的能力(信道容量)的计算;

在有噪信道中能不能实现可靠传输?怎样实现可靠传输?

在这一章要回答前两个问题,在第 6 章介绍第三个问题 .我们将按信道的分类介绍它们的数学模型及信道容量的计算 .

4.1　信道的分类

在通信中,信道按其物理组成常被分成微波信道、光纤信道、电缆信道等,这种分类是因为信号在这些信道中传输时遵循不同的物理规律,而通信技术必须研究这些规律以获得信号在这些信道中的传输特性.信息论不研究怎样获得这些传输特性,而假定传输特性是已知的,并在此基础上研究信息的传输问题.

信息论不研究信号在信道中传输的物理过程,它假定信道的传输特性是已知的,这样信道就可以用图 4.1 所示的抽象的数学模型来描述.由于信道输入随机变量 X 和输出随机变量 Y 往往不是确定关系,在信息论中,信道用在输入已知的情况下输出的条件概率分布 $P(Y|X)$ 来表示,加上输入随机变量 X 和输出随机变量 Y,通常表示成:$\{X, P(Y|X), Y\}$.

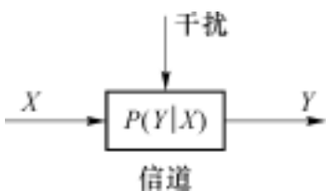


图 4.1　信道模型

根据实际应用的需要,信道有几种分类方法:

(1) 按其输入/输出信号在幅度和时间上的取值是离散或连续来划分

此分类方式如表 4.1 所示:

表 4.1　按输入/输出信号在幅度和时间上是离散或连续划分

幅度	时间	信道名称
离散	离散	离散信道(数字信道)
连续	离散	连续信道
连续	连续	模拟信道(波形信道)
离散	连续	(理论和实用价值均很小)

(2) 按其输入/输出信号之间关系的记忆特性来划分

可分为有记忆信道和无记忆信道.如果信道的输出只与信道该时刻的输入有关而与其他时刻的输入无关,则称此信道是无记忆的,反之称为有记忆的.实际信道一般都是有记忆的,信道中的记忆现象来源于物理信道中的惯性,如电缆信道中的电感电容、无线信道中的电波传播的衰落现象等.有记忆信道的分析比较复杂,有用的研究成果很少,因此,我们主要研究无记忆信道.

(3) 按输入/输出信号之间的关系是否确定来划分

可分为有噪声信道和无噪声信道.一般来说,因为信道中总是存在某种程度的噪声,所以信道输入/输出之间的关系是一种统计依赖的关系.但是当噪声与信号相比很小时可以近似为无噪声信道.有噪声信道是信息论研究的主要对象.信道输入、输出以及信道输入/输出信号之间的统计关系的描述就构成了有噪声信道的数学模型.

(4) 根据信道输入和输出的个数来划分

① 两端信道(单用户信道):只有一个输入端和一个输出端的单向通信的信道 .

② 多端信道(多用户信道):双向通信或三个或更多个用户之间相互通信的情况 .

本课程主要研究两端信道的情况 .

(5) 根据信道的统计特性是否随时间变化来划分

① 恒参信道(平稳信道):信道的统计特性不随时间变化 .卫星通信信道在某种意义上可以近似为恒参信道 .

② 随参信道(非平稳信道):信道的统计特性随时间变化 .如短波通信中,其信道可看成随参信道 .

本课程主要研究恒参信道的情况 .

4.2 离散单符号信道及其信道容量

4.2.1 离散单符号信道的数学模型

信道的输入、输出都取值于离散符号集,且都用一个随机变量来表示的信道就是离散单符号信道 .它是最简单的信道,是实际信道的基本组成单元 .

设离散单符号信道的输入随机变量为 X,其所有可能的取值为 x_i, $i=1,2,\cdots,r$,输出随机变量为 Y,其所有可能的取值为 y_j, $j=1,2,\cdots,s$,由于信道中存在干扰,因此输入符号在传输中会产生错误,这种信道干扰对传输的影响可用传递概率 $p(y_j \mid x_i)$ 来描述:

$$p(y_j \mid x_i) = P_r\{Y=y_j \mid X=x_i\} \qquad i=1,2,\cdots,r\ ;\ j=1,2,\cdots,s$$

信道传递概率实际上是一个传递概率矩阵,称为信道矩阵,记为

$$P=\begin{array}{c} \\ x_1 \\ x_2 \\ \cdots \\ x_r \end{array}\begin{array}{c} \begin{array}{cccc} y_1 & y_2 & \cdots & y_s \end{array} \\ \begin{bmatrix} p(y_1 \mid x_1) & p(y_2 \mid x_1) & \cdots & p(y_s \mid x_1) \\ p(y_1 \mid x_2) & p(y_2 \mid x_2) & \cdots & p(y_s \mid x_2) \\ \cdots & \cdots & & \cdots \\ p(y_1 \mid x_r) & p(y_2 \mid x_r) & \cdots & p(y_s \mid x_r) \end{bmatrix} \end{array}$$

为了表述简便,常常写成

$$P=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \cdots & \cdots & & \cdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix}$$

并且传递概率满足 $p_{ij} \geq 0$，$\sum_{j=1}^{s} p_{ij} = 1$，$i = 1, 2, \cdots, r$，即信道矩阵中每个元素均为非负，每一行元素之和为 1 .

最常见的信道是二元对称信道 BSC(Binary Symmetric Channel)，如图 4 .2 所示，输入符号集和输出符号集分别为 $X = \{0, 1\}$ 和 $Y = \{0, 1\}$.其信道传递概率为

$p(y_1 \mid x_1) = p(0 \mid 0) = 1 - p = $ 琯，$p(y_2 \mid x_1) = p(1 \mid 0) = p$

$p(y_2 \mid x_2) = p(1 \mid 1) = 1 - p = $ 琯，$p(y_1 \mid x_2) = p(0 \mid 1) = p$

其中，琯 表示单个符号无错误传输的概率，p 表示单个符号传输发生错误的概率 所以二元对称信道(BSC)的信道矩阵为 $P = \begin{pmatrix} 琯 & p \\ p & 琯 \end{pmatrix}$，它满足 $\sum_{j=1}^{2} p(y_j \mid x_1) = \sum_{j=1}^{2} p(y_j \mid x_2) = 1$.

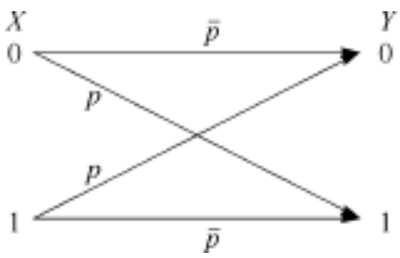


图 4 .2　二元对称信道

下面推导一般离散单符号信道的一些概率关系:

设信道输入随机变量的概率空间为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_2 & \cdots & x_r \\ p(x_1) & \cdots & p(x_2) & \cdots & p(x_r) \end{bmatrix}$$

并且 $\sum_{i=1}^{r} p(x_i) = 1, 0 \leq p(x_i) \leq 1$，$i = 1, 2, \cdots, r$.又设输出符号集为 $Y = \{y_1, y_2, \cdots, y_s\}$，给定信道矩阵为

$$P = \begin{bmatrix} p(y_1 \mid x_1) & p(y_2 \mid x_1) & \cdots & p(y_s \mid x_1) \\ p(y_1 \mid x_2) & p(y_2 \mid x_2) & \cdots & p(y_s \mid x_2) \\ \cdots & \cdots & & \cdots \\ p(y_1 \mid x_r) & p(y_2 \mid x_r) & \cdots & p(y_s \mid x_r) \end{bmatrix}$$

(1) 输入输出随机变量的联合概率分布为 $P_r\{X = x_i, Y = y_i\} = p(x_i y_j)$，则有

$$p(x_i y_j) = p(x_i)\, p(y_j \mid x_i) = p(y_j)\, p(x_i \mid y_j) \tag{4 .1}$$

其中，$p(y_j \mid x_i)$ 是信道传递概率，即输入为 x_i，通过信道传输输出 y_j 的概率，称为前向概率，通常用它描述信道噪声的特性 .而 $p(x_i \mid y_j)$ 是已知信道输出符号 y_j 时，输入符号为 x_i 的概率，称为后向概率 .有时把 $p(x_i)$ 称为输入符号的先验概率(在接收到输出符号之前，判断输入符号为 x_i 的概率)，而对应地把 $p(x_i \mid y_j)$ 称为输入符号的后验概率(接收到输出符号 y_j 之后，判断输入符号为 x_i 的概率) .

(2) 由全概率公式，可从先验概率和信道传递概率求输出符号的概率:

$$p(y_j) = \sum_{i=1}^{r} p(x_i)\, p(y_j \mid x_i) \tag{4 .2}$$

写成向量的形式:$[p(y_1), p(y_2), \cdots, p(y_s)] = [p(x_1), p(x_2), \cdots, p(x_r)] \cdot P$ 或记成 $P_Y = P_X P_{Y|X}$.

(3) 根据贝叶斯公式,可由先验概率和信道的传递概率求后向概率:

$$p(x_i | y_j) = \frac{p(x_i y_j)}{p(y_j)} = \frac{p(x_i)\, p(y_j | x_i)}{\sum_{i=1}^{r} p(x_i)\, p(y_j | x_i)} \quad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s \tag{4.3}$$

且 $\sum_{i=1}^{r} p(x_j | y_i) = 1, \quad j = 1, 2, \cdots, s$.

4.2.2 信道容量的概念

平均互信息 $I(X; Y)$是接收到输出符号集 Y 后所获得的关于输入符号集 X 的信息量.信源的不确定性为 $H(X)$,由于干扰的存在,接收端收到 Y 后对信源仍然存在的不确定性为 $H(X|Y)$, $H(X|Y)$又称为信道疑义度.信宿所消除的关于信源的不确定性,也就是获得的关于信源的信息为 $I(X; Y)$,它是平均意义上每传送一个符号流经信道的信息量,从这个意义上来说,平均互信息 $I(X; Y)$又称为信道的信息传输率,通常用 R 表示.即

$$R = I(X; Y) = H(X) - H(X|Y) \text{ 比特/符号} \tag{4.4}$$

有时我们所关心的是信道在单位时间内平均传输的信息量.如果平均传输一个符号为 t 秒,则信道平均每秒钟传输的信息量为

$$R_t = \frac{1}{t} I(X; Y) \text{ 比特/秒} \tag{4.5}$$

一般称为信息传输速率 .

$I(X; Y)$是信源概率分布 $p(x_i)$和信道转移概率 $p(y_j | x_i)$的二元函数,当信道特性 $p(y_j | x_i)$固定后, $I(X; Y)$是 $p(x_i)$的一元函数,并且由定理 2.1 可知,对于给定的信道转移概率 $p(y_j | x_i)$, $I(X; Y)$是输入分布 $p(x_i)$的上凸函数.因此对于固定的信道,总存在一种信源(某种输入概率分布)使信道传输一个符号接收端获得的平均信息量最大,也就是说对于每个固定信道都有一个最大的信息传输率,这个最大的信息传输率即为信道容量,而相应的输入概率分布称为最佳输入分布 .

因此,对于某一个固定信道,必然有一个最佳输入分布使 $I(X; Y)$得到极大值.因此信道容量是信道转移概率的函数,它是由信道的统计特性决定的,是信道的最大信息传输率 .

定义 4.1 信道容量为平均互信息对于输入概率分布的最大值:

$$C = \max_{p(x)} I(X; Y) \tag{4.6}$$

单位依所用的对数底不同可以是比特/符号、奈特/符号等 .

若平均传输一个符号需要 t 秒钟,则信道在单位时间内平均传输的最大信息量为

$$C_t = \max_{p(x)} I(X;Y) \tag{4.7}$$

信道容量是信道传送信息的最大能力的度量,信道实际传送的信息量必然不大于信道容量 如果待传输的信息量大于信道容量,则在传送过程中将会发生错误 .这是第 6 章信道编码定理即香农第二定理的内容 .

下面以二元对称信道为例来说明信道容量与输入概率分布和信道转移概率的关系 .

【例 4 .1】

输入概率分布 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \omega & 1-\omega \end{bmatrix}$,信道矩阵为 $P = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$, p 为信道错误传递概率 .

求二元对称信道的信道容量 .

解

二元对称信道的平均互信息　$I(X;Y) = H(Y) - H(Y|X) = H(\omega(1-p) + (1-\omega)p) - H(p)$

固定信道时, p 是一个固定常数, $I(X;Y)$是输入概率分布的上凸函数,因此存在一个关于 ω的极大值,当 $\omega = 1-\omega = \frac{1}{2}$时, $H(\omega(1-p) + (1-\omega)p) = H\left(\frac{1}{2}\right) = 1$,因而二元对称信道的信道容量 $C = 1 - H(p)$比特/ 符号 .

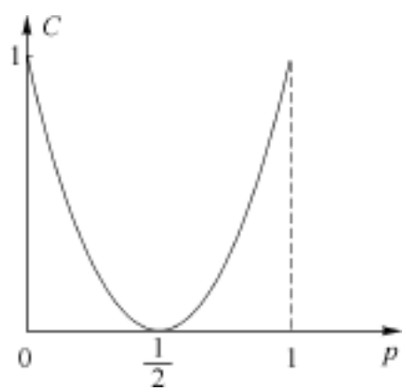


图 4 .3　BSC 的信道容量

由此可见,信道容量 C 仅为信道传递概率 p 的函数,而与信道输入随机变量 X 的概率分布无关 .不同的二元对称信道,其传递概率 p 不同,信道容量也不同 .

图 4 .3 表示不同的二元对称信道,其传递概率 p 不同,信道容量也不同 .

当 $p = 1/2$ 时,是一种最坏的信道,这时 $C = 0$,即该信道不能传递任何信息,信息全部损失在信道中了 .而当 $p = 0$ 或 $p = 1$ 时, $C = 1$,这是最好的情况,信道能够无失真地传送信源信息 .

4 .2 .3　几种特殊信道的信道容量

对于一般信道,求信道容量的计算是非常复杂的,需要对平均互信息 $I(X;Y)$求极大值 .下面先讨论某些特殊类型信道的信道容量,然后再讨论一般离散信道的信道容量的计算 .

1 .具有扩展性能的无损信道

无损信道是一个输入对应多个输出,例如图 4 .4 所示

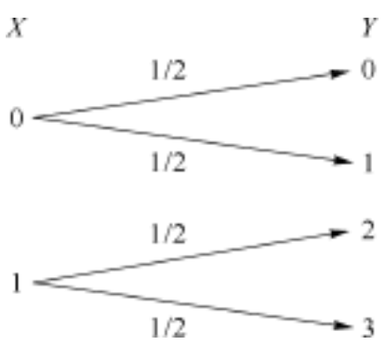


图 4 .4　无损信道

信道,其信道矩阵 $$P=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

无损信道的信道矩阵中每一列只有一个非零元素,接收到信道输出符号后对输入符号将不存在不确定性,即信道疑义度 $H(X|Y)=0$.同时 $H(X|Y)$又表示信源符号通过有噪信道传输后损失的信息量,因为如果没有信息损失的话,信源含有的信息量将全部到达接收端,接收端对信源不再有不确定性,所以 $H(X|Y)$又称为损失熵.对于无损信道,

$$I(X;Y)=H(X)-H(X|Y)=H(X) \tag{4.8}$$

其信道容量

$$C=\max_{p(x)} I(X;Y)=\max_{p(x)} H(X)=\log_2 r \tag{4.9}$$

当信道输入等概分布时,信道达到信道容量.并且由于噪声熵 $H(Y|X)>0$,所以,

$$I(X;Y)=H(X)<H(Y) \tag{4.10}$$

2.具有归并性能的无噪信道

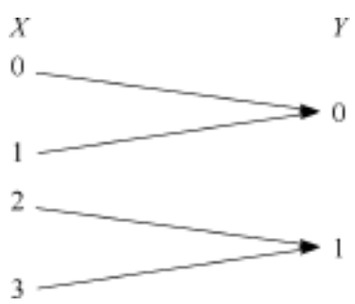


图 4.5 无噪信道

无噪信道是一个输出对应多个输入,例如图 4.5 所示信道,其信道矩阵 $$P=\begin{bmatrix}1 & 0\\ 1 & 0\\ 0 & 1\\ 0 & 1\end{bmatrix}$$

无噪信道每一行只有一个非零元素 1,信道矩阵元素非 0 即 1.已知信道输入符号,必能确定输出符号,因此 $H(Y|X)=0$. $H(Y|X)$又称为噪声熵,因为是信道的噪声使得 $H(Y|X)\neq 0$.

无噪信道的信道容量

$$C=\max_{p(x)} I(X;Y)=\max_{p(x)} H(Y)=\log_2 s \tag{4.11}$$

当信道输出等概分布时,信道达到信道容量.但是输出端接收到某个符号后并不能确定是哪一个输入符号,因而信道疑义度 $H(X|Y)>0$,于是无噪信道的平均互信息

$$I(X;Y)=H(Y)<H(X) \tag{4.12}$$

3.具有一一对应关系的无噪无损信道

无噪无损信道输入、输出之间有确定的一一对应关系,即 $y=f(x)$,其信道传递概率为

$$p(y_j|x_i)=\begin{cases}1 & y_j=f(x_i)\\ 0 & y_j\neq f(x_i)\end{cases}$$

例如图 4.6 所示信道,其信道矩阵 $P=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

图 4.6 无噪无损信道

无噪无损信道每一行、每一列只有一个“1”,已知 X 后对 Y 不存在不确定性,收到 Y 后对 X 也不存在不确定性,所以噪声熵和损失熵均为零.这时

$$I(X;Y)=H(X)=H(Y) \tag{4.13}$$

其信道容量

$$C=\max_{p(x)} I(X;Y)=\max_{p(x)} H(Y)=\max_{p(x)} H(X)=\log_2 s=\log_2 r \tag{4.14}$$

当信道输入等概分布时,输出也为等概分布,信道达到信道容量.

对于以上 3 种信道,求信道容量 C 的问题已经从求 $I(X;Y)$ 的极值问题转化为求 $H(X)$ 或 $H(Y)$ 的极值问题.信道容量 C 只决定于信道的输入符号数 r 或输出符号数 s,与信源无关,它表征信道的统计特性.

4.2.4　离散对称信道的信道容量

离散信道中有一类特殊的信道,其特点是信道矩阵具有行对称性,利用这个对称性可以简化信道容量的计算.

定义 4.2　若信道矩阵中,每行都是第一行元素的不同排列,则称此类信道为行对称信道.

例如 $P=\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{pmatrix}$, $P=\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ 都是行对称信道.

定义 4.3　若信道矩阵中,不仅每行都是第一行元素的不同排列,而且每列都是第一列元素的不同排列,这类信道称为对称信道.

例如 $P=\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ 和 $P=\begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{pmatrix}$ 是对称信道.

定义 4.4　若信道矩阵中,每行都是第一行元素的不同排列,每列并不都是第一列元素的不同排列,但是可以按照信道矩阵的列将信道矩阵划分成若干对称的子矩阵,则称这类信道为准对称信道.

例如,信道矩阵

$$P=\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

可以划分成两个对称的子矩阵

$$P_1=\begin{pmatrix} 0.8 & 0.1 \\ 0.1 & 0.8 \end{pmatrix},\ P_2=\begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}$$

因此它是准对称信道 .

定义 4 .5 若对称信道中输入符号和输出符号个数相同,且信道中总的错误概率为 p,平均分配给 $r-1$ 个输出符号, r 为输入输出符号的个数,即信道矩阵为

$$P=\begin{pmatrix} 1-p & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & 1-p & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \cdots & \cdots & \cdots & & \cdots \\ \frac{p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & 1-p \end{pmatrix}$$

则称此信道为强对称信道或均匀信道 .

二元对称信道就是 $r=2$ 的均匀信道 .一般信道的信道矩阵中各行之和为 1,但各列之和不一定等于 1,而均匀信道中各列之和亦等于 1 .

定理 4 .1 对于对称信道,当输入分布为等概分布时,输出分布必能达到等概分布 .

证明

当输入为等概分布时, $p(x_i)=\frac{1}{r}$, $i=1,2,\cdots,r$,输出

$$p(y_j)=\sum_i p(x_i)\,p(y_j\mid x_i)=\frac{1}{r}\sum_i p(y_j\mid x_i)=\frac{1}{r}H_j \tag{4.15}$$

其中 $H_j=\sum_{i=1}^{r} p(y_j\mid x_i)$ 表示信道矩阵 P 中第 j 列元素之和 .由信道的对称性可知, H_j 是一个与 j 无关的常数,每一列元素之和均为 H_j .由于信道矩阵每一行的元素之和为 1,所以 r 行元素之和为 r,并且 r 行元素之和必等于 s 列元素之和,即 $sH_j=r$, $H_j=r/s$,因而,

$$p(y_j)=\frac{1}{r}H_j=\frac{1}{s}\qquad j=1,2,\cdots,s \tag{4.16}$$

即当信道输入为等概分布 $p(x_j)=1/r$, $i=1,2,\cdots,r$ 时,输出 $p(y_j)=1/s$, $j=1,2,\cdots,s$ 亦为等概分布 .

证毕

定理 4 .2 若一个离散对称信道具有 r 个输入符号, s 个输出符号,则当输入为等概分布时达到信道容量,且

$$C=\log_2 s-H(p_1',p_2',\cdots,p_s') \tag{4.17}$$

其中 $p_1',p_2',\cdots,p_s'$ 为信道矩阵中的任一行元素 .

证明

平均互信息 $$I(X;Y)=H(Y)-H(Y\mid X) \tag{4.18}$$

其中噪声熵:

$$H(Y|X)=\sum_i\sum_j p(x_iy_j)\log_2\frac{1}{p(y_j|x_i)}$$

$$=\sum_i p(x_i)\sum_j p(y_j|x_i)\log_2\frac{1}{p(y_j|x_i)}$$

$$=\sum_i p(x_i)\,H(Y|x_i)$$

由于信道的对称性，$H(Y|x_i)$与 x_i 无关，且 $H(Y|x_i)=H(p_1',p_2',\cdots,p_s')$，所以

$$I(X;Y)=H(Y)-H(p_1',p_2',\cdots,p_s')$$

根据信道容量的定义可得

$$C=\max_{p(x)}I(X;Y)=\max_{p(x)}[H(Y)-H(p_1',p_2',\cdots,p_s')]=\max_{p(x)}H(Y)-H(p_1',p_2',\cdots,p_s')$$

当输出 Y 为等概分布时，$H(Y)$达到最大 $\log_2 s$，所以当信源 X 的概率分布使输出 Y 等概分布时，信道达到信道容量，并且

$$C=\log_2 s-H(p_1',p_2',\cdots,p_s')$$

即离散对称信道的信道容量只与输出符号个数和信道矩阵中的任一行元素 $p_1',p_2',\cdots,p_s'$ 有关．

证毕

推论　均匀信道的信道容量为 $C=\log_2 r-p\log_2(r-1)-H(p)$．　(4.19)

证明

均匀信道中输入、输出符号数相等，$r=s$，所以

$$C=\log_2 r-H(p_1',p_2',\cdots,p_s')$$

$$=\log_2 r-H(1-p,\frac{p}{r-1},\cdots,\frac{p}{r-1})$$

$$=\log_2 r+(1-p)\log_2(1-p)+\frac{p}{r-1}\log_2\frac{p}{r-1}+\cdots+\frac{p}{r-1}\log_2\frac{p}{r-1}$$

$$=\log_2 r+(1-p)\log_2(1-p)+p\log_2\frac{p}{r-1}$$

$$=\log_2 r-p\log_2(r-1)+(1-p)\log_2(1-p)+p\log_2 p$$

$$=\log_2 r-p\log_2(r-1)-H(p)$$

其中，p 是总的错误传递概率，$1-p$ 是正确传递概率．

证毕

当输入为等概分布时，输出也为等概分布，信道达到信道容量．当 $r=2$ 时的均匀信道常称为二元对称信道，这时 $C=1-H(p)$．

对于一般的离散行对称信道，信道容量 C 仍然可以写成：

$$C=\max_{p(x)}H(Y)-H(p_1',p_2',\cdots,p_s')\qquad(4.20)$$

但是不一定存在一种输入分布使输出达到等概分布，此时的信道容量

$$C\leqslant\log_2 s-H(p_1',p_2',\cdots,p_s')\qquad(4.21)$$

而离散对称信道的信道矩阵中每一列都是由同一组元素的不同排列组成,所以保证了当输入符号 X 为等概分布时,输出符号 Y 一定也是等概分布,输出随机变量的熵可以达到 $\log_2 s$.

对于离散准对称信道,由于不一定存在一种输入分布使输出等概,所以,

$$C \leq \log_2 s - H(p_1', p_2', \cdots, p_s') \tag{4.22}$$

但是可以证明当输入为等概分布时,可以达到信道容量

$$C = \log_2 r - \sum_{k=1}^{n} N_k \log_2 M_k - H(p_1', p_2', \cdots, p_s')$$

其中,N_k 是 n 个子矩阵中第 k 个子矩阵中行元素之和,M_k 是第 k 个子矩阵中列元素之和.(证明留给读者作为书后习题)

【例 4.2】

设某离散对称信道的信道矩阵为

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$$

求信道容量.

解

这是一个对称信道,所以

$$\begin{aligned} C &= \log_2 s - H(p_1', p_2', \cdots, p_s') \\ &= \log_2 3 - H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) \\ &= \log_2 3 + \frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{3}\log_2 \frac{1}{3} + \frac{1}{6}\log_2 \frac{1}{6} \\ &= 0.126 \text{ 比特/符号} \end{aligned}$$

在这个对称信道中,每个符号平均能够传输的最大信息量为 0.126 bit.只有当信道输入符号是等概分布时可以达到这个最大值.

【例 4.3】

求二元对称删除信道的信道容量.

$$P = \begin{bmatrix} 1-p-q & q & p \\ p & q & 1-p-q \end{bmatrix}$$

解

这是一个准对称信道

$$N_1 = 1 - q,\ M_1 = 1 - q,\ N_2 = q,\ M_2 = 2q$$

所以

$$\begin{aligned} C &= \log_2 r - \sum_{k=1}^{2} N_k \log_2 M_k - H(p_1, p_2, \ldots, p_s) \\ &= \log_2 2 - (1-q)\log_2 (1-q) - q\log_2 (2q) - H(1-p-q, q, p) \end{aligned}$$

4.2.5　一般离散信道的信道容量

信道容量定义为在信道固定的条件下,平均互信息对所有可能的输入分布的极大值.前面已经导出,平均互信息 $I(X;Y)$是输入概率分布 $p(x)$的上凸函数,因此极大值必定存在.

在信道固定的条件下,平均互信息 $I(X;Y)$是 r 个变量 $p(x_i)$, $i=1,2,\cdots,r$ 的多元函数,且满足约束条件 $\sum_{i=1}^{r} p(x_i)=1$,故可用拉格朗日乘子法来求这个条件极值.即在

$$\begin{cases} p(x_i)\geq 0 & i=1,2,\cdots,r \\ \sum_i p(x_i)=1 \end{cases}$$

的条件下求 $I(X;Y)$的极值.因为 $I(X;Y)$是关于 $p(x_i)$的上凸函数,所以得到的极值是极大值.

设辅助函数:

$$F=I(X;Y)-\lambda\sum_i p(x_i) \tag{4.23}$$

当$\dfrac{F}{p(x_i)}=0$ 时求得的 $I(X;Y)$的值即为信道容量.

整理式(4.23),得

$$\begin{aligned} F &= H(Y)-H(Y\mid X)-\lambda\sum_i p(x) \\ &= \sum_i p(x_i)\sum_j p(y_j\mid x_i)\log_2 p(y_j\mid x_i)-\sum_j p(y_j)\log_2 p(y_j)-\lambda\sum_i p(x_i) \end{aligned} \tag{4.24}$$

因为

$$p(y_j)=\sum_i p(x_i)\,p(y_j\mid x_i) \tag{4.25}$$

所以

$$\frac{p(y_j)}{p(x_i)}=p(y_j\mid x_i) \tag{4.26}$$

又因为

$$\log_2 p(y_j)=\frac{\ln p(y_j)}{\ln 2} \tag{4.27}$$

$$\frac{\log_2 p(y_j)}{p(x_i)}=\frac{\ln p(y_j)}{p(x_i)}\cdot\frac{1}{\ln 2}=\frac{1}{p(y_j)}\frac{p(y_j)}{p(x_i)}\log_2 e=\frac{1}{p(y_j)}p(y_j\mid x_i)\log_2 e=\frac{p(y_j\mid x_i)}{p(y_j)}\log_2 e \tag{4.28}$$

因此

$$\begin{aligned} \frac{F}{p(x_i)} &= \sum_j p(y_j\mid x_i)\log_2 p(y_j\mid x_i)-\sum_j p(y_j\mid x_i)\log_2 p(y_j)-\sum_j p(y_j)\frac{p(y_j\mid x_i)}{p(y_j)}\log_2 e-\lambda \\ &= \sum_j p(y_j\mid x_i)\log_2\frac{p(y_j\mid x_i)}{p(y_j)}-\sum_j p(y_j)\frac{p(y_j\mid x_i)}{p(y_j)}\log_2 e-\lambda \\ &= \sum_j p(y_j\mid x_i)\log_2\frac{p(y_j\mid x_i)}{p(y_j)}-\log_2 e-\lambda \end{aligned} \tag{4.29}$$

令$\dfrac{F}{p(x_i)}=0$,得

$$\sum_j p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{p(y_j)} - \log_2 e - \lambda = 0 \qquad (4.30)$$

即

$$\sum_j p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{p(y_j)} = \log_2 e + \lambda \qquad i = 1,2,\cdots,r \qquad (4.31)$$

将式(4.31)两边同乘以 $p(x_i)$ 并对 i 求和,得

$$\sum_i p(x_i) \sum_j p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{p(y_j)} = \sum_i p(x_i)(\log_2 e + \lambda) \qquad (4.32)$$

式(4.32)左边即为平均互信息的极大值 C,所以得到

$$C = \log_2 e + \lambda \qquad (4.33)$$

这样得到的信道容量有一个参数 λ.在某些情况下可以消去 λ 得到信道容量值.

例如对于信道矩阵为可逆矩阵的信道,可以采用解方程组的方法得解.

在一般信道的信道容量的推导中推出下式:

$$\sum_j p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{p(y_j)} = \log_2 e + \lambda = C \qquad i = 1,2,\cdots,r \qquad (4.34)$$

移项得

$$\sum_j p(y_j \mid x_i)\log_2 p(y_j \mid x_i) = \sum_j p(y_j \mid x_i)\log_2 p(y_j) + C = \sum_j p(y_j \mid x_i)[\log_2 p(y_j) + C] \qquad (4.35)$$

令

$$\beta_j = \log_2 p(y_j) + C \qquad (4.36)$$

则

$$\sum_j p(y_j \mid x_i)\log_2 p(y_j \mid x_i) = \sum_j p(y_j \mid x_i)\beta_j \qquad (4.37)$$

这是含有 s 个未知数 β_j、由 r 个方程组成的方程组.

当 $r = s$,且信道矩阵是可逆矩阵时,该方程组有唯一解.这时就可以求出 β_j,然后根据 $p(y_i) = 2^{\beta_j - C}$(由式(4.36)可得)和 $\sum_j p(y_j) = 1$ 求出信道容量:

$$\sum_j 2^{\beta_j - C} = 1 \qquad (4.38)$$

所以

$$C = \log_2 \sum_j 2^{\beta_j} \qquad (4.39)$$

由 β_j 和 C 求得输出概率分布 $p(y_j)$:

$$p(y_j) = 2^{\beta_j - C} \qquad (4.40)$$

再根据

$$p(y_j) = \sum_i p(x_i)\, p(y_j \mid x_i) \qquad (4.41)$$

列方程组求出 $p(x_i)$.

将计算步骤总结如下:

(1) 由式(4.37)列方程组求出 β_j;

(2) 由式(4.39)求出 C;

(3) 由式(4.40)求出 $p(y_j)$;

(4) 由式(4.41)列方程组求 $p(x_i)$.

需要强调的是,在第(2)步求出信道容量以后,计算并没有结束,还必须解出 $p(x_i)$,如果所有的 $p(x_i)\geq 0$,则求出的信道容量才是正确的.因为用拉格朗日乘子法没有加入 $p(x_i)\geq 0,\ i=1,2,\cdots,r$ 的约束条件,因此算出的 $p(x_i)$有可能是负值.如果 $p(x_i)$有负值,则此解无效,它表明所求得的极限值出现的区域不满足概率条件,那么这时最大值必在边界上,即有某些输入符号的概率 $p(x_i)=0$.因此必须设某些输入符号的概率 $p(x_i)=0$,然后重新进行计算.这样的计算比较复杂,一般要通过迭代来实现.

【例 4.4】

求如下信道的信道容量:

$$P=\begin{pmatrix} \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

解

信道矩阵中 $r=s$,且为可逆矩阵(满秩矩阵),所以以下方程组有唯一解.

$$\begin{cases} \frac{1}{2}\beta_1+\frac{1}{4}\beta_2+\frac{1}{4}\beta_4=\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{4}\log_2\frac{1}{4} \\ \beta_2=0 \\ \beta_3=0 \\ \frac{1}{4}\beta_1+\frac{1}{4}\beta_3+\frac{1}{2}\beta_4=\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{2}\log_2\frac{1}{2} \end{cases}$$

解方程组得　　$\beta_2=\beta_3=0,\ \beta_1=\beta_4=-2$

$$C=\log_2\sum_j 2^{\beta_j}=\log_2(2^{-2}+2^0+2^0+2^{-2})=\log_2 5-1$$

再根据 $p(y_j)=2^{\beta_j-C}$ 求 $p(y_j)$,得

$$p(y_1)=p(y_4)=2^{-2-\log_2 5+1}=\frac{1}{10}$$

$$p(y_2)=p(y_3)=2^{0-\log_2 5+1}=\frac{4}{10}$$

最后根据 $p(y_j)=\sum_i p(x_i)\,p(y_j\mid x_i)$ 列方程组求 $p(x_i)$,求出最佳输入分布:

$$p(x_1)=p(x_4)=\frac{4}{30},\ p(x_2)=p(x_3)=\frac{11}{30}$$

上述求得的 $p(x_i),\ i=1,2,3,4$ 都大于 0,故求得的结果是正确的.

如果输入概率分布只有一个变量.例如 $r=2$ 时,可以设输入概率分布为 α 和 $1-\alpha$,

因此输入概率分布只有一个变量,则可以求 $I(X;Y)$直接求导,这时可以直接对 $I(X;Y)$求导,求出 α,从而得出 $I(X;Y)$的极大值 C.对于 $r\neq 2$ 的情况,可以通过已知条件消去一些变量,使得最后的输入概率分布只有一个变量.

【例 4.5】

已知信道的转移矩阵为 $P=\begin{pmatrix}0.5 & 0.3 & 0.2\\ 0.3 & 0.5 & 0.2\end{pmatrix}$,求信道容量.

解

设输入概率分布 $p(x_1)=\alpha$, $p(x_2)=1-\alpha$,则输出 y_1, y_2, y_3 的概率分布为

$$P_Y=P_XP_{Y|X}=(\alpha,1-\alpha)\begin{pmatrix}0.5 & 0.3 & 0.2\\ 0.3 & 0.5 & 0.2\end{pmatrix}=(0.3+0.2\alpha,0.5-0.2\alpha,0.2)$$

其中, $p(y_3)$固定,与 x_i 的分布无关.

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y|X)\\&=-\sum_j p(y_j)\log_2 p(y_j)+\sum_i p(x_i)\sum_j p(y_j|x_i)\log_2 p(y_j|x_i)\\&=-(0.3+0.2\alpha)\log_2(0.3+0.2\alpha)-(0.5-0.2\alpha)\log_2(0.5-0.2\alpha)\\&\quad-0.2\log_2 0.2+0.5\log_2 0.5+0.3\log_2 0.3+0.2\log_2 0.2\end{aligned}$$

由 $\dfrac{I(X;Y)}{\alpha}=0$,得 $0.2\log_2(0.3+0.2\alpha)-0.2+0.2\log_2(0.5-0.2\alpha)+0.2=0$

解得 $\alpha=1/2$,即输入等概分布时 $I(X;Y)$达到极大值,且

$$C=\max I(X;Y)=0.036 \text{ 比特/符号}$$

4.2.6 信道容量定理

从以上的讨论可知,求信道容量的问题实际上是在约束条件下求多元函数极值的问题,在通常情况下,计算量是非常大的.下面介绍一般离散信道的平均互信息 $I(X;Y)$达到信道容量的充要条件,在某些情况下它可以帮助我们较快地找到极值点.

定理 4.3 设有一般离散信道,它有 r 个输入符号, s 个输出符号.当且仅当存在常数 C,使输入分布 $p(x_i)$满足

(1) $I(x_i;Y)=C$ 当 $p(x_i)\neq 0$

(2) $I(x_i;Y)\leq C$ 当 $p(x_i)=0$

时, $I(X;Y)$达到最大值.其中

$$I(x_i;Y)=\sum_j p(y_j|x_i)\log_2\frac{p(y_j|x_i)}{p(y_j)} \tag{4.42}$$

它表示信道输入 x_i 时,所给出关于输出 Y 的信息量.常数 C 即为所求的信道容量.

信道容量对输入概率分布求偏导可以得出以下关系式:

$$\frac{I(X;Y)}{p(x_i)}=\sum_j p(y_j\mid x_i)\log_2\frac{p(y_j\mid x_i)}{p(y_j)}-\log_2 e=I(x_i;Y)-\log_2 e \qquad (4.43)$$

用式(4.33)和式(4.43)可以将上述充要条件改写成:

(1) $\dfrac{I(X;Y)}{p(x_i)}=\lambda \qquad p(x_i)\neq 0$

(2) $\dfrac{I(X;Y)}{p(x_i)}\leq\lambda \qquad p(x_i)=0$

我们将利用信道容量定理的引理条件(参见附录 B.4)证明这个改写后的充要条件.

证明

充分性,也就是要证明如果输入分布 $p=(p_1,p_2,\cdots,p_r)$ 满足

(1) $\dfrac{I(X;Y)}{p(x_i)}=\lambda \qquad p(x_i)\neq 0$

(2) $\dfrac{I(X;Y)}{p(x_i)}\leq\lambda \qquad p(x_i)=0$

则 p 一定使平均互信息 $I(X;Y)$ 达到极大值 $I(p)$.也就是对于任何其他的输入分布 $q=(q_1,q_2,\cdots,q_r)$,必然有

$$I(q)\leq I(p) \qquad (4.44)$$

因为,平均互信息 $I(X;Y)$ 是输入分布的上凸函数,所以

$$\theta I(q)+(1-\theta)I(p)\leq I(\theta q+(1-\theta)p) \qquad (4.45)$$

其中 $\theta+(1-\theta)=1,0<\theta<1$,移项得

$$I(q)-I(p)\leq[I(\theta q+(1-\theta)p)-I(p)]/\theta \qquad (4.46)$$

上式对一切 $0<\theta<1$ 均成立.取 $\theta\to 0$,根据引理,可得

$$I(q)-I(p)\leq\sum_{i=1}^{r}(q_i-p_i)\frac{I(p)}{p_i} \qquad (4.47)$$

其中,$p_i=p(x_i)$,$q_i=q(x_i)$.

根据假设,输入分布 p 满足

(1) $\dfrac{I(p)}{p_i}=\lambda \qquad p_i\neq 0$

(2) $\dfrac{I(p)}{p_i}\leq\lambda \qquad p_i=0$

所以,$I(q)-I(p)\leq\lambda\sum_{i=1}^{r}(q_i-p_i)=\lambda\left[\sum_{i=1}^{r}q_i-\sum_{i=1}^{r}p_i\right]=0$,即

$$I(q)\leq I(p) \qquad (4.48)$$

充分性得证.

必要性,就是要证明如果输入分布 p 使平均互信息 $I(X;Y)$ 达到极大值 $I(p)$,则输

入分布 p 必然满足

(1) $\frac{I(p)}{p_i}=\lambda \qquad p_i\neq 0$

(2) $\frac{I(p)}{p_i}=\lambda \qquad p_i=0$

如果输入分布 p 使平均互信息 $I(X;Y)$ 达到极大值,取任一其他输入分布 $q=(q_1,q_2,\cdots,q_r)$,必然有

$$I(\theta q+\overline{1-\theta}p)-I(p)\leq 0 \tag{4.49}$$

式(4.49)两边同除以 θ 并取当 $\theta\to 0$ 时的极限

$$\lim_{\theta\to 0}\frac{1}{\theta}\{I[\theta q+(1-\theta)p]-I(p)\}\leq 0 \tag{4.50}$$

根据引理,可得

$$\sum_{i=1}^{r}(q_i-p_i)\frac{I(p)}{p_i}\leq 0 \tag{4.51}$$

对于输入分布 p,因为概率分布的完备性 $\sum_{i=1}^{r}p_i=1$,所以其中至少有一个分量不为零,令 $p_l\neq 0$,再选择另一个输入分布 $q=(q_1,q_2,\cdots,q_r)$,并且满足:

$$\begin{cases} q_l=p_l-\varepsilon \\ q_j=p_j+\varepsilon & \text{(保证输入分布 } q \text{ 的完备性)} \\ q_i=p_i & \text{(其他分量均相同)} \end{cases} \tag{4.52}$$

其中,ε 为任意数.代入式(4.51),得到

$$-\varepsilon\frac{I(p)}{p_l}+\varepsilon\frac{I(p)}{p_j}\leq 0 \tag{4.53}$$

令 $$\frac{I(p)}{p_l}=\lambda$$

所以,有 $$\varepsilon\frac{I(p)}{p_j}\leq\lambda\varepsilon$$

因为概率的非负性,所以 $p_l-\varepsilon\geq 0$, $p_j+\varepsilon\geq 0$, ε 必满足 $-p_j\leq\varepsilon\leq p_l$.

当 $p_j=0$ 时,$0\leq\varepsilon\leq p_l$,ε 为正数,所以,$\frac{I(p)}{p_j}\leq\lambda$.

当 $p_j>0$ 时,$-p_j\leq\varepsilon\leq p_l$,$\varepsilon$ 可为正数,也可为负数.

如果 ε 取正数,$\frac{I(p)}{p_j}\leq\lambda$;如果 ε 取负数,$\frac{I(p)}{p_j}\geq\lambda$.所以,当 $p_j\neq 0$ 时,必然有 $\frac{I(p)}{p_j}=\lambda$,因此输入分布 p 必满足充要条件,必要性得证.

证毕

$I(x_i;Y)$ 表示信道输入 x_i 时,所给出关于输出 Y 的信息量,一般来说,x_i 不同,

$I(x_i;Y)$值不同.信道容量定理告诉我们,平均互信息 $I(X;Y)$取到极大值也就是信道容量时,对于任意 x_i,只要它出现的概率大于 0, $I(x_i;Y)$都相等.

【例 4.6】

证明当输入为等概分布时,离散准对称信道达到信道容量.

证明

根据信道容量定理,需要证明输入为等概分布 $p(x_i)=1/r$ 时, $I(x_i;Y)$为一个与 x_i 无关的常数.

$$
\begin{aligned}
I(x_i;Y) &= \sum_{j=1}^{s} p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{p(y_j)} \\
&= \sum_{j=1}^{s} p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{\sum_{k=1}^{r} p(x_k)\,p(y_j \mid x_k)} \\
&= \sum_{j=1}^{s} p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{\frac{1}{r}\sum_{k=1}^{r} p(y_j \mid x_k)}
\end{aligned}
$$

准对称信道的信道矩阵可以按列分为一些对称的子阵 $P_1, P_2, \cdots, P_l, \cdots, P_n$,在同一子阵中,每一列都是第一列的同一组元素的排列,所以在同一子阵 P_l 中, $p(y_j)=\frac{1}{r}\sum_{k=1}^{r} p(y_j \mid x_k)$, $y_j \in Y_l$ 都相等.而同一子阵中每一行又都是其他行的同一组元素的排列,所以,同一子阵 P_l 中,对于任意 x_i, $\sum_{y_j \in Y_l} p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{\frac{1}{r}\sum_{k=1}^{r} p(y_j \mid x_k)}$ 也都相等.于是对于任意 x_i, $I(x_i;Y)=\sum_{l}\sum_{y_j \in Y_l} p(y_j \mid x_i)\log_2 \frac{p(y_j \mid x_i)}{\frac{1}{r}\sum_{k=1}^{r} p(y_j \mid x_k)}$ 必然相等.所以 $I(x_i;Y)$ 是一个与 x_i 无关的常数,根据信道容量定理,这时信道达到信道容量,即当输入为等概分布时,离散准对称信道达到信道容量.

证毕

信道容量定理只给出了达到信道容量时,最佳输入概率分布应满足的条件,并没有给出最佳输入概率分布值,也没有给出信道容量的值.另外,定理本身也隐含着达到信道容量的最佳分布不是唯一的,只要输入概率分布满足充要条件,就是信道的最佳输入分布.在一些特殊情况下,常常利用这一定理寻求最佳输入分布和信道容量值.

【例 4.7】

设离散信道如图 4.7 所示,输入符号集$\{x_1, x_2, x_3, x_4, x_5\}$,输出符号集$\{y_1, y_2\}$,求 C.

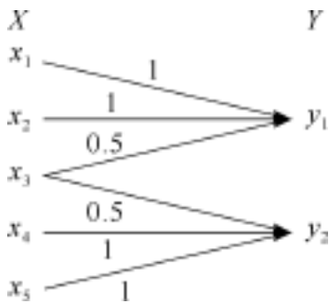


图 4.7 例 4.7 的离散信道

解

该信道信道矩阵 $P=\begin{bmatrix}1 & 0\\ 1 & 0\\ \frac{1}{2} & \frac{1}{2}\\ 0 & 1\\ 0 & 1\end{bmatrix}$，不是对称信道.由于 x_3 传递到 y_1，y_2 是等概率的，如果令 $p(x_3)=0$，则会减少收到 Y 以后对输入 X 的不确定性，这时，x_1，x_2 与 x_4，x_5 分别转移到 y_1，y_2 如果又令 $p(x_2)=p(x_4)=0$，信道就变成了一一对应的信道，接收到 Y 后对输入端 X 是完全确定的，这时再令 $p(x_1)=p(x_5)=1/2$，检查它是否满足信道容量定理的条件，若满足，则该输入分布就是最佳输入分布.

可计算得

$$I(x_1;Y)=I(x_5;Y)=\log_2 2,\ I(x_2;Y)=I(x_4;Y)=\log_2 2,\ I(x_3;Y)=0$$

满足信道容量定理的充要条件，因此信道容量

$$C=\log_2 2=1\ \text{比特/符号}$$

若设 $p(x_1)=p(x_2)=p(x_4)=p(x_5)=\frac{1}{4}$，$p(x_3)=0$ 也满足信道容量定理的充要条件，这时

$$I(x_1;Y)=I(x_2;Y)=I(x_4;Y)=I(x_5;Y)=\log_2 2,\ I(x_3;Y)<\log_2 2$$

所以该分布也是最佳分布.

可见，这个信道的最佳输入分布不是唯一的.由于 $I(x_i;Y)$ 仅直接与信道传递概率及输出符号概率有关，因而达到信道容量的输入概率分布不是唯一的，但是输出概率分布是唯一的.

对于某些比较简单直观的信道，可以利用以上方法求信道容量.

*4.2.7 信道容量的迭代算法

前述几种方法都不能保证对于任意离散信道求出其信道容量.利用计算机的迭代算法可以以任意给定的精度及有限步数求出任意离散信道的信道容量.

$$I(X;Y)=H(X)-H(X|Y)=-\sum_i p(x_i)\ln p(x_i)+\sum_i\sum_j p(x_i)\,p(y_j|x_i)\ln p(x_i|y_j) \tag{4.54}$$

对于某一固定的信道，其转移概率 $p(y_j|x_i)$ 是已定的，所以 $I(X;Y)$ 是关于 $p(x_i)$ 和 $p(x_i|y_j)$ 的函数(上凸函数).虽然事实上，$p(x_i|y_j)=\dfrac{p(x_i)\,p(y_j|x_i)}{\sum_i p(x_i)\,p(y_j|x_i)}$ 也是 $p(x_i)$ 的函

数,但是可以把 $I(X;Y)$看作是关于 $p(x_i)$和 $p(x_i|y_j)$的函数,记为 $I[p(x_i),p(x_i|y_j)]$.

先固定变量 $p(x_i)$,求 $I[p(x_i),p(x_i|y_j)]$关于 $p(x_i|y_j)$的极值.这是在约束条件 $\sum_i p(x_i|y_j)=1,\ j=1,2,\cdots,s$ 下的条件极值.利用拉格朗日乘子法,设辅助函数:

$$F=I[p(x_i),p(x_i|y_j)]-\sum_j\lambda_j\sum_i p(x_i|y_j) \tag{4.55}$$

$$\begin{aligned}\frac{F}{p(x_i|y_j)}&=\frac{}{p(x_i|y_j)}\left[-\sum_i p(x_i)\ln p(x_i)+\sum_i\sum_j p(x_i)p(y_j|x_i)\ln p(x_i|y_j)-\sum_j\lambda_j\sum_i p(x_i|x_j)\right]\\&=\frac{p(x_i)p(y_j|x_i)}{p(x_i|y_j)}-\lambda_j\end{aligned} \tag{4.56}$$

令 $\frac{F}{p(x_i|y_j)}=0$,得

$$\lambda_j=\frac{p(x_i)p(y_j|x_i)}{p(x_i|y_j)} \tag{4.57}$$

$$p(x_i|y_j)=\frac{p(x_i)p(y_j|x_i)}{\lambda_j} \tag{4.58}$$

其中,$i=1,2,\cdots,r;\ j=1,2,\cdots,s$.

利用约束条件 $\sum_i p(x_i|y_j)=1$,得

$$\sum_i p(x_i|y_j)=\sum_i\frac{p(x_i)p(y_j|x_i)}{\lambda_j}=1 \tag{4.59}$$

所以

$$\lambda_j=\sum_i p(x_i)p(y_j|x_i)\qquad j=1,2,\cdots,s \tag{4.60}$$

因此,求得使 $I[p(x_i),p(x_i|y_j)]$达到极值的 $p(x_i|y_j)^*$ 为

$$p(x_i|y_j)^*=\frac{p(x_i)p(y_j|x_i)}{\sum_i p(x_i)p(y_j|x_i)}\qquad i=1,2,\cdots,r;\ j=1,2,\cdots,s \tag{4.61}$$

在求得 $p(x_i|y_j)^*$ 后,再固定 $p(x_i|y_j)$,求 $I[p(x_i),p(x_i|y_j)]$关于 $p(x_i)$的极值.此时的约束条件是 $\sum_i p(x_i)=1$.

设辅助函数:

$$\begin{aligned}Q&=I[p(x_i),p(x_i|y_j)]-\lambda\sum_i p(x_i)\\&=-\sum_i p(x_i)\ln p(x_i)+\sum_i\sum_j p(x_i)p(y_j|x_i)\ln p(x_i|y_j)-\lambda\sum_i p(x_i)\end{aligned} \tag{4.62}$$

$$\frac{Q}{p(x_i)}=-\ln p(x_i)-1+\sum_j p(y_j|x_i)\ln p(x_i|y_j)-\lambda \tag{4.63}$$

令 $\frac{Q}{p(x_i)}=0$,得

$$-\ln p(x_i) - 1 + \sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j) - \lambda = 0 \tag{4.64}$$

所以

$$p(x_i) = \exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j) - \lambda - 1\right] = \frac{\exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j)\right]}{\exp(1+\lambda)} \tag{4.65}$$

利用约束条件 $\sum_i p(x_i) = 1$,得

$$\sum_i p(x_i) = \sum_i \frac{\exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j)\right]}{\exp(1+\lambda)} = 1 \tag{4.66}$$

所以,
$$\exp(1+\lambda) = \sum_i \exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j)\right] \tag{4.67}$$

$$1 + \lambda = \ln \sum_i \exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j)\right] \tag{4.68}$$

所以,使 $I[p(x_i), p(x_i \mid y_j)]$达到极值的 $p(x_i)^*$ 为

$$p(x_i)^* = \frac{\exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j)\right]}{\sum_i \exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j)\right]} \qquad i = 1, 2, \cdots, r \tag{4.69}$$

由式(4.64)移项得

$$-\ln p(x_i) + \sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j) = 1 + \lambda \tag{4.70}$$

式(4.70)两端同乘以 $p(x_i)$,并对 i 求和,有

$$-\sum_i p(x_i)\ln p(x_i) + \sum_i p(x_i)\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j) = 1 + \lambda$$

得
$$I[p(x_i)^*, p(x_i \mid y_j)^*] = 1 + \lambda = \ln \sum_i \exp\left[\sum_j p(y_j \mid x_i)\ln p(x_i \mid y_j)\right] \tag{4.71}$$

利用式(6.61)、(6.69)、(6.71)便可以对信道容量进行迭代计算.

用迭代法计算信道容量 C 的计算步骤如下:

记 $p(y_j \mid x_i) = p_{ij}$, $p(x_i) = p_i$, $p(x_i \mid y_j) = \phi_{ji}$ $\quad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s$

算法:

① 初始化信源分布 $p^{(0)} = (p_1, p_2, \cdots, p_i, \cdots, p_r)$(一般初始化为均匀分布),置迭代计数器 $k = 0$,设信道容量相对误差门限为 δ, $\delta > 0$;

② $$\phi_{ji}^{(k)} = \frac{p_{ij}p_i^{(k)}}{\sum_i p_{ij}p_i^{(k)}} \tag{4.72}$$

③ $$p_i^{(k+1)} = \frac{\exp\left[\sum_j p_{ij}\ln\phi_{ji}^{(k)}\right]}{\sum_i \left\{\exp\left[\sum_j p_{ij}\ln\phi_{ji}^{(k)}\right]\right\}} \tag{4.73}$$

④ $C^{(k+1)}=\ln\left\{\sum_i \exp\left[\sum_j p_{ij}\ln\phi_{ji}^{(k)}\right]\right\}$　　(4.74)

⑤ 如果 $\Delta C^{(k+1)}=\dfrac{|C^{(k+1)}-C^{(k)}|}{C^{(k+1)}}\le\delta$，转向⑦；

⑥ 置迭代序号 $k+1\to k$，转向②；

⑦ 输出 $p_i^{(k+1)}$ 和 $C^{(k+1)}$ 的结果；

⑧ 停止．

可以证明，平均互信息 $I[p(x_i),p(x_i|y_j)]$ 具有收敛性，即 $\lim\limits_{k\to\infty}|C^{(k+1)}-C^{(k)}|=0$，所以迭代算法最终能求出任意精度的解．算法的收敛速度与信源初始概率分布的选择有很大的关系，初始分布选得越接近最佳输入分布，则收敛的速度越快，若初始分布选得正好是最佳输入分布，则一步就可求得信道容量．

4.3　离散多符号信道及其信道容量

在 4.2 节中，讨论了最简单的离散信道，即信道的输入和输出都只是单个随机变量的信道．实际离散信道的输入和输出常常是随机变量序列，用随机矢量来表示，称为离散多符号信道，如图 4.8 所示．实际离散信道往往是有记忆信道，为了简化起见，我们主要研究离散无记忆信道．

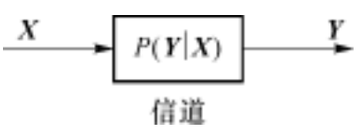


图 4.8　离散多符号信道模型

定义 4.6　若信道在任意时刻的输出只与此时刻信道的输入有关，而与其他时刻的输入和输出无关，则称之为离散无记忆信道，简称为 DMC(Discrete Memoryless Channel)．

输入、输出随机序列的长度为 N 的离散无记忆平稳信道通常称为离散无记忆信道的 N 次扩展信道．

输入随机序列 $X=X_1X_2\cdots X_N$ 中，每一个随机变量 X_i，$i=1,2,\cdots,N$ 都取值于同一输入符号集 X，而符号集 X 共有 r 个符号，所以随机矢量 X 的可能取值共有 r^N 个．输出随机序列 $Y=Y_1Y_2\cdots Y_N$ 中，每一个随机变量 Y_i，$i=1,2,\cdots,N$ 都取值于同一输出符号集 Y，而符号集 Y 共有 s 个符号，所以随机矢量 Y 的可能取值有 s^N 个，因此 N 次扩展信道的信道矩阵是一个 $r^N\times s^N$ 的矩阵．离散无记忆信道的数学模型仍然表示为 $\{X,P(Y|X),Y\}$，注意这时输入、输出均为随机矢量．

根据信道无记忆的特性，其转移概率

$$\begin{aligned}P(Y|X)&=P(Y_1Y_2\cdots Y_N|X_1X_2\cdots X_N)\\&=P(Y_1|X_1)P(Y_2|X_2)\cdots P(Y_N|X_N)\end{aligned}$$

$$= \prod_{k=1}^{N} P(Y_k \mid X_k) \tag{4.75}$$

【例 4.8】

求二元对称信道的二次扩展信道的信道矩阵.

解

二元对称信道的二次扩展信道的输入、输出序列的每一个随机变量均取值于{0,1},输入共有 $r^N = 2^2 = 4$ 个取值,输出共有 $s^N = 2^2 = 4$ 个取值.根据 $P(Y \mid X) = \prod_{k=1}^{N} P(Y_k \mid X_k)$ 可求出

$$p(y_1 \mid x_1) = p(00 \mid 00) = p(0 \mid 0)\, p(0 \mid 0) = \bar{p}^2$$
$$p(y_2 \mid x_1) = p(01 \mid 00) = p(0 \mid 0)\, p(1 \mid 0) = \bar{p}p$$
$$p(y_3 \mid x_1) = p(10 \mid 00) = p(1 \mid 0)\, p(0 \mid 0) = p\bar{p}$$
$$p(y_4 \mid x_1) = p(11 \mid 00) = p(1 \mid 0)\, p(1 \mid 0) = p^2$$

同理可求出其他的转移概率 p_{ij}, $i = 2,3,4$; $j = 1,2,3,4$,得到信道矩阵:

$$P = \begin{bmatrix} \bar{p}^2 & \bar{p}p & p\bar{p} & p^2 \\ \bar{p}p & \bar{p}^2 & p^2 & p\bar{p} \\ p\bar{p} & p^2 & \bar{p}^2 & \bar{p}p \\ p^2 & p\bar{p} & \bar{p}p & \bar{p}^2 \end{bmatrix}$$

二元对称信道的二次扩展信道如图 4.9 所示.

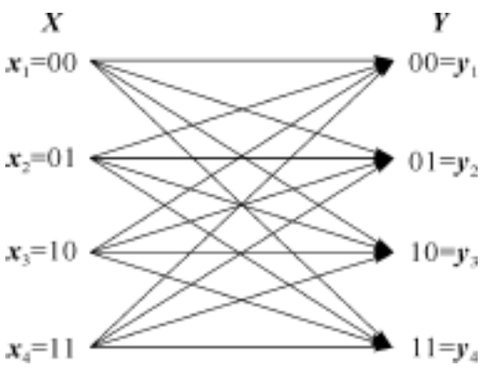


图 4.9 二元对称信道的二次扩展信道

关于离散无记忆信道的平均互信息,有以下定理:

定理 4.4 若信道的输入和输出分别是 N 长序列 X 和 Y,且信道是无记忆的,则

$$I(X;Y) \leq \sum_{k=1}^{N} I(X_k;Y_k) \tag{4.76}$$

这里 X_k、Y_k 分别是序列 X 和 Y 中第 k 位随机变量.

证明

$$I(X;Y) = H(Y) - H(Y \mid X) \tag{4.77}$$

根据熵函数的链规则和条件熵与无条件熵的关系,可得

$$\begin{aligned} H(Y) &= H(Y_1 Y_2 \cdots Y_N) \\ &= H(Y_1) + H(Y_1 \mid Y_2) + \cdots + H(Y_N \mid Y_1 Y_2 \cdots Y_{N-1}) \\ &\leq \sum_{k=1}^{N} H(Y_k) \end{aligned} \tag{4.78}$$

根据熵函数的链规则和离散无记忆信道的定义,可得

$$H(Y \mid X) = H(Y_1 Y_2 \cdots Y_N \mid X_1 X_2 \cdots X_N)$$

$$= H(Y_1 \mid X_1 X_2 \cdots X_N) + H(Y_2 \mid X_1 X_2 \cdots X_N Y_1) + \cdots + H(Y_N \mid X_1 X_2 \cdots X_N Y_1 Y_2 \cdots Y_{N-1})$$

$$= \sum_{k=1}^{N} H(Y_k \mid X_k) \tag{4.79}$$

所以，$$I(X;Y) \leq \sum_{k=1}^{N} H(Y_k) - \sum_{k=1}^{N} H(Y_k \mid X_k) = \sum_{k=1}^{N} I(X_k;Y_k)$$

即对于离散无记忆信道，其平均互信息 $I(X;Y)$ 小于等于序列 X 和 Y 中所有对应时刻的随机变量 X_k，Y_k 的平均互信息 $I(X_k;Y_k)$ 之和．当且仅当信源也是无记忆信源时等号成立．

当信源是无记忆信源时

$$P(X) = \prod_{k=1}^{N} P(X_k)$$

$$P(XY) = P(X)P(Y \mid X) = \prod_{k=1}^{N} P(X_k) \prod_{k=1}^{N} P(Y_k \mid X_k) = \prod_{k=1}^{N} P(X_k)P(Y_k \mid X_k) = \prod_{k=1}^{N} P(X_k Y_k) \tag{4.80}$$

所以 $$p(y_j) = \sum_{i=1}^{r^N} p(x_i y_j) = \sum_{i=1}^{r^N} p(x_{i_1} y_{j_1}) p(x_{i_2} y_{j_2}) \cdots p(x_{i_N} y_{j_N})$$

$$= \sum_{i_1=1}^{r} p(x_{i_1} y_{j_1}) \sum_{i_2=1}^{r} p(x_{i_2} y_{j_2}) \cdots \sum_{i_N=1}^{r} p(x_{i_N} y_{j_N}) = \prod_{k=1}^{N} p(y_{j_k}) \tag{4.81}$$

即 $$P(Y) = \prod_{k=1}^{N} P(Y_k)$$

因此 $$H(Y) = \sum_{k=1}^{N} H(Y_k) \tag{4.82}$$

这时 $$I(X;Y) = \sum_{k=1}^{N} I(X_k;Y_k) \tag{4.83}$$

即信源和信道均为无记忆时，其序列 X 和 Y 的平均互信息 $I(X;Y)$ 等于序列中所有对应时刻随机变量 X_k，Y_k 的平均互信息 $I(X_k;Y_k)$ 之和．

证毕

对于离散无记忆 N 次扩展信道，如果信道的输入序列中的每一个随机变量均取值于同一信源符号集并且具有同一种概率分布(取自于同一概率空间)，通过相同的信道传送到输出端，

则输出序列中的每一个随机变量也取自同一符号集,并且具有相同的概率分布.因此有

$$X_1 = X_2 = \cdots = X_N = X;\ Y_1 = Y_2 = \cdots = Y_N = Y$$

$$I(X_1; Y_1) = I(X_2; Y_2) = \cdots = I(X_N; Y_N) = I(X; Y) \tag{4.84}$$

于是

$$I(X; Y) = \sum_{k=1}^{N} I(X_k; Y_k) = NI(X; Y) \tag{4.85}$$

式(4.85)表明,对于离散无记忆 N 次扩展信道,当信源是平稳无记忆信源时,其平均互信息 $I(X; Y)$等于单符号信道的平均互信息的 N 倍.

离散无记忆信道的 N 次扩展信道的信道容量为

$$C^N = \max_{P(X)} I(X; Y) = \max_{P(X)} \sum_{k=1}^{N} I(X_k; Y_k) = \sum_{k=1}^{N} \max_{P(X_k)} I(X_k; Y_k) = \sum_{k=1}^{N} C_k \tag{4.86}$$

式中,$C_k = \max\limits_{P(X_k)} I(X_k; Y_k)$是时刻 k 通过离散无记忆信道传输的最大信息量,可以用前面介绍的求离散单符号信道的信道容量的方法求解.因为现在输入随机序列 $X = X_1 \cdots X_k \cdots X_N$ 在同一信道中传输,所以任何时刻通过离散无记忆信道传输的最大信息量都相同,即 $C_k = C$, $k = 1, 2, \cdots, N$.所以

$$C^N = NC \tag{4.87}$$

即离散无记忆信道的 N 次扩展信道的信道容量等于单符号离散信道的信道容量的 N 倍,当信源也是无记忆信源并且每一时刻的输入分布各自达到最佳输入分布时,才能达到这个信道容量 NC.

一般情况下,消息序列在离散无记忆 N 次扩展信道中传输时,其平均互信息量 $I(X; Y) \leq NC$.

4.4 组合信道及其信道容量

前面分析了单符号离散信道和离散无记忆信道.实际应用中常常会遇到两个或更多个信道组合在一起使用的情况.例如,待发送的消息比较多时,可能要用两个或更多个信道并行发送,这种组合信道称为并联信道;有时消息会依次地通过几个信道串联发送,例如无线电中继信道,数据处理系统,这种组合信道称为级联信道.在研究较复杂信道时,为使问题简化,往往可以将它们分解成几个简单信道的组合.这一节将讨论这两种组合信道的信道容量与其组成信道的信道容量之间的关系.

4.4.1 独立并联信道

一般独立并联信道如图 4.10 所示.

设有 N 个信道并联,它们的输入分别为 X_1, X_2, $\cdots$, X_N,输出分别是 Y_1, Y_2, $\cdots$,

Y_N，N 个信道的传递概率分别是 $P(Y_1 \mid X_1)$，$P(Y_2 \mid X_2)$，…，$P(Y_N \mid X_N)$.在这 N 个独立信道中,每一个信道的输出 Y_k 只与本信道的输入 X_k 有关,而与其他的信道的输入、输出无关 这 N 个信道的联合传递概率满足以下关系:

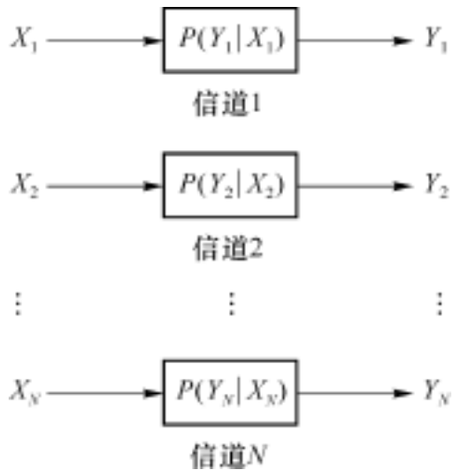


图 4 .10　独立并联信道

$$P(Y_1 Y_2 \cdots Y_N \mid X_1 X_2 \cdots X_N) = P(Y_1 \mid X_1) P(Y_2 \mid X_2) \cdots P(Y_N \mid X_N) \quad (4\ 88)$$

这相当于离散无记忆信道应满足的条件 .因此可以把定理 4 .4 的结论推广到 N 个独立并联信道:

$$I(X_1 X_2 \cdots X_N; Y_1 Y_2 \cdots Y_N) \leq \sum_{k=1}^{N} I(X_k; Y_k)$$

即联合平均互信息不大于各信道的平均互信息之和 .因此独立并联信道的信道容量

$$C_{并} = \max_{P(X_1 \cdots X_N)} I(X_1 \cdots X_N; Y_1 \cdots Y_N) = \sum_{k=1}^{N} C_k \quad (4\ 89)$$

式中，$C_k = \max_{P(X_k)} I(X_k; Y_k)$ 是各个独立信道的信道容量 .

所以独立并联信道的信道容量等于各个信道的信道容量之和 .只有输入随机变量相互独立且当每个输入随机变量的概率分布均达到各自信道的最佳输入分布时,独立并联信道的信道容量才等于各信道容量之和,即

$$C_{并} = \sum_{k=1}^{N} C_k \quad (4\ 90)$$

当 N 个独立并联信道的信道容量都相同时,

$$C_{并} = NC \quad (4\ 91)$$

4 .4 .2　级联信道

级联信道是信道最基本的组合形式,许多实际信道都可以看成是级联信道 .图 4 .11 是由两个离散单符号信道组成的最简单的级联信道 .

信道 Ⅰ,输入随机变量为 X,输出随机变量为 Y .信道Ⅱ,输入随机变量为 Y,输出随机变量为 Z .信道 Ⅰ 的输出恰好是信道Ⅱ的输入 .信道 Ⅰ 的输出 Y 与输入 X 统计相关,而信道Ⅱ的输出 Z 与输入 Y 统计相关,一般来说，Z 将与 X 统计相关 .但是级联的结构又决定了在给定 Y 以后，Z 的取值将不再与 X 有关,而只取决于信道Ⅱ的前向转移概率 $P(Z \mid Y)$,也就是说 $X \to Y \to Z$ 组成一个马尔可夫链 .根据马尔可夫链的性质,级联信道的总的信道矩阵等于这两个串接信道的信道矩阵的乘积 .求得

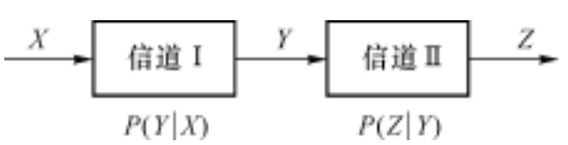


图 4 .11　级联信道

级联信道的总的信道矩阵后,级联信道的信道容量就可以用求离散单符号信道的信道容量的方法计算 .

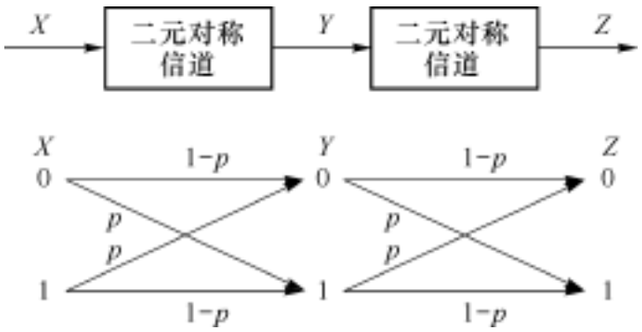


图 4 .12 二元对称信道的级联信道

【例 4 .9】

设有两个离散二元对称信道,其级联信道如图 4 .12 所示,求级联信道的信道容量 .

解

两个二元对称信道的信道矩阵均为

$$P_1 = P_2 = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

因为 X, Y, Z 组成马尔可夫链,则级联信道的总的信道矩阵为

$$P = P_1 P_2 = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} = \begin{bmatrix} (1-p)^2 + p^2 & 2p(1-p) \\ 2p(1-p) & (1-p)^2 + p^2 \end{bmatrix}$$

因此级联信道仍然是一个二元对称信道 .

$$C_{级} = 1 - H[2p(1-p)]$$

*4 .5 连续信道及其信道容量

4 .5 .1 连续随机变量的互信息

连续随机变量 X 和 Y 之间的平均互信息定义为

$$I(X;Y) = \int_{R^2} p(xy) \log_2 \frac{p(xy)}{p(x)\,p(y)} \mathrm{d}x\mathrm{d}y \tag{4 .92}$$

连续随机变量的平均互信息 $I(X;Y)$的计算和离散随机变量一样,只要将离散情况下的概率分布换成概率密度,求和换成积分即可 .连续随机变量的平均互信息具有和离散随机变量的平均互信息一样的性质:

1 . 对称性

$$I(X;Y) = I(Y;X) = h(X) - h(X|Y) = h(Y) - h(Y|X) = h(X) + h(Y) - h(XY) \tag{4 .93}$$

2 .非负性

$$I(X;Y) \geq 0 \tag{4 .94}$$

当且仅当随机变量 X 和 Y 统计独立时等号成立 .

因此虽然连续随机变量的熵不具有非负性,但连续随机变量的熵差 $I(X;Y)$仍然具有非负性 .

【例 4 .10】

设 $p(xy)$是二维高斯随机变量 XY 的概率密度函数

$$p(xy)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2}-\frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y}+\frac{(y-m_Y)^2}{\sigma_Y^2}\right]\right\}$$

其中, m_X、m_Y、σ_X^2、σ_Y^2 分别表示随机变量 X 和 Y 的均值和方差, ρ 是归一化相关函数 $\rho=\frac{E[(X-E(X))(Y-E(Y))]}{\sigma_X\sigma_Y}$.求 $I(X;Y)$.

解

(1) 先求 X 和 Y 的一维概率密度函数

$$p(x)=\int_{-\infty}^{+\infty}p(xy)\mathrm{d}y=\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left[-\frac{(x-m_X)^2}{2\sigma_X^2}\right]$$

$$p(y)=\int_{-\infty}^{+\infty}p(xy)\mathrm{d}x=\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left[-\frac{(y-m_Y)^2}{2\sigma_Y^2}\right]$$

(2) 由平均互信息的定义求得

$$I(X;Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}p(xy)\ln\frac{p(xy)}{p(x)p(y)}\mathrm{d}x\mathrm{d}y$$

$$=\ln\frac{1}{\sqrt{1-\rho^2}}-\frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{(x-m_X)^2}{(1-\rho^2)\sigma_X^2}-\frac{2\rho(x-m_X)(y-m_Y)}{(1-\rho^2)\sigma_X\sigma_Y}+\frac{(y-m_Y)^2}{(1-\rho^2)\sigma_Y^2}-\frac{(x-m_X)^2}{\sigma_X^2}-\frac{(y-m_Y)^2}{\sigma_Y^2}\right]p(xy)\mathrm{d}x\mathrm{d}y$$

$$=-\frac{1}{2}\ln(1-\rho^2)-\frac{1}{2}\left[\frac{1}{1-\rho^2}-\frac{2\rho^2}{1-\rho^2}+\frac{1}{1-\rho^2}-1-1\right]$$

$$=-\frac{1}{2}\ln(1-\rho^2)\ \text{奈特/ 样值}$$

它表明两个高斯随机变量之间的互信息只与相关系数 ρ 有关,而与数学期望 m_X、m_Y 及方差 σ_X^2、σ_Y^2 无关 .因为数学期望 m_X、m_Y 代表变量的直流成分,而直流成分不会含有任何信息 .互信息只与归一化相关函数值或功率的相对大小有关,与功率的绝对大小无关 .

4 .5 .2　高斯加性信道的信道容量

可以证明,连续信道输入、输出随机变量的平均互信息是信源的概率密度 $p(x)$的上凸函数 .我们仍然定义连续信道的信道容量为平均互信息关于信源概率密度函数的极大

值,即

$$C = \max_{p(x)} I(X;\ Y) \tag{4 95}$$

一般连续信道的信道容量并不容易计算,而加性噪声信道则相对简单一些,下面只研究这种信道,其噪声(记为连续随机变量 N)与输入随机变量 X 相互统计独立 这种信道噪声对输入的干扰作用表现为输出是噪声和输入的线性叠加,即 $Y = X + N$.

对于加性噪声信道,由坐标变换理论可以证明 $p(y \mid x) = p(n)$,其中 $p(n)$是噪声 N 的概率密度函数,也就是说信道的条件概率密度函数等于噪声的概率密度函数,这时

$$\begin{aligned} h(Y \mid X) &= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(xy)\log_2 p(y \mid x)\,\mathrm{d}x\mathrm{d}y \\ &= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x)\,p(y \mid x)\log_2 p(y \mid x)\,\mathrm{d}x\mathrm{d}y \\ &= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x)\,p(n)\log_2 p(n)\,\mathrm{d}x\mathrm{d}n \\ &= -\int_{-\infty}^{+\infty} p(n)\log_2 p(n)\,\mathrm{d}n \\ &= h(N) \end{aligned} \tag{4 96}$$

该结论进一步说明条件熵 $h(Y \mid X)$是由信道中噪声引起的,它完全等于噪声的信源熵,所以称为噪声熵 其信道容量

$$C = \max_{p(x)} I(X;\ Y) = \max_{p(x)}[\,h(Y) - h(Y \mid X)\,] = \max_{p(x)}[\,h(Y) - h(N)\,] \tag{4 97}$$

由于加性信道的噪声 N 和信源 X 相互统计独立, X 的概率密度 $p(x)$的变动不会引起噪声熵 $h(N)$的改变,所以通过选择 $p(x)$使输出随机变量熵 $h(Y)$达到最大值时,加性信道即达到信道容量:

$$C = \max_{p(x)}\ h(Y) - h(N) \tag{4 98}$$

对于不同的限制条件,连续随机变量具有不同的最大值,所以连续信道的信道容量取决于输入随机变量 X 所受的限制条件以及噪声 N(即信道)的统计特性 .

可以证明,噪声功率 σ_N^2 给定后,高斯噪声信道的信道容量 C 最小,因此高斯信道是最差的信道 实际应用中往往把噪声视为高斯噪声 .下面研究噪声源为高斯噪声的加性信道 .

如果噪声 N 是均值为 0、方差为 σ_N^2 的高斯随机变量,即满足

$$\int_{-\infty}^{+\infty} p(n)\,\mathrm{d}n = 1$$

$$\int_{-\infty}^{+\infty} np(n)\,\mathrm{d}n = 0$$

$$\int_{-\infty}^{+\infty} n^2 p(n)\,\mathrm{d}n = \sigma_N^2 = P_N$$

其中, P_N 表示噪声 N 的平均功率 .这种信道称为高斯加性信道 .

一般输入随机变量 X 的平均功率是有限的,假设限定为 P_X,而噪声的平均功率限定为 $P_N=\sigma_N^2$,因而输出随机变量 Y 的平均功率也是有限的,设为 P_Y.根据最大连续熵定理,要使 $h(Y)$达到最大,Y 必须是一个高斯随机变量.而当输入 $p(x)$满足什么条件时才能使 Y 为高斯分布呢?

由概率论的知识可知,当 X、N 统计独立且 $Y=X+N$ 时,若输入 X 是均值为 0、方差为 $\sigma_X^2=P_X$ 的高斯随机变量,即 $p(x)=\dfrac{1}{\sqrt{2\pi\sigma_X^2}}\mathrm{e}^{-\frac{x^2}{2\sigma_X^2}}$,则 $p(y)=\dfrac{1}{\sqrt{2\pi\sigma_Y^2}}\mathrm{e}^{-\frac{y^2}{2\sigma_Y^2}}$,$Y$ 为高斯分布,并且

$$\sigma_Y^2=\sigma_X^2+\sigma_N^2=P_Y \tag{4.99}$$

也就是说,当输入随机变量 X 的概率密度是均值为 0、方差为 σ_X^2 的高斯随机变量,加性信道的噪声 N 是均值为 0、方差为 σ_N^2 的高斯随机变量时,输出随机变量 Y 也是一个高斯随机变量,其均值为 0、方差为 $\sigma_Y^2=\sigma_X^2+\sigma_N^2=P_Y$,此时输出随机变量的熵 $h(Y)$达到最大,而信道达到信道容量

$$\begin{aligned}C&=\max_{p(x)}h(Y)-h(N)\\&=\frac{1}{2}\log_2 2\pi\mathrm{e}(\sigma_X^2+\sigma_N^2)-\frac{1}{2}\log_2 2\pi\mathrm{e}\sigma_N^2\\&=\frac{1}{2}\log_2\frac{\sigma_X^2+\sigma_N^2}{\sigma_N^2}\\&=\frac{1}{2}\log_2\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right)\\&=\frac{1}{2}\log_2\left(1+\frac{P_X}{P_N}\right)\end{aligned} \tag{4.100}$$

其中,$\dfrac{P_X}{P_N}$称为信道的信噪比.

4.5.3　多维高斯加性信道的信道容量

对多维高斯加性信道,当噪声是加性噪声时信道必然是一个无记忆信道,所以

$$I(X;Y)\leq\sum_{k=1}^{n}I(X_k;Y_k) \tag{4.101}$$

因此

$$\begin{aligned}C&=\max_{p(x)}I(X;Y)=\max_{p(x)}\sum_{k=1}^{n}I(X_k;Y_k)\\&=\frac{1}{2}\sum_{k=1}^{n}\log_2\left(1+\frac{P_{X_k}}{P_{N_k}}\right)\end{aligned}$$

$$= \frac{n}{2}\log_2\left(1+\frac{P_X}{P_N}\right) \quad i=1,2,\cdots,n \tag{4.102}$$

当且仅当输入随机矢量 X 中各分量统计独立,并且均为高斯变量时达到信道容量.

如果在每个抽样时刻信源和噪声是均值为 0、方差分别为 σ_X^2 和 σ_N^2 的高斯随机变量,则

$$C=\frac{n}{2}\log_2\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right) \text{ 比特/ } n \text{ 个样值} \tag{4.103}$$

*4.6 波形信道及其信道容量

波形信道通常根据抽样定理转化成多维连续信道进行处理.

一般来说,信道的带宽总是有限的.假设某信道的频带限于$(0, B)$,则其输入、输出信号和噪声都是限频的随机过程,频带限于$(0, B)$.根据抽样定理,可把一个时间连续的信道变换成时间离散的信道来处理,即用每隔$1/(2B)$秒时间的采样值来表示输入、输出信号和噪声.我们把一次采样看成信道的一次传输,由于每秒传送 $2B$ 个样值,所以单位时间的信道容量为

$$C_t = B\log_2\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right) \text{ bit/ s} \tag{4.104}$$

当噪声是双边功率谱密度为$\frac{N_0}{2}$的高斯白噪声时,有

$$C_t = B\log_2\left(1+\frac{\sigma_X^2}{N_0 B}\right) \tag{4.105}$$

这就是著名的香农公式,它适用于加性高斯白噪声信道.从前面的讨论可知,只有当输入信号为功率受限的高斯白噪声信号时,才能达到该信道容量.

香农公式说明,当信道容量一定时,增大信道的带宽可以降低对信噪功率比的要求;反之,当信道频带较窄时,可以通过提高信噪功率比来补偿.

当 $B\to\infty$ 时,则

$$C_t=\lim_{B\to\infty} B\log_2\left(1+\frac{\sigma_X^2}{N_0 B}\right)=\lim_{B\to\infty}\frac{\sigma_X^2}{N_0}\frac{N_0 B}{\sigma_X^2}\log_2\left(1+\frac{\sigma_X^2}{N_0 B}\right)=\frac{\sigma_X^2}{N_0}\log_2 e=1.44\frac{\sigma_X^2}{N_0} \tag{4.106}$$

上式表明当频带很宽时,信道容量正比于信号功率与噪声谱密度之比.上式是加性高斯白噪声信道信息传输率的极限值.

【例 4.11】

一般模拟电话信道的带宽为 3 300 Hz,若信噪比为 20 dB(即$\frac{\sigma_X^2}{N_0 B}=100$),则根据香农

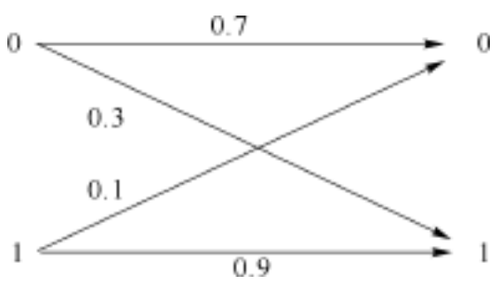
公式可得电话信道的信道容量:

$$C_t = B\log_2\left(1+\frac{\sigma_X^2}{N_0 B}\right) = 22\,000\ \text{bit/s}$$

由于高斯加性信道是实际信道中最差的信道,所以香农公式可用于确定实际信道的信道容量的下限值.香农公式给出了在有噪信道中无失真传输所能达到的极限信息传输率,因此对实际通信系统的设计有非常重要的指导意义.

习　题　4

4.1　设一个二元信道如题图 4.1 所示,其输入概率空间为$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.2 & 0.8 \end{bmatrix}$,试计算 $I(x=0;y=1)$, $I(x=1;Y)$和 $I(X;Y)$.



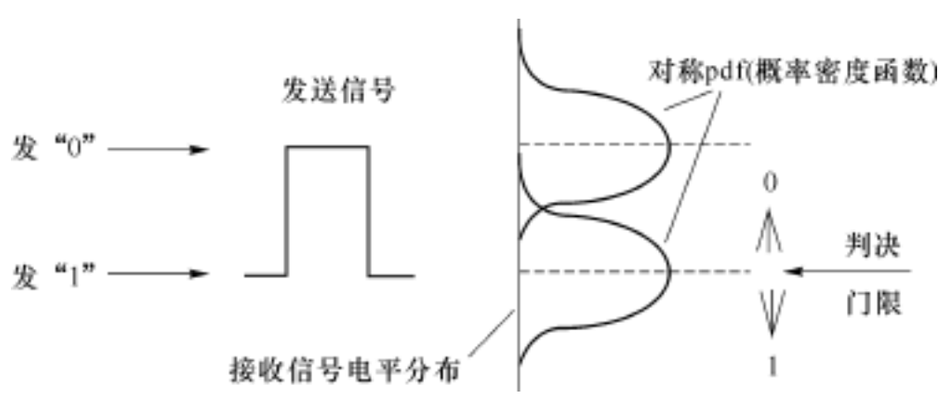
题图 4.1

4.2　二元删除信道有两个输入:0,1 和 3 个输出:0,1, E,其中 E 表示可检出但无法纠正的错误.信道前向转移概率是

$$p(0|0)=1-\alpha \qquad p(E|0)=\alpha \qquad p(1|0)=0$$
$$p(0|1)=0 \qquad p(E|1)=\alpha \qquad p(1|1)=1-\alpha$$

求信道容量 C.

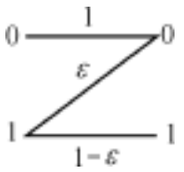
4.3　设某二进制数字传输系统接收判决器的输入信号电平、噪声密度分布、及判决电平如题图 4.2 所示.试求:(1)信道模型;(2)互信息;(3)信道容量.



题图 4.2

4.4 设有扰离散信道的输入端是以等概率出现的 A, B, C, D 4 个字母.该信道的正确传输概率为 1/2,错误传输概率平均分布在其他 3 个字母上.验证在该信道上每个字母传输的平均信息量为 0.21 bit.

4.5 Z 信道及它的输入、输出如题图 4.3 所示.



$$p(y|x)=\begin{bmatrix}1 & 0\\ \varepsilon & 1-\varepsilon\end{bmatrix}\quad x, y\in\{0,1\}$$

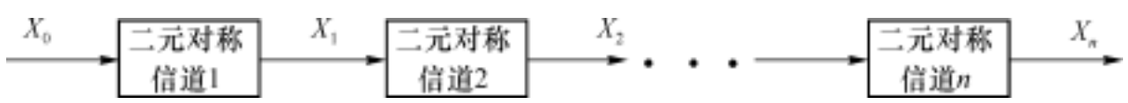
(1) 求最佳输入分布;

(2) 求 $\varepsilon=1/2$ 时的信道容量;

(3) 求当 $\varepsilon\to 0$ 和 $\varepsilon\to 1$ 时的最佳输入分布值.

题图 4.3

4.6 如题图 4.4 所示把 n 个二元对称信道串接起来,每个二元对称信道的错误传递概率为 p.证明这 n 个串接信道可以等效于一个二元对称信道,其错误传递概率为 $\frac{1}{2}[1-(1-2p)^n]$,并证明 $\lim\limits_{n\to\infty} I(X_0; X_n)=0$,设 $p\neq 0$ 或 1.

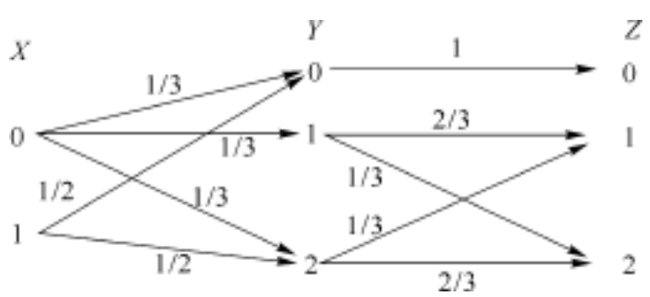


题图 4.4

4.7 试求出准对称信道的信道容量的一般表达式.

4.8 试画出三元对称信道在理想(无噪声)和强噪声(输出不依赖于输入)情况下的信道模型,设信道输入等概分布.

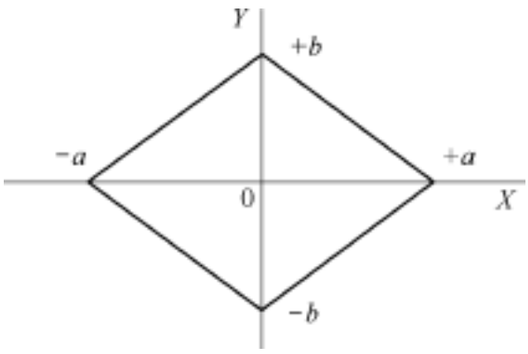
4.9 串联信道如题图 4.5 所示,求总的信道矩阵.



题图 4.5

4.10 设一时间离散、幅度连续的无记忆信道的输入是一个零均值、方差为 E 的高斯随机变量,信道噪声为加性高斯噪声,方差为 $\sigma^2=1$ μW,信道的符号传输速率为 $r=8\,000$符号/秒.如令一路电话通过该信道,电话机产生的信息率为 64 kbit/s,求输入信号功率 E 的最小值.

4.11 连续随机变量 X 和 Y 的联合概率密度函数在由 $\frac{1}{a}|x|+\frac{1}{b}|y|\leq 1$ 确定的菱形内均匀分布,如题图 4.6 所示.

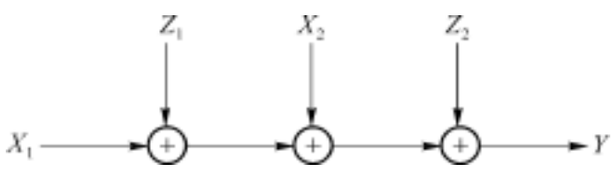


题图 4.6

(1) 求 $I(X;Y)$;

(2) 解释为什么 $I(X;Y)$与 a 和 b 无关.

4.12　高斯加性信道,输入信号 X_1, X_2,噪声信号 Z_1, Z_2,输出信号 $Y = X_1 + Z_1 + X_2 + Z_2$,如题图 4.7 所示.输入和噪声均为相互独立的零均值的高斯随机变量,功率分别为 P_1, P_2 和 N_1, N_2.



题图 4.7

(1) 求 $I(X_1;Y)$和 $I(X_2;Y)$;

(2) 求 $I(X_1X_2;Y)$;

(3) 当输入信号的总功率受限 $P_1 + P_2 \leq P$ 时,求 $I(X_1;Y) + I(X_2;Y)$的最大值.

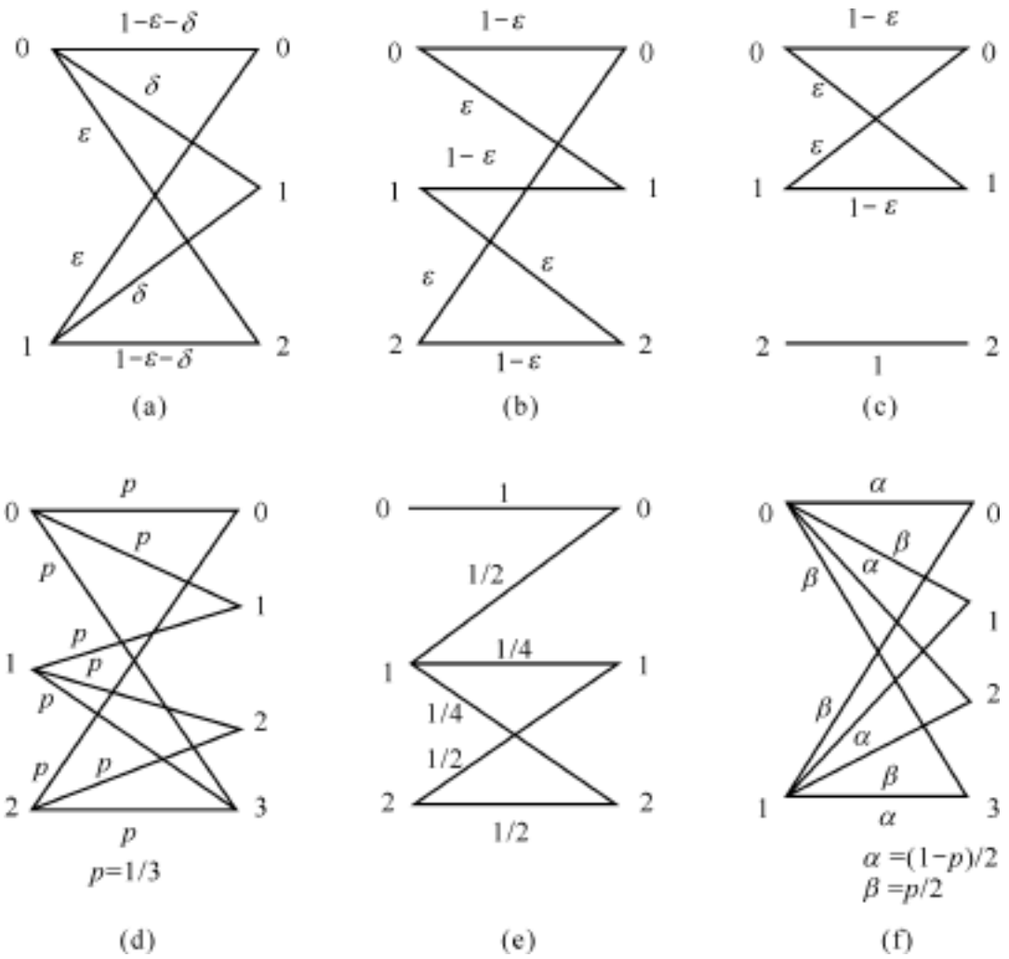
4.13　一个无记忆信道输入为离散随机变量 X,噪声 Z 在区间$[-a, +a]$上均匀分布,因此输出 $Y = X + Z$ 是一个连续随机变量.

(1) 当 $X \in \{-1, +1\}$并且等概分布时,求 $I(X;Y)$(用 a 表示);

(2) 当 $X \in \{-1, 0, +1\}$, $a = 1.2$ 时求最佳输入分布.

4.14　设某一信号的信息输出率为 5.6 kbit/s,噪声功率谱为 $N = 5 \times 10^{-6}$ mW/Hz,在带宽 $B = 4$ kHz 的高斯信道中传输.试求无差错传输需要的最小输入功率 P.

4.15　判断题图 4.8 中各信道是否对称,如对称,求出其信道容量.


1−ε−δ
0
0
δ
ε
1
ε
δ
1
2
1−ε−δ
(a)
1−ε
0
0
ε
1− ε
1
1
ε
ε
2
2
1−ε
(b)
1− ε
0
0
ε
ε
1
1
1− ε
2
2
1
(c)
p
0
0
p
p
1
p
1
p
p
2
p
p
2
3
p
p=1/3
(d)
1
0
0
1/2
1/4
1
1
1/4
1/2
2
2
1/2
(e)
α
0
0
β
β
α
1
β
2
α
β
1
3
α
α =(1−p)/2
β =p/2
(f)

题图 4 *8*

第 5 章

无失真信源编码

5.1 信源编码的相关概念

5.1.1 编码器

对于信源来说有两个重要问题:一个是信源输出信息量的定量度量问题;另一个是如何更有效地表示信源输出问题.第 3 章回答了第一个问题,这一章要讨论的是第二个问题.信源输出的符号序列,需要变换成适合信道传输的符号序列,一般称为码序列,对信源输出的原始符号按照一定的数学规则进行的这种变换称为编码,完成编码功能的器件,称为编码器.接收端有一个译码器完成相反的功能.

信源编码器的输入是信源符号集 $S=\{s_1, s_2, \cdots, s_q\}$,共有 q 个信源符号.同时存在另一个符号集 $X=\{x_1, x_2, \cdots, x_r\}$,称为码符号集,共有 r 个码符号,码符号集 X 中的元素称为码元或码符号,编码器的作用就是将信源符号集 S 中的符号 s_i, $i=1,2,\cdots,q$ 变换成由 l_i 个码符号组成的一一对应的码符号序列.编码器输出的码符号序列称为码字,并用 w_i, $i=1,2,\cdots,q$ 来表示,它与信源符号 s_i, $i=1,2,\cdots,q$ 之间是一一对应的关系,如

图 5.1 所示.

码字的集合 C 称为码,即 $C=\{w_1, w_2, \cdots, w_q\}$.信源符号 s_i 对应的码字 w_i 包含 l_i 个码符号,l_i 称为码字长度,简称码长.

所以,信源编码就是把信源符号序列变换到码符号序列的一种映射.若要实现无失真编码,那么这种映射必须是一一对应的、可逆的.一般来说,人们总是希望把信源所有的信息毫无保留地传递到接收端,即实现无失真传递,所以首先要对信源实现无失真编码.

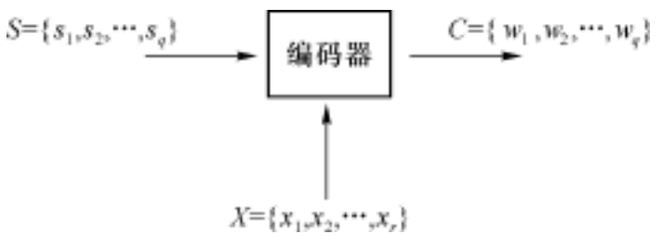


图 5.1 信源编码器

信源编码有以下 3 种主要方法:

(1) 匹配编码

根据信源符号的概率不同,编码的码长不同:概率大的信源符号所编的代码短;概率小的信源符号所编的代码长,这样使平均码长最短.将要讲述的香农编码、哈夫曼编码、费诺码都是概率匹配编码,都是无失真信源编码.

(2) 变换编码

先对信号进行变换,从一种信号空间变换为另一种信号空间,然后针对变换后的信号进行编码.一般是把分布在时空域的信号(如时域的语音信号和平面空间的图像信号)映射到变换域(如频域的频谱信号和其他正交矢量空间域),原来相关性很强的原始信号经过变换后,得到的变换域系数相互独立,并且能量集中在少数低序系数上,这样只需对少量的变换域的系数进行编码,达到数据压缩的目的.常用的变换编码有 DFT、沃尔什变换、K-L 变换、DCT、小波变换等.

(3) 识别编码

识别编码主要用于印刷或打字机等有标准形状的文字、符号和数据的编码,比如中文文字和语音的识别.

后两种信源编码均为有失真的信源编码.

无失真信源编码主要针对离散信源,连续信源在量化编码的过程中必然会有量化失真,所以对连续信源只能近似地再现信源的消息.

由第 3 章的讨论可知,由于信源概率分布的不均匀性和符号之间存在的相关性,使得信源存在冗余度,而实际上传送信源信息只需要传送信源极限熵大小的信息量.信源编码的主要任务就是减少冗余度,提高编码效率.具体地说,就是针对信源输出符号序列的统计特性,寻找一定的方法把信源输出符号序列变换为最短的码字序列.去除冗余度的方法有两个:一是去除相关性,使编码后码序列的各个码符号尽可能地互相独立,这一般利用

对信源的符号序列进行编码而不是对单个的信源符号进行编码的方法实现;二是使编码后各个码符号出现的概率尽可能地相等,即概率分布均匀化,这可以通过概率匹配的方法实现,也就是使小概率消息对应长码,大概率消息对应短码 .

下面举一个例子说明怎样用编码器实现对信源符号的编码 .

【例 5 .1】

信源概率空间为

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & \cdots & s_q \\ p(s_1) & p(s_2) & \cdots & p(s_q) \end{bmatrix}$$

把信源编码为适合二元信道传输的二元码 .二元信道是数字通信中常用的一种信道,它的码符号集为{0,1} .

解

表 5 .1　二元码

信源符号 s_i	$p(s_i)$	码 1	码 2
s_1	$p(s_1)$	00	0
s_2	$p(s_2)$	01	01
s_3	$p(s_3)$	10	001
s_4	$p(s_4)$	11	111

将信源通过一个二元信道传输,就必须把信源符号变换成由 0、1 符号组成的码符号序列 .对每个信源符号,可以编成不同的二元码符号序列,就可得到不同的码 .当 $q=4$ 时,得到表 5 .1 .

一般情况下,码可分两类:一类是固定长度码;另一类是可变长度码 .固定长度码又称为定长码,码中所有码字的长度都相同,如表 5 .1 中的码 1 .可变长度码又称为变长码,码中所有码字的长短不一,即码字中码符号个数不同,如表 5 .1 中的码 2 .

本章讨论的都是同价码,即每个码符号所占的传输时间都相同的码 .显然,对同价码来说,定长码中每个码字 w_i 的传输时间相等,而变长码中每个码字的传输时间不一定相等 .

对离散无记忆信源的 N 次扩展信源进行编码便得到 N 次扩展码 .

假定信源符号集为 $S=\{s_1, s_2, \cdots, s_q\}$,则相应的 N 次扩展信源为 $S=\{s_1, s_2, \cdots, s_q{}^N\}$,对应的 N 次扩展码为 $C=\{w_1, w_2, \cdots, w_q{}^N\}$.

例如表 5 .1 中码 2 的二次扩展码如表 5 .2 所示 .

表 5 .2　二次扩展码

二次扩展信源符号 s_j	二次扩展码字 w_j
$s_1 = s_1 s_1$	00
$s_2 = s_1 s_2$	001
$s_3 = s_1 s_3$	0001
…	…
$s_{16} = s_4 s_4$	111111

5 .1 .2　码的分类

1 . 分组码和非分组码

信源编码过程可以抽象为一种映射,即将信源符号集 $S=\{s_i, i=1,2,\cdots,q\}$ 中的每

一个元素 s_i 映射为一个码中长度为 l_i 的码字 w_i, $i=1,2,\cdots,q$.

定义 5 .1 将信源符号集中的每个信源符号 s_i 固定地映射成一个码字 w_i,这样的码称为分组码 .

例 5 .1 就是把信源的符号序列分成组,从信源符号序列到码字序列的变换是在分组的基础上进行的,即特定的信源符号组唯一地确定了特定的码字组 .

用分组码对信源符号进行编码时,为了使接收端能够迅速准确地将码译出,分组码必须具有一些直观属性 .与分组码对应的是非分组码,又称为树码 .树码编码器输出的码符号通常与编码器的所有信源符号都有关 .

2 .奇异码与非奇异码

定义 5 .2 若一种分组码中的所有码字都不相同,则称此分组码为非奇异码,否则称为奇异码 .

表 5 .3 奇异码和非奇异码

信源符号 s_i	码 1	码 2
s_1	0	0
s_2	11	10
s_3	00	00
s_4	11	01

表 5 .3 中,码 1 是奇异码,码 2 是非奇异码 .非奇异码是分组码能够正确译码的必要条件,而不是充分条件 .例如传送分组码 2 时,如果接收端接收到 00 时,并不能确定发送端的消息是 $s_1 s_1$ 还是 s_3 .

3 .唯一可译码与非唯一可译码

定义 5 .3 任意有限长的码元序列,如果只能唯一地分割成一个个码字,便称为唯一可译码 .

一个分组码若对于任意有限的整数 N,其 N 次扩展码均为非奇异的,则为唯一可译码 .

唯一可译码的物理含义是指不仅要求不同的码字表示不同的信源符号,而且还要求对由信源符号构成的符号序列进行编码时,在接收端仍能正确译码而不发生混淆 .唯一可译码首先是非奇异码,且任意有限长的码字序列不会雷同 .

表 5 .4 即时码和唯一可译码

信源符号 s_i	码 1	码 2
s_1	1	1
s_2	10	01
s_3	100	001
s_4	1000	0001

同是唯一可译码,其译码方法仍有不同 .如表 5 .4 中列出的两组唯一可译码,其译码方法不同 .当传送码 1 时, 信道输出端接收到一个码字后不能立即译码,还需等到下一个码字接收到时才能判断是否可以译码 .若传送码 2,则无此限制,接收到一个完整码字后立即可以译码,我们称后一种码为逗点码,它是一种即时码,是唯一可译码的一种 .

4 .即时码与非即时码

定义 5 .4 无需考虑后续的码符号就可以从码符号序列中译出码字,这样的唯一可译码称为即时码 .

表 5 .4 给出了唯一可译码和即时码的简单示例 .

下面讨论唯一可译码成为即时码的条件 .

定义 5 5　设 $w_i = x_{i_1} x_{i_2} \cdots x_{i_l}$ 为一码字,对于任意的 $1 \leq j \leq l$,称码符号序列的前 j 个元素 $x_{i_1} x_{i_2} \cdots x_{i_j}$ 为码字 w_i 的前缀 .

按照上述的前缀的定义,有下述结论:

定理 5 .1　一个唯一可译码成为即时码的充要条件是其中任何一个码字都不是其他码字的前缀 .

证明

充分性:

如果没有一个码字是其他码字的前缀,则在接收到一个相当于一个完整码字的码符号序列后便可以立即译码,而无需考虑其后的码符号 .

必要性:

如果设 w_i 是 w_j 的前缀,则在收到相当于 w_i 的码符号序列后还不能立即判定它是一个完整的码字,若想正确译码,还必须参考后续的码符号,这与即时码的定义相矛盾,所以即时码的必要条件是其中任何一个码字都不是其他的码字的前缀 .

证毕

即时码可以用树图来构造 .图 5 .2 是一个二元即时码的树图 .

树是没有回路的图,所以它也是由节点和弧构成的 .树中最顶部的节点称为根节点,没有子节点的节点称为叶子节点 .在构造即时码的树图中,每个节点最多有 r 个子节点,在从此节点到其若干个子节点的弧上分别标柱着 $x_1, x_2, \cdots, x_n$, 这里的 $n \leq r$, r 为码符号的个数 .将从根节点到叶子节点的各段弧上的码符号顺次连接,就可得到相应的码字 .

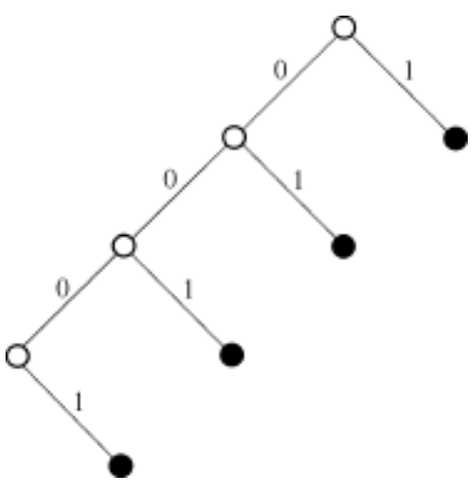


图 5 .2　二元即时码的树图

所有根节点的子节点称为一阶节点,所有一阶节点的子节点称为二阶节点,依此类推 .n 阶节点最多有 r^n 个 .节点的阶次又称为节点的深度 .

即时码的树图还可以用来译码,当接收到一串码符号序列后,首先从树的根节点出发,根据接收到的第一个码符号来选择应走的第一条路径,若沿着所选支路走到中间节点,那就再根据接收到的第二个码符号来选择应走的第二条路径,若又走到中间节点,就再依次继续下去,直到叶子节点为止 .走到叶子节点,就可根据所走的枝路立即判断出所接收的码字,同时使系统重新返回根节点,再做下一个接收码字的判断 .这样就可以将接收到的一串码符号序列译成对应的信源符号序列 .

综上所述,可将信源编码作如下分类:

- 码
  - 非分组码(树码)
  - 分组码(块码)
    - 奇异码
    - 非奇异码
      - 非唯一可译码
      - 唯一可译码
        - 即时码
        - 非即时码

5.2 定长码及定长编码定理

一般来说,若要实现无失真的编码,所编的码必须是唯一可译码,否则,就会因译码带来错误与失真.

对于定长码来说,若定长码是非奇异码,则它的任意有限长 N 次扩展码一定也是非奇异码,因此定长非奇异码一定是唯一可译码.

表 5.5 中码 2 是奇异码.当接收到码符号 11 后既可译成 s_2 也可译成 s_4,所以不能唯一地译码.而码 1 是等长非奇异码,因此它是一个唯一可译码.

表 5.5 定长码

信源符号 s_i	码 1	码 2
s_1	00	00
s_2	01	11
s_3	10	10
s_4	11	11

若对一个有 q 个信源符号的信源 S 进行定长编码,那么信源 S 存在唯一可译定长码的条件是

$$q \leq r^l \tag{5.1}$$

其中,r 是码符号集中的码元数,l 是定长码的码长.例如表 5.5 中,信源 S 共有 $q=4$ 个信源符号,进行二元等长编码,$r=2$,则信源存在唯一可译定长码的条件是 $l \geq 2$.

如果对信源 S 的 N 次扩展信源 S^N 进行定长编码,若要编得的定长码是唯一可译码,则必须满足

$$q^N \leq r^l \tag{5.2}$$

其中,q 是信源 S 的符号个数,q^N 是信源 S 的 N 次扩展信源 S^N 的符号个数,r 是码符号集 X 的码符号数.当把 N 次扩展信源 S^N 的信源符号 s_j, $j=1,2,\cdots,q^N$ 变换成长度为 l 的码字 $w_j = x_{i_1} x_{i_2} \cdots x_{i_l}$ 时,如果要求编得的定长码是唯一可译码,则每一个信源符号 s_j 都有一个码字对应,所以必须满足式(5.2).换句话说,只有当 l 长的码符号序列数 r^l 不小于 N 次扩展信源的符号序列个数 q^N 时,才能存在定长非奇异码.

对式(5.2)两边取对数得 $N\log_2 q \leq l\log_2 r$,则

$$\frac{l}{N} \geq \frac{\log_2 q}{\log_2 r}$$

其中,$\frac{l}{N}$ 表示 S^N 中平均每个原始信源符号所需要的码符号个数,对于定长唯一可译码,

平均每个原始信源符号至少需要 $\log_r q$ 个码符号来表示 .

当 $r=2$ 时, $\frac{l}{N}\geq\log_2 q$, 表示对于二元定长唯一可译码, 平均每个原始信源符号至少需要用 $\log_2 q$ 个二元符号来表示 .

当 $N=1$ 时, 则有 $l\geq\log_2 q$.

例如英文电报有 32 个符号(26 个英文字母加 6 个标点符号), 即 $q=32$, 若利用二元码, 则 $r=2$, 若对信源 S 的每个符号 s_i, $i=1,2,\cdots,32$ 进行二元编码, 则 $l\geq\log_{2_r} q=\log_{2_2}32=5$, 也就是说, 每个英文电报符号至少要用 5 位二元码符号进行编码才能得到唯一可译码 .

由第 4 章已知, 实际英文电报信源, 在考虑了符号出现的概率以及符号之间的相关性以后, 平均每个英文电报符号所提供的信息量约等于 1 .4 bit, 远小于 5 bit .因此, 这种定长编码后, 每个码字只载荷约 1 .4 bit 信息量, 也就是 5 个二元码符号只携带约 1 .4 bit 信息量, 而 5 个二元码符号最大能载荷 5 bit 的信息量 .因此, 这种定长编码的信息传输效率是很低的 .那么怎么才能提高信息传输效率呢 ?

方法 1

考虑符号之间的依赖关系, 对信源 S 的扩展信源进行编码, 考虑符号间的依赖关系以后, 有些信源符号序列不会出现, 这样可能出现的信源符号序列个数会小于 q^N .

方法 2

对于概率等于 0 或非常小的符号序列不予编码, 这样可能会造成一定的误差, 但是, 当 N 足够长以后, 这种误差概率可以任意小, 即可做到几乎无失真编码 .

下面将要讨论的定长编码定理给出了定长编码所需码长的理论极限值 .

由渐近等同分割性和 ε 典型序列(参见附录 B .4)的性质, 可以很容易地推出以下定长信源编码定理 .

定理 5 .2　离散无记忆信源的熵为 $H(S)$, 若对信源长为 N 的序列进行定长编码, 码符号集 X 中有 r 个码符号, 码长为 l, 则对于任意 $\varepsilon>0$, 只要满足 $\frac{l}{N}\geq\frac{H(S)+\varepsilon}{\log_2 r}$, 则当 N 足够大时, 可实现几乎无失真编码, 即译码错误概率 δ 任意小; 反之, 如果 $\frac{l}{N}<\frac{H(S)-2\varepsilon}{\log_2 r}$, 则不可能实现几乎无失真编码, 即当 N 足够大时, 译码错误概率为 1 .

证明

离散无记忆信源的 N 次扩展信源的输出序列可以分为两类, 一类是高概率的 ε 典型序列, 另一类是低概率的非 ε 典型序列, 当 $N\to\infty$ 时, ε 典型序列出现的概率趋于 1, 非 ε 典型序列出现的概率趋于零 .由于 ε 典型序列在全部序列中的比例很小, 因此只对少数的 ε 典型序列进行一一对应的等长编码, 这就要求码字总数不小于 M_G, 即

$$r^l\geq M_G \tag{5.3}$$

$$r^l \geq 2^{N[H(S)+\varepsilon]} \geq M_G \tag{5.4}$$

$$l\log_2 r \geq N[H(S)+\varepsilon] \tag{5.5}$$

$$\frac{l}{N} \geq \frac{H(S)+\varepsilon}{\log_2 r} \tag{5.6}$$

这样就能使 S^N 中所有的 ε 典型序列都有不同的码字与其对应，而非 ε 典型序列的总概率是很小的，但这些非 ε 典型序列仍可能出现，因而会造成译码错误，其错误概率就是 $\overline{G_\varepsilon}$ 出现的概率．因此

$$P_E = P_r\{\overline{G_\varepsilon}\} \leq \delta(N,\varepsilon) = \frac{D[I(s_i)]}{N\varepsilon^2} \tag{5.7}$$

当 $N\to\infty$ 时，$P_E \to 0$．

反之，如果 $\frac{l}{N} \leq \frac{H(S)-2\varepsilon}{\log_2 r}$，则

$$r^l \leq 2^{N[H(S)-2\varepsilon]} \leq 2^{-N\varepsilon} 2^{N[H(S)-\varepsilon]} \leq [1-\delta(N,\varepsilon)]2^{N[H(S)-\varepsilon]}$$

所以此时选取的码字总数小于 ε 典型序列集中序列数，因而 ε 典型序列集中将有部分序列没有码字与其对应，把有对应码字的序列的概率和记作 $P_r\{G_\varepsilon$ 中 r^l 个 $s_j\}$，它必然满足

$$P_r\{G_\varepsilon \text{ 中 } r^l \text{ 个 } s_j\} \leq r^l 2^{-N[H(N)-\varepsilon]} \leq 2^{N[H(S)-2\varepsilon]} 2^{-N[H(S)-\varepsilon]} = 2^{-N\varepsilon} \tag{5.8}$$

G_ε 中 r^l 个信源符号序列由于有不同的码字与其对应，所以在译码时能正确恢复，其他没有对应码字的信源序列在译码时都会产生错误，因而正确译码概率 $\overline{P_E} = 1 - P_E = P_r\{G_\varepsilon$ 中 r^l 个 $s_j\}$．

$1 - P_E \leq 2^{-N\varepsilon}$，$P_E \geq 1 - 2^{-N\varepsilon}$，当 $N\to\infty$ 时，$P_E \to 1$．所以定长编码时，如果码长 l 满足 $\frac{l}{N} < \frac{H(S)-2\varepsilon}{\log_2 r}$，当 N 很大时，将使许多经常出现的 ε 典型序列被舍弃而没有编码，这样就造成很大的译码错误，不可能实现几乎无失真编码．

证毕

定理 5.2 是在离散平稳无记忆信源的条件下证明的，但它同样适合于平稳有记忆信源，只要将式中 $H(S)$ 改为极限熵 $H_\infty(S)$ 即可，条件是有记忆信源的极限熵 $H_\infty(S)$ 和极限方差 $\sigma^2_\infty(S)$ 存在．

定长信源编码定理给出了定长编码时每个信源符号最少所需的码符号的理论极限，该极限值由 $H(S)$ 决定．

当二元编码时，$r=2$，有

$$\frac{l}{N} \geq H(S)+\varepsilon \tag{5.9}$$

这是定长编码时平均每个信源符号所需的二元码符号数的理论极限．

比较式(5.9)与 $\frac{l}{N} \geq \log_2 q$，当信源符号等概分布时两式就完全一致，但一般情况下信源符号并非等概率分布，而且符号之间有很强的关联性，故信源的熵 $H_\infty(S)$（极限熵）大

大小于 $\log_2 q$.根据定长信源编码定理,每个信源符号平均所需的二元码符号可大大减少,从而使编码效率提高 .

仍以英文电报为例,由于英文信源的极限熵 $H_\infty(S) \approx 1.4$ 比特/ 信源符号,所以,码长 l 要满足 $\frac{l}{N} > 1.4$ 二元码符号/ 信源符号,即平均每个英文符号只需近似用 1 .4 个二元符号来编码,这比由 $\frac{l}{N} \geq \log_2 q$ 计算的需要 5 位二元符号减少了许多,从而提高了信息传输效率 .

定理 5 .2 的条件又可写成 $l\log_2 r > NH(S)$,这个式子表明只要 l 长码符号序列所能携带的最大信息量大于 N 长信源序列所携带的信息量,就可以实现无失真编码,当然条件是 N 足够大 .

定义 5 .6　设熵为 $H(S)$的离散无记忆信源,若对信源的长为 N 的符号序列进行定长编码,码符号集中码符号个数为 r,设码字长为 l,定义 $R' = \frac{l}{N}\log_2 r$ 比特/ 信源符号为编码速率,它表示平均每个信源符号编码后能载荷的最大信息量 .

这时,定理 5 .2 的条件可表述为 $R' \geq H(S) + \varepsilon$,即编码速率大于信源的熵才能实现几乎无失真编码 .为了衡量编码效果,引进编码效率 .

定义 5 .7　定义 $\eta = \frac{H(S)}{R'} = \frac{H(S)}{\frac{l}{N}\log_2 r}$ 为编码效率 .

由定理 5 .2 可得最佳编码效率为 $\eta = \frac{H(S)}{H(S) + \varepsilon}$, $\varepsilon > 0$,所以 $\varepsilon = \frac{1 - \eta}{\eta}H(S)$.由于 $P_E = P_r\{\overline{G_\varepsilon}\} \leq \delta(N, \varepsilon) = \frac{D[I(s_i)]}{N\varepsilon^2}$,当方差 $D[I(s_i)]$和 ε 均为定值时,只要 N 足够大,$\delta(N, \varepsilon)$就可以任意小,错误概率 P_E 就可以小于任一正数 δ .当允许错误概率 P_E 不大于某给定的 δ 时,信源序列长度 N 必满足

$$N \geq \frac{D[I(s_i)]}{\varepsilon^2\delta} = \frac{D[I(s_i)]}{H^2(S)} \frac{\eta^2}{(1 - \eta)^2\delta} \tag{5.10}$$

式(5 .10)给出了在已知方差和信源熵的条件下,信源符号序列长度 N 与最佳编码效率 η 和允许错误概率 P_E 的关系,当允许错误概率越小,编码效率又要高,那么信源符号序列长度 N 必须越长 .在实际情况下,要实现几乎无失真的定长编码, N 需要的长度将会大到难以实现 .

【例 5 .2】

设有离散无记忆信源

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 \\ 0.40 & 0.18 & 0.10 & 0.10 & 0.07 & 0.06 & 0.05 & 0.04 \end{bmatrix}$$

如果对信源符号采用定长二元编码,要求编码效率 $\eta = 90\%$,允许错误概率 $P_E \leq 10^{-6}$,求

所需信源序列长度 N.

解

信息熵 $H(S)=E[-\log_2 p(s_i)]=-\sum_{i=1}^{8}p(s_i)\log_2 p(s_i)=2.55$ 比特/信源符号

自信息的方差

$$D[I(s_i)]=\sum_{i=1}^{8}p(s_i)[-\log_2 p(s_i)]^2-H^2(S)=7.82$$

$$\varepsilon=\frac{1-\eta}{\eta}H(S)=0.28$$

所以
$$N\geq\frac{D[I(s_i)]}{\varepsilon^2\delta}=\frac{7.82}{0.28^2\times10^{-6}}\approx9.8\times10^7\approx10^8$$

即信源序列长度 N 需长达 10^8 以上才能实现上述给定的要求.

【例 5.3】

设离散无记忆信源

$$\begin{bmatrix}S\\P\end{bmatrix}=\begin{bmatrix}s_1 & s_2\\3/4 & 1/4\end{bmatrix}$$

要求 $\eta=0.96,\delta\leq10^{-5}$, 求 N.

解

信源熵 $H(S)=\frac{1}{4}\log_2 4+\frac{3}{4}\log_2\frac{4}{3}=0.811$ 比特/信源符号

自信息的方差

$$\begin{aligned}D[I(s_i)]&=\sum_{i=1}^{2}p(s_i)[-\log_2 p(s_i)]^2-H^2(S)\\&=\frac{3}{4}\times\left(\log_2\frac{3}{4}\right)^2+\frac{1}{4}\times\left(\log_2\frac{1}{4}\right)^2-(0.811)^2\\&=0.4715\end{aligned}$$

因为
$$\varepsilon=\frac{1-\eta}{\eta}H(S)$$

所以

$$\begin{aligned}N&\geq\frac{D[I(s_i)]}{\varepsilon^2\delta}=\frac{D[I(s_i)]}{\delta}\frac{\eta^2}{(1-\eta)^2H^2(S)}\\&=\frac{0.4715}{10^5}\times\frac{(0.96)^2}{(0.04)^2\times(0.811)^2}\\&\approx4.13\times10^7\end{aligned}$$

即信源序列长度长达 4×10^7 以上, 才能实现给定的要求, 这在实际中是很难实现的.

因此,一般来说,当 N 有限时,高传输效率的定长码往往要引入一定的失真和错误,但是变长码则可以在 N 不大时实现无失真编码.

5.3　变长码及变长编码定理

变长码要成为唯一可译码不仅本身应是非奇异的,而且它的有限长 N 次扩展码也应是非奇异的.例如表 5.6 中码 2 是非奇异码,但是当信宿收到"00"时,它不能判断是 s_1s_1 还是 s_3,所以不是唯一可译码.

表 5.6　变长码

信源符号 s_i	概率分布	码 1	码 2	码 3	码 4
s_1	1/2	0	0	1	1
s_2	1/4	11	10	10	01
s_3	1/8	00	00	100	001
s_4	1/8	11	01	1000	0001

5.3.1　Kraft 不等式和 McMillan 不等式

信源符号数、码符号数、码字长度之间满足什么条件才可以构成即时码和唯一可译码呢?Kraft 不等式和 McMillan 不等式回答了上述问题.这两个不等式在形式上是完全一样.

定理 5.3　设信源符号集为 $S=\{s_1,s_2,\cdots,s_q\}$,码符号集为 $X=\{x_1,x_2,\cdots,x_r\}$,对信源进行编码,得到的码为 $C=\{w_1,w_2,\cdots,w_q\}$,码长分别为 $l_1,l_2,\cdots,l_q$.即时码存在的充要条件是

$$\sum_{i=1}^{q} r^{-l_i} \leq 1 \tag{5.11}$$

这称为 Kraft 不等式.

证明

充分性:

证明满足不等式(5.11)便可得到即时码.

现在假设码字长度满足 $l_1 \leq l_2 \leq \cdots \leq l_q \leq l$,由于这只影响码字的编号,所以是可以的.深度为 l 的 r 叉树,如果所有分支都延伸到最后一节则有 r^l 个叶子节点,可以编 r^l 个码字.可以从根节点出发,在 l_1 阶上取一个节点作为码字,则后续的树枝被砍去,共有 r^{l-l_1} 个 l 阶节点不能得到使用.

同样,在 l_i 阶节点上的终端节点使得 r^{l-l_i} 个 l 阶节点不能得到使用.最后 l 阶节点只剩下 $r^l - r^{l-l_1} - r^{l-l_2} - \cdots - r^{l-l_q} = r^l - \sum_{i=1}^{q} r^{l-l_i}$ 个可以使用的终端节点.

由定理给出的条件 $\sum_{i=1}^{q} r^{-l_i} \leq 1$ 可以得出 $1 - \sum_{i=1}^{q} r^{-l_i} \geq 0$, $r^l - \sum_{i=1}^{q} r^{l-l_i} \geq 0$,即可以使用的 r 阶节点数大于等于 0,所以可以按照上述方法构造得到一个即时码.

必要性:

证即时码必然满足不等式(5.10).

由于即时码必然可以用树图来构造,并且叶子节点不会再生出树枝.我们可取一个有 l 阶的 r 叉树且 $l \geq \max_i l_i$,树的第 0 阶为根,在第 l 阶上共有 r^l 个节点,于是长为 l_i 的码字相当于砍去了该 r 叉树第 i 阶上的 r^{l-l_i} 个节点,q 个码字共砍去第 l 阶的节点数必小于 r^l 即 $\sum_{i=1}^{q} r^{l-l_i} \leq r^l$, $\sum_{i=1}^{q} r^{-l_i} \leq l$.

证毕

由 Kraft 不等式可知,给定 r 和 q,只要允许码字长度可以足够长,则码长总可以满足 Kraft 不等式而得到即时码,Kraft 不等式指出了即时码的码长必须满足的条件.后来,McMillan 证明对于唯一可译码的码长也必须满足此不等式.在码长的选择上唯一可译码并不比即时码有更宽松的条件.对于唯一可译码,该不等式又称为 McMillan 不等式.

定理 5.4 唯一可译码存在的充要条件是

$$\sum_{i=1}^{q} r^{-l_i} \leq 1 \tag{5.12}$$

r 为码符号个数,l_i 为码字长度,q 为信源符号个数.

证明

充分性:

由于即时码就是唯一可译码,所以由定理 5.3 可知,满足该不等式的条件便可以构造一个即时码,因此,也就可以构造一个唯一可译码.

必要性:

即证明唯一可译码必须满足不等式 $\sum_{i=1}^{q} r^{-l_i} \leq 1$.

一个唯一可译码码组中的任意长度相同的码字序列必定不相同.设一个唯一可译码码组中码字的长度为 l_1, l_2, $\cdots$, l_q,考虑它的 N 次扩展码,N 是一个任意的正整数,考虑等式

$$\left[\sum_{i=1}^{q} r^{-l_i}\right]^N = \left(r^{-l_1} + r^{-l_2} + \cdots + r^{-l_q}\right)^N$$

$$= \sum_{i_1=1}^{q} r^{-l_{i_1}} \sum_{i_2=1}^{q} r^{-l_{i_2}} \cdots \sum_{i_N=1}^{q} r^{-l_{i_N}}$$

$$= \sum_{i_1=1}^{q}\sum_{i_2=1}^{q}\cdots\sum_{i_N=1}^{q} r^{-(l_{i_1}+l_{i_2}+\cdots+l_{i_N})} \tag{5.13}$$

上式右边当 $i_1, i_2, \cdots, i_N$ 分别取 1～q 之间的各种值时就产生了 N 个码字，可能构成的全部序列共有 q^N 项，每项对应于 N 个码字组成的一个码字序列，如图 5.3 所示．

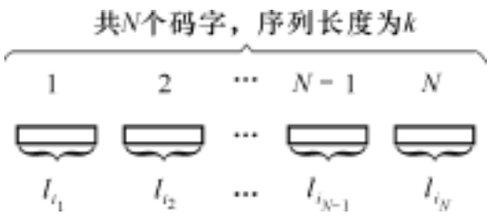


图 5.3　码的 N 次扩展

图中 $1, 2, \cdots, N$ 表示码字的序号，$l_{i1}, l_{i2}, \cdots, l_{iN}$ 分别为对应码字的码长．

$$k = l_{i_1} + l_{i_2} + \cdots + l_{i_N} \qquad i_1, i_2, \cdots, i_N \in \{1, 2, \cdots, q\} \tag{5.14}$$

其中，k 为 N 个长度分别为 $l_{i1}, l_{i2}, \cdots, l_{iN}$ 的码字组成的码字序列的总长度，也就是扩展码的码长．因为是变长码，故单个码字 w_i 对应的长度 l_i 的取值范围为 $1 \le l_i \le l_{\max}$，因此有 $N \le k \le Nl_{\max}$．

在 $r^{-(l_{i_1}+l_{i_2}+\cdots+l_{i_N})}$ 中，因为 $l_{i1}, l_{i2}, \cdots, l_{iN}$ 都可以取 $l_1, l_2, \cdots, l_q$ 中任一值，而 $l_1, l_2, \cdots, l_q$ 又都可以取 $1, 2, \cdots, l_{\max}$ 中之一，故相同 k 值的项可能出现不止一次，也就是在 q^N 个码字序列中码符号总长相等的码字序列不止一个．令总长为 k 的码字序列共有 M_k 个．例如在表 5.7 中，考虑码 $C = \{0, 10, 00, 01\}$ 的码长分别为 $l_1 = 1$，$l_2 = l_3 = l_4 = 2$，它的二次扩展码的码长 k 分别为 2，3，4．当 $k = 3$ 时，共有 6 个不同的码字序列，所以 $M_k = 6$．合并同类项后有

表 5.7　唯一可译码

信源符号	码字	信源符号	码字
$s_1 s_1$	00	$s_3 s_1$	000
$s_1 s_2$	010	$s_3 s_2$	0010
$s_1 s_3$	000	$s_3 s_3$	0000
$s_1 s_4$	001	$s_3 s_4$	0001
$s_2 s_1$	100	$s_4 s_1$	010
$s_2 s_2$	1010	$s_4 s_2$	0110
$s_2 s_3$	1000	$s_4 s_3$	0100
$s_2 s_4$	1001	$s_4 s_4$	0101

$$\left[\sum_{i=1}^{q} r^{-l_i}\right]^N = \sum_{k=N}^{Nl_{\max}} M_k r^{-k} \tag{5.15}$$

因为是唯一可译码，故总长为 k 的所有码字序列必定是不相同的，即非奇异的，故必存在以下关系：

$$\left[\sum_{i=1}^{q} r^{-l_i}\right]^N \le \sum_{k=N}^{Nl_{\max}} r^{-k} r^k = Nl_{\max} - N + 1 \le Nl_{\max} \tag{5.16}$$

其中，$M_k \le r^k$，r^k 是 k 长的 r 元码序列的总数．因此

$$\sum_{i=1}^{q} r^{-l_i} < (Nl_{\max})^{\frac{1}{N}} \tag{5.17}$$

对于一切正整数 N,上式均成立,所以当 $N\to\infty$时

$$\lim_{N\to\infty}(Nl_{\max})^{\frac{1}{N}}=1 \tag{5.18}$$

因此

$$\sum_{i=1}^{q}r^{-l_i}\leq 1 \tag{5.19}$$

由此证得,唯一可译码一定满足不等式(5.12).

证毕

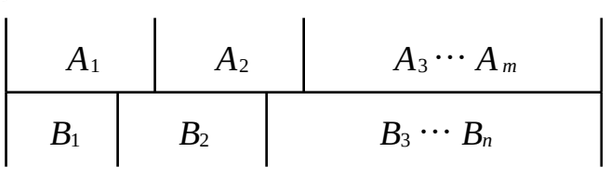
定理5.4指出了唯一可译码中 r、q、l_i 之间的关系,如果满足这个不等式的条件,则一定能够构成至少一种唯一可译码,否则,无法构成唯一可译码.它给出了唯一可译变长码的存在性.

例如在表5.6中,码1、码2码长为 $l_1=1$, $l_2=l_3=l_4=2$, $\sum_{i=1}^{4}2^{-l_i}=\frac{5}{4}\geq 1$,所以不能构成唯一可译码.而码3、码4的码长为 $l_1=1$, $l_2=2$, $l_3=3$, $l_4=4$, $\sum_{i=1}^{4}2^{-l_i}=\frac{15}{16}\leq 1$,可以构成唯一可译码.当然满足此条件的也不一定就是唯一可译码,例如 $C=\{1,01,011,0001\}$,但至少可以找到一种唯一可译码.同样满足此码长条件的码也可能不是即时码,但至少可以找到一种即时码.

另外,从定理5.3和定理5.4可以得到一个重要的结论,即任何一个唯一可译码均可用一个相同码长的即时码来代替,因为即时码很容易用树图法构造,因此要构造唯一可译码,只要构造即时码就可以了.

5.3.2 唯一可译码的判别准则

定理5.4只给出了唯一可译码的码长条件,不能作为一个具体的码是否是唯一可译码的判别方法,因为满足McMillan不等式的码不一定是唯一可译码,因此只能根据唯一可译码的定义来判断,在实际应用中因为不可能一一检查所有 N 次扩展码的奇异性,为此,A.A.Sardinas和G.W.Patterson于1957年提出下述算法用于判断码 C 的唯一可译性.此算法的原理如下所示:



其中 A_i, B_i 都是码字.可知,当且仅当某个有限长的码符号序列能译成两种不同的码字序列时,此码不是唯一可译码,此时 B_1 一定是 A_1 的前缀,而 A_1 的尾随后缀一定是另一

码字 B_2 的前缀;而 B_2 的尾随后缀又是其他码字的前缀.最后,码符号序列的尾部一定是一个码字.

设 C 为码字集合,按以下步骤构造此码的尾随后缀集合 F:

(1) 考查 C 中所有的码字,若 W_i 是 W_j 的前缀,则将相应的后缀作为一个尾随后缀码放入集合 F_1 中;

(2) 考查 C 和 F_i 两个集合,若 $W_i \in C$ 是 $W_i \in F_i$ 的前缀或 $W_i \in F_i$ 是 $W_j \in C$ 的前缀,则将相应的后缀作为尾随后缀码放入集合 F_{i+1} 中;

(3) $F = \bigcup_i F_i$ 即为码 C 的尾随后缀集合;

(4) 若 F 中出现了 C 中的元素,则算法终止,返回假(C 不是唯一可译码);否则若 F 中没有出现新的元素,则返回真.

定理 5.5　一个码是唯一可译码的充要条件是 $F_1, F_2, \cdots$ 的并集中没有 C 中的码字.

【例 5.4】

设消息集合共有 7 个元素{$s_1, s_2, s_3, s_4, s_5, s_6, s_7$},它们分别被编码为{$a, c, ad, abb, bad, deb, bbcde$},判断是否是唯一可译码.

解

按照上述方法构造表 5.8 所列的尾随后缀集 F_5 中的第一个元素正好是 C 的第三个元素,因此 C 不是唯一可译码.

该命题没有证明,直观上理解是当码字组成码字序列时,所有的后缀不能是一个码字,否则不是唯一可译码.

表 5.8　唯一可译码判决准则

C	F_1	F_2	F_3	F_4	F_5
a	d	eb	de	b	ad
c	bb	cde			$bcde$
ad					
abb					
bad					
deb					
$bbcde$					

5.3.3　无失真变长编码定理

由 5.3.2 节讨论可知,对于已知信源 S 可用码符号集合 X 进行变长编码,而且对同一信源用同一码符号集编成的即时码或唯一可译码可有很多种,究竟哪一种最好呢?从高效传输信息的角度希望选择由短的码符号组成的码字,就是用码长作为选择标准,为此引入码的平均长度.

定义 5.8　设有信源

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & \cdots & s_q \\ p(s_1) & p(s_2) & \cdots & p(s_q) \end{bmatrix}$$

编码后的码字分别为 $w_1, w_2, \cdots, w_q$,各码字相应的码长分别为 $l_1, l_2, \cdots, l_q$.因为是唯一可译码,信源符号 s_i 和码字 w_i 一一对应,则定义此码的平均码长为

$$\overline{L} = \sum_{i=1}^{q} p(s_i)\, l_i \text{码符号/ 信源符号} \tag{5.20}$$

其中，$\overline{L}$ 表示每个信源符号平均需用的码元数．

当信源给定时，信源熵 $H(S)$ 就确定了，而编码后每个信源符号平均用 $\overline{L}$ 个码元来变换，故平均每个码元载荷的信息量即编码后信源的信息传输率．

定义 5.9　编码后信源的信息传输率为

$$R = H(X) = \frac{H(S)}{\overline{L}} \text{比特/ 码符号} \tag{5.21}$$

如果传输一个码符号平均需要 t 秒时间，则编码后信源每秒钟提供的信息量为

$$R_t = \frac{H(S)}{\overline{L}t} \tag{5.22}$$

可以看出，$\overline{L}$ 越小，则 R_t 越大，信息传输率就越高，因此我们感兴趣的码是使平均码长 $\overline{L}$ 最短的码．

定义 5.10　对于给定的信源和码符号集，若有一个唯一可译码，其平均码长 $\overline{L}$ 小于所有其他唯一可译码，则称这种码为紧致码或最佳码．

无失真信源编码的核心问题就是寻找紧致码．

下面的定理给出了紧致码的平均码长可能达到的理论极限．

定理 5.6　若一个离散无记忆信源 S，熵为 $H(S)$，用拥有 r 个码符号的码符号集 $X = \{x_1, x_2, \cdots, x_r\}$ 对 S 进行无失真编码，则总可以找到一种唯一可译码，其平均码长满足

$$\frac{H(S)}{\log_2 r} \leq \overline{L} < 1 + \frac{H(S)}{\log_2 r} \tag{5.23}$$

证明

先证下界成立，即证明 $H(S) - \overline{L}\log_2 r \leq 0$

$$\begin{aligned} H(S) - \overline{L}\log_2 r &= -\sum_{i=1}^{q} p(s_i)\log_2 p(s_i) - (\log_2 r)\sum_{i=1}^{q} p(s_i)\, l_i \\ &= -\sum_{i=1}^{q} p(s_i)\log_2 p(s_i) + \sum_{i=1}^{q} p(s_i)\log_2 r^{-l_i} \\ &= \sum_{i=1}^{q} p(s_i)\log_2 \frac{r^{-l_i}}{p(s_i)} \\ &\leq \log_2\left[\sum_{i=1}^{q} p(s_i)\frac{r^{-l_i}}{p(s_i)}\right] \\ &= \log_2\left(\sum_{i=1}^{q} r^{-l_i}\right) \end{aligned} \tag{5.24}$$

根据 McMillan 不等式，有

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

所以

$$\overline{L} \geq \frac{H(S)}{\log_2 r} \tag{5.25}$$

等号成立的条件是 $\forall i, r^{-l_i} = p(s_i)$，这时 $H(S) - \overline{L}\log_2 r = \sum_{i=1}^{q} p(s_i)\log_2 1 = 0$，因此有

$$\overline{L} = \frac{H(S)}{\log_2 r} \tag{5.26}$$

也就是说只有当每个码字的相应码长 $l_i = \frac{-\log_2 p(s_i)}{\log_2 r} = -\log_r p(s_i)$ 时，$\overline{L}$ 才能达到下界 $\frac{H(S)}{\log_2 r}$.

证明上界成立，即证明满足条件 $\overline{L} < 1 + \frac{H(S)}{\log_2 r}$ 的唯一可译码是存在的.

选择码长 l_i 使满足

$$r^{-l_i} \leq p(s_i) < r^{-(l_i - 1)} \qquad i = 1, 2, \cdots, q \tag{5.27}$$

对上式左边的不等式求和得

$$\sum_{i=1}^{q} r^{-l_i} \leq \sum_{i=1}^{q} p(s_i) = 1 \tag{5.28}$$

即码长 $l_1, l_2, \cdots, l_q$ 满足 McMillan 不等式，所以可以构造一个唯一可译码.

对右边的不等式略加变化可得

$$\sum_{i=1}^{q} p(s_i)\log_2 p(s_i) < \sum_{i=1}^{q} p(s_i)\log_2 r^{-(l_i - 1)} \tag{5.29}$$

$$H(S) > (\log_2 r)\sum_{i=1}^{q} p(s_i)(l_i - 1) = \log_2 r(\overline{L} - 1) \tag{5.30}$$

所以

$$\overline{L} < \frac{H(S)}{\log_2 r} + 1 \tag{5.31}$$

综合上下界得到

$$\frac{H(S)}{\log_2 r} \leq \overline{L} < \frac{H(S)}{\log_2 r} + 1 \tag{5.32}$$

证毕

若熵以 r 进制为单位，则式(5.32)可写成

$$H_r(S) \leq \overline{L} < H_r(S) + 1 \tag{5.33}$$

定理 5.6 说明，码字的平均长度 $\overline{L}$ 不能小于极限值 $\frac{H(S)}{\log_2 r}$，否则唯一可译码不存在，

同时定理又给出了平均码长的上界，这不是说大于这个上界就不能构成唯一可译码，而是说即使平均码长小于这个上界也一定存在唯一可译码．

另外还可以看到平均码长的极限值与无失真定长信源编码定理中的极限值是一致的．

5.3.4 香农第一编码定理

与无失真定长信源编码定理一样，无失真变长信源编码定理（香农第一定理）也是一个极限定理，是一个存在性定理．极限和定长编码的极限是一样的．

定理 5.7 设离散无记忆信源 S 的信源熵 $H(S)$，它的 N 次扩展信源 $S^N=\{s_1, s_2, \cdots, s_{q^N}\}$，其熵 $H(S^N)$．并用码符号 $X=\{x_1, x_2, \cdots, x_r\}$ 对信源 S^N 进行编码，总可以找到一种唯一可译码，使信源 S 中每个信源符号所需的平均码长满足

$$\frac{H(S)}{\log_2 r}\leq\frac{\overline{L}_N}{N}<\frac{H(S)}{\log_2 r}+\frac{1}{N} \tag{5.34}$$

或者写成

$$H_r(S)\leq\frac{\overline{L}_N}{N}<H_r(S)+\frac{1}{N} \tag{5.35}$$

当 $N\to\infty$ 时，$\frac{\overline{L}_N}{N}=H_r(S)$，式中，$\overline{L}_N=\sum_{i=1}^{q^N}p(s_i)\lambda_i$，其中 λ_i 是扩展信源的信源符号 s_i 所对应的码字长度，因此 $\overline{L}_N$ 是扩展信源中每个符号的平均码长，而 $\frac{\overline{L}_N}{N}$ 仍是信源 S 中单个信源符号所需的平均码长．这里要注意 $\frac{\overline{L}_N}{N}$ 与 $\overline{L}$ 的区别：它们两个都是单个信源符号所需的码符号的平均数，但是 $\frac{\overline{L}_N}{N}$ 的含义是，为了得到这个平均值，不是对单个信源符号 s_i 进行编码，而是对 N 个信源符号的序列 s_i 进行编码，然后对 N 求平均．

定理 5.6 可以看作定理 5.7 在 $N=1$ 时的特殊情况．

证明

将 S^N 视为一个新的离散无记忆信源，它的熵为 $H_r(S^N)$，平均码长 $\overline{L}_N$．

根据定理 5.6 可得

$$H_r(S^N)\leq\overline{L}_N\leq H_r(S^N)+1 \tag{5.36}$$

由于离散无记忆信源的 N 次扩展信源 S^N 的熵 $H_r(S^N)$ 是信源 S 的熵 $H(S)$ 的 N 倍，即

$$H_r(S^N)=NH_r(S) \tag{5.37}$$

代入式(5.36)得

$$NH_r(S)\leq\overline{L}_N\leq NH_r(S)+1 \tag{5.38}$$

两边除以 N 得

$$H_r(S) \leq \frac{\overline{L_N}}{N} \leq H_r(S) + \frac{1}{N} \tag{5.39}$$

证毕

当 $N \to \infty$ 时，有

$$\lim_{N\to\infty} \frac{\overline{L_N}}{N} = H_r(S) \tag{5.40}$$

定理 5.7 的结论推广到平稳遍历的有记忆信源(如马尔可夫信源)便有

$$\lim_{N\to\infty} \frac{\overline{L_N}}{N} = \frac{H_\infty}{\log_2 r} \tag{5.41}$$

式中，H_∞ 为有记忆信源的极限熵 .

定理 5.7 是香农信息论的主要定理之一 .定理指出，要做到无失真信源编码，每个信源符号平均所需最少的 r 元码元数就是信源的熵值(以 r 进制单位为信息量单位) .若编码的平均码长小于信源的熵值，则唯一可译码不存在，在译码或反变换时必然要带来失真或差错，同时定理还指出，通过对扩展信源进行变长编码，当 $N \to \infty$ 时，平均码长 $\overline{L}$(这时它等于 $\frac{\overline{L_N}}{N}$)可达到这个极限值 .可见，信源的信息熵 $H(S)$ 是无失真信源编码码长的极限值，也可以认为信源的信息熵($H(S)$ 或 H_∞)是表示每个信源符号平均所需最少的二元码符号数 .

类似于定长码的编码速率，定义变长码的编码速率 $R = \frac{\overline{L_N}}{N}\log_2 r$，表示平均每个信源符号编码后能载荷的最大信息量，也就是编码信息率 .

香农第一定理也可以写成

$$H(S) \leq \frac{\overline{L_N}}{N}\log_2 r \leq H(S) + \frac{1}{N}\log_2 r = H(S) + \varepsilon \tag{5.42}$$

因此也可以表述为：若 $R \geq H(S)$，则存在唯一可译变长码；若 $R < H(S)$，则唯一可译码不存在，不能实现无失真的信源编码 .可以看到，定长码和变长码的编码信息率的理论极限是一致的，而且要达到这个极限上，也就是平均单个信源符号所需的码符号数最少，所用的方法都是对信源的 N 次扩展信源进行编码，但是变长码与定长码的区别在于变长码在 N 不需很大时就能达到这个极限，而定长码的 N 值通常大到设备难以实现，而且定长码在达到这个码长极限时往往还会引入一定的失真，但是变长码则不会引入失真 .

定义 5.11　编码后信道的信息传输率为

$$R = \frac{H(S)}{\overline{L}} \frac{\text{比特/ 信源符号}}{\text{码符号/ 信源符号}} = \frac{H(S)}{\overline{L}} \text{比特/码符号} \tag{5.43}$$

这里，$\overline{L} = \frac{\overline{L_N}}{N} \geq \frac{H(S)}{\log_2 r}$，所以 $R \leq \log_2 r$.

当平均码长 $\overline{L}$ 达到极限值 $\frac{H(S)}{\log_2 r}$ 时，编码后信道的信息传输率为 $R=\log_2 r$，这时信道的信息传输率等于无噪无损信道的信道容量 C，信息传输效率最高.因此，无失真信源编码的实质就是对离散信源进行适当的变换，使变换后新的码符号信源(信道的输入)尽可能为等概率分布，使新信源的每个码符号平均所含的信息量达到最大，从而使信道的信息传输率 R 达到信道容量，实现信源与信道理想的统计匹配.这就是香农第一定理的物理意义.

无失真信源编码定理通常又称为无噪信道编码定理，此定理可以表述为：总能对信源的输出进行适当的编码，使得在无噪无损信道上能无差错地以最大信息传输率 C 传输信息，但要使信道的信息传输率 R 大于 C 而无差错地传输则是不可能的.

为了衡量各种编码是否已达到极限情况，我们定义变长码的编码效率.

定义 5.12 设对信源 S 进行无失真编码所得到的平均码长为 $\overline{L}$，则 $\overline{L}\geq H_r(S)$，定义

$$\eta=\frac{H_r(S)}{\overline{L}} \tag{5.44}$$

为编码效率，$\eta\leq 1$.

对同一信源来说，码的平均码长 $\overline{L}$ 越短，越接近极限 $H_r(S)$，信道的信息传输率越高，越接近无噪无损信道的信道容量，这时 η 也越接近于 1，所以用码的编码效率 η 来衡量各种编码的优劣.

另外，为了衡量各种编码与最佳码的差距，引入码的剩余度的概念.

定义 5.13 定义

$$\gamma=1-\eta=1-\frac{H_r(S)}{\overline{L}} \tag{5.45}$$

为码的剩余度.

在二元无噪无损信道中 $r=2$，$\eta=\frac{H(S)}{\overline{L}}$，所以在二元无噪无损信道中信息传输率 $R=\frac{H(S)}{\overline{L}}=\eta$.

注意它们数值相同，单位不同.η 是个无单位的比值，而 R 的单位是比特/码符号.因此在二元信道中可直接用码的效率来衡量编码后信道的信息传输率是否提高了.当 $\eta=1$ 时，即 $R=1$，达到二元无噪无损信道的信道容量，编码效率最高，码剩余度为零.

与定长码一样，通过对扩展信源进行编码，可以提高编码后信道的信息传输率.

【例 5.5】

有一离散无记忆信源

$$\begin{bmatrix} S \\ P \end{bmatrix}=\begin{bmatrix} s_1 & s_2 \\ 3/4 & 1/4 \end{bmatrix}$$

求其信息传输率及二次、三次、四次扩展信源的信息传输率 .

解

信源熵

$$H(S)=\frac{1}{4}\log_2 4+\frac{3}{4}\log_2\frac{4}{3}=0.811$$

用二元码符号{0,1}来构造一个即时码: $s_1\to 0$, $s_2\to 1$.

这时, $\overline{L}=1$, 编码效率 $\eta=\frac{H(S)}{\overline{L}}=0.811$, 信源的信息传输率 $R=0.811$.

表 5.9　变长编码的编码效率

s_j	$p(s_j)$	即时码
$s_1 s_1$	9/16	0
$s_1 s_2$	3/16	10
$s_2 s_1$	3/16	110
$s_2 s_2$	1/16	111

如果对信源 S 的二次扩展信源 S^2 进行编码 . 得到它的一种即时码如表 5.9 所示 .

这时码的平均长度

$$\overline{L_2}=\frac{9}{16}\times 1+\frac{3}{16}\times 2+\frac{3}{16}\times 3+\frac{1}{16}\times 3=\frac{27}{16}\text{二元码符号/ 二个信源符号}$$

信源符号的平均码长 $\overline{L}=\frac{\overline{L_2}}{2}=\frac{27}{32}$二元码符号/ 信源符号

编码效率 $\eta_2=\frac{0.811\times 32}{27}=0.961$, $R_2=0.961$ 比特/ 二元码符号

可见编码复杂了些, 但信息传输效率有了提高 .

用同样方法进一步对信源 S 的三次和四次扩展信源进行编码, 并求出其编码效率为 $\eta_3=0.985$, $\eta_4=0.991$.信道的信息传输率分别为 $R_3=0.985$, $R_4=0.991$.

将此例与例 5.3 相比较, 对于同一信源, 要求编码效率达到 96% 时, 变长码只需对二次扩展信源($N=2$)进行编码, 而等长码则要求 $N\geq 4.13\times 10^7$.因此用变长码编码时, N 不需很大就可以达到相当高的编码效率, 而且可实现绝对无失真编码, 随着扩展信源次数 N 的增加, 编码的效率越来越接近于 1 .编码后信道的信息传输率 R 也越来越接近于无噪无损二元信道的信道容量 $C=1$ 比特/ 二元码符号, 从而达到信源与信道匹配, 使信道得到充分利用 .

5.4　变长码的编码方法

本章介绍的变长码的常见编码方法, 如香农编码、香农-费诺-埃利斯编码、霍夫曼编码、费诺编码均为匹配编码, 也称统计编码, 都是通过使用较短的码字来给出现概率较高的信源符号编码 .而出现概率较小的信源符号用较长的码字来编码, 从而使平均码长最短, 达到最佳编码的目的 .下面介绍这几种方法:

5.4.1 香农编码

香农第一定理指出了平均码长与信源熵之间的关系,同时也指出了可以通过编码使码长达到极限值,如何构造这种码?香农码的方法是选择每个码字长度 l_i 满足

$$l_i = \left\lceil \log_2 \frac{1}{p(s_i)} \right\rceil \qquad i = 1, 2, \cdots, q \tag{5.46}$$

$\lceil x$ 表示不小于 x 的整数,即 x 为整数时等于 x,x 不是整数时,等于 x 取整加 1.

由定理 5.7 可知,这样选择的码长一定满足 Kraft 不等式,所以一定存在即时码,然后按照这个码长 l_i 用树图法就可编出一组即时码.

按照香农编码方法构造的码,其平均码长 $\overline{L}$ 不超过上界,即 $\overline{L} \leq H_r(S) + 1$.

只有当信源符号的概率分布满足 $\left(\frac{1}{r}\right)^{\alpha_i}$($\alpha_i$ 是正整数)形式时,$\overline{L}$ 才能达到极限值 $H_r(S)$.一般情况下,香农码的 $\overline{L}$ 不是最短,即编出来的不是紧致码(最佳码),其具体方法如下:

(1) 将信源发出的 q 个消息符号按其概率的递减次序排列

$$p(s_1) \geq p(s_2) \geq \cdots \geq p(s_q)$$

(2) 计算出各个信源符号的累加概率

$$F(s_i) = \sum_{k=1}^{i-1} p(s_k) \tag{5.47}$$

(3) 按下式计算第 i 个消息的二元代码组的码长 l_i

$$l_i = \left\lceil \log_2 \frac{1}{p(s_i)} \right\rceil$$

(4) 将累加概率 $F(s_i)$(十进制小数)变换成二进制小数,根据码长 l_i 取小数点后 l_i 个二进制符号作为第 i 个消息的码字.

【例 5.6】

参见表 5.10,例如当 $i = 4$ 时,先求第 4 个信源符号的二元码的码长 $l_4 = \lceil -\log_2 p(s_4) \rceil = 3$,码长取为 3.

表 5.10 香农编码

信源符号 s_i	概率 $p(s_i)$	累加概率 $F(s_i)$	$-\log_2 p(s_i)$	码长 l_i	二元码
s_1	0.20	0	2.34	3	000
s_2	0.19	0.2	2.41	3	001
s_3	0.18	0.39	2.48	3	011
s_4	0.17	0.57	2.56	3	100
s_5	0.15	0.74	2.74	3	101
s_6	0.10	0.59	3.34	4	1110
s_7	0.01	0.99	6.66	7	1111110

再计算累加概率 $F(s_4)=\sum_{k=1}^{3}p(s_k)=p(s_1)+p(s_2)+p(s_3)=0.57$

将累加概率 $F(s_4)$ 变成二进制小数 $F(s_4)=(0.57)_{10}=(0.1001\cdots)_2$

根据码长 $l_4=3$ 取小数点后三位作为第 4 个信源符号的二元码,即“100”,其他信源符号的编码可依次求得.

由表 5.10 可以看出,一共有 5 个三位的二元码,各码字至少有一位码符号不同,这个码是唯一可译码,而且是即时码.

平均码长 $\overline{L}=\sum_{i=1}^{7}p(s_i)\,l_i=3.14$ 码符号/信源符号

编码后信道信息传输率 $R=\dfrac{H(S)}{\overline{L}}=\dfrac{2.61}{3.14}=0.831$ 比特/码符号

5.4.2　香农-费诺-埃利斯编码

将香农编码中的累加概率换成修正累加概率即可得到相应的香农-费诺-埃利斯编码:

(1) 计算出各个信源符号的修正累加概率

$$\overline{F}(s_i)=\sum_{k=1}^{i-1}p(s_k)+\frac{1}{2}p(s_i) \tag{5.48}$$

(2) 按下式计算第 i 个消息的二元代码组的码长 l_i

$$l_i=\left\lceil \log_2\frac{1}{p(s_i)}\right\rceil+1$$

(3) 将累加概率 $\overline{F}(s_i)$(十进制小数)变换成二进制小数.根据码长 l_i 取小数点后 l_i 个二进制符号作为第 i 个消息的码字.

香农-费诺-埃利斯编码与香农编码不同,它不需要对信源符号进行排序,直接计算修正累加概率即可.

5.4.3　霍夫曼编码

这是霍夫曼于 1952 年提出的一种构造紧致码的方法.具体方法如下:

(1) 将 q 个信源符号按概率大小递减排列 $p(s_1)\geq p(s_2)\geq\cdots\geq p(s_q)$;

(2) 用“0,1”码符号分别代表概率最小的两个信源符号,并将这两个概率最小的信源符号合并成一个,从而得到只包含 $q-1$ 个符号的新信源,称为缩减信源 S_1;

(3) 把缩减信源 S_1 的符号仍按概率大小递减次序排列,再将其最后两个概率最小的信源符号分别用“0”和“1”码符号表示,并且合并成一个符号,这样又形成了 $q-2$ 个信源

符号的缩减信源 S_2;

(4) 依次继续下去,直至信源最后只剩下两个信源符号为止,将这最后两个信源符号分别用二元码符号“0”和“1”表示;

(5) 然后从最后一级缩减信源开始,进行回溯,就得到各信源符号所对应的码符号序列,即相应的码字 .

【例 5 .7】

以例 5 .6 信源为例编制霍夫曼码 .

解

编码过程如表 5 .11 所示 .

表 5 .11 霍夫曼编码

信源符号s_i	概率$p(s_i)$	编码过程					码字 w_i	码长 l_i
		S_1	S_2	S_3	S_4	S_5		
						0.61		
					0.39	0.39		
				0.35	0.35			
			0.26	0.26	0.26			
s_1	0.20	0.20	0.20	0.20			10	2
s_2	0.19	0.19	0.19	0.19			11	2
s_3	0.18	0.18	0.18				010	3
s_4	0.17	0.17	0.17				011	3
s_5	0.15	0.15					000	3
s_6	0.10	0.11					0010	4
s_7	0.01						0011	4

该码的平均码长为

$$\overline{L} = \sum_{i=1}^{7} p(s_i)\, l_i$$

$$= 0.20 \times 2 + 0.19 \times 2 + 0.18 \times 3 + 0.17 \times 3 + 0.15 \times 3 + 0.10 \times 4 + 0.01 \times 4$$

$$= 2.72 \text{ 比特/码符号}$$

其编码效率为

$$\eta = \frac{H_r(S)}{\overline{L}} = \frac{2.61}{2.72} = 0.960$$

霍夫曼编码方法得到的码并非是唯一的 造成非唯一的原因有两个:

(1) 每次对信源缩减时,赋予最后两个概率最小的信源符号的码符号“0”和“1”是可以互换的,所以可得到不同的霍夫曼码;

(2) 对信源进行缩减时,如果两个概率最小的信源符号合并后的概率与其他信源符号的概率相同,则在进行概率排序时的次序可以是任意的,因此会得到不同的霍夫曼码 .

表 5 .12 给出了一个信源的两种霍夫曼码这两种码的平均码长、编码效率都相同 .

表 5 .12　霍夫曼编码之间的比较

信源符号 s_i	概率 $p(s_i)$	码 1	码 1 的码长	码 2	码 2 的码长
s_1	0 .4	1	1	00	2
s_2	0 .2	01	2	10	2
s_3	0 .2	000	3	11	2
s_4	0 .1	0010	4	010	3
s_5	0 .1	0011	4	011	3

$$\overline{L} = \sum_{i=1}^{7} p(s_i)\, l_i = 2 .2 \text{ 码符号/ 信源符号}$$

$$\eta = \frac{H_r(S)}{\overline{L}} = 0 .965$$

但两种码的质量不完全相同,因为它们码长的方差不同 .

表 5 .12 中码 1 的码方差:

$$\sigma_1^2 = \sum_{i=1}^{5} p(s_i)(l_i - \overline{L})^2 = 1 .36$$

表 5 .12 中码 2 的码方差:

$$\sigma_2^2 = \sum_{i=1}^{5} p(s_i)(l_i - \overline{L})^2 = 0 .16$$

由此可见,第二种霍夫曼码的码方差要比第一种霍夫曼码的码方差小很多,因此第二种霍夫曼码的质量要好 .

从此例可以看出,进行霍夫曼编码,信源缩减排列时,应使合并的信源符号位于缩减信源中尽可能高的位置上,这样可以使合并的信源符号码长减少充分利用短码,而非合并的信源符号码长变长所以得到方差最小的码 .霍夫曼码的编码过程也可以用树图来表示,如图 5 .4 所示 .

可以看出霍夫曼码是即时码 .

霍夫曼码是用概率匹配的方法进行信源编码 .它有两个明显特点:

(1) 霍夫曼码的编码方法保证了概率大的信源符号对应的码长小,概率小的信源符号对应的码长大,充分利用了短码;

(2) 每次缩减信源的最长两个码字有相同的码长,并且仅仅只有最后一位码符号不同 .

这两个特点保证了所得的霍夫曼码一定是最佳码 .下面来证明这一点 .

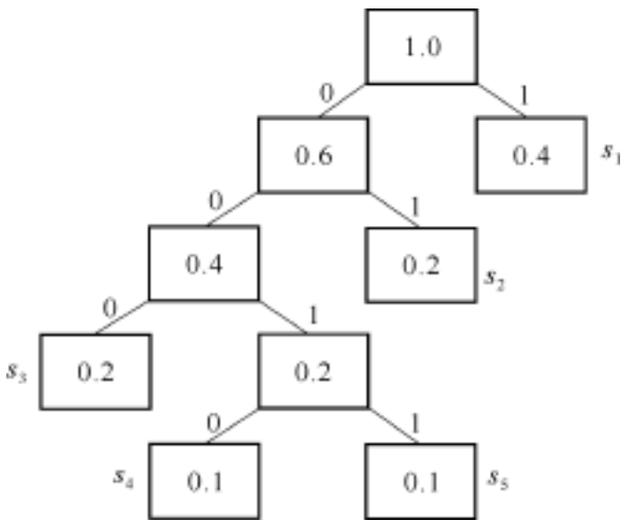


图 5.4 霍夫曼编码的树图

定理 5.8 霍夫曼码是紧致码.

证明

在这里只证明二元霍夫曼码是紧致码，其结论可以推广到 r 元霍夫曼编码.

由于霍夫曼码最后一步得到的缩减信源只有两个信源符号，编码为“0”和“1”.它们是紧致码，所以可以假设缩减后的编码是紧致码，由此假设证明缩减前的编码是紧致码，这样最终可以证明霍夫曼码是紧致码.

设霍夫曼码中第 j 步缩减信源为 S_j，缩减信源被编码为 C_j，其平均码长为 $\overline{L}_j$，S_j 有 m 个信源符号，则 $\overline{L}_j = \sum_{i=1}^{m} p(s_i) l_i$，$S_i$ 中的某一元素 s_k 是由前一次缩减信源 S_{j-1} 中的两个概率最小的信源符号 s_{k_0} 和 s_{k_1} 合并而来，即 $p(s_k) = p(s_{k_0}) + p(s_{k_1})$.

设 C_{j-1} 为第 $j-1$ 步缩减信源 S_{j-1} 的编码，其平均码长为 $\overline{L}_{j-1}$，S_{j-1} 有 $m+1$ 个元素

$$\begin{aligned}\overline{L}_{j-1} &= \sum_{i=1}^{m-1} p(s_i) l_i + p(s_{m_0})(l_m + 1) + p(s_{m_1})(l_m + 1) \\ &= \sum_{i=1}^{m} p(s_i) l_i + p(s_{m_0}) + p(s_{m_1}) \\ &= \overline{L}_j + p(s_{m_0}) + p(s_{m_1})\end{aligned}$$

缩减信源 S_j 和 S_{j-1} 的平均码长之差是一个与码长 l_i 无关的固定常数，所以如果平均码长 $\overline{L}_j$ 最小则 $\overline{L}_{j-1}$ 最小.也就是如果 C_j 是缩减后信源 S_j 的紧致码，则 C_{j-1} 是缩减前信源 S_{j-1} 的紧致码.

由于最后一级缩减信源是最佳码，则由霍夫曼编码方法使得它前面一级缩减信源的编码也一定是最佳码.由递推可知信源 S 所得的编码是最佳码.

证毕

5.4.4　r 元霍夫曼编码

二进制霍夫曼码的编码方法很容易推广到 r 进制的情况，只是编码过程中构成缩减信源时，每次都是将 r 个概率最小的信源符号合并，并分别用码符号 $0,1,\cdots,(r-1)$ 表示．

为了充分利用短码，使霍夫曼码的平均码长最短，必须使最后一个缩减信源恰好有 r 个信源符号．因此对于 r 元霍夫曼编码，信源 S 符号个数 q 必须满足 $q=(r-1)\theta+r$．θ 表示信源缩减次数．如果不满足上式，则可以在最后增补一些概率为 0 的信源符号，因此上式又可以写成 $q+i=(r-1)\theta+r$．i 为增加的信源符号数，并且是满足上式的最小正整数或 0．当 $r=2$ 时的二元码．信源 S 的符号个数 q 满足 $q=\theta+2$．

【例 5.8】

构造一个三元霍夫曼码

解

编码结果如表 5.13 所示．

表 5.13　三元霍夫曼编码

信源符号	概率 $p(s_i)$	码字	码长
s_1	0.4	1	1
s_2	0.2	2	1
s_3	0.2	00	2
s_4	0.1	01	2
s_5	0.1	02	2

这里 $q+i=2\theta+3$，i 的最小值为 0，所以不需增加信源符号．该码的平均码长

$$\overline{L}=\sum_{i=1}^{5}p(s_i)\,l_i=1.4 \text{ 三进制码符号/信源符号}$$

$$R=\frac{H_2(S)}{\overline{L}}=\frac{2.122}{1.4}=1.515 \text{ 比特/三进制码符号}$$

$$\eta=\frac{H_3(S)}{\overline{L}}=\frac{1.339}{1.4}=0.9564$$

当需要增加信源符号时，原来的码树一定是非整树，因为要充分利用短码，从码树的角度看，就是充分利用一阶节点生成码字，一阶节点不够用时，再考虑从某个节点伸出若干树枝，引出二阶节点出成码字．如此类推，这样生成的码平均码长 $\overline{L}$ 最短．

信源的 N 次扩展信源同样可以使用霍夫曼编码方法．

由于霍夫曼码是紧致码，所以编码后单个信源符号平均码长随 N 的增加很快接近于极限值——信源熵．

5.4.5　费诺编码

费诺码与香农码一样，也不是最佳的编码方法，但是某些情况也能得到紧致码的性能．

费诺码编码过程如下：

(1) 将信源符号 s_i，$i=1,2,\cdots,q$ 以概率递减次序排列，即

$$p(s_1)\geq p(s_2)\geq\cdots\geq p(s_q)$$

(2) 将依次排列的信源符号以概率分为两组,使两个组的概率和基本相等,并对各组赋予二元码符号“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和近于相等,又分别赋予各组二元码符号“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符号序列即为费诺码.

【例 5.9】

信源与例 5.6 和例 5.7 中相同.请编制费诺码.

解

费诺码编码过程如表 5.14 所示.

表 5.14 费诺编码

信源符号	概率	第 1 次分组	第 2 次分组	第 3 次分组	第 4 次分组	码字	码长
s_1	0.2	0	0			00	2
s_2	0.19		1	0		010	3
s_3	0.18			1		011	3
s_4	0.17	1	0			10	2
s_5	0.15		1	0		110	3
s_6	0.10			1	0	1110	4
s_7	0.01				1	1111	4

该码的平均码长为

$$\overline{L} = \sum_{i=1}^{7} p(s_i)\, l_i$$

$$= 0.20 \times 2 + 0.19 \times 3 + 0.18 \times 3 + 0.17 \times 2 + 0.15 \times 3 + 0.10 \times 4 + 0.01 \times 4$$

$$= 2.74 \text{ 码符号/信源符号}$$

信息传输率为

$$R = \frac{H(S)}{\overline{L}} = \frac{2.61}{2.74} = 0.953 \text{ 比特码符号}$$

几种编码方法的性能比较如表 5.15 所示.

表 5.15 三种编码方法的比较

编码	平均码长 $\overline{L}$	信息传输率 R
香农码	3.14	0.831
霍夫曼	2.72	0.959
费诺码	2.74	0.953

如果信源概率满足 $p(s_i) = \left(\frac{1}{r}\right)^{n_i}$, $i = 1, 2, \cdots, q$, n_i 为正整数,则 3 种码都能得到紧致码.

上面讲述的几种编码都是针对离散无记忆信源的单个信源符号的编码,因此在编码时没有考虑其信源符号之间的相关性.仅仅考虑了信源符号分布的不均

匀性 .对于有记忆信源, 用单个符号编制变长码时不可能使编码效率接近于 1,因为编码信息率只能接近一维熵 $H_1(S)$, 而 $H_\infty(S)$一定小于 $H_1(S)$.所以需要多个符号一起编码,并且需要知道多个符号的联合概率分布,才能进一步提高编码效率 .有一些编码方法如预测编码、变换编码考虑了信源符号间的相关性 .

以上讨论了霍夫曼码和其他一些编码方法,并且证明了霍夫曼码是最佳码,当 N 不很大时,它能使无失真编码的效率接近于 1,但是在实际使用时设备较复杂 .

首先,每个信源符号所对应的码长不同,一般情况下,信源符号以匀速输出,信道也是匀速传输的 .通过霍夫曼编码后,会造成编码输出的信息速率不是常量,因而不能直接由信道来传送 .为了适应信道,必须增加缓冲寄存器,将编码输出暂存在缓冲器中,然后再匀速向信道输出 .但当缓冲器容量有限时,会出现缓冲器溢出或取空的现象 .例如当信源连续输出低概率的信源符号时,因为码长较长,有可能使缓冲器存不下而溢出;反之,当信源连续输出高概率符号,则有可能使缓冲区输入比特数小于向信道输出的比特数,以致缓冲器取空 .所以一般变长码只适用于有限长的信息传输,或者在输出一批消息后能暂停一下 .

其次,差错扩散的问题 .变长码的一个码元的差错可能造成译码错误,并且还会造成同步错误,结果后面一系列码字也会译错 .

最后,霍夫曼码的编译码需要用查表的方法来进行 .在信息传输过程中必须先存储与传输这一霍夫曼编码表 .这会影响信息的传输效率,特别是当 N 增大时,所需存储的码表也增大,使得设备复杂化,且查表搜索的开销增大 .我们还须根据信源的统计特性,事先建立霍夫曼编码表,因此这种方法只适用于已知其统计特性的信源 .

尽管如此,霍夫曼方法还是一种有效的无失真信源编码方法,它便于硬件实现和计算机上的软件实现,适用于文件传真、语声处理和图像处理的数据压缩 .

5 .5　实用的无失真信源编码方法

5 .5 .1　游程编码

游程编码主要用于黑白二值文件、传真的数据压缩 .由于传真文件中连“ 0 ”和连“ 1 ”较多 .这些连“ 0 ”或连“ 1 ”的字符串称为游程 .对游程长度进行霍夫曼编码或其他的编码处理就可以达到压缩数据的目的 .

图 5 .5 是一幅 10×50 黑白二值图像 .

如果将黑像素编码为“ 1 ”,白像素编码为“ 0 ”,则一幅黑白二值图像所需的比特数等于其像素的个数 .在图 5 .5 中这个值为 500 .

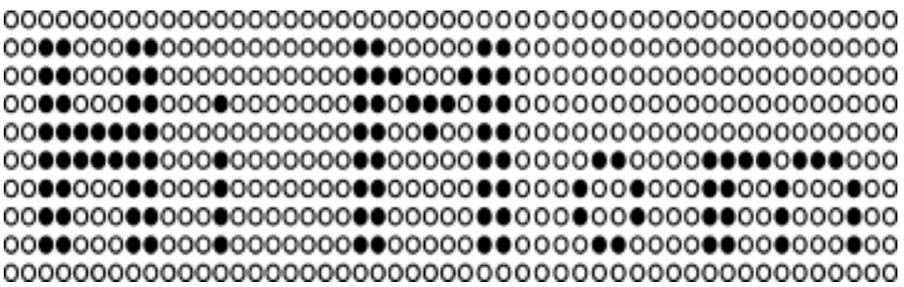


图 5.5 二值图像

通过观察,发现白像素的数目比黑像素多,并且无论是黑像素还是白像素都经常连续地出现,形成一个一个的游程.

可以通过将整个游程用一个字节进行编码来提高效率.即用每个字节的第一个比特表示游程的颜色,剩下的 7 个比特表示游程的长度(0 到 127).在这种方案下,一个 127 像素长的白游程只需 8 个比特来表示,压缩率为 127/8 = 15.9,然而并不是所有游程都是这样的长度.

对图 5.5 而言,总共有 55 个白游程以及 54 个黑游程,每个游程的长度都不超过 127,所以总共需要 8×(55 + 54) = 872 bit.不仅没有达到压缩的效果,所需的比特数反而增加了.

幸运的是,上述方案还有许多值得改进的地方.

首先,黑游程和白游程一定是交替出现的,所以只需给第一个游程的颜色进行编码,而不必为每一个后续游程的颜色都进行编码.需要注意的是当游程的长度超过 127 时,可以指定这样的规则:一个全为"1"的码字表明后续的码字对应的游程的颜色不发生改变.此时共需要 1 + 7×(55 + 54) = 764 bit,依然比每个像素采用一个比特编码的方案差.

其次,尽管上述方案对长度较大的游程有很高的压缩比,但是对长度很小的游程就显得不那么好了.解决的方法之一是减少用于表示游程长度的比特数.经过计算,当用 3 bit 表示游程长度(0～7)时,最后需要 459 bit,压缩率大约为 8%.

第三,在图 5.5 中白游程普遍要比黑游程长,因此可以考虑对黑、白游程的长度采用不同的比特数分别编码,并且对黑、白游程长度的编码可以采用霍夫曼码等性能较好的方案.

第四,码字序列 011 000 010 实际上表示的是 5 个连续的白像素,而 5 个连续的白像素将由码字 101 来表示,因此编码器永远不会产生上述序列,所以浪费了很多码字.

MH 编码是一种实用的游程编码算法,应用于黑、白传真数据的压缩编码,根据不同的黑、白游程长度有两张结尾码表和两张组合码表.其基本的编码规范为

(1) 游程长度在 0～63 时,直接查表用相应的结尾码作为码字;

(2) 游程长度在 64～1 728 范围内时,用组合码加上结尾码作为相应的码字;

(3) 每行的第一个游程规定为白游程(长度可以为零),每行用一个结束码(EOL)终止;

(4) 在传输时,每页数据之前加一个结束码,每页尾部连续使用 6 个结束码 .

【例 5 .10】

设有一页传真数据中某一行上的像素点如下分布:

|←73 白→|←7 黑→|←11 白→|←18 黑→|←1619 白→|

求:(1) 该行图像的 MH 编码;

(2) 编码后该行的总比特数;

(3) 该行的编码压缩率 .

解

(1) 根据 MH 码表可得该行图像的 MH 码为

64 白 + 9 白	7 黑	11 白	18 黑	1600 白 + 19 白	EOL
11011 10100	00011	01000	0000001000	010011010 0001100	000000000001

(2) 编码后该行的总比特数为 58 位 .

(3) 该行的压缩率为:1 728∶58≈29 .8 .

5 .5 .2　算术编码

对于符号表很小的信源来说,只有采用较长的分组长度才能得到较高的编码效率,因此需要一个有效的算法来处理较长的信源符号序列,霍夫曼编码在这方面做的不够好,因为它必须预先计算所有可能的信源符号序列的概率,然后才能进行编码 .

算术编码是一种非分组码,无需计算出所有 N 长信源序列的概率分布及码表,可以直接对输入的信源符号序列编码输出 .这种方法是由香农-费诺-埃利斯编码直接扩展得到的 .

香农-费诺-埃利斯的编码方法是将累积概率分布函数的[0, 1]分成许多互不重叠的小区间,每个信源符号对应于各个小区间,每一区间的长度等于这个信源符号的概率分布 .在此小区间内取一点, 取该点二进制小数点后 l 位作为这个信源符号的码字 .把这基本思想适用到信源符号序列中来,计算出信源符号序列的累积分布函数和对应的区间 .再在区间内取一点将其二进数小数点后 l 位作为这个符号序列的码字 .

算术编码算法的中心思想是高效地计算 n 长信源符号序列 x 的分布概率 $p(x)$ 和累积概率 $F(x)$,然后用区间$(F(x) - p(x), F(x)]$中的一个值来作为 x 的码字 .

假定信源是二进制的,并且分组的长度 n 已知 .

定义 5 .14　两个 n 长信源符号序列 x 和 y,定义 x 大于 y,当第一个使得 $x_i \neq y_i$ 的 i 有 $x_i = 1, y_i = 0$.

我们可将上述信源符号序列安排为一棵深度为 n 的树上的叶子节点 .并使得 $x > y$ 对应着 x 在 y 的右边 .

根据累积概率的定义,需要计算 x 节点左侧所有叶子节点的概率 .所有这些概率的

和等于 x 节点左侧所有子树的概率和 .

我们用 $T_{x_1 x_2 \cdots x_{k-1}0}$ 来表示节点 $x_1\ x_2 \cdots x_{k-1}0$ 对应的子树,其概率为

$$p(T_{x_1 x_2 \cdots x_{k-1}0}) = \sum_{y_{k+1} \cdots y_n} p(x_1\ x_2 \cdots x_{k-1}0\ y_{k+1} \cdots y_n)$$

$$= p(x_1\ x_2 \cdots x_{k-1}0) \tag{5.49}$$

因此其计算非常简单 .

我们将累积概率 $F(x)$ 改写为

$$F(x) = \sum_{y \leq x} p(y) = \sum_{T:\ T\text{位于}x\text{的左边}} p(T) = \sum_{k:\ x_k = 1} p(x_1\ x_2 \cdots x_{k-1}0) \tag{5.50}$$

因此可以通过 $p(x)$ 快速地计算出 $F(x)$.

上述过程的出发点是可以简单地计算出 $p(x)$ 对于独立同分布的信源,有: $p(x) = \prod_{i=1}^{n} p(x_i)$;

对于马尔可夫信源,有: $p(x) = p(x_1) \prod_{i=2}^{n} p(x_i \mid x_{i-1})$.

5.5.3 LZW 码

对于统计特性已知的平稳信源,霍夫曼码和算术码的编码效率已非常高,而且实现也不算太困难,在信源统计特性不知时需要用到具有自适应性能的通用编码方法 .LZW 码就是一种高效的通用编码,现在已成为计算机文件压缩的标准算法,我们常用的 ZIP、ARC 压缩解压程序就是 LZW 码的改进算法 .

LZW 码也称基于字典的编码方法,它是定长码 .

(1) 基于字典编码的基本原理

计算机文件是以字节为单位组成的 .每个字节的取值从 0 到 255 .每个字节都看成字符,共 256 种字符 .再把连续的若干个“字符”看成是一个“单词”,全部字符、单词及它们对应的序号(码字)组成字典 .编码时,把字符、单词用对应的码字来代替 .通常序号(码字)的长度为 12 bit(即字典的容量为 4 096),而“单词”的平均长度远大于 12 bit .从而达到压缩的目的 .例如在一个文件中,有一个片段的内容为“ABCD 空空空空空 20000”共 14 个字节,长度为 112 bit,编码被分割成 4 个单词,即“ABCD”、“空空空空空”、“2”和“0000”,编码长度为 48 bit,则压缩比为 2.33 .LZW 码是一种自适应编码,它的字典是直接由被压缩文件在编码过程中生成的 .

(2) 字典的构成

字典的容量为 4 096(0～4 095),序号用 12 bit 表示 .最后一个单词(第 4 095 个单词)为空 .单词看成由前缀字符串和尾字符串两部分组成 .其中前缀字符串是字典中已经存在的某个“单词”,用序号表示,尾字符是本单词的最后一个字符 .比如,单词 ABC 可以变换

成假定 AB 这个单词在字典中已经存在,且序号为 100,这样,任何单词的内容都可以用 3 个字节表示.其中两个字节表示前缀单词的序号.一个字节为尾字符.将压缩文件恢复为原始文件(解压缩)时,根据单词的这种格式,使用递归算法来恢复单词的内容.

(3) 算法

① 字典初始化:将被压缩文件中所有使用到的单字节字符放入字典中,为了压缩任何类型的文件,可以将字典的前 256 个位置(0x000 到 0x0FF)依次分配给 0x00 到 0xFF 的 256 个单字节字符.

② 动态数据初始化:初始化新单词存放位置指针 P,将它指向字典的第一个空位置.例如 $P=256$(即 0x100),读入被压缩文件的第一个字符 *cha*,作为待处理单词 W.单词的前缀 Q 为空,即 $Q=4\,095$,尾字符就是 *cha*,序号(码字)就是 *cha* 的序号.

③ 如果文件再没有字符了,输出当前单词 W 的序号.编码结束.如果文件中还有字符,把当前单词 W 作为前缀,再从被压缩文件中读入一个字符 *CH*,把 *CH* 作为尾字符,得到一个单词 W_1.

④ 如果字典中已有 W_1,则将 W_1 看作当前单词 W,返回③.如果字典中没有 W_1(发现一个新单词),先将原单词 W 的序号输出,再加新单词 W_1,增加到字典中,然后把刚刚读入的字符 *CH* 作为当前单词 W,返回③.

【例 5.11】

信源符号序列为 *ABCABDABCAAAABBBABCABCA*,其编解码过程见表 5.16 和 5.17.

表 5.16　LZW 压缩编码的压缩编码过程

步骤	读入字符	尾字符	查找对象	新单词	输出码字
0	*A*				
1	*B*	*A*	*AB*	100	041
2	*C*	*B*	*BC*	101	042
3	*A*	*C*	*CA*	102	043
4	*B*	*A*	*AB*		
5	*D*	*B*	*ABD*	103	100
6	*A*	*D*	*DA*	104	044
7	*B*	*A*	*AB*		
8	*C*	*B*	*ABC*	105	100
9	*A*	*C*	*CA*		
10	*A*	*A*	*CAA*	106	102
11	*A*	*A*	*AA*	107	041
12	*A*	*A*	*AA*		

续 表

步骤	读入字符	尾字符	查找对象	新单词	输出码字
13	*B*	*A*	*AAB*	108	107
14	*B*	*B*	*BB*	109	042
15	*B*	*B*	*BB*		
16	*A*	*B*	*BBA*	10 *A*	105
17	*B*	*A*	*AB*		
18	*C*	*B*	*ABC*		
19	*A*	*C*	*ABCA*	10 *B*	105
20	*B*	*A*	*AB*		
21	*C*	*B*	*ABC*		
22	*A*	*C*	*ABCA*		
23		*A*			10 *B*

表 5 .17 LZW 压缩编码的解压缩过程

步骤	读入码字	记忆码字	解码输出	字典位置	新单词
0	041	-	A	-	-
1	042	041	B	100	AB
2	043	042	C	101	BC
3	100	043	AB	102	CA
4	044	100	D	103	ABD
5	100	044	AB	104	DA
6	102	100	CA	105	ABC
7	041	102	A	106	CAA
8	107	041	AA	107	AA
9	042	107	B	108	AAB
10	149	042	BB	109	BB
11	105	149	ABC	10A	BBA
12	10B	105	ABCA	10B	ABCA
13	FFF	-	-	-	-

(4) 适用文件类型

不适合小文件的压缩(因为压缩编码初期,由于字典中的单词很少,字典对压缩效果的贡献也很少,主要是进行字典的扩充),也不适合太大的文件(因字典容量有限,文件太大时字典满了,效率将受到制约) .适合内容有明显单词结构的文件(如文本文件、程序文件) .

(5) 译码

① 字典初始化:将字典的前 256 个位置(0x000 到 0x0FF)依次分配给 0x00 到 0xFF 这个 256 个单字节字符 .

② 动态数据初始化:初始化新单词存放位置指针 P,将它指向字典的第一个空位置,例如 $P = 256$(即 0x100),读入压缩文件的第一个码字,由于第一个码字必定是一个单字符,可以从初始字典中查表得到,译码输出,并记忆它的码字 .

③ 如果压缩中已经没有码字,解码结束 .否则继续读入一个码字 .

④ 如果读入的码字是无效码字 FFF,则解码结束,否则进入下一步 .

⑤ 如果在字典中已经有该码字对应的单词,则采用递归算法,输出该单词的内容 .并将单词的第一个有效字符作为尾字符,将已经记忆的前一个码字作为前缀,组成一个新单词,写入字典中,然后将当前码字记忆下来,返回③;否则,首先在字典中生成新的单词,然后再输出这个单词,将新单词的码字记忆下来,返回③ .这时的新单词一定是首尾相同的单词 .

(6) LZW 编码算法的优化

为了进一步提高压缩效果和适应超大型件的压缩需要,LZW 压缩算法不断被改进 .字典的大小根据需要可以扩充,码字的长度也可以不断调整 .PKZIP、ART、ARC、LHA、WINZIP 等压缩软件都是在 LZW 压缩编码的基础上各自采用了不同的技术改进而成的 .

习　题　5

5 .1 设 S 为一离散无记忆信源,其符号集合为{0, 1},概率分布为 $p(0) = 0.995$, $p(1) = 0.005$.令信源符号序列的长度为 $n = 100$,假定对所有只包含 3 个以下符号“1”的序列编制长度为 k 的非奇异二进制码 .求:

(1) 信源的熵 $H(S)$及其冗余度;

(2) k 的最小值应该为多少?试比较 k/n 和 $H(S)$;

(3) 信源产生的序列没有码字与其对应的概率 .

5 .2 用 Fibonacci 数列的前 8 个非零元素构成相应的概率分布为 $P_{\mathrm{Fib}}^{8} = \left\{\frac{13}{34}, \frac{8}{34}, \frac{5}{34}, \frac{3}{34}, \frac{2}{34}, \frac{1}{34}, \frac{1}{34}, \frac{1}{34}\right\}$,对于这样一个概率分布存在着许多拥有不同码长分布的最优编码方案 .

(1) 为 P_{Fib}^{8}构造一个 Huffman 编码并求其平均码长;

(2) 求出最大码长最小的码的码长分布;

(3) 求出最大码长最大的码的码长分布;

(4) 请问一共有多少种码长分布不同的最优编码 .

5.3 设 X_1, X_2, X_3 为独立的二进制随机变量,并且有 $P_r\{X_1=1\}=\frac{1}{2}$, $P_r\{X_2=1\}=\frac{1}{3}$, $P_r\{X_3=1\}=\frac{1}{4}$,请给出联合随机变量(X_1, X_2, X_3)的 Huffman 编码并求其平均码长.

5.4 下面以码字集合的形式给出5种不同的编码,第一个码的码符号集合为{x, y, z},其他4个码都是二进制码.

{xx, xz, y, zz, xyz};

{000,10,00,11};

{100,101,0,11};

{01,100,011,00,111,1010,1011,1101};

{01,111,011,00,010,110}.

对于上面列出的五种编码,分别回答下述问题:

(1) 此码的码长分布是否满足 Kraft-McMillan 不等式?

(2) 此码是否是即时码?如果不是,请给出反例.

(3) 此码是否是唯一可译码?如果不是,请给出反例.

5.5 请问下述编码中哪些不可能是任何概率分布对应的 Huffman 编码?

(1) {0,10,11};

(2) {00,01,10,110};

(3) {01,10}.

5.6 设一信源有 2^k 种不同的符号,其中 k 为任意正整数.对此信源进行二进制 Huffman 编码.假设此信源的分布概率满足 $p_i/p_j<2$ " $i,j\in\{1,2,\cdots,2^k\}$.试证明此 Huffman 编码中所有的码长都为 k.

5.7 请找出一个唯一可译码,既不满足前缀条件也不满足后缀条件.

5.8 前缀码又称为即时码是因为在译码时产生的译码延时为零.变长码的译码延时定义为在译码时所需查看的下一个码字的码符号个数的最大值.例如,码 $C=\{0,01\}$的译码延时为1.请给出一个译码延时为3的变长码.

5.9 等概信源.

(1) 试证明对于一个有 n 个符号的等概信源,其最佳前缀码的各个码长之间最多相差1个码符号;

(2) 变长码的冗余度定义为 $L-H$.设一个随机变量 S 有 n 个等概的输出,其中 $2^m\leq n<2^{m+1}$,对此随机变量进行二进制的变长编码,得到的码的冗余度为 $L-\log_{2_2} n$,试问 n 取何值时冗余度最大?当 $n\to\infty$时码的冗余度的极限值是多少?(提示:0.0861)

5.10 令符号表 $D=\{1,2\}$的符号代价为1,2.

(1) 对于一个有8个等概输出的信源,找出一个在 D 上的最佳前缀码;

(2) 对于分布$\left\{\frac{1}{n},\cdots,\frac{1}{n}\right\}$,求相应最佳码的平均码字代价的近似值.

5.11 证明一个码 C 是唯一可译码当且仅当其 k 次扩展

$$C^k(x_1,x_2,\cdots,x_k)=C(x_1)C(x_2)\cdots C(x_k)$$

对任意的 $k\geq 1$ 都是一个从 χ^k 到的 D^* 一一映射.

5.12 香农的信源编码定理告诉我们,一个随机变量 X 的最佳码的平均码长满足 $L<H(X)+1$,这只是一个最佳的上界,而事实上在很多情况下编码的效率要差得多.例如 X 是一个二进制的随机变量,$p_X(1)=\varepsilon$,则当 $\varepsilon\to 0$ 时 $H(X)=H(\varepsilon)\to 0$,而此时 $L=1$.请找出一个随机变量,其熵为 2,同时其相应最佳码的平均码长为 3.

5.13 对于分布$\left\{\frac{1}{5\,050},\frac{2}{5\,050},\cdots,\frac{100}{5\,050}\right\}$,试求其熵和相应的二进制编码的最小平均码长.

5.14 考虑这样一个信源分布$\left\{\frac{1}{3},\frac{1}{3},\frac{1}{4},\frac{1}{12}\right\}$.

(1) 为此信源构造一个 Huffman 码;

(2) 为此信源找出两组不同的最佳码长方案;

(3) 用实例说明在最佳码中某些码字的码长将会大于相应信源符号对应的香农编码的码字长度 $l_i=\left\lceil\log_2\frac{1}{p_i}\right\rceil$.

5.15 枚举编码.

(1) 求序列 101010001001 在由长度为 12、权重为 5 的二进制序列构成的字典中的索引号;

(2) 假设一二进制信源,其输出中不可能出现两个相连的“1”,而序列 101010001001 为此信源的一个输出序列.求此序列对应的字典索引号.(提示:长度为 n 的允许序列有 F_{n+2} 个,其中 $F_0=0$, $F_1=1$,且 $F_n=F_{n-1}+F_{n-2}$, " $n\geq 2$)

5.16 一离散信源的符号表为$\{a,b,c,d,e\}$,而 $x=daadcadbea$ 为此信源的观察序列.假设此信源为具体分布未知的独立同分布随机过程.

(1) 通过观察序列,我们可以得到信源概率分布函数的一个估计,根据这个估计求信源的熵.

(2) 根据估计出来的信源概率分布函数构造一个 Huffman 码,计算平均码长并指出对序列 x 编码所需的比特数与平均码长的关系.

(3) 对序列 x 的前四个符号进行自适应的 Huffman 编码.初始的估计分布为$\left\{\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right\}$,此后每收到一个符号,就用当前的码本对其进行编码,然后根据此符号更新估计分布和码本.例如,当收到第一个符号 d 后,更新的概率分布为$\left\{\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{2}{6},\frac{1}{6}\right\}$,更新的码本为$\{00,010,011,10,11\}$.

5.17 一个不合格的骰子,6 个面的出现概率为$\left\{\frac{1}{3},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{12},\frac{1}{12}\right\}$.

(1) 求此骰子的熵.($\log_2 3 = 1.584\,96$)

(2) 令 $A_0^{(n)}$ 是长度为 n 的精确典型列集合.

$$A_0^{(n)} = \{(x_1, \cdots, x_n): p(x_1, \cdots, x_n) = 2^{-nH(X)}\}$$

当 $n = 12$ 时求上述骰子对应的 $A_0^{(n)}$ 中元素的个数.

(3) 求此骰子对应的输出序列 x 的累积分布函数的取值范围[$F(x)$, $F(x) + p(x)$).其中, x 的前四个输出符号为$(x_1, x_2, x_3, x_4) = (1, 2, 3, 4)$.(注:此问中 $F(x) = P(X < x)$).

5.18 Kraft 不等式.

(1) 当 $r = 2$ 时,试问无限长的即时码 $l_1 = 1$, $l_2 = 2, \cdots, l_k = k, \cdots$是否满足 Kraft 不等式?

(2) 将上问推广到任意 r 的情况.

5.19 试证明在即时码 $W_1 = 0$, $W_2 = 10$, $W_3 = 11$ 中一个码符号的差错有可能传播到无限远处.(提示:考虑 $W_1(W_3)^n W_2$)

第 6 章

有噪信道编码

从第 5 章已知,在理论上,在无噪无损信道中,只要对信源的输出进行恰当的编码,总能以最大信息传输率 C(信道容量)无错误地传输信息.但一般信道中总存在噪声或干扰,信息传输会造成损失,那么有噪信道中怎么能使消息通过传输后发生的错误最少?在有噪信道中无错误传输的可达的最大信息传输率是多少?这就是本章要研究的内容,即通信的可靠性问题.香农在 1948 年的文章中提出并证明了这个极限信息传输率的存在,这个定理叫信道编码定理,也称香农第二定理.

在第 4 章中,利用平均互信息的概念得到信道容量,知道信道容量是信道可以传输的最大信息量,这个信息量是在已知信道输出的情况下对输入不确定性消除的量.在有噪信道中,输入输出之间是统计依赖关系而不是确定关系,因此,由信道输出要唯一地译成输入一般将无法避免差错,这时,根据信道输出确定信道输入的可靠程度就反映为错误概率.对于有噪信道,这一错误概率完全取决于信道的特性,且不可能为零.但是香农的研究表明,如果把要传送的消息在传送前事先进行编码,并在接收端采用适当的译码,则消息有可能得到无错误的传输,也就是说,通过不可靠的信道可以实现可靠的信息传输.

6.1 信道编码的相关概念

信道编码又称为数据传输码或差错控制码,虽然和信源编码一样都是一种编码,但信

源编码的作用是压缩冗余度以得到信息的有效表示,提高传输时的信息传输率.而信道编码的作用是提高信息传输时的抗干扰能力以增加信息传输的可靠性.在研究信道编码时,信源编码器和信源译码器分别归于信源和信宿.等效通信系统模型如图 6.1 所示.下面研究编码信道中的信道编码和译码.

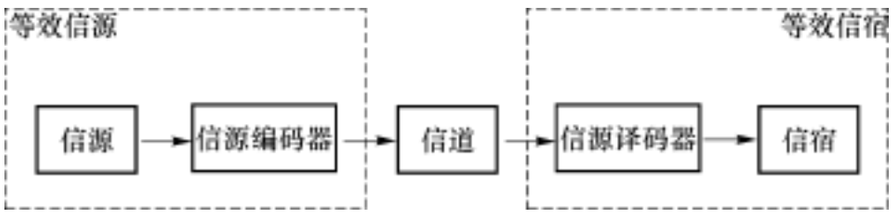


图 6.1 等效通信系统模型

信源符号 s_i, $i=1,2,\cdots,M$ 一般是已经经过信源编码的 M 种信源码字.信道编码的编码对象就是这 M 种信源码字.这 M 种信源码字通常是由二元符号"0,1"构成的码字序列,也叫信息序列,而且符号"0"和"1"是独立等概的.所谓信道编码,就是按一定的规则给信息序列增加一些多余的码元,使不具有规律性的信息序列变为具有某种规律性的信道码字序列 X,也就是说码字序列 X 的码元之间是相关的.在接收端,信道译码器利用这种相关性(也就是已知的编码规则)来译码,检验接收到的码字序列 Y 中是否有错,并且纠正其中的差错.根据相关性来检测和纠正传输过程中产生的差错就是信道编码的基本思想.

在有噪信道中,传输信息发生错误的错误概率与信道的统计特性、编码方法和译码规则都有关系.下面分别讨论这些因素,看看能不能对这些因素加以控制以提高通信的可靠性.

6.1.1 错误概率和译码规则

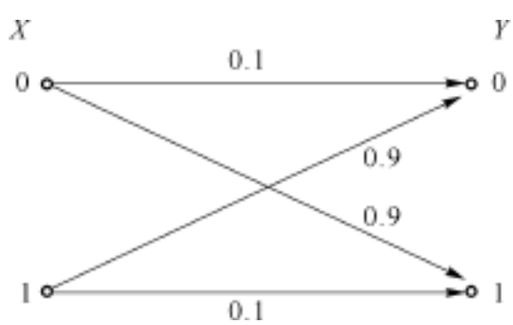


图 6.2 二元对称信道

我们已经知道,错误概率与信道的统计特性有关.信道的统计特性可由信道的传递矩阵来表示,由信道矩阵可以求出错误概率.例如在图 6.2 的二元对称信道中,单个符号的错误传递概率是 p,单个符号的正确传递概率为 $\overline{p}=1-p$,因此错误概率与信道的统计特征有关.但是通信的过程并不是信息传输到信道输出端就结束了,还要经过译码过程才能到达信宿,译码过程和译码规则对系统的错误概率影响很大.

下面举一个例子来说明译码规则对系统的错误概率的影响.

【例 6.1】

设有一个二元对称信道,其输入符号为等概分布.如果给定在信道输出端接收到符号

“0”时,译码器把它译成“0”,接收到“1”时,把它译成“1”,那么译码错误概率 $P_E = 0.9$;反之,如果规定在信道输出端接收到符号“0”时,译码器把它译成“1”,接收到“1”时,把它译成“0”,则译码错误概率 $P_E = 0.1$.可见,错误概率既与信道统计特性有关,也与译码规则有关.

1.译码规则

定义 6.1　设信道的输入符号集 $X = \{x_i, i = 1,2,\cdots,r\}$,输出符号集 $Y = \{y_j, j = 1,2,\cdots,s\}$,若对每一个输出符号 y_j,都有一个确定的函数 $F(y_j)$,使 y_j 对应于唯一的一个输入符号 x_i,则称这样的函数为译码规则,记为

$$F(y_j) = x_i \qquad i = 1,2,\cdots,r;\ j = 1,2,\cdots,s \tag{6.1}$$

对于有 r 个输入,s 个输出的信道而言,输出 y_j 可以对应 r 个输入中的任何一个,所以译码规则共有 r^s 种.

【例 6.2】

设有一信道,信道矩阵为

$$P = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

根据此信道矩阵,可以设计一个译码规则如下:

$$A:\begin{cases} F(y_1) = x_1 \\ F(y_2) = x_2 \\ F(y_3) = x_3 \end{cases}$$

也可以设计另一个译码规则

$$B:\begin{cases} F(y_1) = x_1 \\ F(y_2) = x_3 \\ F(y_3) = x_2 \end{cases}$$

由于 $r = 3$, $s = 3$,总共可以设计出 $r^s = 27$ 种译码规则,应该怎样选择译码规则呢?一个很自然的准则就是使平均错误概率最小.

因此我们先介绍平均错误概率.

(1) 错误概率

在确定译码规则 $F(y_j) = x_i$ 后,若信道输出端接收到符号 y_j,则一定译成 x_i,如果发送端发送的确实就是 x_i,就是正确译码;反之,若发送端发送的不是 x_i 就认为是错误译码.于是收到符号 y_j 条件下,译码的条件正确概率为

$$p[F(y_j) \mid y_j] = p(x_i \mid y_j) \tag{6.2}$$

而条件错误概率

$$p(e \mid y_j) = 1 - p(x_i \mid y_j) = 1 - p[F(y_j) \mid y_j] \tag{6.3}$$

其中，e 表示除了 $F(y_j)=x_i$ 以外的所有输入符号的集合．

译码后的平均错误概率 P_E 是条件错误概率 $p(e|y_j)$ 对 Y 空间取平均值，即

$$P_E = E[p(e|y_j)] = \sum_{j=1}^{s} p(y_j)\,p(e|y_j) \tag{6.4}$$

它表示经过译码后平均接收到一个符号所产生的错误大小．

(2) 译码规则

如何设计译码规则 $F(y_j)=x_i$ 使 P_E 最小呢？由于式(6.4)右边是非负项之和，所以可以选择译码规则使每一项为最小，则所得 P_E 为最小．因为 $p(y_j)$ 与译码规则无关，所以只要设计译码规则 $F(y_j)=x_i$ 使条件错误概率 $p(e|y_j)$ 最小，也就是要选择 $p[F(y_j)|y_j]$ 最大．这就是最大后验概率准则．

定义 6.2　选择译码函数 $F(y_j)=x^*$，使之满足条件

$$p(x^*|y_j) \geq p(x_i|y_j) \qquad \forall\, i,\ x^* \in X \tag{6.5}$$

则称为最大后验概率译码规则，又称为最小错误概率准则、最优译码、最佳译码、理想观测者规则．

它对于每一个输出符号 y_j，$j=1,2,\cdots,s$ 均译成具有最大后验概率的那个输入符号 x^*，则信道译码平均错误概率最小．

因为一般是已知信道的前向概率 $p(y_j|x_i)$ 和输入符号的先验概率 $p(x_i)$，所以根据贝叶斯定律，式(6.5)又可写成

$$\frac{p(y_j|x^*)\,p(x^*)}{p(y_j)} \geq \frac{p(y_j|x_i)\,p(x_i)}{p(y_j)} \qquad \forall\, i \tag{6.6}$$

一般 $p(y_j)\neq 0$，这样，最大后验概率译码规则就可表示为：选择译码函数 $F(y_j)=x^*$，$x^*\in X$，$y_j\in Y$，使其满足 $p(y_j|x^*)\,p(x^*)\geq p(y_j|x_i)\,p(x_i)$，$x_i\in X$．

当输入符号的先验概率 $p(x_i)$ 相等时，上式又可写成 $p(y_j|x^*)\geq p(y_i|x_i)$，因此又定义了一个极大似然译码规则．

定义 6.3　选择译码函数 $F(y_i)=x^*$，使之满足条件

$$p(y_j|x^*) \geq p(y_i|x_i) \qquad \forall\, i,\ x^* \in X \tag{6.7}$$

称为极大似然译码规则．

当输入符号等概时，这两个译码规是等价的，根据极大似然译码准则可以直接从信道矩阵的传递概率中去选定译码函数：当收到 y_j 后，译成信道矩阵 P 第 j 列中最大的转移概率所对应的 x_i．

极大似然译码准则本身不依赖于先验概率 $p(x_i)$，当先验概率为等概率分布时，它与最大后验概率译码准则是等价的，可以使平均错误概率 P_E 达到最小．如果先验概率不相等或不知道时，仍可以采用这个准则，但不一定能使 P_E 最小．

根据上述译码规则，可以计算平均错误概率．

2 . 平均错误概率

$$P_{\mathrm{E}}=\sum_{j=1}^{s}\sum_{i\neq *}p(y_j\mid x_i)\,p(x_i)=\sum_{Y}\sum_{X-x^*}p(xy) \tag{6.8}$$

共$(r-1)s$项求和 .求和号下面的 $X-x^*$ 表示对输入符号集 X 中除 x^* 以外的所有元素求和 .上式表示对联合概率矩阵 $P=[p(x_iy_j)]$ 中除 $p(x^*y_j)$以外的所有元素求和,而平均正确概率为

$$\overline{P_{\mathrm{E}}}=1-P_{\mathrm{E}}=\sum_{j=1}^{s}p[F(y_j),y_j]=\sum_{j=1}^{s}p(x^*y_j) \tag{6.9}$$

式(6.8)又可写成

$$P_{\mathrm{E}}=\sum_{Y,\,X-x^*}p(y_j\mid x_i)\,p(x_i) \tag{6.10}$$

如果输入为等概分布,即 $p(x_i)=\dfrac{1}{r}$,则

$$P_{\mathrm{E}}=\frac{1}{r}\sum_{Y,\,X-x^*}p(y_j\mid x_i) \tag{6.11}$$

式(6.11)表明,在输入等概的情况下,译码错误概率可用信道矩阵中的元素 $p(y_j\mid x_i)$的求和来表示,求和是除去每列中对应于 $F(y_j)=x^*$ 的那一项后,矩阵中其余元素之和 .

【例 6.3】

求例 6.2 中两种译码规则对应的平均错误概率 .

解

当输入为等概分布时,译码规则 B 就是极大似然译码规则,两种译码规则所对应的平均错误概率分别为

$$P_{\mathrm{E}}(A)=\frac{1}{3}\sum_{Y,\,X-x^*}p(y_j\mid x_i)$$

$$=\frac{1}{3}\times[(0.2+0.3)+(0.3+0.3)+(0.2+0.5)]=0.6$$

$$P_{\mathrm{E}}(B)=\frac{1}{3}\sum_{Y,\,X-x^*}p(y_j\mid x_i)$$

$$=\frac{1}{3}\times[(0.2+0.3)+(0.3+0.3)+(0.2+0.4)]=0.5667$$

所以在输入为等概分布时,极大似然译码规则可使信道平均错误概率最小 .

当输入为不等概分布时,假设某个输入概率分布 $p(x_1)=1/4$,, $p(x_2)=1/4$, $p(x_3)=1/2$,则因为极大似然译码规则与输入概率分布无关,所以与上例相同,

$$P_{\mathrm{E}}(B)=\frac{1}{3}\sum_{Y,\,X-x^*}p(y_j\mid x_i)$$

$$= \frac{1}{3} \times [(0.2 + 0.3) + (0.3 + 0.3) + (0.2 + 0.4)] = 0.566\,7$$

根据最小错误概率译码准则,由其联合概率矩阵

$$P(XY) = \begin{pmatrix} 0.125 & 0.075 & 0.050 \\ 0.050 & 0.075 & 0.125 \\ 0.150 & 0.150 & 0.200 \end{pmatrix}$$

可得译码函数

$$C: \begin{cases} F(y_1) = x_3 \\ F(y_2) = x_3 \\ F(y_3) = x_3 \end{cases}$$

所以此时错误概率为

$$P_E(C) = [(0.125 + 0.075 + 0.050) + (0.050 + 0.075 + 0.125)] = 0.5$$

所以输入不是等概分布时最大似然译码准则的平均错误概率不是最小.

译码时发生错误是由信道中的噪声引起的,而信道噪声的影响使在接收端收到输出符号 Y 后对发送端发送的符号仍然存在不确定性,因此平均错误概率与信道疑义度存在着一定的关系,这个关系用费诺不等式表示.

定理 6.1 平均错误概率 P_E 与信道疑义度 $H(X|Y)$ 满足以下关系:

$$H(X|Y) \leq H(P_E) + P_E \log_2 (r - 1) \tag{6.12}$$

证明

$$\overline{P_E} = 1 - P_E = \sum_{Y} p(x^* y) \tag{6.13}$$

$$P_E = \sum_{Y, X - x^*} p(xy) \tag{6.14}$$

$$\begin{aligned} & H(P_E) + P_E \log_2 (r - 1) \\ &= P_E \log_2 \frac{1}{P_E} + (1 - P_E) \log_2 \frac{1}{1 - P_E} + P_E \log_2 (r - 1) \\ &= P_E \log_2 \frac{r - 1}{P_E} + (1 - P_E) \log_2 \frac{1}{1 - P_E} \\ &= \sum_{Y, X - x^*} p(xy) \log_2 \frac{r - 1}{P_E} + \sum_{Y} p(x^* y) \log_2 \frac{1}{1 - P_E} \end{aligned} \tag{6.15}$$

而信道疑义度

$$\begin{aligned} H(X|Y) &= \sum_{XY} p(xy) \log_2 \frac{1}{p(x|y)} \\ &= \sum_{Y, X - x^*} p(xy) \log_2 \frac{1}{p(x|y)} + \sum_{Y} p(x^* y) \log_2 \frac{1}{p(x^*|y)} \end{aligned} \tag{6.16}$$

所以

$$H(X|Y) - H(P_E) - P_E \log_2 (r - 1)$$

$$= \sum_{Y,\ X\neq x^*} p(xy)\log_2 \frac{P_E}{(r-1)\,p(x\mid y)} + \sum_{Y} p(x^* y)\log_2 \frac{1-P_E}{p(x^*\mid y)} \tag{6.17}$$

因为

$$\log_2 x \leq x-1 \tag{6.18}$$

所以

$$\begin{aligned}
&H(X\mid Y) - H(P_E) - P_E\log_2(r-1)\\
&\leq \sum_{Y,\ X\neq x^*} p(xy)\left[\frac{P_E}{(r-1)\,p(x\mid y)} - 1\right] + \sum_{Y} p(x^* y)\left[\frac{1-P_E}{p(x^*\mid y)} - 1\right]\\
&= \frac{P_E}{r-1}\sum_{Y,\ X\neq x^*} p(y) - \sum_{Y,\ X\neq x^*} p(xy) + (1-P_E)\sum_{Y} p(y) - \sum_{Y} p(x^* y)\\
&= \frac{P_E}{r-1}\sum_{X\neq x^*}\sum_{Y} p(y) - P_E + (1-P_E) - (1-P_E)\\
&= \frac{P_E}{r-1}(r-1) - P_E\\
&= 0
\end{aligned} \tag{6.19}$$

证毕

虽然 P_E 与译码规则有关,但是不管采用什么译码规则该不等式都是成立的.费诺不等式表明,接收到 Y 后关于 X 的平均不确定性可以分为两部分:第一部分 $H(P_E)$是指接收到 Y 后是否产生错误的不确定性;第二部分 $P_E\log_2(r-1)$是当错误 P_E 发生后,判断是哪个输入符号造成错误的最大不确定性,是$(r-1)$个符号不确定性的最大值与 P_E 的乘积.若以 $H(X\mid Y)$为纵坐标,P_E 为横坐标,函数 $H(P_E)+P_E\log_2(r-1)$随 P_E 变化的曲线如图 6.3 所示.

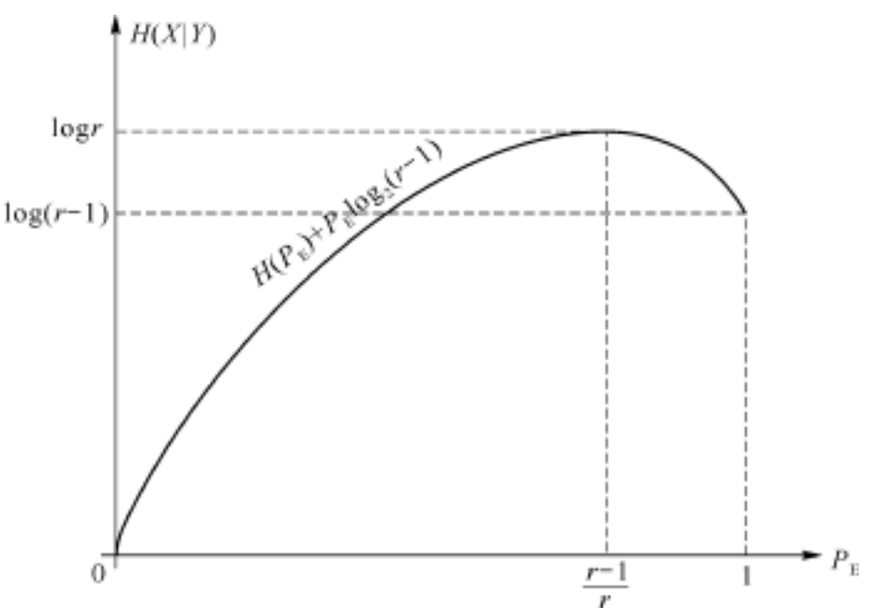


图 6.3　费诺不等式曲线图

P_E 的最大值为 1,这时 $H(X|Y)=\log_2(r-1)$,当 $P_E=(r-1)/r$ 时,$H(X|Y)=\log_2 r$,取到最大值 .

6 .1 .2 错误概率与编码方法

在 6 .1 节中讨论了平均错误概率 P_E 与译码规则的关系,选择最佳译码规则只能使错误概率 P_E 有限地减小,无法使 P_E 任意地小,要想进一步减小错误概率 P_E,还必须选择恰当的编码方法 .下面举例说明编码方法是怎样影响错误概率的 .

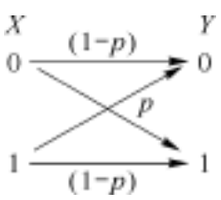


图 6 .4 二元对称信道

设有二元对称信道如图 6 .4 所示 .相应的信道矩阵为

$$P=\begin{pmatrix}0.99 & 0.01\\ 0.01 & 0.99\end{pmatrix}$$

选择最佳译码规则为

$$\begin{cases}F(y_1)=x_1\\ F(y_2)=x_2\end{cases}$$

总的平均错误概率在输入分布为等概的条件下为

$$P_E=\frac{1}{r}\sum_{Y,\,X\neq x^*}p(y|x)=\frac{1}{2}\times(0.01+0.01)=10^{-2}$$

对于一般数字通信系统,这个错误概率是非常大的,一般数字通信要求错误概率在 $10^{-6}\sim10^{-9}$ 的范围内,有的甚至要求更低的错误概率 .

那么,在上述统计特性的二元信道中,是否有办法使错误概率降低呢?实际经验告诉我们:只要在发送端把消息重复发几遍,就可使接收端接收消息时错误减少,从而提高通信的可靠性 .

例如,发信源符号“0”时,重复发送 3 个“0”,发“1”时,重复发送 3 个“1”,这可以看成离散无记忆信道的 3 次扩展信道 .在信道输入端有两个码字“000”和“111”,在输出端由于信道干扰,各个码元都可能发生错误,则有 8 个可能的输出序列 .输入是 3 次扩展信道的 8 个可能出现的二元序列中选两个作为消息,而输出端这 8 个可能的二元序列都是接收序列,如表 6 .1 所示 .

这时信道矩阵为

$$P=\begin{pmatrix}\bar{p}^3 & \bar{p}^2p & \bar{p}^2p & \bar{p}p^2 & \bar{p}^2p & \bar{p}p^2 & \bar{p}p^2 & p^3\\ p^3 & \bar{p}p^2 & \bar{p}p^2 & \bar{p}^2p & \bar{p}p^2 & \bar{p}^2p & \bar{p}^2p & \bar{p}^3\end{pmatrix}$$

表 6 .1　简单重复编码举例

输入状态	没有使用的码字	用作消息的码字	输出端接收序列	输出状态
x_1		000	000	y_1
	001		001	y_2
	010		010	y_3
	011		011	y_4
	100		100	y_5
	101		101	y_6
	110		110	y_7
x_2		111	111	y_8

假设输入符号为等概分布,采用极大似然译码规则,即取信道矩阵中每列数值最大的元素所对应的译码函数为

$$\begin{cases} F(y_1) = x_1 \\ F(y_2) = x_1 \\ F(y_3) = x_1 \\ F(y_4) = x_2 \\ F(y_5) = x_1 \\ F(y_6) = x_2 \\ F(y_7) = x_2 \\ F(y_8) = x_2 \end{cases}$$

译码后的平均错误概率为

$$\begin{aligned} P_{\mathrm{E}} &= \sum_{Y,\ X-x^*} p(y \mid x)\, p(x) \\ &= \frac{1}{M} \sum_{Y,\ X-x^*} p(y \mid x) \\ &= \frac{1}{2}\left(p^3 + \overline{p}p^2 + \overline{p}p^2 + \overline{p}p^2 + \overline{p}p^2 + \overline{p}p^2 + p^3\right) \\ &= p^3 + 3\overline{p}p^2 \\ &\approx 3 \times 10^{-4} \end{aligned}$$

我们也可以直观地采用择多译码,即根据输出端接收序列是“0”多还是“1”多来译码.如果有两个以上是“0”则译码器就判决为“0”,如果有二个以上是“1”则判决为“1”.根据择多译码规则,同样可得到

$$\begin{aligned} P_E &= P_r\{\text{错 3 个码元的概率} + \text{错 2 个码元的概率}\} \\ &= C_3^3 p^3 + C_3^2 \overline{p} p^2 \\ &= p^3 + 3\overline{p} p^2 \\ &\approx 3\times 10^{-4} \end{aligned}$$

可见,择多译码的平均错误概率与极大似然译码规则的平均错误概率是一样的.与原来的二元对称信道的平均错误概率 10^{-2} 相比,这种简单重复编码(重复 3 次)的平均错误概率降低了近两个数量级.这是因为输入消息码字 x_0 与 4 个接收序列 y_0, y_1, y_2, y_4 对应,而 x_1 与 y_3, y_5, y_6, y_7 对应,若接收码字中有一位码元发生错误,译码器还能正确译出所发送的码字,若传输中两位或 3 位码元发生错误,译码器就会译错.所以这种简单重复编码能纠正一位码元的错误,使得错误概率降低.

显然,如果进一步增大重复次数 n,则会继续降低平均错误概率.可算得

$n = 5$ $P_E = 10^{-5}$

$n = 7$ $P_E = 4\times 10^{-7}$

$n = 9$ $P_E = 10^{-8}$

$n = 11$ $P_E = 5\times 10^{-10}$

可见,当 n 很大时,使 P_E 很小是可能的.但这时带来了一个新问题,当 n 很大时,信息传输会降低很多.我们把编码后的信息传输率表示为

$$R = \frac{\log_2 M}{n} \tag{6.20}$$

含义为:M 个信源符号(一般已接近等概分布),平均每个信源符号所携带的信息量为 $\log_2 M$ bit,用 n 个码元的信道编码来传输,平均每个码符号所携带的信息量即为信息传输率 R.

如果传输每个符号平均需要 t 秒,则编码后的信息传输速率为

$$R_t = \frac{R}{t} \tag{6.21}$$

当 $M = 2$ 时可依次求得简单重复编码的信息传输率和信息传输速率:

$n = 1$ $R = 1$

$n = 3$ $R = 1/3$

$n = 5$ $R = 1/5$

… …

$n = 11$ $R = 1/11$

由此可见,利用简单重复编码减小平均错误概率 P_E 是以降低信息传输率 R 为代价的,那么有没有可能找到一种编码方法,使平均错误概率 P_E 充分小而信息传输率又不至于太低呢?

1. 消息符号个数

首先看一下简单重复编码为什么使信息传输率降低?在未重复以前,输入端有 2 个消息,假设为等概率分布则每个消息携带的信息量是 $\log_2 M = 1$ bit.简单重复编码($n = 3$)后,可以把信道看成是无记忆信道的三次扩展信道,这时输入端有 8 个二元序列可以作为消息,但是我们只选择了两个二元序列作为消息 $M = 2$,每个消息携带的平均信息量仍为 1 bit,而传送一个消息需要 3 个二元码符号,所以 R 就降低到 1/3.

如果在扩展信道的输入端把 8 个可能作为消息的二元序列都用上作为消息,则 $M = 8$,每个消息携带的平均信息量就是 $\log_2 M = \log_2 8 = 3$ bit,而传递一个消息所需的码符号仍为 3 个二元码,这样 R 就提高到 1 比特/码符号.

译码时接收端 8 个接收序列译成与它对应的发送序列,只要接收序列中有一个码元发生错误就会变成其他的码字序列,使译码造成错误.只有接收序列中每个码元都不发生错误才能正确传输,所以得到正确传输概率为 $1 - P_E = \overline{p}^3$.于是错误概率为

$$P_E = 1 - \overline{p}^3 \approx 3 \times 10^{-2} \qquad (p = 0.01)$$

这时 P_E 反而比单符号信道传输的 P_E 大 3 倍.

因此看到这样一个现象:在一个二元信道的 n 次无记忆扩展信道中,输入端有 2^n 个符号序列可以作为消息.如果选出其中的 M 个作为消息传递,则:M 大一些,P_E 也大些,R 也大;M 取小一些,P_E 就降低,而 R 也要降低.

若在 3 次无记忆扩展信道中,取 $M = 4$,用如下 4 个符号序列作为消息:

000　　011　　101　　110

按照最大似然译码规则,可计算出错误概率为

$$P_E \approx 2 \times 10^{-2}$$

信息传输率为

$$R = \frac{\log_2 4}{3} = \frac{2}{3}$$

与 $M = 8$ 的情况相比,错误概率降低了,而信息传输率也降低了.因此仅仅依靠改变消息符号个数不能解决 P_E 和 R 的矛盾.再进一步看,从 2^n 个符号序列中取 $M = 4$ 个作为消息可以有 C_8^4 共 70 种选择方法,选取的方法也就是编码方法不同,错误概率不同,现在来比较下面两种取法:

$M = 4$　　第 1 种选法　000,011,101,110

$M = 4$　　第 2 种选法　000,001,010,100

可求得第 1 种选法的错误概率为

$$P_E = 2\times10^{-2} \qquad R = \frac{2}{3}$$

第 2 种选法的错误概率为

$$P_E = 2.28\times10^{-2} \qquad R = \frac{2}{3}$$

两者 R 相同.第 2 种码的平均错误概率大.对于第 1 种码来说,当接收到的发送的 4 个符号序列中任一码元发生错误,就可判断消息在传输中发生了错误,但无法判断由哪个消息发生错误而来;对第 2 种码,当发送消息"000"时,传输中任一码元发生错误就变成了其他 3 个可能发送的消息,根本无法判断传输时有无发生错误.可见错误概率与编码方法有很大关系.

2.(5,2)线性码

考察这样一个例子:信道输入端所选的消息数不变,任取 $M=4$,而增加码字的长度,取 $n=5$.这时信道为二元对称信道的五次扩展信道,这个信道输入端有 $2^5=32$ 个不同的二元序列.我们选取其中 4 个作为发送消息.这时信息传输率为

$$R = \frac{\log_2 4}{5} = \frac{2}{5}$$

设输入序列 $x_i = x_{i_1}\,x_{i_2}\,x_{i_3}\,x_{i_4}\,x_{i_5}$, $x_{i_k}\in\{0,1\}$, $i=1,2,3,4,5$,其中 x_{i_k} 为 x_i 序列中第 k 个分量,若 x_i 中各分量满足方程

$$\begin{cases} x_{i_1} = x_{i_1} \\ x_{i_2} = x_{i_2} \\ x_{i_3} = x_{i_1} \bar{\mathbf{1}} \quad x_{i_2} \\ x_{i_4} = x_{i_4} \\ x_{i_5} = x_{i_1} \bar{\mathbf{1}} \quad x_{i_2} \end{cases}$$

其中,$\bar{1}$ 为模二和运算,也叫异或.

或者写成矩阵形式

$$x_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_{i_1} \\ x_{i_2} \\ x_{i_3} \\ x_{i_4} \\ x_{i_5} \end{pmatrix}$$

由上述编码方法得到一种(5,2)线性码.如果译码采用极大似然译码规则,它的译码规则如表 6.2 所示.

选用此码,接收端译码能纠正码字中所有发生一位码元的错误,也能纠正其中两个两位码元的错误,所以可计算得

正确译码概率 $\overline{P_E} = \overline{p}^5 + 5\overline{p}^4 p + 2\overline{p}^3 p^2$

错误译码概率 $P_E = 1 - \overline{P_E} \approx 7.8 \times 10^{-4}$ ($p = 0.01$)

将这种编码方法与前述 $M = 4$, $n = 3$ 的两种编码方法相比,虽然信息传输率略降低了些,但错误概率减少很多.再与 $n = 3$, $M = 2$ 的简单重复码相比较,它们的错误概率接近于同一个数量级,但(5,2)线性码的信息传输率却比 $n = 3$ 的重复码的信息传输率大.因此采用增大 n,并且适当增大 M 并采用恰当的编码方法,既能使 P_E 降低,又能使信息传输率不会减少太多.

表 6.2　(5,2)线性码译码规则

接收码字	译码输出	接收码字	译码输出
00000	00000	10000	00000
00001	00000	10001	00000
00010	00000	10010	11010
00011	00000	10011	10111
00100	00000	10100	10111
00101	01101	10101	10111
00110	10111	10110	10111
00111	10111	10111	10111
01000	00000	11000	11010
01001	01101	11001	11010
01010	11010	11010	11010
01011	11010	11011	11010
01100	01101	11100	01101
01101	01101	11101	01101
01110	01101	11110	11010
01111	01101	11111	10111

3.汉明距离

我们先引入一个概念——码字距离,然后再解释(5,2)线性码能获得较高 P_E 的原因.

定义 6.4　长度为 n 的两个符号序列(码字) x_i 与 y_j 之间的距离是指序列 x_i 和 y_j 对应位置上码元不同的位置的个数,通常又称为汉明距离.

例如,二元序列 $x_i = 101111$, $y_j = 111100$,则得 $D(x_i, y_j) = 3$;四元序列 $x_i = 1320120$, $y_j = 1220310$,则得 $D(x_i, y_j) = 3$ 对于二元码序列,令 $x_i = x_{i_1} x_{i_2} \cdots x_{i_n}$, $x_{i_k} \in \{0,1\}$, $y_j = y_{j_1} y_{j_2} \cdots y_{j_n}$, $y_{j_k} \in \{0,1\}$,则 x_i 和 y_j 的汉明距离为

$$D(x_i, y_j) = \sum_{k=1}^{N} x_{i_k} \bar{1} y_{j_k} \tag{6.22}$$

这样定义的码字距离满足距离公理,即汉明距离满足以下性质:

(1) 非负性　$D(x_i, y_j) \geq 0$,当且仅当 $x_i = y_j$ 时等号成立;

(2) 对称性　$D(x_i, y_j) = D(y_j, x_i)$;

(3) 三角不等式　$D(x_i, z_k) + D(z_k, y_j) \geq D(x_i, y_j)$.

定义 6.5　码 C 中,任意两个码字的汉明距离的最小值称为该码 C 的最小距离.

$$d_{\min} = \min D(w_i, w_j) \qquad w_i \neq w_j;\ w_i, w_j \in C \tag{6.23}$$

【例 6.4】

设有 $n=3$ 的两组码,对于码 C_1 有 $d_{\min}=2$,对于码 C_2 有 $d_{\min}=1$.

码的最小距离 $d_{\min}$ 与译码错误概率密切相关.我们用距离概念来考察以下 5 个码,如表 6.3 所示.

表 6.3 码的最小距离与平均译码错误概率的关系

	码 1	码 2	码 3	码 4	码 5
码字	000	000	000	00000	000
	111	011	001	01101	001
		101	010	10111	010
		110	100	11010	011
					100
					101
					110
消息数 M	2	4	4	4	8
信息传输率 R	1/3	2/3	2/3	2/5	1
码的最小距离 $d_{\min}$	3	2	1	3	1
错误概率 P_E(极大似然译码)	3×10^{-4}	2×10^{-2}	2.28×10^{-2}	7.8×10^{-4}	3×10^{-2}

显然,$d_{\min}$ 越大,P_E 越小.码的最小距离 $d_{\min}$ 越大,受干扰后,越不容易把一个码字错成另一个码字,因而错误概率小;$d_{\min}$ 越小,受干扰后越容易把一个码字变成另一个码字,因而错误概率大.这就说明:在编码选择码字时,要使码字之间的距离尽可能地大.

现在,把汉明距离与最大似然译码规则联系起来,用汉明距离来表述极大似然译码准则.

极大似然译码准则为:选择译码函数 $F(y_j)=x^*$,使 $p(y_j|x^*)\geq p(y_j|x_i)$,"$i$.设码字 $x_i=x_{i_1}x_{i_2}\cdots x_{i_n}$,$y_j=y_{j_1}y_{j_2}\cdots y_{j_n}$,在传输过程中输入码字 x_i 中有 D_{ij} 个位置发生错误,接收端接收序列为 y_j,即 $D(x_i,y_j)=D_{ij}$,没有发生错误的位置有 $n-D_{ij}$ 个.

当二元对称信道是无记忆信道时,

$$p(y_j|x_i)=p(y_{j_1}|x_{i_1})\,p(y_{j_2}|x_{i_2})\cdots p(y_{j_n}|x_{i_n})=p^{D_{ij}}(1-p)^{n-D_{ij}} \tag{6.24}$$

只要 $p<1/2$(这是正常情况,例如 $p=10^{-2}$),则 D_{ij} 越大,$p(y_j|x_i)$ 越小;D_{ij} 越小,$p(y_j|x_i)$ 越大.因此二元对称信道中最大似然译码规则可用汉明距离表示为:

选择译码函数 $F(y_j)=x^*$ 使 $D(x^*,y_j)\leq D(x_i,y_j)$ 即

$$D(x^*,y_j)=\min_i D(x_i,y_j) \tag{6.25}$$

也就是在接收到码字 y_j 后,在输入码字集$\{x_i,\ i=1,2,\cdots,r\}$中寻找一个与 y_j 的汉明距

离最小的码字 x^*,这又称为最小距离译码准则 .

这时平均译码错误概率也可用汉明距离来表示 .设输入码字数为 M(并设输入等概分布),则

$$P_E = \frac{1}{M} \sum_{Y,\ X - x^*} p(y_j \mid x_i) = \frac{1}{M} \sum_j \sum_{i \neq *} p^{D_{ij}} (1 - p)^{n - D_{ij}} \tag{6.26}$$

或者

$$P_E = 1 - \frac{1}{M} \sum_Y p(y_j \mid x^*) = 1 - \frac{1}{M} \sum_j p^{D_{*j}} (1 - p)^{n - D_{*j}} \tag{6.27}$$

其中,

$$D_{*j} = D(x^*, y_j) \tag{6.28}$$

在非二元对称信道中也可采用最小距离译码准则,但它不一定等于极大似然译码准则,也就不一定能使 P_E 最小 .

上面的讨论在 M 和 n 相同的情况下,即保持一定的信息传输率 R 时,选择不同的编码方法可取得具有不同最小距离的码,我们选择最小距离最大的那一个码;在译码时,则将接收序列译成与其距离最小的码字,这样得到的 P_E 最小 .那么只要码长 n 足够长,总可以通过恰当的选择 M 个消息所对应的码字使 P_E 很小,而 R 保持一定 .下面我们将证明这一点 .

6.2　有噪信道编码定理

6.2.1　有噪信道编码定理

定理 6.2　设有一个离散无记忆平稳信道,其信道容量为 C.当信息传输率 $R < C$,只要码长 n 足够长,则总存在一种编码,可以使平均译码错误概率任意小 .

这个定理称为有噪信道编码定理,又称为香农第二定理 .通过一个有噪信道可以实现几乎无失真传输,这与人们的直观想象似乎是大相径庭的,而定理的证明也是非常巧妙的 .按照通常的思路证明此结论可能先要构造一个理想的好码,然后计算此码用于传输时的平均错误译码概率,但这两点都难以实现 .首先,构造具有理想性能的好码是一个非常复杂的问题,在当时的计算条件下还无法解决;其次,想在 n 很大时计算这一理想好码在最佳译码规则或极大似然译码规则下的 P_E 也是极其困难的 .香农巧妙地避开了这两个难题,首先,他不去构造理想的好码,而是用随机编码的方法得到所有可能生成的码的集合,然后在码集合中随机选择一个码作为输入码序列,最后计算这样随机选择的一个码在码集合上的平均性能 .由于这样计算出的 P_E 可以达到任意小,可以证明一定存在一种编码,它的性能 P_E 达到或者超过随机编码的平均性能 .由于所求的是平均性能,就可以用

大数定律且不必考虑码的结构.在译码时,利用了联合典型序列的概念,即将接收序列译成与其联合典型的码字.这种译码方法不是最优译码,但便于理论分析.

下面介绍联合 ε 典型序列和联合渐进等分割性.

在定长信源编码定理中我们介绍过 ε 典型序列.联合 ε 典型序列的概念是涉及两个随机序列时的自然扩展.其定义如下:

定义 6.6 设(x_i, y_j)是长为 n 的随机序列对,并且 $p(x_iy_j)=\prod_{k=1}^{n} p(x_{i_k}y_{j_k})$,则在这些随机序列对中同时满足以下条件的序列对称为联合 ε 典型序列.

(1) $\left| -\frac{1}{n}\log_2 p(x_i) - H(X) \right| < \varepsilon$,即 x_i 是 X 的 ε 典型序列;

(2) $\left| -\frac{1}{n}\log_2 p(y_i) - H(X) \right| < \varepsilon$,即 y_i 是 X 的 ε 典型序列;

(3) $\left| -\frac{1}{n}\log_2 p(x_i, y_i) - H(XY) \right| < \varepsilon$, ε 是任意小的正数.

联合 ε 典型序列的全体构成的集合称为联合 ε 典型序列集,记作 $G_\varepsilon(XY)$;而把 X 的 ε 典型序列集记作 $G_\varepsilon(X)$, Y 的 ε 典型序列集记作 $G_\varepsilon(Y)$.

它们包含的典型序列的数目分别记作 $M_G(XY)$, $M_G(X)$, $M_G(Y)$.按照以上定义可以得到联合渐进等分割性.

对于任意小的正数 $\varepsilon\geq 0$, $\delta\geq 0$,当 n 足够大时,有

(1) 典型列集和联合典型列集的概率满足:

$$P_r\{G_\varepsilon(X)\}\geq 1-\delta \tag{6.29}$$

$$P_r\{G_\varepsilon(Y)\}\geq 1-\delta \tag{6.30}$$

$$P_r\{G_\varepsilon(XY)\}\geq 1-\delta \tag{6.31}$$

(2) 典型列和联合典型列出现的概率满足:

$$2^{-n[H(X)+\varepsilon]} < p(x_i) < 2^{-n[H(X)-\varepsilon]} \tag{6.32}$$

$$2^{-n[H(Y)+\varepsilon]} < p(y_j) < 2^{-n[H(Y)-\varepsilon]} \tag{6.33}$$

$$2^{-n[H(XY)+\varepsilon]} < p(x_iy_j) < 2^{-n[H(XY)-\varepsilon]} \tag{6.34}$$

(3) 典型列集和联合典型列集中元素的个数满足:

$$(1-\delta)2^{n[H(X)-\varepsilon]}\leq M_G(X)\leq 2^{n[H(X)+\varepsilon]} \tag{6.35}$$

$$(1-\delta)2^{n[H(Y)-\varepsilon]}\leq M_G(Y)\leq 2^{n[H(Y)+\varepsilon]} \tag{6.36}$$

$$(1-\delta)2^{n[H(XY)-\varepsilon]}\leq M_G(XY)\leq 2^{n[H(XY)+\varepsilon]} \tag{6.37}$$

典型序列 x_i 是输入端高概率出现的序列,典型序列 y_j 是输出端高概率出现的序列,而联合典型序列对(x_i, y_j)则是信道输入和输出之间关联密切且经常出现的序列对.也就是说,当某一输入典型序列 x_i 发送时,必定高概率地以和它构成联合 ε 典型序列对的接收序列 y_j 接收.

如果 (x, y) 和 (x_i, y_j) 统计独立并与 $p(x_i y_j)$ 有相同的边缘分布，即

$$p(x, y) = p(x)\, p(y) \tag{6.38}$$

且

$$p(x) = p(x_i) \tag{6.39}$$

$$p(y) = p(y_j) \tag{6.40}$$

则

$$\begin{aligned} P_r\{(x, y) \in G_\varepsilon(XY)\} &= \sum_{(x_i, y_j) \in G_\varepsilon(XY)} p(x_i)\, p(y_j) \\ &\leq 2^{n[H(XY)+\varepsilon]} \cdot 2^{-n[H(X)-\varepsilon]} \cdot 2^{-n[H(Y)-\varepsilon]} \\ &= 2^{-n[I(X;Y)-3\varepsilon]} \end{aligned} \tag{6.41}$$

并且

$$\begin{aligned} P_r\{(x, y) \in G_\varepsilon(XY)\} &= \sum_{(x_i, y_j) \in G_\varepsilon(XY)} p(x_i)\, p(y_j) \\ &\geq (1-\delta) 2^{n[H(XY)-\varepsilon]} \cdot 2^{-n[H(X)+\varepsilon]} \cdot 2^{-n[H(Y)+\varepsilon]} \\ &= (1-\delta) 2^{-n[I(X;Y)+3\varepsilon]} \end{aligned} \tag{6.42}$$

也就是说，在联合 ε 典型序列集中出现相互独立的随机序列对的概率是非常小的 .

下面在二元信道中证明有噪信道编码定理 .证明的基本思路是:

(1) 允许平均错误概率 P_E 任意小而非零;

(2) 在 n 次无记忆扩展信道中讨论，且 n 足够大，这样可以使用大数定理;

(3) 随机编码，在随机编码的基础上计算整个码集的码的平均错误概率，由此证明至少有一种好码存在，因为是随机编码，所以求错误译码概率时与特定的码字无关 .

所谓随机编码，是指在 n 长的输入序列中，随机选择 M 个作为输入码字序列组成一个码 C_k, $C_k = \{x_1, x_2, \cdots, x_M\}$, M 为信源消息数 .每次选择一个码字序列有 2^n 种可能的选择，得到一个码需选择 M 次码字，共有 $(2^n)^M$ 种可能的选择，也就是说通过随机编码可以得到 2^{nM} 个码 .这是一个很大的数，比如 $M = 2^8$, $n = 16$ 时为 $2^{4\,096} \approx 10^{1\,233}$，是一个非常大的数 .

当然，在这些码中有一部分是无法用的，比如某些码的码字有重复码，但由于码中码字数为 $M = 2^{nR}$，这只占全部可能的码字序列 2^n 中的很小的一部分，因此同一码中码字相同的概率很小，可以忽略这个问题 .(同一码中码字不同共有 $2^n(2^n - 1)\cdots(2^n - M + 1)$ 种可能性)

证明

假设通过随机编码，我们得到一个码集合 $\{C_k, k = 1, 2, \cdots, 2^{nM}\}$.令输入为等概分布，则码 C_k 的平均错误概率为(对码 C_k 中的 M 个码字求平均)

$$P_E(C_k) = \frac{1}{M} \sum_{i=1}^{M} p(e \mid x_i) \tag{6.43}$$

这里 e 是指输出序列中所有能够引起译码错误的序列 .

在码集$\{C_k\}$上对 $P_E(C_k)$求平均

$$\overline{P_E(C_k)} = \sum_{k=1}^{2^{nM}} p(C_k) P_E(C_k) = P_E(C_k) \tag{6.44}$$

这是由于对于随机编码而言,每个码的平均错误概率都相等,即 $P_E(C_k)$对所有 $k = 1, 2, \cdots, 2^{nM}$都相等,而且由于随机编码 C_k 中任何一个码字的 $p(e|x_i)$也都相等 .以 x_1 为例,得 $P_E = p(e|x_i) = p(e|x_1)$.

设 y_j 是发送码字序列 x_1 时信道输出处得到的接收序列,定义 y_j 与 x_i 构成联合典型序列的事件为 A_i,即 $A_i = \{(x_i, y_j) \in G_\varepsilon(XY), i = 1, 2, \cdots, M\}$.

按照联合典型译码法,也就是把接收序列 y_j 译成与它构成联合典型序列的那个码字,译码错误的发生有两种可能:

(1) y_j 不与 x_1 构成联合典型序列;

(2) y_j 与 x_1 以外的其他码字构成联合典型序列(这些事件有可能相交) .

所以

$$P_E = p(e|x_1) = P_r\{\overline{A_1} \cup A_2 \cup \cdots \cup A_M\} \tag{6.45}$$

根据和事件的概率关系可得

$$P_E \leq P_r\{\overline{A_1}\} + \sum_{i=2}^{M} P_r\{A_i\} \tag{6.46}$$

由于 $P_r\{A_1\} = p[G_\varepsilon(XY)] \geq 1 - \delta$,所以 $P_r\{\overline{A_1}\} \leq \delta$.而 y_j 与 $x_2, x_3, \cdots, x_M$ 是相互独立的序列,所以

$$P_r\{A_i\} \leq 2^{-n[I(X;Y) - 3\varepsilon]} \tag{6.47}$$

$$\begin{aligned} P_E &\leq \delta + (M - 1) \cdot 2^{-n[I(X;Y) - 3\varepsilon]} \\ &\leq \delta + 2^{nR} \cdot 2^{-n[I(X;Y) - 3\varepsilon]} \\ &= \delta + 2^{-n[I(X;Y) - R - 3\varepsilon]} \end{aligned} \tag{6.48}$$

其中,

$$M = 2^{nR} \tag{6.49}$$

因此选择任意小正数 δ 和 ε,当 $R < I(X; Y)$且 n 足够大时, P_E 可以为任意小 .

信道传递信息时,我们总希望信息传输率 R 尽可能大 .我们在证明中可以选择使 $I(X; Y)$达到信道容量的输入分布,于是可达条件 $R < I(X; Y)$成为 $R < C$.事实上,等概分布也确实是使对称信道的 $I(X; Y)$达到信道容量的输入分布 .

因此证得当 n 足够大时, $R < C$ 时,在整个码集合$\{C_k\}$上求 Y 的平均错误概率 P_E 可以达到任意小,因此至少存在一个码,其错误概率小于或等于平均值 P_E,定理得证 .

证毕

6.2.2 有噪信道编码逆定理

定理 6.3 设有一个离散无记忆平稳信道,其信道容量为 C .对于任意 $\varepsilon > 0$,若选用

码字总数 $M = 2^{n(C+\varepsilon)}$，则无论 n 取多大，也找不到一种编码，使译码错误概率 P_E 任意小．

证明

设选用 $M = 2^{n(C+\varepsilon)}$ 个码字组成一个码，不失一般性，认为码字为等概分布，即 $p(x_i) = 1/M$，$i = 1, 2, \cdots, M$．于是，信源熵 $H(X^n) = \log_2 M$．一般无记忆信道的扩展信道的平均互信息为

$$I(X^n; Y^n) = H(X^n) - H(X^n \mid Y^n) \leq nC \tag{6.50}$$

而

$$H(X^n) = \log_2 M = \log_2 2^{n(C+\varepsilon)} = n(C+\varepsilon) \tag{6.51}$$

$$H(X^n \mid Y^n) \geq n\varepsilon \tag{6.52}$$

根据费诺不等式

$$H(X^n \mid Y^n) \leq H(P_E) + P_E \log_2 (M-1) \tag{6.53}$$

$$H(P_E) = H(P_E, 1-P_E) \leq \log_2 2 = 1 \tag{6.54}$$

$$M - 1 < M = 2^{n(C+\varepsilon)} \tag{6.55}$$

$$n\varepsilon \leq H(X^n \mid Y^n) \leq 1 + n(C+\varepsilon) P_E \tag{6.56}$$

得

$$P_E \geq \frac{n\varepsilon - 1}{n(C+\varepsilon)} = \frac{\varepsilon - \frac{1}{n}}{C+\varepsilon} \tag{6.57}$$

当 n 增大时，P_E 不会趋于零．

这时信息传输率

$$R \geq \frac{H(S)}{\overline{L}} = \frac{\log_2 M}{n} = C + \varepsilon \tag{6.58}$$

即

$$R > C \tag{6.59}$$

因此，当 $R > C$ 时，平均错误概率 P_E 不可能趋于零．要想使 $R > C$ 而又无错误地传输消息是不可能的．

证毕

以上定理只是在离散无记忆信道的情况下证明的，但是对连续信道和有记忆信道同样成立．

6.2.3　错误概率的上界

离散无记忆信道中 P_E 趋于零的速度与 n 成指数关系，即当 $R < C$ 时，平均错误概率

$$P_E \leq \exp[-nE_r(R)] \tag{6.60}$$

式中，$E_r(R)$称为随机编码指数，又称为可靠性函数或加拉格(Gallaqer)函数．一般可靠性函数 $E_r(R)$与信息传输率 R 的关系曲线如图 6.5 所示，它是一条下凸函数曲线．在 $R<C$范围内 $E_r>0$，所以随 n 增大 P_E 以指数趋于零．实际编码的码长 n 不需选得很大．

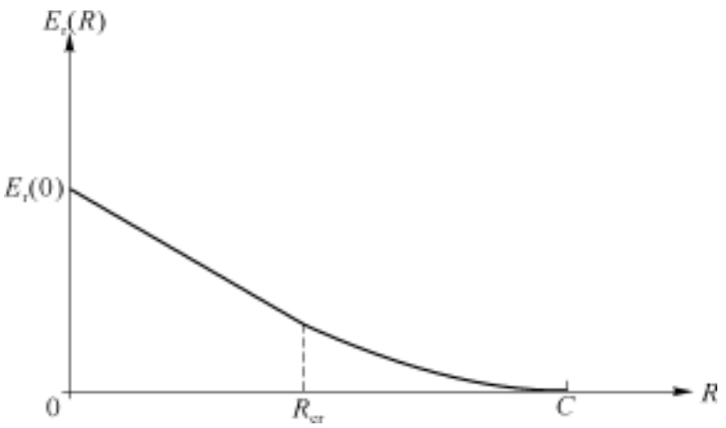


图 6.5 $E_r(R)$曲线图

可靠性函数 $E_r(R)$在信道编码中有极其重要的意义，它表示在码长 n 已定时，P_E 的上界．在实际问题中，为了达到一定的可靠性，要求 P_E 小于某个值(例如 10^{-6})，可靠性函数 $E_r(R)$可以帮助我们选择信息传输率 R 和编码长度 n．综合上述定理证明和论述可知，信道的信道容量是可靠传输的分界点：当 $R<C$ 时，P_E 以指数趋于零，当 $R>C$ 时，P_E 很快地趋于 1．因此，在任何信道中，C 是可达的最大的可靠信息传输率．

香农第二定理也只是一个存在定理，它说明错误概率趋于零的好码是存在的，但是没有说明如何构造这个好码．尽管如此，香农第二定理仍然具有重要的理论意义和实践指导作用，它可以指导各种通信系统的设计，有助于评价各种通信系统及编码效率．

从香农第一定理和香农第二定理可以看出，要做到有效和可靠地传输信息，可以将编码分成信源编码和信道编码两部分．首先，通过信源编码，用尽可能少的信道符号来表达信源，也就是对信源数据用最有效的表达方式表达，尽可能减少编码后数据的冗余度．然后对信源编码后的数据设计信道编码，也就是适当增加一些冗余度以纠正和克服信道中干扰引起的错误．这两部分是分别独立考虑的．

这种分两部分编码的方法在实际通信系统中有着重要的意义．近代大多数通信系统都是数字通信系统，与模拟通信系统相比有着许多优点．在实际数字通信系统中，信道常常是共用的数字信道(二元信道)，而无论语音、音乐、图像、数据都用同一通信信道来传输．因此，可以将语音、图像先数字化，再对数字化的语音、图像等信源进行不同的信源编码，针对各自信源的不同特点用不同的数据压缩方法．而对于共同的数字信道，输入端只是二元序列，所以信道编码只需针对信道特性进行，纠正信道中带来的错误，这样，可以大大降低通信系统设计的复杂度．可以证明，这种分两步编码处理的方法与一步编码处理方法是一样有效的．

6.3　纠错编码

自香农给出信道编码定理以来,引起了人们对信道编码的极大兴趣,但是香农只是证明了满足这种特性($R < C$ 时 $P_E \to 0$)的码的存在,还不能按其证明方法得到这种好码.证明过程是采用随机编码的方法,由于随机编码所得的码集很大,通过搜索得到好码的方法在实际上很难实现,而且即使找到其中的好码,这种码的码字也是毫无结构的,这意味着译码时只能用查表的方法,而在 N 很大时译码表所需的存储量也是很难被接受的,因此真正实用的信道编码还需通过各种数学工具来构造,使码具有很好的结构性以便于译码.抽象代数(也称近世代数)就是编码理论的最重要的数学工具,它包括群论、环论、域论、格论、线性代数等许多分支.

信道编码的目的是提高信号传输的可靠性,广义的信道编码还包括为特定信道设计的传输信号,如 NRZ(不归零)码、HDB3 码、伪随机序列码都属于信道编码,而纠错编码作为提高传输可靠性的最主要措施之一,是我们研究的主要内容.

6.3.1　纠错码分类

由于信道中干扰和噪声的存在,使得经信道传输后的接收码字与原来的发送码字存在着差异,也就是差错.一般信道中噪声干扰越大,码字产生错误的概率也越大.

信道中的干扰和噪声一般分为两种形式:一是随机噪声,它主要来源于设备的热噪声和散弹噪声以及传播媒介的热噪声,是通信系统中的主要噪声;二是脉冲干扰和信道衰落,它的特点是突发出现,主要来源于雷电、通电开关、负荷突变或设备故障等.

根据信道中干扰和噪声的形式,可将信道分为 3 类:随机信道、突发信道和混合信道.随机噪声产生的错误是独立随机出现的,称为随机错误.它的特点是各码元是否发生错误是随机的,且相互独立,因而不会出现成片的错误.只产生随机错误的信道称为随机信道,这是一类比较常见的信道.以高斯白噪声为主体的信道属于这类信道,比如卫星信道、同轴电缆信道、光缆信道以及大多数微波中继信道.

脉冲干扰和信道衰落产生的错误是成串出现的,错误之间具有相关性,这类错误称为突发错误.产生突发错误的信道称为突发信道.实际的短波信道、移动通信信道、由于物理损伤造成成串差错的光盘和磁盘,均为突发信道.

有些实际信道既有随机错误又有突发错误,称为混合信道.

对不同类型的信道要设计不同类型的信道编码才能收到良好效果.根据不同的信道类型设计的信道编码分为纠随机错误码、纠突发错误码和纠混合错误码.在通信系统中,纠检错的工作方式有反馈重传纠错、前向纠错和混合纠错等.

1．反馈重传纠错

图 6.6 即反馈重传纠错图示．

图 6.6 反馈重传纠错

发送端发出的是能够发现错误的检错码，接收端对接收到的码字进行译码，发现有错时，通过反馈系统向发送端请求重传已发送的全部或部分码字，直到接收端认为没有错误为止，我国的电报系统就是一种反馈重传纠错系统．

2．前向纠错

图 6.7 即前向纠错图示．

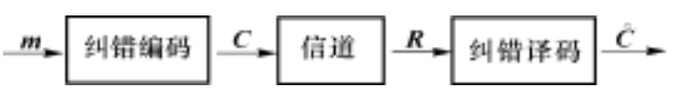


图 6.7 前向纠错

前向纠错也称为自动纠错．发送端发出的是具有纠错能力的纠错码，接收端根据编码规则进行解码．当误码个数在码的纠错能力范围之内时，译码器可以自动纠正错误．

3．混合纠错

对发送端进行适当的编码，当错误不严重时，在码的纠错能力之内，采用自动纠错，当产生的差错超出码的纠错能力时，则通过反馈系统向发送端要求重发，这种同时具有反馈重传纠错和自动纠错工作方式的纠错称为混合纠错．

检错码和纠错码在不加区别时统称为纠错码．纠错编码的目的是通过引入剩余度，也就是在传输的信息码元后增加一些多余的码元(称为校验元)，以使信息损失或发生传输错误后仍然能在接收端恢复．信息序列记为 m，纠错编码输出码序列记为 C，通过信道传输后接收码序列为 R，我们希望通过纠错以后恢复的码序列 $\hat{C}=C$．

编码实际上实现的是信息序列到纠错码序列的映射．在具体实现时，由于时延和计算复杂度的限制，只能将信息序列分组后按一定的映射关系变换成输出码序列．

纠错码的分类如下：

(1) 根据信息码元与校验码元之间是否存在线性关系可分为线性码和非线性码．线性码的校验码元是若干位信息码元的线性组合，而非线性码的校验位与信息位不满足线性关系．线性码具有良好的数学结构，编译码比较简单，性能优于同样纠错能力的非线性码．所以我们主要介绍线性码．

(2) 根据不同的分组及映射方式，纠错码可以分成分组码和树码．

① 分组码

把信息序列以每 k 个码元分组,然后把每组 k 个信息元按一定规律产生 r 个多余的校验元,输出码序列每组长为 $n = k + r$.每一码字的 r 个校验元只与本组的 k 个信息元有关,与别的组的信息位无关,记为分组码(n, k) .

本章主要介绍线性分组码,线性分组码中有一类特别重要的是循环码,它具有完整的代数结构,编译码都比较简单和易于实现 .

② 树码

信息序列以每 k_0 (通常较小)个码元分段,编码器输出该段的校验元 $r = n - k_0$ 不仅与本段的 k_0 个信息元有关,而且还与其前面若干段的信息元有关,称为树码或链码 .树码中最重要的一类是卷积码,它的校验元与信息之间是线性关系 .

纠错码按结构分类如下:

- 纠错码
  - 分组码
    - 线性分组码
      - 循环码
      - 非循环码
    - 非线性分组码
  - 树码
    - 线性树码
    - 非线性树码

目前的通信系统大多采用二进制的数字系统,所以以下提到的纠错码都是指二元码 .

6 .3 .2　纠错码的基本概念

信源编码把信源符号用二元序列来表示,这个二元序列称为信息序列,信源编码主要解决的是通信的有效性问题,即用尽可能少的码符号来表示信源符号或信源符号序列 .信道编码要解决的问题是通信的可靠性问题 .信息序列送到信道进行传输之前还需要经过信道编码变成具有纠检错能力的码序列 .通过在信息序列中插入冗余码元(称为校验元或监督元),使新序列的码元之间具有相关特性,然后再送入信道进行传输 .在接收端,信道译码器根据这个相关特性对接收序列进行译码,在纠错能力范围内可以对差错进行自动纠正,恢复原发送码序列 .

将信源编码器的输出序列进行分组,分组长度为 k,则可以有 $M = 2^k$ 个不同的分组信息序列 .每个分组信息序列用一个 n 长的码字来表示$(n > k)$, $C = C_1\ C_2 \cdots C_n$,这样的 2^k 个码字的集合称为二元分组码 .通常我们认为经过信源编码以后的码元符号已经是等概分布 .

对信道编码的一般要求是:

(1) 纠错检错能力强,可发现和纠正多个错误;

(2) 信息传输率高,信息传输率也称为码率: $R = \dfrac{\log_2 M}{n}$, M 为信息序列数,它表示每个码元符号所携带的信息量 .信息传输率越高,码长 n 应越短;

(3) 编码规律简单,实现设备简单且费用合理;

(4) 与信道的差错统计特性相匹配 .

信道编码就是综合考虑以上因素的情况下选择和设计合理的编译码实现方案 .

每个码字 $C = C_1\ C_2 \cdots C_n$ 中 k 位称为信息位,其余 $n - k$ 位为校验位或监督位 .信息序列的个数为 $M = 2^k$,而长为 n 的二元码一共有 2^n 个,选出其中的 M 个作为码字,称为许用序列,而其他序列为禁用序列 .

【例 6 .5】

下面给出一种编码如表 6 .4 所示 .

表 6 .4 (7,3)码

消息序列	码字	消息序列	码字
000	0000000	100	1001110
001	0011101	101	1010011
010	0100111	110	1101001
011	0111010	111	1110100

上例中信息位为 $k = 3$,码长为 $n = 7$,监督位为 $r = 4$,用 $\eta = k/n$ 表示码字中信息位所占的比重,称编码效率,它表明了信道的利用率 .η 越大,编码效率越高,它是衡量码性能的一个重要参数 .上例中 $\eta = \frac{3}{7} \approx 43\%$,而 $R = \frac{\log_2 M}{n} = \frac{\log_2 2^k}{n} = \frac{k}{n} = \eta$.

如果 n 长码字的每一位与原始信息序列 $d = d_1\ d_2 \cdots d_k$ 的 k 个信息位之间的函数关系 $c_i = f_i(d_1, d_2, \cdots, d_k)$, $i = 1, 2, \cdots, n$ 是线性关系,则称该分组码为线性分组码,否则称为非线性分组码 .

若 (n, k) 分组码字中 k 个信息位与原始信息序列的 k 个信息位相同,且位于 n 长码字的前(或后) k 位,而校验位位于其后(或前) $n - k$ 位,则该分组码为系统码,否则为非系统码 .

在 6 .1 .2 节我们介绍了汉明距离和码的最小距离 .码的最小汉明距离是衡量码的纠检错能力的重要参数,码的最小距离越大,其纠检错能力越强 .

对于某一码字,其非零元素的个数称为该码字的汉明重量 .

对于二元码,其码字的重量是码字中 1 的个数 .若码字为 $c_i = c_{i_1}\ c_{i_2} \cdots c_{i_n}$,则其重量可以表示为 $W(c_i) = \sum_{k=1}^{n} c_{i_k}$.例如,码字 $c_1 = 1010101$,其重量为 4 .

*6 .4 几种重要的纠错码

6 .4 .1 线性分组码

线性分组码是最有实用价值的一类码 .比如汉明码、Golay 码、RS 码、BCH 码等都属

于线性分组码 .

线性分组码的编码方式是将信源输出序列分组,每组是长为 k 的信息序列,然后按照一定的编码规则插入 $n-k$ 位的校验位,校验位是信息位的线性组合,组成 n 长的码元序列 .

1 . 校验矩阵与生成矩阵

线性分组码可以由一组信息元的模 2 线性方程生成 .

例如,一个 $(7,3)$ 线性分组码, $C = C_1\ C_2\ C_3\ C_4\ C_5\ C_6\ C_7$, 其中 $C_1\ C_2\ C_3$ 为信息元, $C_4\ C_5\ C_6\ C_7$ 为校验元,假设校验元可用下面的方程组得到:

$$\begin{cases} C_4 = C_1 + C_3 \\ C_5 = C_1 + C_2 + C_3 \\ C_6 = C_1 + C_2 \\ C_7 = C_2 + C_3 \end{cases} \tag{6.61}$$

这里的“ + ”为模 2 加 这是一组线性方程,它确定了由信息元得到校验元的规则,所以称为校验方程或监督方程 .

方程组还可以写成矩阵形式:

首先将方程组改写一下,使方程的等号右边为 0 .

$$\begin{cases} C_1 + C_3 + C_4 = 0 \\ C_1 + C_2 + C_3 + C_5 = 0 \\ C_1 + C_2 + C_6 = 0 \\ C_2 + C_3 + C_7 = 0 \end{cases} \tag{6.62}$$

然后,写成矩阵相乘的形式:

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.63}$$

令

$$H = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.64}$$

则式(6.62)可写成 $HC^{\mathrm{T}}=0$ 或 $CH^{\mathrm{T}}=0$.

式中 H 称为一致校验矩阵.一旦建立了校验矩阵,校验元与信息元的关系就确定了,码字也随之确定.

校验方程(6.61),还可以改写成:

$$\begin{cases} C_1 = C_1 \\ C_2 = C_2 \\ C_3 = C_3 \\ C_4 = C_1 + C_3 \\ C_5 = C_1 + C_2 + C_3 \\ C_6 = C_1 + C_2 \\ C_7 = C_2 + C_3 \end{cases} \tag{6.65}$$

令

$$m = [\, C_1 \; C_2 \; C_3 \,] \tag{6.66}$$

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \tag{6.67}$$

则式(6.65)可写成 $C = mG$.

对于 k 位信息位 n 长码元序列的线性方程组,可有下面关系式存在:

$$C = mG \tag{6.68}$$

其中,C 为 n 维行向量,m 为 k 维行向量,G 为 $k\times n$ 矩阵,称为线性分组码 C 的生成矩阵.

利用生成矩阵可将信息序列 m 变成码字序列 C.例如当 $m=(0\quad 1\quad 1)$时,可以得到

$$\begin{aligned} C &= mG \\ &= (0\quad 1\quad 1)\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \\ &= (0\quad 1\quad 1\quad 1\quad 0\quad 1\quad 0) \end{aligned}$$

为方便起见,可以将生成矩阵写成:

$$G = \begin{bmatrix} G_1 \\ G_2 \\ \cdots \\ G_k \end{bmatrix} \tag{6.69}$$

式中,G_i,$i=1,2,\cdots,k$ 为 n 维行向量 $G_i=(g_{i_1}, g_{i_2}, \cdots, g_{i_n})$.

信息序列 $m=(m_1, m_2, \cdots, m_k)$则 $C=mG=\sum_{i=1}^{k} m_i G_i$.

式(6 .68)表明,码字 C 为信息序列 m 和生成矩阵 G 的行向量的线性组合,和为“模 2 加”.当信息组 m 中只有一个非零元素时,码字为生成矩阵的某一行,因此生成矩阵的每一行都是一个码字,k 个不相同的码字可以构成码的生成矩阵 .而由这 k 个码字的不同线性组合便生成了整个码组 .所选取的 k 个码字必须是线性无关的,也就是生成矩阵的秩为 k,才能由生成矩阵组合出 2^k 种不同的码字,这样生成的全部码字组成了 n 维矢量空间的一个 k 维子空间,矩阵 G 的 k 个行向量是这一子空间的 k 个基矢量,而 k 维子空间的任意 k 个线性无关的矢量都可以作为这一子空间的基矢量 .这些基矢量组成的矩阵都可以看作这一子空间或这一码的生成矩阵,同一码的所有可能的生成矩阵都是等价的 .

在本例中,G 可以写成分块矩阵 .即 $G=(I \quad P)$.式中 I 为 $k\times k$ 的单位矩阵,P 为 $k\times(n-k)$的一般矩阵 .这样生成的码 C 是系统码,也就是信息位在码字的前 k 位或后 k 位 .当生成矩阵不能写成由 k 阶单位矩阵构成的分块矩阵时,生成的码 C 不是系统码 .但根据矩阵理论,可以将一般形式的矩阵通过行初等变换转换成标准形式的矩阵,两个矩阵是等价的,而这样产生的码与系统码是等价的,每一个线性码对应唯一一个标准形式的生成矩阵,因此以系统码为研究对象不失一般性 .

一般地,构造一个(n, k)线性分组码,只需找出一个秩为 $n-k$ 的 $n-k$ 行、n 列矩阵 H,则可由齐次线性方程组 $HC^{\mathrm{T}}=0^{\mathrm{T}}$ 的解空间的全部向量作为许用码字,得到一个(n, k)线性分组码 .因此线性分组码可以用齐次线性方程组这样方便的数学工具来研究 .

由于标准形式的生成矩阵有 k 阶单位子阵,因此组成标准形式的生成矩阵 k 个码字是线性无关的 .化成标准形式之后,容易验证生成矩阵的各行是否线性无关 .

由于生成矩阵 G 的每一行都是一个码字,所以生成矩阵和校验矩阵有如下关系:

$$HG^{\mathrm{T}}=0 \text{ 或 } GH^{\mathrm{T}}=0 \tag{6.70}$$

即线性分组码的生成矩阵和校验矩阵的行矢量彼此正交 .以上结果表明,线性分组码既可以由生成矩阵确定,也可以由校验矩阵确定 .(n, k)线性分组码是 n 维向量构成的线性空间中的一个 k 维线性子空间,它可以由 G 或 H 确定 .同时 H 矩阵的行矢量在 n 维矢量空间中张成一个 $n-k$ 维子空间,这两个子空间的矢量是互相垂直,$n-k$ 维子空间也对应一个线性码 .此码与 G 生成的码互为对偶码 .G 为 $n-k$ 维子空间的校验矩阵 .

标准形式的校验矩阵可以写成 $H=(Q \quad I)$的形式,式中 Q 为$(n-k)\times k$ 矩阵,I 为 $n-k$维单位矩阵 .而标准形式的生成矩阵可以写成 $G=(I \quad P)$.

所以

$$HG^{\mathrm{T}}=(Q \quad I)\begin{pmatrix} I \\ P^{\mathrm{T}} \end{pmatrix}=Q+P^{\mathrm{T}}=0 \tag{6.71}$$

可得

$$P^{\mathrm{T}}=Q \text{ 或 } P=Q^{\mathrm{T}} \tag{6.72}$$

因此标准形式的校验矩阵和生成矩阵可以很方便地实现转换 .

线性分组码的性质是:

(1) 码中任意两个码字之和仍为一码字;

(2) 任意码字是 G 的行向量 G_1, G_2,…, G_k 的线性组合;

(3) 零向量 $0=(0,0,\cdots,0)$是一个码字,称为零码字;

(4) 线性分组码的最小距离等于非零码字的最小重量 .

线性分组码最重要的性质是其线性特性以及在此基础上的对称性 .所谓线性特性是指线性码中任意两个码字的和或差仍为一码字,对称性是指在一个码的所有码字上减去一个特定的码字,结果仍是同一码的全部码字 .这样在求码字间的距离分布只需求出任一码字与其他所有码字的距离分布就可以了,因为所有码字的距离分布都相同,又因为任两个码字的差仍为一码字,所以边距分布相同 .

【例 6 .6】

3 重复码是一个$(3,1)$线性分组码,其生成矩阵为

$$G=(1\quad 1\quad 1)$$

$$C=C_1\,C_2\,C_3=(m_1)(1\quad 1\quad 1)=(m_1\quad m_1\quad m_1)$$

【例 6 .7】

$(4,3)$偶校验码是一个$(4,3)$线性分组码,其生成矩阵为

$$G=\begin{pmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{pmatrix}$$

$$C=C_1\,C_2\,C_3\,C_4$$

$$=(m_1\quad m_2\quad m_3)\begin{pmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{pmatrix}$$

$$=(m_1\quad m_2\quad m_3\quad m_1+m_2+m_3)$$

【例 6 .8】

已知生成矩阵为

$$G=\begin{pmatrix}1&0&0&1&1&1&0\\0&1&0&0&1&1&1\\0&0&1&1&1&0&1\end{pmatrix}$$

求生成的线性分组码及由 H 生成的线性分组码 .

解

由于 $G=(I\quad P)$,则有

$$P=\begin{pmatrix}1&1&1&0\\0&1&1&1\\1&1&0&1\end{pmatrix}$$

又因为 $Q = P^{T}$,则

$$H = (Q \quad I) = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

由生成矩阵 G 生成的(7,3)码为

m	C	m	C
000	0000000	100	1001110
001	0011101	101	1010011
010	0100111	110	1101001
011	0111010	111	1110100

把校验矩阵 H 当作生成矩阵,可生成(7,4)码 .

m	C'	m	C'
0000	0000000	1000	1011000
0001	0110001	1001	1101001
0010	1100010	1010	0111010
0011	1010011	1011	0001011
0100	1110100	1100	0101100
0101	1000101	1101	0001101
0110	0010110	1110	1001110
0111	0100111	1111	1111111

这样,生成的 C 和 C'是互相正交的 .

2 . 线性分组码的纠检错能力

由生成矩阵产生的码字在信道传输过程中,由于干扰的存在,使得一些码元发生错误 接收端通过编码规则进行译码,如能发现错误,则称为检错,如果再能纠正错误,称为纠错 .码能纠检错误码元的个数称为该码的纠检错能力 .发现错误和纠正错误的个数越多,则说明该码的纠检错能力越强 .只要接收码字 R 没有错到变成其他发送码字,就可以

发现错误.因此设计的码字之间应有较大的区别,即它们的汉明距离要大.

在第 6.1.2 节中知道,极大似然译码规则与最小汉明距离译码是等价的,而当输入为等概时,最小汉明距离译码与最大后验概率译码也是等价的.

定理 6.4 对于一个二进制对称信道,当输入为 2^k 个等可能的 n 长码字时,最大后验概率准则等效于最小汉明距离译码准则.

证明

对于二元对称信道,设单个码元的错误概率为 $p(p<0.5)$,正确概率为 $1-p$.当发送码字为 x,接收码字为 y 时,其汉明距离为 $d(x,y)$,即有 $d(x,y)$个码元发生错误.而 $n-d(x,y)$个码元正确传送.

转移概率 $p(y|x)=p^{d(x,y)}(1-p)^{n-d(x,y)}$.根据最大后验概率译码规则,选择译码函数 $F(y_j)=x^*$ 使

$$p(x^*|y_j)\geq p(x_i|y_j) \quad \forall i$$

利用贝叶斯公式,得

$$\frac{p(y_j|x^*)\,p(x^*)}{p(y_j)}\geq\frac{p(y_j|x_i)\,p(x_i)}{p(y_j)}$$

当输入等概率时,上式等价于

$$p(y_j|x^*)\geq p(y_j|x_i)$$

令 $\lambda=\dfrac{p(y_j|x^*)}{p(y_j|x_i)}$,对于无记忆信道,由于

$$\lambda=\frac{p^{d(x^*,y_j)}(1-p)^{n-d(x^*,y_j)}}{p^{d(x_i,y_j)}(1-p)^{n-d(x_i,y_j)}}=\left(\frac{p}{1-p}\right)^{d(x^*,y_j)-d(x_i,y_j)}$$

而 $p<0.5$,$\dfrac{p}{1-p}<1$,所以,当 $d(x^*,y_j)<d(x_i,y_j)$ 时,$\lambda>1$,因此,$p(y_j|x^*)>p(y_j|x_i)$等价于 $\lambda>1$ 等价于 $d(x^*,y_j)<d(x_i,y_j)$,即最大后验概率准则等效于最小汉明距离译码准则.

证毕

当输入等概时,最大后验概率译码等价于最小汉明距离译码,译码时把接收码字译成与它距离最小的发送码字,而编码时则应使发送码字之间的距离尽可能大.定义码中任意两个码字的汉明距离的最小值为码的最小距离,应选择最小汉明距离尽可能大的码作为发送码.线性分组码的最小汉明距离与码中码字的重量有如下关系:

定理 6.5 线性分组码的最小距离等于非零码字的最小重量.

证明

根据线性分组码的封闭性可知,任意两个码字的和仍为码中的码字.根据码字之间的距离的定义,两个码字和的非零个数即为它们之间的距离,而两个码字和的非零个数又是新码字的重量.所以,线性分组码的最小距离必为它的非零码字的最小重量.

证毕

码的最小距离越大,即码中任意两个码字之间的差别越大 .越不容易把一个码字误传成其他的码字,译码时也容易正确译码,因此码的纠检错能力越大 .即码的最小距离和最小重量决定了码的纠错能力 .关于码的最小距离与纠检错能力的关系有以下结论:

推论 6 .6　对于一个(n, k)线性分组码,设 $d_{\min}$为其最小汉明距离,则

(1) 这组码能纠正 u 个错误的充要条件是 $d_{\min} = 2u + 1$;

(2) 检测 l 个错误的充要条件是 $d_{\min} = l + 1$;

(3) 能纠正 u 个错误,同时可以发现 $l(l > u)$个错误的充分必要条件为 $d_{\min} = u + l + 1$.

证明

(1) 分组码中任意两个码字之间的汉明距离 $d(X_i, X_j) \geq d_{\min} = 2u + 1$.

若信道输入码字 X_i,在信道传输过程中发生小于等于 u 位的差错,信道输出端接收序列为 Y,则有 $d(X_i, Y) \leq u$.任一其他码字 X_j,有

$$d(X_i, Y) + d(Y, X_j) \geq d(X_i, X_j)$$

$$d(Y, X_j) \geq d(X_i, X_j) - d(X_i, Y) \geq 2u + 1 - u = u + 1$$

若接收码字 Y 与信道输入码字 X_i 的汉明距离 $d(X_i, Y) \leq u$,则与其他码字的汉明距离 $d(Y, X_j) \geq u + 1$,按最小汉明距离译码准则,它必定被正确地译成输入码字 X_i .因此具有 $d_{\min} = 2u + 1$ 的分组码具有纠正 u 个差错的能力 .可作一直观的解释 .两个码字 X_i, X_j 的距离不小于 $2u + 1$.凡由码字发生小于等于 u 位差错而变成的接收序列 Y 均落在以 X_i 为球心, u 为半径的闭球体内 .故 Y 与 X_i 的距离不大于 u,而 Y 与任一其他码字 X_j 的汉明距离 $d(Y_1, X_j) \geq u + 1$, Y 必定正确地被译成 X_i,不会被译为其他码字 .同理,由码字 X_j 发生小于等于 u 个差错而变成的接收序列可以被正确的译成 X_j .由于 $d(X_i, X_j) \geq 2u + 1$,两个球体不可能相交,所以不可能出现接收序列 Y 既在以 X_i 为球心、u 为半径的闭球体内,又在以 X_j 为球心、u 为半径的闭球体内 .凡是小于等于 u 位差错的接收序列都可以唯一正确地译为相应的发送码字 .

(2) 任意两个码字 $d(X_i, X_j) \geq d_{\min} = l + 1$.当发送码字 X_i 在传输发生小于等于 l 位差错时,接收序列为 Y,则有 $d(X_i, Y) \leq l$,因此 Y 不可能变成其他的码字 .当发生的错误位数等于 $l + 1$ 时则可能变成其他的码字,因此 $d_{\min} = l + 1$ 的分组码能检测 l 个错误 .而当 $d(X_i, Y) \leq l$ 时,也不能译为 X_i,因为不能判断是否 $d(X_j, Y) > d(X_i, Y)$,假设 $d(X_i, Y) = l$,如果要正确译码为 X_i,根据最小距离译码准则,则需 $d(X_j, Y) > d(X_i, Y) = l$,而 $d(X_j, Y) \geq d(X_i, X_j) - d(X_i, Y) \geq d(X_i, X_j) - l > l$,即要求 $d(X_i, X_j) > 2l$,这与已知条件 $d(X_i, X_j) \geq d_{\min} = l + 1$ 矛盾 .

所以 $d_{\min} = l + 1$ 的分组码只能发现 l 个错误而不能纠正它们 .

(3) 这是一种纠检结合的工作方式 .对于错码个数较少也是出现较频繁的码字,按前向纠错方式工作,以节省反馈重传的时间 .而对错码较多的码字,在超过该码的纠错能力

后能自动按检错重传方式工作.若接收码字与某一许用码组的距离在纠错能力 u 范围内,按纠错方式工作;如果与任何许用码字的距离都超过 u,则按检错方式工作.

对于任意码字 $d(X_i, X_j) \geq d_{\min} = u + l + 1$,其中 $l > u$,由前面的讨论可知,该码具有纠正小于等于 u 个差错,发现小于等于 l 个差错的能力.

码字 X_i 发生小于等于 u 个差错的接收序列 R_1 落在以 X_i 为球心、u 为半径的闭球体内.码字 X_j 发生小于等于 u 个差错的接收序列 R_2 落在以 X_j 为球心、u 为半径的闭球体内.由于 $d_{\min} = u + l + 1$,且 $l > u$,所以这两个球不会相交,按照最小汉明距离译码准则,R_1、R_2 必然会被正确译码,即该码具有纠正小于等于 u 个差错的能力.而 X_i 发生大于 u 且小于等于 l 个差错的接收码字 R_1' 和 X_j 发生大于 u 且小于等于 l 个差错的接收码字 R_2',则分别落在以 X_i, X_j 为球心以 l 为半径的闭球体内.这两个球可能相交,所以 R_1' 和 R_2' 都能既在以 X_i 为球心,l 为半径的闭球内,又在以 X_j 为球心,l 为半径的闭球内,都不能保证与发送码字的距离最小,所以不能正确译码,因此不能按纠错方式工作.但是因为还没有落入另一个码字的纠错范围,所以可以发现 l 个错误而不能纠正.

证毕

例如,$d_{\min} = 5$ 时,按检错方式工作时 $l = 4$;按纠错方式工作时 $u = 2$;按纠检结合方式工作时,若设计 $u = 1$,则 $l = 3$,当发送码字 X_i 发生 1 位或 2 位错时,可纠;发生 3 位错时,可发现;发生 4 个错码时,将落入另一个发送码 X_j 的纠错范围内,被错纠为 X_j 了.

【例 6.9】 重复码,(3,1)分组码 $k = 1$, $n = 3$.

m	C	R
0	000	000
		001
		010
		100
1	111	011
		101
		110
		111

按照最小距离译码,$d_{\min} = 3$ 可纠一个错误.

以上定理说明了码字间的距离与纠错能力的关系.选择编码方式时选择极大最小距离码可以得到较高的纠检错能力.

推论 6.7 码的纠错能力 u 与码字的长度 n 和消息数 M 也满足一定的关系:

$$M \sum_{i=0}^{u} C_n^i \leq 2^n \tag{6.73}$$

证明

n 长码字的总数为 2^n,即接收序列会有 2^n 种可能 .

当发送某一 n 长码字时,若错误码元小于或等于 u,则接收码字有

$$C_n^0 + C_n^1 + \cdots + C_n^u = \sum_{i=0}^{u} C_n^i$$

种可能 .

发送的消息数为 M,即发送 M 个码字时,接收码字共有 $M\sum_{i=0}^{u} C_n^i$ 种可能 .因为这些接收码字不能有重复,所以必有 $M\sum_{i=0}^{u} C_n^i \leq 2^n$.

证毕

若 n 长码字中信息位为 k,则 $M = 2^k$,所以有 $\sum_{i=0}^{u} C_n^i \leq 2^{n-k}$.

当上式取等号时,即$(n-k)$个监督位正好全部用来提示错误数小于等于 u 的码字时称为完备码 .后面要介绍的汉明码就是完备码 .上式是具有纠正小于或等于 u 位错误能力码的必要条件而不是充分条件,即不是所有消息数 m 满足该不等式的码都具有这样的纠错能力 .

3 . 校验矩阵与最小距离的关系

校验矩阵 H 除了用来校验码字外,还与码的最小距离进而与码的纠错能力发生一定的关系 .

定理 6 .8　对于(n, k)线性分组码:校验矩阵 H 中的任意 t 列线性无关而 $t+1$ 列线性相关,则码的最小距离 $d_{\min}$或码字的最小重量为 $t+1$.反之,若码字的最小重量或码的最小距离为 $t+1$ 则 H 的任意 t 列线性无关而 $t+1$ 列线性相关 .

证明

把 H 写成列向量的形式 $H = (H_1\ H_2\ \cdots\ H_n)$,其中 H_j 为列向量 .

对于任意的码字 C,有

$$\begin{aligned} HC^{\mathrm{T}} &= (H_1\ H_2 \cdots\ H_n)(C_{i_1} C_{i_2} \cdots C_{i_n})^{\mathrm{T}} \\ &= C_{i_1} H_1 + C_{i_2} H_2 + \cdots + C_{i_n} H_n \\ &= 0 \end{aligned}$$

可见 HC^{T} 代表了 n 个$(n-k)$维列向量 H_j 的线性组合,并且由 $HC^{\mathrm{T}} = 0$, n 个向量 H_j 是线性相关的 .

若 H 中任意 t 列线性无关而 $t+1$ 列线性相关,则说明 C_{i_1}, C_{i_2}, $\cdots$, C_{i_n} 这 n 个码符号中必有 $t+1$ 个1,其余为0,使得 $t+1$ 个向量 H_j 的线性组合等于0,即 C_i 的码重为 $t+1$.而且 C_i 的码重不可能为 t,否则可使 t 个向量 H_j 的线性组合等于0,即 H 中有 t 列线性相关 .

反过来,如果码字的最小重量为 $t+1$,则重量最小的码字有 $t+1$ 个非零码元.代入 $HC^{\mathrm{T}}=0$,有 $t+1$ 个 H_j 的线型组合等于0,则说明必有 $t+1$ 个向量线性相关,并且任意 t 个向量必定线性无关,因为如果有 t 个 H 的列向量线性相关,必然存在一个重量为 t 的码字,这与最小码重为 $t+1$ 是矛盾的.

证毕

由于 H 是 $(n-k)\times k$ 矩阵,其秩至多为 $(n-k)$,即最多有 $(n-k)$ 个列向量线性无关.在寻找好码时,$d_{\min}$ 越大越好,即希望 H 中线性无关的列向量越多越好,而线性无关的列向量最多为 $(n-k)$ 个,所以 $d_{\min}\leq n-k+1$.

如果设计的 (n,k) 线性分组码达到了 $d_{\min}=n-k+1$,则称为极大最小距离码.

由于码的最小汉明距离与纠错能力又有以下关系:

$$d_{\min}=2u+1$$

即如果线性分组码能纠正 u 个错误,则 H 中必有 $2u$ 个列向量线性无关,而 $2u+1$ 个列向量线性相关.

* 4.线性分组码的伴随式及伴随式译码

由于 $CH^{\mathrm{T}}=0$,校验矩阵可以用来验证接收码字是否为许用码字.设发送码字为 C,接收码字为 R,令 $S=RH^{\mathrm{T}}$.当 R 为许用码字时满足校验方程 $CH^{\mathrm{T}}=0$,因此 $S=0$.若 $S\neq 0$,则说明 R 不是发送码字,码字在传输过程中产生了错误.所以 S 是码字在传输过程中是否出现错误的标志,称为伴随式(或称监督子、校验子等).

设接收码字 R 是发送码字 C 在传输过程中产生差错得到的,可以将 R 写成 $R=C+E$,$E=(e_1\,e_2\cdots e_n)$ 称为差错图样.当码字的第 i 位发生错误时 $e_i=1$,否则 $e_i=0$.

所以伴随式仅与错误图样有关,与码字无关,即伴随式中仅含有错误图样信息,$S=0$ 表示传输中要么无差错发生,要么错误图样恰好为一个码字,而错误图样恰好为一个码字的机率是很小的.因此可以通过伴随式得到错误图样信息,然后对接收码字进行修正,以得到正确的译码.

伴随式 S 是伴随接收码字 R 的一个 $(n-k)$ 维向量,但是从 $S=RH^{\mathrm{T}}$ 可以看出,S 并不反映发送的码字是什么,而只是反映信道对码字造成“怎样的干扰”.差错图样 E 是 n 维矢量,共有 2^n 种可能的组合,而伴随式 S 是 $(n-k)$ 维矢量,只有 2^{n-k} 种可能的组合,因此不同的差错图样可能有相同的伴随式.

在接收端,我们并不知道发送码字 C 是什么,但可以知道 H 和 R 是什么.通过伴随式译码找到 C 的估值,其过程是:$S=RH^{\mathrm{T}}=EH^{\mathrm{T}}$ E $C=R+E$,即先算出 S,再由 S 算出 E,最后令 $C=R+E$,求出 C.这里关键是如何从 S 算出 E.

$S^{\mathrm{T}}=HE^{\mathrm{T}}=\sum_{i=1}^{n}e_iH_i$,$H_i$ 为 H 的列向量.伴随式是接收码字中发生错误的码元在 H 中对应列的矢量和.由于同一个伴随式会对应多个差错图样,所以根据最小距离译码规则,求到 S 后应该译成所有对应 E 中重量最小的图样.伴随式有 2^{n-k} 个,其中1个对应没

有差错的图样，n 个对应 n 个码元中有一个发生错误的图样，C_n^2 个对应 n 个码元中有两个发生差错的图样，C_n^3 个对应 n 个码元中有三个发生差错的图样，C_n^u 个对应 n 个码元中有 u 个发生差错的图样，直到把 2^{n-k} 个伴随式用尽．$2^{n-k} \geq \sum_{i=0}^{u} C_n^i$，即该线性分组码可以纠正 u 个差错码元．

把 S 和 E 以及 R、C 的对应关系列成一个表，称为标准阵列．根据标准阵列译码通过查表可以很快得到发送码字 C．

下面举一个例子来说明这个译码过程．

【例 5.10】

某(5, 2)系统线性码的生成矩阵是 $G=\begin{bmatrix}1&0&1&1&1\\0&1&1&0&1\end{bmatrix}$，设接收码字是 $R=(10101)$．求发送码字的估值 C．

解

(1) 把信息组 $m=(00),(01),(10),(11)$ 及已知的 G，代入 $C=mG$，求出 4 个许可的码字：

$$C_1=(00000)\quad C_2=(10111)\quad C_3=(01101)\quad C_4=(11010)$$

(2) 由 $G\rightarrow H$

$$H=(P^{\mathrm{T}}\quad I)=\begin{bmatrix}1&1&1&0&0\\1&0&0&1&0\\1&1&0&0&1\end{bmatrix}=[H_1\quad H_2\quad H_3\quad H_4\quad H_5]$$

(3) 求标准阵列

因为

$$S^{\mathrm{T}}=HE^{\mathrm{T}}=\sum_{i=1}^{n} e_i H_i \tag{6.74}$$

所以，无差错时 $S=0$．当只有一个码元发生差错时，S^{T} 等于对应的 H 列；当两个以上码元发生差错时，S^{T} 等于对应列的矢量和．

由于 $n-k=3$，所以伴随式共有 $2^3=8$ 种，而错误图样中表示无差错的有一种，表示一个差错的图样有 $C_5^1=5$ 种．把 8 个伴随式对应这 6 个重量最小的错误图样还多出 2 个伴随式，让它们对应两个差错为 2 的图样．而差错为 2 的图样共有 $C_5^2=10$ 种，在其中选出两个，选的方法可有多种，不是唯一的．

将

$$E=(00000),(10000),(01000),(00100),(00010),(00001)$$

代入式 $S=EH^{\mathrm{T}}$，求出对应的 S 为

$$(000),(111),(101),(100),(010),(001)$$

剩下的伴随式中(011)对应的图样有 2^k 个(n 个未知数, $n-k$ 个方程, $S^{\mathrm{T}}=\sum_{i=1}^{n} e_i H_i$), 即(00011), (10100), (01110), (11001), 其中(00011)和(10100)并列重量最轻, 任选其中一个比如(00011) .同样伴随式(110)选一个对应的重量最轻的差错图案(00110) .

根据以上讨论, 画出标准阵列如下 .

S	R / C / E	$C_1=00000$	$C_2=10111$	$C_3=01101$	$C_4=11010$
000	00000	00000	10111	01101	11010
111	10000	10000	00111	10110	01010
101	01000	01000	11111	00101	10010
100	00100	00100	10011	01001	11110
010	00010	00010	10101	01111	11000
001	00001	00001	10110	01100	11011
011	00011	00011	10100	01110	11001
110	00110	00110	10001	01011	11100

(4) 由 $R\rightarrow\hat{C}$, 例如, 若 $S=RH^{\mathrm{T}}=(010)$, 查出对应的差错图案为 $E=(00010)$, 所以

$$C=R+E=(10101)+(00010)=(10111)$$

进一步分析可知, 该码的 $d_{\min}=3$, 所以纠错能力 $u=1$.因此译码阵列中只有前 6 行具有唯一性、可靠性 .而第 7、8 行的错误图样表示有 2 个差错, 已超出了 $u=1$ 的纠错能力, 译码已不可靠 .第 7、8 行在选择错误图样时有多种选法, 选法不同, 造成最终的译码结果不同 .

由于表示单个错误的错误图样的伴随式等于 H 矩阵中错误码元的对应列, 所以为了使不同错误码元对应不同的伴随式以便译码, 就应该使 H 中的 n 列互不相同且不能为 0 (若某列为 0 就不能表示对应码元出现错误的情况) .伴随式的个数 2^{n-k} 与 n、k 及纠错能力 u 之间满足以下的关系 .

定理 5 .9 若(n, k)线性分组码能够纠正 u 个错误, 则其校验位的数目必须满足

$$2^{n-k}\geq\sum_{i=0}^{u}C_n^i \tag{6 .75}$$

证明

由于产生 $i(i\leq n)$ 个错误的错误图样有 C_n^i 种, 能够产生不多于 u 个错误的错误图样总共有 $\sum_{i=0}^{u}C_n^i$ 个 .而 $n-k$ 位校验元共有 2^{n-k} 种不同的组合, 即有 2^{n-k} 个伴随式, 如果能够纠正 u 个错误, 则 $\sum_{i=0}^{u}C_n^i$ 个错误图样都有一个对应的伴随式, 即伴随式的数目满

足条件 $2^{n-k} \geq \sum_{i=0}^{u} C_n^i$.

证毕

式(6 .75)等号成立时的线性分组码称为完备码 .即完备码的伴随式数目不多不少恰好等于不大于 u 个差错的错误图样的数目, 这时校验位得到最充分的利用,所有的伴随式都唯一可靠地对应其中的一个错误图样 .

从多维矢量空间的角度来看完备码 .假定我们围绕每一个码字 C_i,作一个半径为 u 的球,每个球内包含了与该码字的汉明距离小于等于 u 的所有接收码字 R 的集合,所有落在此闭球内的接收码字都被译码为该发送码字 这样在这个半径为 $u = \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor$ 的球内的接收码字的个数为 $\sum_{i=0}^{u} C_n^i$.因为有 2^k 个可能的发送码字,也就有 2^k 个不相重叠的半径为 u 的球 .因为包含在这 2^k 个球中的码字总数不会超过 2^n 个可能的接收码字,所以一个纠 u 个差错的码必然满足不等式

$$2^k \cdot \sum_{i=0}^{u} C_n^i \leq 2^2 \tag{6 .76}$$

即

$$2^{n-k} \geq \sum_{i=0}^{u} C_n^i \tag{6 .77}$$

当 $u = 1$ 时,有 $2^{n-k} \geq C_n^0 + C_n^1 = 1 + n$.

当式(6 .76)等号成立时,表示所有的接收码字都落在 2^k 个球内而球外没有一个码字,这就是完备码 .完备码具有下述特性:

每一个接收码字都落在这些球中之一,因此接收码字与发送码字的距离至多为 u;

所有差错数小于等于 u 的接收码字都能得到纠正;

差错数大于等于 $u + 1$ 的接收码字,因为落在另一个球内被纠正为其他的发送码字 .

完备码并不多见,我们知道的有 $u = 1$ 的汉明码、$u = 3$ 的高莱码,以及 $(n, 1)$ 中 n 为奇数的重复码等 .

6 .4 .2　汉明码

汉明码是指能够纠正一个错误的线性分组码 .它是香农的信道编码定理提出后最早发现的码,有二进制的,也有非二进制的 这里讨论二进制的汉明码 .

汉明码是能够纠正一个错误的线性分组码,因此码长 n 和信息位数 k 服从以下规律:

$$(n, k) = (2^m - 1, 2^m - m - 1) \tag{6.78}$$

其中，$m = n - k$ 为校验位数．

当 $m = 3, 4, 5, 6, 7, 8, \cdots$ 时，有 $(7, 4)$，$(15, 11)$，$(31, 26)$，$(63, 57)$，$(127, 120)$，$(255, 247)$，…汉明码．一个 (n, k) 汉明码的校验矩阵有 $(n - k)$ 行和 n 列．二进制时，$(n - k)$ 个码元所能组成的列矢量总数(全零矢量除外)是 $2^{n-k} - 1$，恰好和校验矩阵的列数 $n = 2^m - 1$ 相等，只要把 $(n - k)$ 个码元组成的列矢量按二进制数大小顺序从左到右排列，就可以得到它的 H．这样得到的 H 是非系统码．当发生单个错误时，伴随式是 H 中与错误位置对应的列，并且伴随式的二进制数的值就是错误位置的序号，因此它的编译码规则很简单．

如果想得到系统形式的 H，可以通过列置换把非系统形式的 H 变成系统形式的 H．

【例 5.11】

构造一个 $m = 3$ 的二元 $(7, 4)$ 汉明码．

解

所谓构造就是求一个 $(7, 4)$ 汉明码的生成矩阵．先利用汉明码的特性构造一个校验矩阵 H，再通过列置换将它变为系统形式．

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{\text{列置换}} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} = (P^{\mathrm{T}} \quad I_3)$$

$$\longrightarrow G = (I_4 \quad P) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

由于 G 中包含了单位矩阵 I_4，矩阵的秩是 4，所以矩阵 G 的 4 行是 4 个线性无关的基底，可以张成一个包含 $2^4 = 16$ 个码字的码空间．

必须指出，完备码是标准阵列最规则因而译码最简单的码，但它并不一定是纠错能力最强的码，因为它不一定是极大最小距离码，即满足 $d_{\min} = n - k + 1$．

$(7, 4)$ 汉明码 $d_{\min} = 3$，而同样 n、k 的极大最小距离为 4．

如果给 (n, k) 汉明码添加一位奇偶校验位，可得到一个 $d_{\min} = 4$ 的 $(n + 1, k)$ 扩展汉明码，码长为 $(n + 1)$，校验位为 $n - k + 1$ 位．这时

$$H' = \begin{bmatrix} & & & 0 \\ & H & & 0 \\ & & & \cdots \\ & & & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

其中，H 为 $(n - k) \times n$ 矩阵，而 H' 为 $(n - k + 1) \times (n + 1)$ 矩阵．

这个码除能纠正单个错误外,还能发现两个错误 .因为当码字在传输过程中发生两个错误时,其伴随式为对应校验矩阵中的两列之和 .为了能够发现两个错误,必须使得校验矩阵的任意两列之和不为其他列,即要求校验矩阵中的任意三列线性无关,即码字的最小距离为 4 .

$$d_{\min} = u + l + 1 = 1 + 2 + 1 = 4$$

当出现 1 位错误时,伴随式为 H 的某一列,即 $\times\times\cdots\times 1$ 的形式,最后一位数为 1 .当出现两个错误时,伴随式为某两列之和,最后一位为 0 .因为它与 H' 中的任何一列都不相同,所以与单个错误的伴随式区别开来,故可以检查两个错误 .

在同样的纠错能力下,汉明码的码率是最高的:

$$R = \frac{2^m - 1 - m}{2^m - 1} = 1 - \frac{m}{2^m - 1}$$

当 m 很大时, $R \to 1$.

6 .4 .3　循环码

循环码是线性分组码的一个子集,它具有完整的代数结构,编码和译码可以用具有反馈级联的移位寄存器来实现 .它满足循环移位特性:码 C 中任何一个码字的循环移位仍是码 C 中的一个码字 .

定义 6 .14　对于一个 (n, k) 线性分组码,若某一码字为 $C = (C_{n-1} C_{n-2} \cdots C_1 C_0)$,该码字向左循环一位后为 $C^{(1)} = (C_{n-2} C_{n-3} \cdots C_0 C_{n-1})$,向左循环移动 i 位后为 $C^{(i)} = (C_{n-i-1} C_{n-i-2} \cdots C_{n-i+1} C_{n-i})$.若 $C^{(i)}$ $i = 1, 2, \cdots, n-1$ 均为码字,则称这个 (n, k) 线性分组码为循环码 .这里循环移位也可以定义为向右移位 .

一般 (n, k) 线性分组码的 k 个基底之间不存在规则的联系 .因此需用 k 个基底组成生成矩阵来表示一个码的特征,而循环码的 k 个基底可以是同一个基底循环移位 k 次得到 .因此用一个基底就可以表示一个码的特征 .我们可以用一个 $(n-1)$ 次的码多项式 $C(x)$ 来表示一个码字:

$$C(x) = C_{n-1} x^{n-1} + C_{n-2} x^{n-2} + \cdots + C_2 x^2 + C_1 x + C_0$$

这里,码元序号从 0～$(n-1)$ 而不用 1～n 是为了在以后的多项式运算中系数序号与 x 的幂次一致 .

循环码的循环特性可以用码多项式表示为

移 1 位: $C^{(1)}(x) = xC(x) = C_{n-2} x^{n-1} + \cdots + C_1 x^2 + C_0 x + C_{n-1}$

移 2 位: $C^{(2)}(x) = x^2 C(x) = C_{n-3} x^{n-1} + \cdots + C_0 x^2 + C_{n-1} x + C_{n-2}$

…

移 $n-1$ 位: $C^{(n-1)}(x) = x^{n-1} C(x) = C_0 x^{n-1} + \cdots + C_3 x^2 + C_2 x + C_1$

$C(x)$ 移 n 位后又回到 $C(x)$,一个码字的移位最多能得到 $n-1$ 个新码字,因此循环

码字的循环仍是码字并不意味着循环码可以仅从一个码字循环而得.一个(n, k)循环码有2^k个码字,可能是由几个码字循环得到的几组码字,但它们都是同一基底的线性组合.

根据线性码空间的封闭性,码字的线性组合仍是码字.

在2^k个码字的码多项式中取一个次数最低(即$n-k$次)的多项式作为生成多项式,用$g(x)$表示.可以证明$g(x)$是码多项式中唯一一个次的多项式且常数项不为0.即它的g_{n-k}及g_0均为1,由生成多项式$g(x)$可以得到循环码的生成矩阵$G(x)$,作为生成矩阵.因为$x^i g(x)$, $i=0,1,\cdots,k-1$均是码字且线性无关,故可用G的k行构成一个生成矩阵.

令
$$g(x)=x^{n-k}+\cdots+g_2x^2+g_1x+1$$

$$G(x)=\begin{bmatrix} x^{k-1}g(x) \\ \cdots \\ xg(x) \\ g(x) \end{bmatrix}$$

$$=\begin{bmatrix} 1 & g_{n-k-1} & g_{n-k-2} & \cdots & g_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & g_{n-k-1} & g_{n-k-2} & \cdots & g_1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & g_{n-k-1} & g_{n-k-2} & \cdots & g_1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & \cdots \\ 0 & 0 & \cdots & 0 & 1 & g_{n-k-1} & g_{n-k-2} & \cdots & \cdots & 1 \end{bmatrix} \tag{6.79}$$

$g(x)$的系数是降幂排列.

由生成多项式$g(x)$和信息多项式可以得到循环码的多项式:

$$\begin{aligned} C(x) &= (m_{k-1}\ m_{k-2}\cdots m_1\ m_0)\,G(x) \\ &= \sum_{i=0}^{k-1} m_i x^i g(x) \\ &= m(x)\,g(x) \end{aligned} \tag{6.80}$$

每个码多项式$C(x)$都是$g(x)$的倍式,并且每个次数小于等于$n-1$的$g(x)$的倍式必定是一个码多项式.

生成多项式$g(x)$一定是(x^n+1)的因子,即$x^n+1=g(x)\,h(x)$.这是用$g(x)$构造循环码的充要条件,即循环码的$g(x)$一定是(x^n+1)的因子;反过来,如果$g(x)$是(x^n+1)的$(n-k)$次因子,一定可以构造一个(n, k)循环码.这时可以保证码的循环移位特性,即码字的循环仍是码字.

构造循环码的步骤:

(1) 对(x^n+1)作因式分解,找出$(n-k)$次因式;

(2) 以该$(n-k)$次因式作为生成多项式,与不高于$(k-1)$次的信息多项式相乘得码

多项式 $C(x)=m(x)g(x)$, $C(x)$的次数不高于$(k-1)+(n-k)=(n-1)$次 .

【例 6 .12】

构造一个(7,4)循环码 .

解

(1) 对(x^7+1)作因式分解得: $x^7+1=(x+1)(x^3+x^2+1)(x^3+x+1)$ 3 次因式有两个(x^3+x^2+1)和(x^3+x+1),均可以作为(7,4)循环码的生成多项式 .

(2) 选 $g(x)=x^3+x^2+1$, 共有 16 种可能的组合, 对应 16 个码字 .利用 $C(x)=m(x)g(x)$可得到 16 个码字(选择不同的 $g(x)$会得到不同的(7,4)循环码) .例如 $m=(0110)$对应码字:

$$\begin{aligned}C(x)&=m(x)g(x)\\&=(m_3x^3+m_2x^2+m_1x+m_0)g(x)\\&=x^5+x^3+x^2+x\rightarrow(0101110)\end{aligned}$$

由表 6 5 看出,任何码字的循环仍是码字 .整个码组有 4 组码字的循环,但都是 $g(x)=x^3+x^2+1$ 的线性组合 .

表 6 5　(7,4)循环码

信息比特 $m_3\ m_2\ m_1\ m_0$	码字(循环 1) $C_6\ C_5\ C_4\ C_3\ C_2\ C_1\ C_0$	信息比特 $m_3\ m_2\ m_1\ m_0$	码字(循环 2) $C_6\ C_5\ C_4\ C_3\ C_2\ C_1\ C_0$	信息比特 $m_3\ m_2\ m_1\ m_0$	码字(循环 3 和 4) $C_6\ C_5\ C_4\ C_3\ C_2\ C_1\ C_0$
0001	0001101	0011	0010111	0000	0000000
0010	0011010	0110	0101110	1111	1111111
0100	0110100	1100	1011100		
1000	1101000	0101	0111001		
1101	1010001	1010	1110010		
0111	0100011	1001	1100101		
1110	1000110	1111	1001011		

由本例可以验证码字的循环仍是码字,码字的线性组合也仍是码字 .

在式 $x^n+1=g(x)h(x)$中, $h(x)$称为该循环码的一致校验多项式, 其阶次为 k. $h(x)$的校验作用表现为:任何码多项式 $C(x)$与 $h(x)$的模(x^n+1)乘积一定等于 0,而非码字与 $h(x)$的乘积必不为 0 .因为

$$C(x)h(x)=m(x)g(x)h(x)=m(x)(x^n+1)\bmod(x^n+1)$$

在 $x^n+1=g(x)h(x)$的因式分解中, $g(x)$和 $h(x)$处于同等地位 .既然可以用 $g(x)$生成一个循环码,也就可以用 $h(x)$生成另一个循环码 .此时 $h(x)$用作生成多项式,而$g(x)$用作一致校验多项式 .由 $g(x)$生成的(n,k)循环码和 $h(x)$生成的$(n,n-k)$循环码互为对偶码 .

由校验多项式可以得到校验矩阵.令

$$h(x) = h_k x^k + \cdots + h_1 x + h_0 = \sum_{i=0}^{k} h_i x^i$$

$$g(x) = g_{n-k} x^{n-k} + \cdots + g_2 x^2 + g_1 x + g_0$$

根据

$$\begin{aligned} x^n + 1 &= h(x)\, g(x) \\ &= (h_k x^k + \cdots + h_1 x + h_0)(x^{n-k} + \cdots + g_2 x^2 + g_1 x + 1) \end{aligned}$$

将右边展开,对应项相加,并注意到除常数项和最高次项 x^n 系数为 1 以外,其他项系数为 0.

$$\begin{cases} g_0 h_0 = 1 & x^0 \\ g_0 h_1 + g_1 h_0 = 0 & x^1 \\ g_0 h_i + g_1 h_{i-1} + \cdots + g_i h_0 = 0 & x^i \quad 1 \leq i \leq k-1 \\ \cdots & \\ g_{i-k} h_k + g_{i-k+1} h_{k-1} + \cdots + g_{n-k} h_{i-n+k} = 0 & x^i \quad k \leq i \leq n-1 \\ g_{n-k} h_k = 1 & x^n \end{cases}$$

令 $h^*(x) = h_0 x^k + \cdots + h_{k-1} x + h_k$, $h^*(x)$,称为 $h(x)$的倒多项式,即把 $h(x)$的系数倒过来排列.可得

$$H(x) = \begin{pmatrix} h_0 & h_1 & \cdots & h_k & 0 & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_k & 0 & \cdots & 0 \\ 0 & 0 & h_0 & h_1 & \cdots & h_k & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & \cdots \\ 0 & 0 & 0 & \cdots & h_0 & h_1 & \cdots & h_k \end{pmatrix}$$

$$H = \begin{pmatrix} x^{n-k-1} h^*(x) \\ \cdots \\ x h^*(x) \\ h^*(x) \end{pmatrix}$$

可以证明, $x^i h(x)$, $i = 1, 2, \cdots, n-k-1$ 是线性无关的向量,并且 $GH^{\mathrm{T}} = 0$.这样得到循环码的生成矩阵和 H 均不是系统形式的.通过矩阵的初等变换可以将其变成系统形式的 G 和 H.

【例 6.13】

(7,4)循环码的生成多项式 $g(x) = x^3 + x + 1$.求它的系统形式的生成矩阵.

解

$$G=\begin{bmatrix}1&0&1&1&0&0&0\\0&1&0&1&1&0&0\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}\xrightarrow[\text{将矩阵第三、四行加到第一行}]{\text{将矩阵第四行加到第二行,}}G=\begin{bmatrix}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$$

循环码的译码是利用伴随多项式来进行的.

设发送码的多项式 $C(x)=\sum_{i=0}^{n-1}C_ix^i$,错误图样多项式 $e(x)=\sum_{i=0}^{n-1}e_ix^i$,接收码多项式 $R(x)=\sum_{i=0}^{n-1}r_ix^i$,则有 $R(x)=C(x)+e(x)$.

设 $g(x)$为码的生成多项式,则发送码多项式 $C(x)$能被 $g(x)$除尽.

$$\frac{R(x)}{g(x)}=\frac{C(x)+e(x)}{g(x)}=\frac{e(x)}{g(x)}\ \bmod\ g(x)$$

定义伴随多项式(简称伴随式)为

$$s(x)=\frac{e(x)}{g(x)}=e(x)\ \bmod\ g(x)\tag{6.81}$$

若无错误传输,则 $s(x)=0$,否则 $s(x)\neq 0$.

由伴随多项式可得到发送码多项式的估值:

$$C(x)=R(x)+e(x)\ \bmod\ g(x)\tag{6.82}$$

因为 $g(x)$的次数为$(n-k)$, $e(x)$的次数为$(n-1)$,则伴随式 $s(x)$的最高次数为$(n-k-1)$,即 $s(x)$共有 $n-k$ 项,故有 2^{n-k}种可能的表示式,即有 2^{n-k}个伴随式.若 $2^{n-k}\geq n+1$,则具有至少纠正一位错误的能力.

【例 6.14】

已知(7,4)循环码的生成多项式 $g(x)=x^3+x+1$.若已知接收码的最高位码元发生错误,求其伴随多项式;若已知接收码字为(0111000),求发送码字.

解

求 $s(x)$的计算实际上是对 $g(x)$做除法求余运算.

已知 $E=(1000000)$,对应的错误图样多项式为 $e(x)=x^6$,则

$$s(x)=\frac{e(x)}{g(x)}=\frac{x^6}{x^3+x+1}=(x^2+1)\ \bmod\ g(x)$$

对应的伴随式为 $S=(101)$.

若接收码字为(0001110),它的码多项式为

$$R(x)=x^3+x^2+x$$

$$s(x)=\frac{e(x)}{g(x)}=\frac{R(x)}{g(x)}=\frac{x^3+x^2+x}{x^3+x+1}=(x^2+1)\ \bmod\ g(x)$$

根据 $s(x)$从译码表中找出对应的错误图样多项式,可得到发送码的估值:

$$C(x) = R(x) + e(x) \bmod g(x) = x^6 + x^3 + x^2 + x = (1001110)$$

循环码是线性分组码中非常重要的一个子类,要设计一个码率 $R = k/n$ 的循环码,只要将 $x^n + 1$ 解出一个 $(n - k)$ 次因式 $g(x)$ 就可以生成 (n, k) 循环码.目前有实用价值的纠错码大部分都属于循环码的范围,比如在无线信道上应用最广泛的 BCH、RS 等.

BCH 码是一类重要的循环码.它把生成多项式与码的最小距离和纠错能力联系起来,根据所需要的纠错能力,选取适当的 $g(x)$,可以方便地得到非常有效地纠正多个独立错误的码.

6.4.4 卷积码

前面研究过的各种分组码都是将序列切割后分组进行编译码.信息序列被分组后分组之间的相关信息就损失了,并且分组长度越小,损失的信息就越多;如果把分组长度取得很大,则译码复杂度随之呈指数上升.于是我们考虑在码长 n 有限的情况下,将有限个分组的相关性消息添加到码字里,从而等效地增加码长.译码时利用前后码字的相关性将前面的译码信息反馈到后面作译码参考.这样编码器在某个时间段产生的 n 个码元,不但取决于该时间段进入编码器的信息组的 k 个信息位,还与前面的 $N - 1$ 个时间段内的信息组有关,这就是树码或称链码.

而卷积码是树码中最重要的一类,它的码字与 N 个时间段的信息组的映射关系是时不变的线性关系,卷积码与分组码相似,具有纠正随机错误、突发错误或同时纠正这两类错误的能力.通常,它更适用于前向纠错法,因为其纠错性能对于许多实际情况常优于分组码,而且设备较简单,可用移位寄存起来完成编解码.

卷积码也可以用生成矩阵和校验矩阵来研究它的编解码.下面以(3,1)卷积码为例,讨论卷积码的生成和校验矩阵.

把给定的信息序列 $(m_1, m_2, m_3, \cdots)$ 进行分组,使每组只包含一个信息位,m_i 的校验位有两位 p_{i1}, p_{i2}.对应的码序列为 $(m_1\ p_{11}\ p_{12}, m_2\ p_{21}\ p_{22}, \cdots)$,并设校验位与信息位满足以下关系:

$$p_{i1} = m_i + m_{i-1} + m_{i-3}$$

$$p_{i2} = m_i + m_{i-1} + m_{i-2}$$

当前的校验位与当前的信息位和过去的 3 个信息位有关,且满足线性关系.某信息位影响 4 个分组,即该卷积码的约束长度为 4.

编码器的输入输出关系.令

$$m = (m_1, m_2, m_3, \cdots) \tag{6.83}$$

$$C = (m_1\ p_{11}\ p_{12}, m_2\ p_{21}\ p_{22}, m_3\ p_{31}\ p_{32}, \cdots) \tag{6.84}$$

把监督位与信息位的关系代入得

$$C=(m_1\ m_1\ m_1,\ m_2(m_1+m_2)(m_1+m_2),\ m_3(m_2+m_3)(m_1+m_2+m_3),$$
$$m_4(m_1+m_3+m_4)(m_2+m_3+m_4),\cdots) \tag{6.85}$$

由 m 可以利用下式来生成 C:

$$C=mG=(m_1\quad m_2\quad m_3\quad \cdots)\begin{pmatrix} 111 & 011 & 001 & 010 & 000 & \cdots \\ 000 & 111 & 011 & 001 & 010 & \cdots \\ 000 & 000 & 111 & 011 & 001 & \cdots \\ 000 & 000 & 000 & 111 & 011 & \cdots \\ 000 & 000 & 000 & 000 & 111 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

上式中的矩阵记为 G,称为卷积码的生成矩阵.在第一行中有 m_1 的位置记上 1,第二行中有 m_2 的位置记上 1,……,由信息序列和 G 可得到输出码字.

生成矩阵可以写成下列的分块矩阵的形式:

$$G=\begin{pmatrix} I & P_1 & 0 & P_2 & 0 & P_3 & 0 & P_4 & 0 & 0 & \cdots \\ & & I & P_1 & 0 & P_2 & 0 & P_3 & 0 & P_4 & \cdots \\ & & & & I & P_1 & 0 & P_2 & 0 & P_3 & \cdots \\ & & & & & & I & P_1 & 0 & P_2 & \cdots \\ & & & & & & & & I & P_1 & \cdots \\ & & & & & & & & & & \cdots \end{pmatrix}$$

式中,I 为 $k\times k=1\times 1$ 阶单位阵;0 为 $k\times k=1\times 1$ 阶全 0 方阵;P_i 为 $k\times(n-k)=1\times 2$ 阶矩阵;

$$P_1=(11),\ P_2=(11),\ P_3=(01),\ P_4=(10),\cdots$$

可以看到,生成矩阵每一行都相同,只不过每行是上一行向右移动 3 列.

当输入信息序列 $m=(0100101011\cdots)$时,对应的码字为

$$C=mG=(000111011001\cdots)$$

对于所生成的码序列,每 3 个数字组成一个码字,其中包括一个信息位和两个校验位.

由于生成矩阵每一行都相同,所以生成矩阵完全可以由第一行确定.把第一行 $G_0=(I\quad P_1\quad 0\quad P_2\quad 0\quad P_3\quad 0\quad P_4\quad \cdots)$称为基本生成矩阵.这里 G_0 的设定具有一般性.以时刻 i 为基准(一般可将 i 理解为编码时刻或当前时刻),这个时刻在编码过程中是不断向前移动的.设编码器的初始状态为零(移存器清 0),则随着一个个 k 比特信息组的输入,编码器不断地输出码字.由校验位与信息位的关系还可以确定基本一致校验矩阵.

把有约束关系的 4 个码字写成:

$$C_0=(m_{i-3}\ p_{i-3}\ p_{i-3,2},\ m_{i-2}\ p_{i-2,1}\ p_{i-2,2},\ m_{i-1}\ p_{i-1,11}\ p_{i-1,2},\ m_i p_{i1}\ p_{i2})$$

由

$$\begin{cases} m_{i-3} + m_{i-1} + m_i + p_{i_1} = 0 \\ m_{i-2} + m_{i-1} + m_i + p_{i_2} = 0 \end{cases}$$

可得

$$\begin{pmatrix} 100 & 000 & 100 & 110 \\ 000 & 100 & 100 & 101 \end{pmatrix} C_0^{\mathrm{T}} = 0$$

令 $H_0 = \begin{pmatrix} 100 & 000 & 100 & 110 \\ 000 & 100 & 100 & 101 \end{pmatrix}$，$H_0$ 称为该卷积码的基本一致校验矩阵.它可以判断有约束关系的 4 个接收码字是否发送码字.与线性分组码的一致校验矩阵一样起着校验作用.由于输入序列 m 是一个半无限长序列,生成的卷积码是一个半无限长的码序列.这个半无限长的码序列 $C = (m_1\ p_{11}\ p_{12},\ m_2\ p_{21}\ p_{22},\ m_3\ p_{31}\ p_{32}, \cdots)$ 的校验矩阵是一个有头无尾的半无穷矩阵.

由校验位和信息位的关系可以得到:

$$\begin{cases} m_1 + p_{11} = 0 \\ m_1 + p_{12} = 0 \end{cases}$$

$$\begin{cases} m_1 + m_2 + p_{21} = 0 \\ m_1 + m_2 + p_{22} = 0 \end{cases}$$

$$\begin{cases} m_2 + m_3 + p_{31} = 0 \\ m_1 + m_2 + m_3 + p_{32} = 0 \end{cases}$$

$$\begin{cases} m_1 + m_3 + m_4 + p_{41} = 0 \\ m_2 + m_3 + m_4 + p_{42} = 0 \end{cases}$$

$$\begin{cases} m_2 + m_4 + m_5 + p_{51} = 0 \\ m_3 + m_4 + m_5 + p_{52} = 0 \end{cases}$$

$$\begin{pmatrix} 110 & 000 & \cdots & & & & \\ 101 & 000 & \cdots & & & & \\ 100 & 110 & 000 & \cdots & & & \\ 100 & 101 & 000 & \cdots & & & \\ 000 & 100 & 110 & 000 & \cdots & & \\ 100 & 100 & 101 & 000 & \cdots & & \\ 100 & 000 & 100 & 110 & 000 & \cdots & \\ 000 & 100 & 100 & 101 & 000 & \cdots & \\ 000 & 100 & 000 & 100 & 110 & 000 & \cdots \\ 000 & 000 & 100 & 100 & 101 & 000 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} C^{\mathrm{T}} = 0$$

记系数矩阵为 H,称为$(3,1)$卷积码的一致校验矩阵 .

可以看到,这个有头无尾的半无穷矩阵每三列的结构相同,但后三列比前三列向下移两行,从第七行开始每两行结构相同 .可以看出这两行就是基本一致校验矩阵 .

上式中的校验矩阵还可写为

$$H=\begin{pmatrix}
P_1^{\mathrm{T}} & I & \cdots & & & & & & & \\
P_2^{\mathrm{T}} & 0 & P_1^{\mathrm{T}} & I & \cdots & & & & & \\
P_3^{\mathrm{T}} & 0 & P_2^{\mathrm{T}} & 0 & P_1^{\mathrm{T}} & I & \cdots & & & \\
P_4^{\mathrm{T}} & 0 & P_3^{\mathrm{T}} & 0 & P_2^{\mathrm{T}} & 0 & P_1^{\mathrm{T}} & I & \cdots & \\
0 & 0 & P_4^{\mathrm{T}} & 0 & P_3^{\mathrm{T}} & 0 & P_2^{\mathrm{T}} & 0 & P_1^{\mathrm{T}} & I \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}$$

其中, P_i^{T} 为$(n-k)\times k=2\times 1$ 维矩阵;0 为$(n-k)\times(n-k)=2\times 2$ 全 0 方阵; I 为$(n-k)\times(n-k)=2\times 2$ 维单位阵 .

由循环码的生成矩阵 G 和校验矩阵 H 的表示式可知, G 和 H 有一定的关系:由 G 可以得到 H;反之由 H 也可以得到 G .

令 C 为发送码字序列, E 为错误图样序列,则接收序列为 $R=C+E$.定义接收序列的伴随式为 $S=RH^{\mathrm{T}}$.由于 $CH^{\mathrm{T}}=0$,有 $S=RH^{\mathrm{T}}=(C+E)H^{\mathrm{T}}=EH^{\mathrm{T}}$.接收序列的伴随式包含了错误序列信息,可用于译码 .

以上是用矩阵方法来研究卷积码的编译码 .卷积码更多地是用状态转移图和网络图来描述,感兴趣的同学可以参阅有关文献 .

习　题　6

6 .1 考虑将 9 次重复码 R_9 看成是先后进行两次 R_3 编码,即先进行一次 R_3 编码,然后对每个码字的三个码符号分别再进行一次 R_3 编码,我们将将其称之为“ R_3^2 ”.由此可得另外一种 R_9 的译码方法:先对每组三个符号进行 R_3 译码,然后对三个译码结果再进行一次 R_3 译码 .

计算这种译码方法的错误概率并与 R_9 的择多译码对应的错误概率进行比较 .

6 .2 试证明线性分组码的最小 Hamming 距离等于其非零码字的最小 Hamming 重量 .

6 .3 假设一个(n,k)二进制线性分组码的生成矩阵 G 不包含全零行,例如$(7,4)$ Hamming 码的生成矩阵为

$$G=\begin{bmatrix} I_{4\times 4} \\ P_{3\times 4} \end{bmatrix}=\begin{bmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 0&1&1&1 \\ 1&0&1&1 \\ 1&1&0&1 \end{bmatrix}$$

令 C 为以所有 2^k 个码字为列矢量构成的 $n\times 2^k$ 矩阵:

$$C=G\times\begin{bmatrix} 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1 \\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1 \\ 0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1 \end{bmatrix}$$

$$=\begin{bmatrix} 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1 \\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1 \\ 0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1 \\ 0&1&1&0&1&0&0&1&0&1&1&0&1&0&0&1 \\ 0&1&1&0&0&1&1&0&1&0&0&1&1&0&0&1 \\ 0&1&0&1&1&0&1&0&1&0&1&0&0&1&0&1 \end{bmatrix}$$

试证明对于任意的(n, k)二进制线性分组码,其 C 矩阵中的每一行正好有 2^{k-1} 个 0 和 2^{k-1} 个 1 .(提示: C 的第(i, j)个元素与 G 的第 i 行的关系是什么?当 j 从 0 变到2^{k-1}时上述关系会产生什么结果?)

6 .4 求证: 一个(n, k)二进制线性分组码的最小距离 $d_{\min}$满足下述不等式:

$$d_{\min}\leq\frac{n\cdot 2^{k-1}}{2^k-1}$$

(提示:利用上两题的结论)

6 .5 假设想要设计一个(n, k)码,且要求其最小码距至少为 $2t+1$,其中 k 为信息比特的个数,$(n-k)$为附加的冗余比特数,我们当然希望$(n-k)$越小越好 .

(1) 试证明附加的冗余比特数$(n-k)$满足下述不等式:

$$n-k\geq\log_2(1+C_n^1+C_n^2+\cdots+C_n^t)$$

(2) 试证明 Hamming 码满足上述不等式 .

6 .6 证明:对于任何线性分组码,码字的重量或全部为偶数,或奇数重量的码字数等于偶数重量的码字数 .

6 .7 求: (1) $d_{\min}=3$ 且至多只有 3 位校验码元的二进制码的码长;

(2) $d_{\min}=5$ 且至多只有 8 位校验码元的二进制码的码长 .

6.8 假设信道编码为 Hamming(7,4)码,请将下述接收序列解码:

(1) $r = 1101011$

(2) $r = 0110110$

(3) $r = 0100111$

(4) $r = 1111111$

6.9 Hamming(7,4)码可以纠正 1 比特的差错.请问是否存在一个(14,8)码能够纠正两个比特的差错?

6.10 下述问题针对(15,11)Hamming 码:

(1) 写出此码的校验矩阵 H;

(2) 写出此码的生成矩阵 G;

(3) 写出此码的一个译码函数 $g: \{0,1\}^{15} \to \{0,1\}^{11}$,要求它能纠正所有的 1 比特错误;(提示:用有效码字 x^{15} 与接收序列 y^{15} 之间的 Hamming 距离来构造 g)

(4) 如果将此码放在一个错误概率为 $p = 0.01$ 的 BSC 上传输,请计算一个 11 比特的消息发生译码错误的概率.

6.11 证明:如果两个错误图样 e_1, e_2 的和是一个有效的码字,那么它们具有相同的伴随式.

6.12 设(7,4)循环码的生成多项式为 $g(x) = x^3 + x + 1$,当接收码字为 0010011 时,试问接收码字是否有错.

6.13 选用一个最短的生成多项式设计一个(6,2)循环码.

(1) 计算该码的生成矩阵(系统形式);

(2) 找出所有可能的码字;

(3) 该码能纠多少差错?

6.14 试证明:当且仅当多项式系数之和等于零(模 2)时,$x + 1$ 才是它的一个因子.

6.15 如果已知某一次数下多项式的个数以及低次素多项式之积的个数,试推算出该次数的素多项式的个数.

6.16 设有一码以素多项式 $x^3 + x^2 + 1$ 为模.试找出此码的校验矩阵.

6.17 已知(2,1)卷积码的一致校验方程为 $p_i = m_i + m_{i-1}$,求校验矩阵和生成矩阵.

第 7 章

限失真信源编码

在第 5 章,讨论了离散信源的无失真信源编码,它是一种冗余度压缩编码,可以对信源输出的信息进行有效地表示,并且可以保证信源输出的信号在编译码前后不会有任何失真 .从信号携带信息的角度看,还可以保证编译码前后的信号具有相同的信息熵,因此冗余度压缩编码是无失真的保熵的编码 .

但是无失真的保熵的编码并非是必需的,有时候也不可能实现 .比如,在传送语音信号时,由于人耳接受的带宽和分辨是有限的,因此可以把频谱范围从 20 Hz～20 kHz 的语音信号去掉低端和高端的频率,看成带宽只有 300～3 400 Hz 的信号 .这样虽然会有一些失真,但是这种失真是允许的 .再比如,在传送活动图像时,由于人眼的视觉暂留特性,我们只需每秒种传送 25 帧的静止图像,人们看到的就是连贯的活动图像 .所以在实际生活中,通常总是要求在保证一定质量的前提下,在信宿近似地再现信源输出的信息 .因此,实际的信息传输率可以降低 .

另一方面,由于受到信息存储,处理或传输设备的限制而不得不对信源输出的信号作某种近似的表示 .比如实际信源的输出常常是连续的消息,连续信源的绝对熵 $H(S)$是无限大 .若要求无失真地传送连续信源的消息,则信息传输率 R 也为无限大 .在信道中,由于带宽总是有限的,所以信道容量总是受到限制,而实际信源输出的信息率总是大大超过信道容量($R > C$) ,因此也就不可能实现完全无失真地传输信源的消息 .

如果要把连续信源的消息离散化,由于信源熵为无穷大,根据无失真信源编码定理,

需要用无穷多个比特数才能完全无失真地描述它,这在实际中是做不到的,因此必然会带来一定程度的失真 .

在允许一定失真程度的条件下,怎样用尽可能少的信道符号来表达信源的信息,也就是信源熵所能压缩的极限或者说编码后信源输出的信息率压缩的极限值,这就是本章要讨论的问题——限失真信源编码问题 .限失真信源编码也称保真度准则下的信源编码、熵压缩编码或者称信息率失真理论,它是量化、数模转换、频带压缩和数据压缩的理论基础 .

如果无失真的冗余度压缩编码主要是针对离散信源的,那么,有失真的熵压缩编码主要是针对连续信源 .本章讨论的是离散无记忆信源的限失真信源编码理论,这样便于理解率失真理论的基本概念 .

我们讨论的物理模型仍然是信源编码器,编码器的输入符号集 $X=\{x_1, x_2, \cdots, x_r\}$,输出符号集 $Y=\{y_1, y_2, \cdots, y_s\}$.编码器可以看作一个广义的信道, X 是信道的输入, Y 是信道的输出 .与无失真信源编码不同,这时从输入到输出的映射是一个多对一的映射,它是不可逆的,信源符号序列和码符号序列之间的差异就是编码时引入的失真 .

7 .1　失真测度

我们要研究在给定允许失真的条件下,是否可以设计一种信源编码使信息传输率最低 .为了定量地描述信息率和失真的关系,必须先规定失真的测度 .

7 .1 .1　失真函数

设离散无记忆信源

$$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_r \\ p(x_1) & p(x_2) & \cdots & p(x_r) \end{bmatrix}$$

经过信道传输后接收端的离散变量 Y 的概率空间

$$\begin{bmatrix} Y \\ P \end{bmatrix}=\begin{bmatrix} y_1 & y_2 & \cdots & y_s \\ p(y_1) & p(y_2) & \cdots & p(y_s) \end{bmatrix}$$

对于每一对(x_i, y_j),指定一个非负的函数 $d(x_i, y_j)\geq 0$, $i=1,2\cdots, r$; $j=1,2,\cdots, s$,称 $d(x_i, y_j)$为单个符号的失真度或失真函数,用它来表示信源发出一个符号 x_i,而在接收端再现为 y_j 所引起的误差或失真的大小 .通常较小的 d 值代表较小的失真,而 $d(x_i, y_j)=0$表示没有失真 .由于信源 X 有 r 个符号,信道输出 Y 有 s 个符号,所以 $d(x_i, y_j)$有 $r\times s$ 个,这 $r\times s$ 个非负的函数可以排列成矩阵形式,即

$$D=\begin{bmatrix} d(x_1,y_1) & d(x_1,y_2) & \cdots & d(x_1,y_s) \\ d(x_2,y_1) & d(x_2,y_2) & \cdots & d(x_2,y_s) \\ \cdots & \cdots & & \cdots \\ d(x_r,y_1) & d(x_r,y_2) & \cdots & d(x_r,y_s) \end{bmatrix} \tag{7.1}$$

D 称为失真矩阵,它是一个 $r\times s$ 阶矩阵 .

失真函数是根据人们的实际需要和失真引起的损失、风险大小等人为规定的 .常用的失真函数有

(1) 汉明失真

$$d(x_i,y_i)=\begin{cases} 0 & x_i=y_j \\ 1 & x_i\neq y_j \end{cases} \tag{7.2}$$

在离散对称信道($r=s$)中,定义单个符号失真度为汉明失真 .它表示当再现的接收符号与发送的信源符号相同时,就不存在失真和错误,所以,失真度 $d(x_i,y_j)=0$.当再现的接收符号与发送符号不同时,就有失真存在,而且,认为发送符号与再现符号不同时所引起的失真都相同,所以失真度 $d(x_i,y_j)$, $x_i\neq y_j$ 为常数,通常取值为 1,这种失真称为汉明失真 .汉明失真矩阵 D 通常为方阵,且对角线上的元素为 0 即

$$D=\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & & \cdots \\ 1 & 1 & 1 & \cdots & 0 \end{bmatrix} \tag{7.3}$$

D 是 $r\times r$ 阶方阵 .

(2) 平方误差失真函数

$$d(x_i,y_j)=(x_i-y_j)^2 \qquad \forall i,\forall j \tag{7.4}$$

如果信源符号代表信源输出信号的幅度值,则上式意味着较大的幅度差值要比较小的幅度差值引起的失真更为严重,严重程度用平方表示 .

例如,当信道 $r=s=3$,输入 $X=\{0,1,2\}$,输出 $Y=\{0,1,2\}$时

$$D=\begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ 4 & 1 & 0 \end{bmatrix}$$

【例 7.1】

设信道输入 $X=\{0,1\}$,输出 $Y=\{0,1,2\}$,规定失真函数 $d(0,0)=d(1,1)=0$, $d(0,1)=d(1,0)=1$, $d(0,2)=d(1,2)=0.5$,求 D .

解

$$D=\begin{bmatrix} 0 & 1 & 0 & 5 \\ 1 & 0 & 0 & 5 \end{bmatrix}$$

这是一个二元删除信道 .

以上是单个符号的失真函数,可以推广得到长度为 N 的信源符号序列的失真函数.设信源输出的符号序列 $X = X_1 X_2 \cdots X_N$,其中每一个随机变量 X_i, $i = 1,2,\cdots, N$ 取自于同一符号集 $X = \{x_1, x_2, \cdots, x_r\}$,所以共有 r^N 个不同的信源符号序列,而接收端的符号序列为 $Y = Y_1 Y_2 \cdots Y_N$,其中每一个随机变量 Y_j, $j = 1,2,\cdots, N$ 取值于同一符号集 $Y = \{y_1, y_2, \cdots, y_s\}$,共有 s^N 个不同的接收符号序列.

定义 7.1　设发送序列为 $x_i = x_{i_1} x_{i_2} \cdots x_{i_N}$,接收序列为 $y_j = y_{j_1} y_{j_2} \cdots y_{j_N}$,定义序列的失真度为

$$\begin{aligned} d(x_i, y_i) &= d(x_{i_1}, x_{i_2} \cdots x_{j_N}, y_{j_1}, y_{j_2 \cdots y_{j_N}}) \\ &= d(x_{i_1}, y_{j_1}) + d(x_{i_2}, y_{j_2}) + \cdots d(x_{i_N}, y_{j_N}) \\ &= \sum_{k=1}^{N} d(x_{i_k}, y_{j_k}) \end{aligned} \tag{7.5}$$

也就是说信源序列的失真度等于序列中对应单个符号失真度之和,取不同的 x_i, y_j,其 $d(x_i, y_j)$ 不同,写成矩阵形式时,是 $r^N \times s^N$ 阶矩阵.

【例 7.2】

假设信源输出序列 $X = X_1 X_2 X_3$,其中每个随机变量均取值于 $X = \{0,1\}$,经信道传输后的输出为 $Y = Y_1 Y_2 Y_3$,其中每个随机变量均取值于 $Y = \{0,1\}$.定义失真函数 $d(0,0) = d(1,1) = 0$, $d(0,1) = d(1,0) = 1$,求失真矩阵 $D(N)$.

解

由序列的失真函数的定义,有

$$d(000,000) = d(0,0) + d(0,0) + d(0,0) = 0$$
$$d(000,001) = d(0,0) + d(0,0) + d(0,1) = 1$$

同理可得到矩阵其他元素的数值:

$$D(N) = \begin{bmatrix} 0 & 1 & 1 & 2 & 1 & 2 & 2 & 3 \\ 1 & 0 & 2 & 1 & 2 & 1 & 3 & 2 \\ 1 & 2 & 0 & 1 & 2 & 3 & 1 & 2 \\ 2 & 1 & 1 & 0 & 3 & 2 & 2 & 1 \\ 1 & 2 & 2 & 3 & 0 & 1 & 1 & 2 \\ 2 & 1 & 3 & 2 & 1 & 0 & 2 & 1 \\ 2 & 3 & 1 & 2 & 1 & 2 & 0 & 1 \\ 3 & 2 & 2 & 1 & 2 & 1 & 1 & 0 \end{bmatrix}$$

7.1.2　平均失真

由于 X, Y 都是随机变量,故单个符号失真度 $d(x_i, y_j)$ 也是随机变量.显然,规定了单个符号失真度 $d(x_i, y_j)$ 之后,传输一个符号引起的平均失真,即信源的平均失真度为

$$\begin{aligned}\overline{D} &= E[\,d(x_i, y_j)\,] \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(x_i y_j)\, d(x_i, y_j) \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(x_i)\, p(y_j \mid x_i)\, d(x_i, y_j) \qquad (7.6)\end{aligned}$$

它是在 X, Y 的联合概率空间求平均.

平均失真度已对信源和信道进行了统计平均,所以此值描述了某一信源在某一信道下的失真大小.

N 维信源符号序列的平均失真度为

$$\begin{aligned}\overline{D}(N) &= E[\,d(x_i, y_j)\,] \\ &= \sum_{i=1}^{r^N}\sum_{j=1}^{s^N} p(x_i)\, p(y_j \mid x_i)\, d(x_i, y_j) \\ &= \sum_{i=1}^{r^N}\sum_{j=1}^{s^N} p(x_i)\, p(y_j \mid x_i) \sum_{k=1}^{N} d(x_{i_k}, y_{j_k}) \qquad (7.7)\end{aligned}$$

当信源与信道都是无记忆时,有

$$\overline{D}(N) = \sum_{k=1}^{N} \overline{D}_k \qquad (7.8)$$

这里 $\overline{D}_k$ 是指信源序列第 k 个分量的平均失真度.而信源的平均失真度(单个符号的平均失真度)

$$\overline{D}_N = \frac{1}{N}\overline{D}(N) = \frac{1}{N}\sum_{i=1}^{r^N}\sum_{j=1}^{s^N} p(x_i)\, p(y_j \mid x_i)\, d(x_i, y_j) \qquad (7.9)$$

当信源与信道无记忆时, $\overline{D}_N = \frac{1}{N}\overline{D}(N) = \frac{1}{N}\sum_{k=1}^{N}\overline{D}_k$.注意 $\overline{D}_N$ 和 $\overline{D}_k$ 表示的意义不同.

如果离散信源又是平稳信源,即 $p(x_{i_k}) = p(x_i)$,信道又是平稳信道, $p(y_{j_k} \mid x_{i_k}) = p(y_j \mid x_i)$,则

$$\overline{D}_k = \overline{D} \qquad (7.10)$$

$$\overline{D}(N) = N\overline{D} \qquad (7.11)$$

$$\overline{D}_N = \overline{D} \qquad (7.12)$$

即离散无记忆平稳信源通过离散无记忆平稳信道,其信源序列的平均失真度等于单个符号平均失真度的 N 倍.

7.2　信息率失真函数

7.2.1　D 失真许可信道

如果要求信源符号的平均失真度小于我们所允许的失真 D,即 $\overline{D} \leq D$,称为保真度准则.N 维信源序列的保真度准则是 $\overline{D}(N) \leq ND$.

平均失真度 $\overline{D}$ 不仅与单个符号的失真度有关,还与信源的概率分布和信道的转移概率有关.当信源和单个符号失真度固定,即 $P(X)$ 和 $d(x_i, y_j)$ 给定时,选择不同的试验信道相当于选择不同的编码方法,所得的平均失真度 $\overline{D}$ 不同,有些实验信道 $\overline{D} \leq D$,而有些实验信道 $\overline{D} > D$.凡满足保真度准则的信道称为 D 失真许可的试验信道,所有 D 失真许可的试验信道的集合用 B_D 表示,即

$$B_D = \{ p(y_j \mid x_i) : \overline{D} \leq D \quad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s \} \tag{7.13}$$

对于离散无记忆信源的 N 次扩展信源和离散无记忆信道的 N 次扩展信道,相应的 D 失真许可的试验信道为

$$B_{D(N)} = \{ p(y_j \mid x_i) : \overline{D}(N) \leq ND \quad i = 1, 2, \cdots, r^N;\ j = 1, 2, \cdots, s^N \} \tag{7.14}$$

7.2.2　信息率失真函数的定义

如果信源输出的信息率为 R,在信道容量为 C 的信道上传输,如果 $R > C$,就会引起失真,需要对信源进行压缩,使压缩后信源输出的信息率 R^* 小于信道容量 C.压缩的过程中也会引入失真,但可以控制失真在一个可控的范围内,即满足保真度准则.从另一方面来说,我们总希望在满足保真度准则以后,压缩后的信息传输率 R^* 尽可能地小.

把信源的压缩编码的过程看成一个信道,从这个信道的接收端来说,R^* 可以用平均互信息 $I(X; Y)$ 来表示,压缩过程中引入的失真可以用 $H(X \mid Y)$ 表示.

我们的任务就是在满足保真度准则的 D 失真许可的试验信道集合 B_D 中寻找某一个信道 $p(y_j \mid x_i)$,使 $I(X; Y)$ 达到最小,即

$$R(D) = \min_{p(y_j \mid x_i) \in B_D} I(X; Y) \tag{7.15}$$

这个最小值 $R(D)$ 就是信息率失真函数,也称率失真函数,它的单位是比特/信源符号、哈特莱/信源符号或奈特/信源符号.

对于 N 维信源符号序列,同样可以得其信息率失真函数:

$$R_N(D) = \min_{p(y_j \mid x_i) \in B_D} I(X; Y) \tag{7.16}$$

当信源和信道均为无记忆时，$I(X;Y)=NI(X;Y)$，所以有

$$R_N(D)=NR(D) \tag{7.17}$$

它是在所有满足平均失真度 $\overline{D}(N)\leq ND$ 的 N 维试验信道集合中，寻找某个信道使 $I(X;Y)$ 取极小值．因为平均失真度 $\overline{D}(N)$ 与长度 N 有关，所以，在其他条件(信源概率分布、单个符号的失真度)相同的条件下，对于不同的 N，$R_N(D)$ 是不同的．

对于给定的信源，在满足保真度准则 $\overline{D}\leq D$ 的前提下，信息率失真函数 $R(D)$ 是信源输出的信息率允许压缩到的最小值．由于 $I(X;Y)$ 是 $p(y_j|x_i)$ 的下凸函数，所以在 B_D 集合中，$I(X;Y)$ 的最小值一定存在．

从数学上来看，平均互信息 $I(X;Y)$ 既是信源概率分布 $p(x_i)$ 的上凸函数，又是信道传递概率 $p(y_j|x_i)$ 的下凸函数，因此信道容量 C 和信息率失真函数 $R(D)$ 具有对偶性．

信道容量 $C=\max\limits_{p(x_i)} I(X;Y)$ 是指在信道固定前提下，选择一种信源概率分布使信息传输率最大(求极大值)．它反映了信道传输信息的能力，是信道可靠传输的最大信息传输率．信道容量与信源无关，是信道特性的参量，不同的信道其信道容量不同．

信息率失真函数 $R(D)=\min\limits_{p(y_j|x_i):\overline{D}\leq D} I(X;Y)$ 是在信源固定，满足保真度准则的条件下的信息传输率的最小值，它反映了满足一定失真度的条件下信源可以压缩的程度，也就是满足失真要求而再现信源消息所必须获得的最少平均信息量．$R(D)$ 是信源特性的参量，$R(D)$ 一旦求出就与求极值过程中选择的试验信道无关，不同的信源 $R(D)$ 不同．

这两个概念的适用范围是不一样的，研究信道容量 C 是为了在已知信道中尽可能多地传送信息，是为了充分利用已给定的信道，使传输的信息量最大而错误概率任意小，以提高通信的可靠性，这是信道编码的问题．

研究信息率失真函数是为了在已知信源和允许失真度条件下，使信源输出的信息率尽可能小，也就是在允许的一定失真度 D 的条件下，使信源必须传送给信宿的信息量最少，尽可能用最少的码符号来传送信源信息，使信源的信息可以尽快地传送出去，以提高通信的有效性，这是信源编码问题．

7.2.3 信息率失真函数 $R(D)$ 的性质

1．$R(D)$ 的定义域

D 是允许的平均失真度，$R(D)$ 是对应于 D 的一个确定的信息传输率．对于给定的信源，允许失真 D 不同，$R(D)$ 就不同，它是允许失真度 D 的函数．$R(D)$ 的定义域，也就是 D 的取值范围，必须根据信源的概率分布和选定的失真函数 $d(x_i,y_j)$ 来确定，在不同的试验信道下，求得 $\overline{D}$ 的可能取值范围．平均失真度 $\overline{D}$ 是非负实函数 $d(x_i,y_j)$ 的数学期望，因此 $\overline{D}$ 也是一个非负的实数，$\overline{D}$ 的下限是 0，那么允许失真度 D 的下限也必然是 0，

这就是不允许任何失真的情况 .

平均失真度 $\overline{D}$ 能否达到下限值 0,与单个符号的失真函数的定义有关 .

$$D_{\min} = \min \sum_i \sum_j p(x_i)\, p(y_j \mid x_i)\, d(x_i, y_j)$$

$$= \sum_i p(x_i) \min \sum_j p(y_j \mid x_i)\, d(x_i, y_j) \tag{7.18}$$

选择试验信道,也就是选择转移概率:对于每一个 x_i,找出一个使 $d(x_i, y_j)$ 最小的 y_j,令 $p(y_j \mid x_i) = 1$,而其他的转移概率为 0 这样可以得到

$$D_{\min} = \sum_i p(x_i) \min\ d(x_i, y_j) \tag{7.19}$$

只有当失真矩阵每一行至少有一个零元素时,信源的平均失真度才能达到下限值 0,否则 $D_{\min} > 0$.在实际情况中,一般 $D_{\min} = 0$,它表示信源不允许任何失真存在,直观地理解,这时信息率至少应等于信源输出的平均信息量——信源熵,即 $R(0) = H(X)$.但是当失真矩阵除了满足 $D_{\min} = 0$ 的条件,即每行至少有一个零以外,某些列还有不止一个 0 时,说明信源符号集有些符号可以压缩、合并而不带来任何失真,压缩后的信息率必然减小,这时 $R(0)$ 可以小于 $H(X)$.

【例 7.3】

删除信道 $X = \{0, 1\}$, $Y = \{0, 1, 2\}$, $D = \begin{bmatrix} 0 & 1 & 1/2 \\ 1 & 0 & 1/2 \end{bmatrix}$, 求 $D_{\min}$.

解

最小允许失真度为

$$D_{\min} = \sum_{i=1}^{2} p(x_i) \min\ d(x_i, y_j) = \sum_{i=1}^{2} p(x_i) \cdot 0 = 0$$

当 $D_{\min} = 0$ 时,不管何种信源分布都能达到最小允许失真度 .满足这个最小允许失真度的试验信道是一个无噪无损的试验信道 $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$,并且 B_D 中只有这样一个信道 这时

$$I(X; Y) = H(X)$$

$$R(0) = \min_{p(y_j \mid x_i) \in B_D} I(X; Y) = H(X)$$

【例 7.4】

设信源 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$, 信宿 $Y = \{0, 1\}$, 失真矩阵为 $D = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$, 求 $D_{\min}$.

解

$$D_{\min}=\frac{1}{3}\cdot 0+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 0=\frac{1}{6}$$

满足这个最小允许失真度的试验信道是

$$\begin{cases} p(y_1\mid x_1) = 1 \\ p(y_2\mid x_1) = 0 \\ p(y_1\mid x_2) + p(y_2\mid x_2)=1 \\ p(y_1\mid x_3) = 0 \\ p(y_2\mid x_3) = 1 \end{cases}$$

$B_{D_{\min}}$的试验信道有无穷多个,因为 $p(y_1\mid x_2)$和 $p(y_2\mid x_2)$可以为无穷多个,只要满足和为 1 就行 这些信道的共同特征是信道矩阵中每列有不止一个非零元素,所以信道疑义度 $H(X\mid Y)\neq 0$.

$$R(D_{\min})=R\left(\frac{1}{6}\right)=\min_{p(y_j\mid x_i)\in B_{D_{\min}}} I(X;Y)<H(X)$$

平均失真度也有最大值 $D_{\max}$.根据率失真函数的定义, $R(D)$是在一定约束条件下,平均互信息 $I(X;Y)$的极小值 .由于 $I(X;Y)$是非负的,所以 $R(D)$也必然是非负的,其下限值必为 0 .从直观上理解,不允许任何失真时,平均传送一个信源符号所需的信息率最大,即 $R(D)$此时必须等于信源熵,这就是平均互信息的上限值 .当允许一定的失真存在时,传送信源符号所需的信息率就小些 .反过来说,信息率越小,失真就越大,当 $R(D)$等于 0 时,对应的平均失真最大,这就是 $R(D)$函数定义域的上限值 $D_{\max}$.

事实上,满足 $R(D)=0$ 的 D 可以有无穷多个,只要 $D\geq D_{\max}$,我们取最小的一个定义为 $D_{\max}$.

当 $R(D)=0$ 时,最小的 $I(X;Y)=0$,相当于 X 和 Y 统计独立的情况 .这意味着在接收端收不到信源发送的任何信息,与信源不发送任何信息是等效的 .换句话说,传送信源符号的信息率可以压缩为 0 .

当 $D\geq D_{\max}$时, $R(D)=0$, X 和 Y 相互独立的试验信道就可以使得 $R(D)=0$,这时 $p(y_j\mid x_i)=p(y_j)$ 这样,不同的 $p(y_j)$都可以使得 $R(D)=0$,但是所造成的 $\overline{D}$ 不同,选取其中的最小值定义为 $D_{\max}$.

$$D_{\max}=\min_{p(y_j)}\sum_j p(y_j)\sum_i p(x_i)\,d(x_i,y_j) \tag{7 20}$$

由于信源概率分布 $p(x_i)$和失真函数 $d(x_i,y_j)$已经给定,因此求 $D_{\max}$相当于寻找一种信道输出分布 $p(y_j)$使上式右端最小,如果选取 $\sum_i p(x_i)\,d(x_i,y_j)$ 最小的 $p(y_j)=1$,而令其他的 $\sum_i p(x_i)\,d(x_i,y_j)$ 对应的 y_j 选取 $p(y_j)=0$,则有

$$D_{\max} = \min_{p(y_j)} \sum_i p(x_i)\, d(x_i, y_j) \tag{7.21}$$

【例 7.5】

二元信源 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 0.4 & 0.6 \end{bmatrix}$，$D = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}$，计算 $D_{\max}$.

解

$$D_{\max} = \min(0.4\alpha, 0.6\alpha) = 0.4\alpha$$

综上所述，率失真函数 $R(D)$ 的定义域为 $(D_{\min}, D_{\max})$，一般情况下 $D_{\min} = 0$，$R(0) = H(X)$，$R(D_{\max}) = 0$.而 $D_{\min} < D < D_{\max}$ 时，$H(X) > R(D) > 0$.

2．$R(D)$ 是关于 D 的下凸函数

$R(D)$ 是关于 D 的下凸函数，即对于任意 $0 \leq \alpha \leq 1$ 和 $D_1, D_2 \leq D_{\max}$ 有

$$R[\alpha D_1 + (1-\alpha) D_2] \leq \alpha R(D_1) + (1-\alpha) R(D_2) \tag{7.22}$$

证明

设给定信源 X 和失真函数 $d(x_i, y_j)$，$i = 1, 2, \cdots, r$; $j = 1, 2, \cdots, s$，在 $R(D)$ 函数的定义域内选取两个允许失真度 D_1 和 D_2，并设两个试验信道 $p_1(y_j \mid x_i)$ 和 $p_2(y_j \mid x_i)$ 分别达到相应的信息率失真函数 $R(D_1)$ 和 $R(D_2)$，即分别满足保真度准则

$$\overline{D_1} = \sum_i \sum_j p(x_i)\, p_1(y_j \mid x_i)\, d(x_i, y_j) \leq D_1 \tag{7.23}$$

$$\overline{D_2} = \sum_i \sum_j p(x_i)\, p_2(y_j \mid x_i)\, d(x_i, y_j) \leq D_2 \tag{7.24}$$

并且

$$I[p_1(y_j \mid x_i)] = R(D_1) \tag{7.25}$$

$$I[p_2(y_j \mid x_i)] = R(D_2) \tag{7.26}$$

另设

$$p(y_j \mid x_i) = \alpha p_1(y_j \mid x_i) + (1-\alpha) p_2(y_j \mid x_i) \tag{7.27}$$

对应的

$$\begin{aligned} \overline{D} &= \sum_i \sum_j p(x_i)\, p(y_j \mid x_i)\, d(x_i, y_j) \\ &= \alpha \sum_i \sum_j p(x_i)\, p_1(y_j \mid x_i)\, d(x_i, y_j) + (1-\alpha) \sum_i \sum_j p(x_i)\, p_2(y_j \mid x_i)\, d(x_i, y_j) \\ &\leq \alpha D_1 + (1-\alpha) D_2 \end{aligned} \tag{7.28}$$

所以 $p(y_j \mid x_i)$ 是满足保真度准则 $\overline{D} \leq \alpha D_1 + (1-\alpha) D_2$ 的试验信道 .

根据率失真函数的定义，有

$$I[p(y_j \mid x_i)] \geq R[\alpha D_1 + (1-\alpha) D_2] \tag{7.29}$$

对于固定信源 X 来说，平均互信息是信道传递概率 $p(y_j \mid x_i)$ 的下凸函数，所以

$$I[p(y_j|x_i)] \leq \alpha I[p_1(y_j|x_i)] + (1-\alpha) I[p_2(y_j|x_i)] = \alpha R(D_1) + (1-\alpha) R(D_2) \tag{7.30}$$

综合上面两式,有

$$R[\alpha D_1 + (1-\alpha) D_2] \leq \alpha R(D_1) + (1-\alpha) R(D_2) \tag{7.31}$$

即率失真函数 $R(D)$在定义域内是允许失真度 D 的下凸函数 .

证毕

3 . $R(D)$在定义域内是严格递减函数

由于 $R(D)$具有凸状性,这意味着它在定义域内是连续的 并且 $R(D)$在定义域内是递减的,因为允许的失真越大,需要的信息率可以越小,根据 $R(D)$的定义,它是在平均失真度小于或等于允许失真度 D 的所有信道集合 B_D 中,取 $I(X;Y)$的最小值,当允许失真度 D 扩大,那么 B_D 的集合也扩大,当然包含满足原来条件的所有信道 .这时再在扩大的 B_D 集合中找 $I(X;Y)$的最小值,那么结果或者不变或者比原来的小,因此 $R(D)$是递减的,即在$0 < D < D_{max}$范围内,若 $D_1 < D_2$,则 $R(D_1) \geq R(D_2)$.下面证明它是严格递减的,即上式中等号不成立 .

证明

如果上式中等号成立,则在(D_1, D_2)中 $R(D)$为常数 .下面证明(D_1, D_2)中 $R(D)$不是常数 .

假设 $0 < D_1 < D_2 < D_{max}$, $p_1(y_j|x_i)$和 $p_m(y_j|x_i)$是分别达到相应的信息率失真函数 $R(D_1)$和 $R(D_{max})$的两个试验信道,即

$$\overline{D_1} = \sum_i \sum_j p(x_i)\, p_1(y_j|x_i)\, d(x_i, y_j) \leq D_1 \tag{7.32}$$

$$\overline{D_m} = \sum_i \sum_j p(x_i)\, p_m(y_j|x_i)\, d(x_i, y_j) \leq D_m \tag{7.33}$$

$$I[p_1(y_j|x_i)] = R(D_1) \tag{7.34}$$

$$I[p_m(y_j|x_i)] = R(D_{max}) = 0 \tag{7.35}$$

总能找到足够小的 $\alpha > 0$ 满足

$$D_1 < \alpha D_{max} + (1-\alpha) D_1 < D_2 \tag{7.36}$$

不等式左边 $D_1 + \alpha(D_{max} - D_1) > D_1$ 肯定成立 .不等式右边总能找到一个 α使 $\alpha(D_{max} - D_1) < (D_2 - D_1)$.令 $D_0 = \alpha D_{max} + (1-\alpha) D_1$,则 $D_1 < D_0 < D_2$.

现在定义一个新的试验信道,设其信道传递概率为

$$p(y_j|x_i) = \alpha p_m(y_j|x_i) + (1-\alpha) p_1(y_j|x_i) \tag{7.37}$$

对应的

$$\begin{aligned}\overline{D} &= \sum_i \sum_j p(x_i)\, p(y_j \mid x_i)\, d(x_i, y_j) \\ &= \alpha \sum_i \sum_j p(x_i)\, p_m(y_j \mid x_i)\, d(x_i, y_j) + (1-\alpha) \sum_i \sum_j p(x_i)\, p_1(y_j \mid x_i)\, d(x_i, y_j) \\ &= \alpha \overline{D}_m + (1-\alpha)\, \overline{D}_1 \\ &\leq \alpha D_{\max} + (1-\alpha)\, D_1 \\ &= D_0 \end{aligned} \tag{7.38}$$

可见新试验信道满足保真度准则,因此

$$I[\, p_1(y_j \mid x_i)\,] \geq R(D_0) \tag{7.39}$$

由于平均互信息是信道传递概率 $p(y_j \mid x_i)$的下凸函数,所以

$$\begin{aligned} I(p(y_j \mid x_i)) &\leq \alpha I[\, p_m(y_j \mid x_i)\,] + (1-\alpha)\, I[\, p_1(y_j \mid x_i)\,] \\ &= (1-\alpha)\, R(D_1) < R(D_1) \end{aligned} \tag{7.40}$$

所以 $R(D_0) < R(D_1)$,而 $D_1 < D_0 < D_2$,所以 $R(D)$在(D_1, D_2)内不为常数,即 $R(D_1) \geq R(D_2)$中等号不成立,$R(D)$是定义域内的严格递减函数.

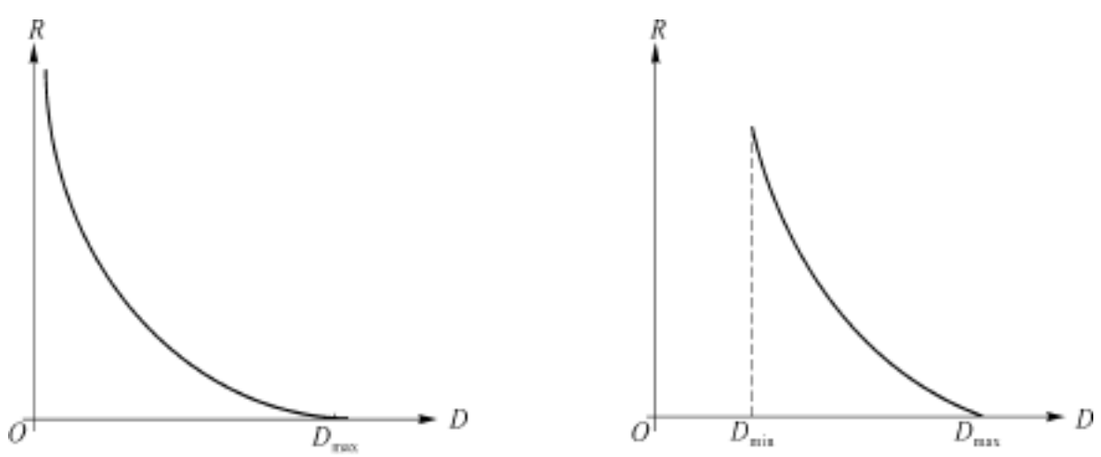
证毕

由于信息率失真函数 $R(D)$是严格的单调递减函数,因此在 B_D 中 $I(X; Y)$为最小的试验信道 $p(y_j \mid x_i)$必在 B_D 的边界上,即

$$\overline{D} = \sum_i \sum_j p(x_i)\, p(y_j \mid x_i)\, d(x_i, y_j) = D \tag{7.41}$$

所以我们通常选择在 $\overline{D} = D$ 的条件下来计算信息率失真函数.

根据以上性质,可以画出率失真函数 $R(D)$的曲线图.由 $R(0) = H(X)$, $R(D_{\max}) = 0$ 决定了曲线的两个端点,在 0 和 $D_{\max}$ 之间 $R(D)$是单调递减的下凸函数.在连续信源的情况下,当 $R \to \infty$时,曲线不与纵轴相交,如图中虚线所示.如果 $D_{\min} \neq 0$,可以得到图 7.1.



(a) D_{min}=0时连续信源　　(b) D_{min}≠0时离散信源

图 7.1　信息率失真函数

7.3 信息率失真函数的计算

已知信源的概率分布和失真函数就可以确定信源的信息率失真函数 $R(D)$.它是在约束条件即保真度准则下求 $I(X;Y)$的极小值的问题.应用拉格朗日乘数法,原则上可以求出解来,但是如果要得到明显的解析表达式是比较困难的,通常只能用参量形式来表示,或采用迭代算法用计算机求解 $R(D)$函数.

7.3.1 应用参量表示式计算 $R(D)$

下面采用拉格朗日乘数法求解 $R(D)$函数,采用 S 作为参量来表示信息率失真函数 $R(D)$和失真函数 $D(S)$.即在

$$p(y_j \mid x_i) \geq 0 \quad i = 1,2,\cdots,r, j = 1,2,\cdots,s \tag{7.42}$$

$$\sum_j p(y_j \mid x_i) = 1 \quad i = 1,2,\cdots,r \tag{7.43}$$

$$\sum_i \sum_j p(x_i)\, p(y_j \mid x_i)\, d_{ij} = D \tag{7.44}$$

的约束条件下求

$$I(X;Y) = \sum_i \sum_j p(x_i)\, p(y_j \mid x_i) \log_2 \frac{p(y_j \mid x_i)}{\sum_k p(x_k)\, p(y_j \mid x_k)} \tag{7.45}$$

的极小值.取拉格朗日乘数 μ_i 与约束(7.43)中的 r 个等式对应,并取拉格朗日乘数 S 与式(7.44)对应,构成辅助函数:

$$F = I(X;Y) - \mu_i \sum_j p(y_j \mid x_i) - SD \tag{7.46}$$

然后对 $p(y_j \mid x_i)$ 求偏导并令其为零,

$$\frac{F}{p(y_j \mid x_i)} = p(x_i) \log_2 \frac{p(y_j \mid x_i)}{p(y_j)} - \mu_i - Sp(x_i)\, d_{ij} = 0 \tag{7.47}$$

为了求解方便,令

$$\log_2 \lambda_i = \frac{\mu_i}{p(x_i)} \tag{7.48}$$

整理式(7.47)可得

$$\log_2 p(y_j \mid x_i) - \log_2 p(y_j) - Sd_{ij} - \log_2 \lambda_i = 0 \tag{7.49}$$

解方程组(7.49)可得 $r \times s$ 个信道转移概率:

$$p(y_j \mid x_i) = p(y_j)\, \lambda_i \exp(Sd_{ij}) \tag{7.50}$$

将上式对 j 求和,得

$$\sum_j p(y_j \mid x_i) = \sum_j \lambda_i p(y_j)\exp(Sd_{ij}) = 1 \tag{7.51}$$

$$\lambda_i = \frac{1}{\sum_j p(y_j)\exp(Sd_{ij})} \quad i = 1,2,\cdots,r \tag{7.52}$$

将式(7.50)乘以 $p(x_i)$,再对 i 求和,得

$$\sum_i p(x_i)p(y_j \mid x_i) = p(y_j)\sum_i \lambda_i p(x_i)\exp(Sd_{ij}) \tag{7.53}$$

若 $p(y_j) \neq 0$,则有

$$\sum_i \lambda_i p(x_i)\exp(Sd_{ij}) = 1 \tag{7.54}$$

将式(7.52)代入式(7.54),可解出用参量 S 表示的 $p(y_j)$,然后将求得的 $p(y_j)$ 代入式(7.52)可求出 λ_i,又可根据式(7.50)求出最佳的试验信道的转移概率 $p(y_j \mid x_i)$.

这时得到的结果是以 S 为参量的表达式,而不是显式的表达式.参量 S 的限制条件为式(7.44).将式(7.50)代入式(7.44)和式(7.45),得到以 S 为参量的信息率失真函数 $R(S)$ 和失真函数 $D(S)$:

$$D(S) = \sum_i \sum_j p(x_i)p(y_j)\lambda_i d_{ij}\exp(Sd_{ij}) \tag{7.55}$$

$$R(S) = SD(S) + \sum_i p(x_i)\log_2 \lambda_i \tag{7.56}$$

由于 $p(y_j)$, $j = 1,2,\cdots$, S 不能为负值,所以参量 S 的取值有一定的限制.由于 D 是参量 S 的函数,λ_i 也是 S 的函数,因此可以把 S 看成是 D 的函数,因此 λ_i 也是 D 的函数.利用全微分公式对 $R(D)$ 求导,可得

$$\frac{\mathrm{d}R(D)}{\mathrm{d}D} = \frac{R(D)}{D} + \frac{R(D)}{S}\cdot\frac{\mathrm{d}S}{\mathrm{d}D} + \sum_{i=1}^{r}\frac{R(D)}{\lambda_i}\cdot\frac{\mathrm{d}\lambda_i}{\mathrm{d}D}$$

$$= S + D\frac{\mathrm{d}S}{\mathrm{d}D} + \sum_{i=1}^{r}\frac{p(x_i)}{\lambda_S}\cdot\frac{\mathrm{d}\lambda_i}{\mathrm{d}D}$$

$$= S + \left[D + \sum_{i=1}^{r}\frac{p(x_i)}{\lambda_i}\cdot\frac{\mathrm{d}\lambda_i}{\mathrm{d}D}\right]\frac{\mathrm{d}S}{\mathrm{d}D} \tag{7.57}$$

通常推导可得出

$$\frac{\mathrm{d}R(D)}{\mathrm{d}D} = S \tag{7.58}$$

这表明参量 S 是信息率失真函数 $R(D)$ 的斜率.由 $R(D)$ 在 $0 < D < D_{\max}$ 之间是严格的单调递减函数可知,S 必是负值.并且由于 $R(D)$ 是下凸函数,所以 S 将随 D 的增加而增加.在 $D = 0$ 处,$S \to -\infty$,当 $D > D_{\max}$ 时,$R(D) = 0$ 斜率为 0,所以 $S = 0$;而在 $D = D_{\max}$ 处,S 一般是从一个非常小的负数跳变到 0.

【例 7.6】

二元信源的信息率失真函数

信源为 $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$，$p \le \frac{1}{2}$.输出符号集为 $(0, 1)$，失真函数定义为 $d_{ij} = \begin{cases} 0 & i = j \\ 1 & i \ne j \end{cases}$，$i, j = 1, 2$，求 $R(D)$.

解

(1) 由式(7.54)计算 λ_1 和 λ_2

记 $p_1 = p(0) = p$，$p_2 = p(1) = 1 - p$.

$$\begin{cases} \lambda_1 p_1 e^{sd_{11}} + \lambda_2 p_2 e^{sd_{21}} = 1 \\ \lambda_1 p_1 e^{sd_{12}} + \lambda_2 p_2 e^{sd_{22}} = 1 \end{cases}$$

把已知量代入

$$\begin{cases} \lambda_1 p + \lambda_2 (1-p) e^{s} = 1 \\ \lambda_1 p e^{s} + \lambda_2 (1-p) = 1 \end{cases}$$

可得

$$\lambda_1 = \frac{1}{p(1+e^{s})}$$

$$\lambda_2 = \frac{1}{(1-p)(1+e^{s})}$$

(2) 由式(7.52)计算 $p(y_1)$ 和 $p(y_2)$

$$\begin{cases} p(y_1) e^{sd_{11}} + p(y_2) e^{sd_{12}} = \frac{1}{\lambda_1} \\ p(y_1) e^{sd_{21}} + p(y_2) e^{sd_{22}} = \frac{1}{\lambda_2} \end{cases}$$

求出 $p(y_1)$ 和 $p(y_2)$，并将 λ_1, λ_2 用 s 表示

$$p(y_1) = \frac{\frac{1}{\lambda_1} e^{sd_{22}} - \frac{1}{\lambda_2} e^{sd_{12}}}{e^{sd_{11}+sd_{22}} - e^{sd_{12}+sd_{21}}} = \frac{p - (1-p)e^{s}}{1 - e^{s}}$$

$$p(y_2) = \frac{\frac{1}{\lambda_2} e^{sd_{11}} - \frac{1}{\lambda_1} e^{sd_{21}}}{e^{sd_{11}+sd_{22}} - e^{sd_{12}+sd_{21}}} = \frac{1 - p - pe^{s}}{1 - e^{s}}$$

(3) 将求得的 λ_1, λ_2 和 $p(y_1)$，$p(y_2)$，代入式(7.55)得到平均失真度 D

$$\begin{aligned} D(s) = {} & \lambda_1 p_1 p(y_1) d_{11} e^{sd_{11}} + \lambda_1 p_1 p(y_2) d_{12} e^{sd_{12}} \\ & + \lambda_2 p_2 p(y_1) d_{21} e^{sd_{21}} + \lambda_2 p_2 p(y_2) d_{22} e^{sd_{22}} \\ = {} & \frac{e^{s}}{1 + e^{s}} \end{aligned}$$

解出参量 s 为 $s=\ln\dfrac{D}{1-D}$.

(4) 将参量 s 代入式(7.56)得到率失真函数 $R(D)$

因为

$$1+\mathrm{e}^{s}=\frac{1}{1-D}$$

所以有

$$\begin{aligned}R(D)&=sD+p\log_2\lambda_1+(1-p)\log_2\lambda_2\\&=D\log_2\frac{D}{1-D}-p\log_2 p(1+\mathrm{e}^{s})-(1-p)\log_2(1-p)(1+\mathrm{e}^{s})\\&=H(p,1-p)-H(D,1-D)\\&=H(p)-H(D)\end{aligned}$$

第一项是信源熵,第二项则是因容忍一定的失真而可以压缩的信息率.

当 $D=0$ 时, $R(D)=H(p)$;

当 $D=D_{\max}=p$ 时, $R(D_{\max})=0$.

对于不同 p 值可以得出一组 $R(D)$ 的曲线,如图 7.2 所示,可以看出,对于给定的平均失真度 D,信源分布越均匀(p 值接近 1/2), $R(D)$ 越大,可压缩性越小;反之,信源分布越不均匀, $R(D)$ 越小,可压缩性越大.因为在 D 固定的条件下,由最大离散熵定理,信源越趋于等概分布,其熵越大,即不确定越大.要去除这不确定性所需的信息传输率越大.而 $R(D)$ 正是去除信源不确定性所必须的信息传输率(容忍一定的失真).当 $D=D_{\max}$ 时, $R(D_{\max})$,即完全不传输信息.例如不管信源发 0 还是 1,都把它编成 1,只选一种符号当然也就不用传送信息,即 $R=0$.

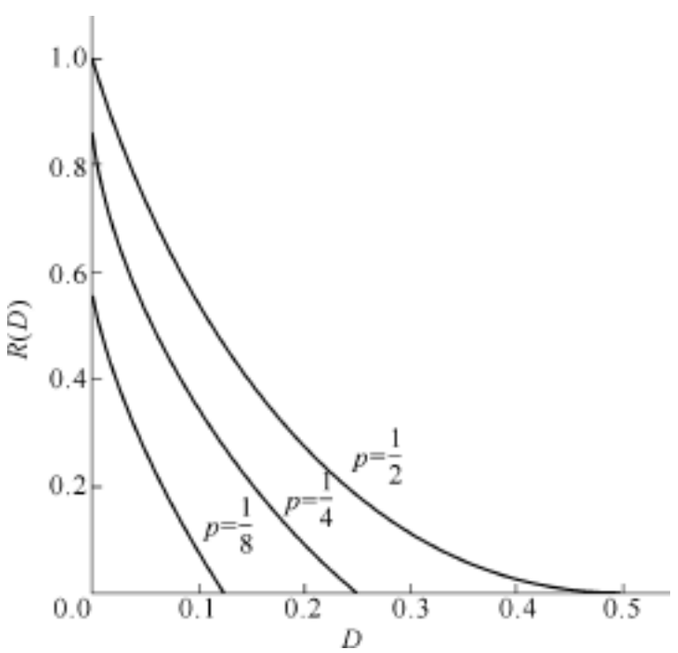


图 7.2　不同 p 值对应的信息率失真函数

【例 7.7】

等概信源的信息率失真函数

信源输出符号集 $X=\{x_1, x_2, \cdots, x_r\}$，等概分布 $p(x_i)=1/r, 1, 2, \cdots, r$，信道输出符号集 $Y=\{y_1, y_2, \cdots, y_r\}$，失真函数定义为 $d(x_i, y_j)=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases}(i, j=1, 2, \cdots, r)$.

求 $R(D)$.

解

引入记号:

$$p_i = p(x_i) = \frac{1}{r},\ i=1, 2, \cdots, r$$

$$q_j = p(y_j),\ j=1, 2, \cdots, r$$

$$d_{ij} = d(x_i, y_j),\ i, j=1, 2, \cdots, r$$

(1) 由式(7.54)确定出 λ_i

$$\begin{cases}\lambda_1 + \lambda_2 e^{s} + \cdots + \lambda_r e^{s} = r\\ \lambda_1 e^{s} + \lambda_2 + \cdots + \lambda_r e^{s} = r\\ \cdots\\ \lambda_1 e^{s} + \lambda_2 e^{s} + \cdots + \lambda_r = r\end{cases}$$

解得 $\lambda_i = \dfrac{r}{1+(r-1)e^{s}},\quad i=1, 2, \cdots, r$.

(2) 由式(7.52)确定出 q_j

$$\begin{cases}q_1 + q_2 e^{s} + \cdots + q_r e^{s} = \dfrac{1+(r-1)e^{s}}{r}\\ q_1 e^{s} + q_2 + \cdots + q_r e^{s} = \dfrac{1+(r-1)e^{s}}{r}\\ \cdots\\ q_1 e^{s} + q_2 e^{s} + \cdots + q_r = \dfrac{1+(r-1)e^{s}}{r}\end{cases}$$

解得 $p(y_j) = q_j = \dfrac{1}{r},\ j=1, 2, \cdots, r$.

(3) 将 λ_i, q_j 代入式(7.55)得 $D(s)$

$$D(s) = \frac{(r-1)e^{s}}{1+(r-1)e^{s}}$$

所以, $s = \log_2 \dfrac{D}{(r-1)(1-D)}$.

(4) 将 s 代入式(7.56)得 $R(D)$

$$R(D) = sD + \sum_{i=1}^{r} p_i \log_2 \lambda_i$$

$$= \log_2 r - D\log_2(r-1) + D\log_2 D + (1-D)\log_2(1-D)$$

$$= \log_2 r - D\log_2(r-1) - H(D)$$

$R(D)$ 的定义域为 $0 \leq D \leq 1 - 1/r$, 值域为 $0 \leq R(D) \leq \log_2 r$.

对于不同的 r 值可以得到一组 $R(D)$ 曲线. 对于给定的平均失真度 D, r 越大, $R(D)$ 越大, 信源可压缩性越小; 反之, r 越小, $R(D)$ 越小, 信源可压缩性越大, 如图 7.3 所示.

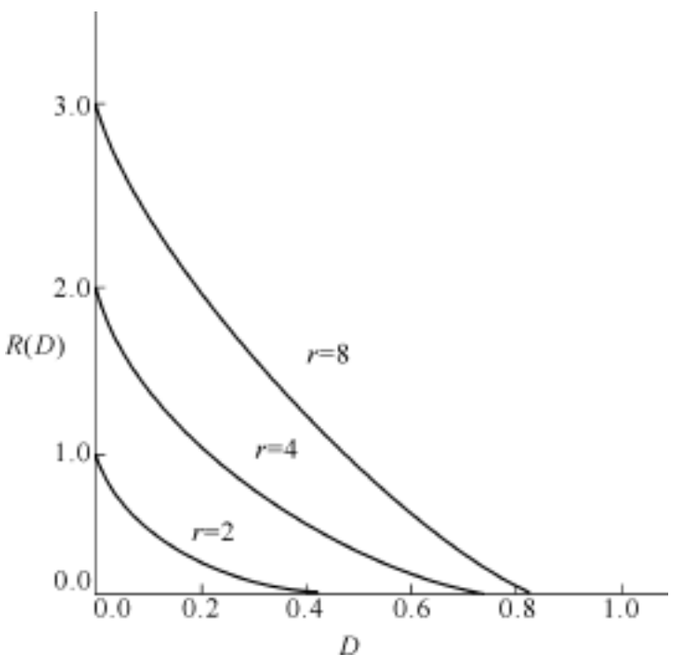


图 7.3　不同 r 值对应的信息率失真函数

因此信息率失真函数可以用于比较编码方法的压缩效果.

【例 7.8】

设信源符号集 $X = \{x_1, x_2, \cdots, x_{2r}\}$, 概率分布为 $p(x_i) = 1/2r$, $i = 1, 2, \cdots, 2r$, 失真函数选为 $d(x_i, y_j) = \begin{cases} 0 & x_i = y_j \\ 1 & x_i \neq y_j \end{cases}$. 当允许的失真度为 1/2 时求相应的信息率.

解

由信源概率分布求得信源熵

$$H(X) = H\left(\frac{1}{2r}, \frac{1}{2r}, \cdots, \frac{1}{2r}\right) = \log_2 2r$$

如果对信源进行二元无失真编码, 平均每个符号至少需要 $\log_2 2r$ 个二元码.

当允许的失真度为 1/2 时, 平均每个符号需要的码元个数可以减少到什么程度呢?

假设采用如下编码方法:

(1) 当信源输出符号为 $x_1, x_2, \cdots, x_r$ 时,分别赋予一个码字 $y_1, y_2, \cdots, y_r$;

(2) 当信源输出符号为 $x_{r+1}, x_{r+2}, \cdots, x_{2r}$ 时,都赋予相同的一个码字 y_r .

即试验信道的输入符号集 $X=\{x_1, x_2, \cdots, x_{2r}\}$,输出符号集 $Y=\{y_1, y_2, \cdots, y_r\}$,转移概率矩阵

$$P=\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & & \cdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \cdots & & & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

由平均失真的定义可求得平均失真度为

$$\begin{aligned} \overline{D} &= E[d(x_i, y_j)] \\ &= \sum_{i=1}^{2r}\sum_{j=1}^{r} p(x_i)\,p(y_j \mid x_i)\,d(x_i, y_j) \\ &= \sum_{i=r+1}^{2r} p(x_i)\,p(y_r \mid x_i)\,d(x_i, y_r) \\ &= \sum_{i=r+1}^{2r} p(x_i) \\ &= \frac{1}{2} \end{aligned}$$

所以上述编码方法满足保真度准则 $\overline{D} \leq D$.假设该信道具有归并性能的无噪性能,故 $H(Y \mid X)=0$,所以, $I(X;Y)=H(Y)-H(Y \mid X)=H(Y)$.

由输入概率和转移概率求输出概率分布:

$$\begin{cases} p(y_j)=\dfrac{1}{2r}, j=1,2,\cdots,r-1 \\ p(y_r)=\dfrac{r+1}{2r} \end{cases}$$

所以, $H(Y)=H\left(\dfrac{1}{2r}, \dfrac{1}{2r}, \cdots, \dfrac{1}{2r}, \dfrac{r+1}{2r}\right)=\log_2 2r-\dfrac{r+1}{2r}\log_2(r+1)$即采用上述编码方案时,平均每个符号所需的二元码符号个数由原来的 $\log_2 2r$ 减少到 $\log_2 2r-\dfrac{r+1}{2r}\log_2(r+1)$,减少了 $\dfrac{r+1}{2r}\log_2(r+1)$ 个码元 .换句话说,信源的信息率由原来的 $\log_2 2r$ 压缩到了 $\log_2 2r-\dfrac{r+1}{2r}\log_2(r+1)$,即信息率压缩了 $\dfrac{r+1}{2r}\log_2(r+1)$.这是不是达到 $R(1/2)$ 了呢?

可以求出 $D=1/2$ 时，

$$R\left(\frac{1}{2}\right)=\log_2 2r-\frac{1}{2}\log_2(2r-1)-1$$

当 $r>1$ 时，$I(X;Y)=H(Y)>R(1/2)$，所以存在更好的压缩编码方案能够进一步进行压缩，达到更好的压缩效果．

信息率失真理论给出了在给定的失真度 D 条件下，信源输出的信息率所能压缩的极限 $R(D)$，没有给出具体的压缩方法．但是它可以作为一种尺度，衡量一种压缩编码的方法的压缩效果．

7.3.2　二元信源和离散等概信源的 $R(D)$ 函数

对于汉明失真的离散信源，除了参量表示式以外还可以运用一些技巧来求解 $R(D)$．下面分别针对二元信源和离散等概信源的计算来说明如何应用这些技巧．

【例 7.9】

二元信源 $\begin{bmatrix}X\\P(X)\end{bmatrix}=\begin{bmatrix}0 & 1\\p & 1-p\end{bmatrix}$，$p\leq 1/2$，接收变量 $Y=\{0,1\}$，汉明失真矩阵 $D=\begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}$，求 $R(D)$．

解

$D_{\min}=0$ 且满足该最小失真的试验信道是一个无噪无损信道．

$$P=\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$$

$$R(D_{\min})=R(0)=H(X)=H(p)$$

$$D_{\max}=\min_j\sum_i p(x_i)d(x_i,y_j)=\min(1-p,p)=p$$

达到最大允许失真度的试验信道为

$$P=\begin{pmatrix}0 & 1\\0 & 1\end{pmatrix}$$

即这个试验信道能正确传送信源符号 $X=1$，当传送 $X=0$ 时，接收符号为 1，有失真．$X=0$ 出现的概率为 p，所以信道的平均失真度为 p．

这时 $R(D_{\max})=R(p)=0$．

当 $0<D<D_{\max}=p$ 时，平均失真度为

$$\overline{D}=E[d(x_i,y_j)]=\sum_{ij}p(x_iy_j)d(x_i,y_j)$$

$$= p(x=0, y=1) + p(x=1, y=0) = P_E$$

即在汉明失真度的情况下,平均失真度等于信道传输的平均错误概率 .

此时,选取一信道使 $\overline{D} = D$

$$I(X;Y) = H(X) - H(X|Y) = H(p) - H(X|Y)$$

根据费诺不等式,当 $r=2$ 时

$$H(X|Y) \leq H(P_E) = H(D) .$$

所以, $I(X;Y) \geq H(p) - H(D)$.

根据 $R(D)$的定义,满足保真度的试验信道的 $I(X;Y)$的最小值就是 $R(D)$.为了证实这一点,必须找到一个试验信道 .使其平均失真度 $\overline{D} \leq D$,而 $I(X;Y)$达到这个最小值,即 $I(X;Y) = R(D) = H(p) - H(D)$.

也就是 $H(X|Y) = H(D)$根据信道疑义度的定义可选取一个试验信道,它的反向信道矩阵如下:

$$\begin{bmatrix} 1-D & D \\ D & 1-D \end{bmatrix}$$

$$p(y_1) = \frac{p-D}{1-2D} \qquad p(y_2) = \frac{1-p-D}{1-2D}$$

$$0 < D < p \leq \frac{1}{2}$$

$$0 < p(y_j) < 1$$

所设的试验信道是存在的 .并且

$$\begin{aligned} \overline{D} &= E[d(x_i, y_j)] \\ &= \sum_{ij} p(x_i y_j)\, d(x_i, y_j) \\ &= \frac{D(1-p-D)}{1-2D} + \frac{D(p-D)}{1-2D} \\ &= D \end{aligned}$$

$$I(X;Y) = H(X) - H(X|Y) = H(p) - H(D) = R(D)$$

所以所选的试验信道正是满足平均失真 $\overline{D} \leq D$,而平均互信息达到最小值的信道 .

在汉明失真下,二元信源的

$$R(D) = \begin{cases} H(p) - H(D) & 0 \leq D \leq p \\ 0 & D > p \end{cases}$$

【例 7 .10】

离散等概信源的 $R(D)$函数:

信源 $X = \{x_1, x_2, \cdots, x_r\}$等概分布 $p(x_i) = 1/r$,信源输出符号为 $Y = \{y_1, y_2, \cdots, y_r\}$,汉明失真度定义为 $d(x_i, y_j) = \begin{cases} 0 & x_i = y_j \\ 1 & x_i \neq y_j \end{cases}$.求 $R(D)$.

解

因为在汉明失真度定义下,平均失真度

$$\overline{D}=E[d(x_i,y_j)]=\sum_{ij}p(x_iy_j)d(x_i,y_j)=\sum_{x_i\neq y_j}p(x_iy_j)=P_E$$

即平均失真度等于平均错误概率 P_E,可计算得

$$D_{min}=0 \qquad R(0)=H(X)$$

$$D_{max}=1-\frac{1}{r} \qquad R(D_{max})=0$$

$R(D)$的定义域为 $0\leq D\leq 1-\frac{1}{r}$.

我们选择一试验信道,使 $\overline{D}=D$而 $I(X;Y)=H(X)-H(X|Y)=\log_2 r-H(X|Y)$.

根据费诺不等式

$$H(X|Y)\leq H(P_E)+P_E\log_2(r-1)=H(D)+D\log_2(r-1)$$

$$I(X;Y)\geq\log_2 r-H(D)-D\log_2(r-1)$$

找到一个试验信道满足 $\overline{D}\leq D$ 而 $I(X;Y)$达到这个最小值,就可以证明这个值就是 $R(D)$ 根据信道疑义度的概念我们选择一个试验信道,它的反向信道如下:

$$p(x_i|y_j)=\begin{cases}1-D & i=j\\ \dfrac{D}{r-1} & i\neq j\end{cases}$$

可算出 $p(y_j)=\frac{1}{r}$, " j.

在该试验信道中

$$\overline{D}=E[d(x_i,y_j)]=\sum_{x_i\neq y_j}p(x_iy_j)$$

而

$$H(X|Y)=H(1-D,\frac{D}{r-1},\cdots,\frac{D}{r-1})$$

$$=-\left[(1-D)\log_2(1-D)+D\log_2\frac{D}{r-1}\right]$$

$$=H(D)+D\log_2(r-1)$$

所以

$$I(X;Y)=H(X)-H(X|Y)$$

$$=\log_2 r-H(D)-D\log_2(r-1)$$

$$=R(D)$$

这正是 $\overline{D}\leq D$ 而 $I(X;Y)$达到最小值的信道.

因此在汉明失真度下, r 元等概信源的信息率失真函数为

$$R(D)=\begin{cases}\log_2 r - D\log_2(r-1) - H(D) & 0\leq D\leq 1-\dfrac{1}{r}\\ 0 & D>1-\dfrac{1}{r}\end{cases}$$

从以上两个例子看出,由于采用了汉明失真度,平均失真度就等于平均错误概率,使得求解 $R(D)$的过程简化,结果与参量表示式中求得的结果是一样的.但汉明失真度不是一种最合理的失真函数.实际应用中,设计一种符合信源主观要求的,合理的失真函数是很重要的.

7.4 限失真信源编码定理和逆定理

7.4.1 限失真信源编码定理

对于无失真信源编码来说,每一个信源符号(或符号序列)必须对应一个码字(或码字序列),信源输出信息率不能减少.而在允许一定失真的情况下,信源输出信息率最少可减少到信息率失真函数 $R(D)$,有可能是多个信源符号(符号序列)对应一个码字(码字序列).限失真信源编码定理就是关于信息率和失真关系的一个极限定理,也称香农第三定理,保真度准则下的离散信源编码定理.

定理 7.1 设 $R(D)$是离散无记忆信源的信息率失真函数并且失真函数为有限值.对于任意的允许失真度 $D\geq 0$ 和任意小的正数 $\varepsilon>0$,当信源序列长度 N 足够长时,一定存在一种编码 C_K,其码字个数 $M\leq\exp\{N[R(D)+\varepsilon]\}$,而编码后的平均失真度 $\overline{D}(C_K)\leq D+\varepsilon$.

证明

现定义一个含有 M 个码字的信源编码 C_K,这个码的每一个码字都是随机产生的.

$$C_K=\{y_1,y_2,\cdots,y_M\} \tag{7.59}$$

它是离散无记忆信道的 N 次扩展信道的输出序列集 $Y^N=Y_1Y_2\cdots Y_N=\{y_1,y_2,\cdots,y_s\}$的一个子集.用码 C_K 对离散无记忆信源的 N 次扩展信源 $X^N=X_1X_2\cdots X_N\in\{x_1,x_2,\cdots,x_{r^N}\}$中的每一个输入符号序列 x_i 进行编码.把 $x_i(i=1,2,\cdots,r^N)$变换成与之失真最小的 y_j, $j=1,2,\cdots,M$,即

$$f(x_i)=y_j \quad x_i\in\{x_1,x_2,\cdots,x_{r^N}\},\ j=1,2,\cdots,M \tag{7.60}$$

并且

$$d(x_i,\Delta)=\min_{y_j\in C_K} d(x_i,y_j) \tag{7.61}$$

把这种编码方法看作一个假设信道.信道输入为 $x_i(i=1,2,\cdots,r^N)$,输出码字 y_j,其传递概率

$$p_O(x_i y_j)=\begin{cases}1 & \text{当 } y_j\in C_K \text{ 且 } d(x_i,y_j)=d[x_i,f(x_i)]\\ 0 & \text{其他}\end{cases} \tag{7.62}$$

这时,码 C_K,中每一个码符号与信源符号之间的平均失真度

$$\begin{aligned}\overline{D}_N(C_K)&=\frac{1}{N}\overline{D}(N)\\&=\frac{1}{N}\sum_{i=1}^{r^N}\sum_{j=1}^{s^N}p(x_i)p_O(y_j\mid x_i)d(x_i,y_j)\\&=\frac{1}{N}\sum_{i=1}^{r^N}p(x_i)d[x_i,f(x_i)]\end{aligned} \tag{7.63}$$

因为码 C_K 是由随机选择的码字组成,并且每个码字的选取是彼此独立的,所以码 C_K 的发出概率为

$$P_r\{C_K\}=p(y_1y_2\cdots y_M)=p(y_1)p(y_2)\cdots p(y_M)=\prod_{j=1}^{M}p(y_j) \tag{7.64}$$

$$\begin{aligned}\overline{D}_N(C)&=\sum_K p(C_K)\overline{D}_N(C_K)\\&=\sum_K p(C_K)\frac{1}{N}\sum_{i=1}^{r^N}p(x_i)d[x_i,f(x_i)]\end{aligned} \tag{7.65}$$

把信源符号序列集分成两个子集:

$$\begin{aligned}U&:\{x_i:d[x_i,f(x_i)]\le N(D+\delta)\}\\\overline{U}&\{x_i:d[x_i,f(x_i)]>N(D+\delta)\}\end{aligned} \tag{7.66}$$

其中,δ是任意小的正数.这样

$$\overline{D}_N(C)=\sum_K p(C_K)\cdot\frac{1}{N}\left\{\sum_{x_i\in U}p(x_i)d[x_i,f(x_i)]+\sum_{x_i\in\overline{U}}p(x_i)d[x_i,f(x_i)]\right\} \tag{7.67}$$

根据子集 U 的定义,显然有

$$\sum_{x_i\in U}p(x_i)d[x_i,f(x_i)]\le N(D+\delta) \tag{7.68}$$

令 $d_{\max}$是失真函数 $d(x_i,y_j)$的最大值,则有

$$\max[d(x_i,y_j)]=\max\left[\sum_{k=1}^{N}d(x_{i_k},y_{j_k})\right]\le\sum_{k=1}^{N}\max[d(x_{i_k},y_{j_k})]\le Nd_{\max} \tag{7.69}$$

在子集 $\overline{U}$ 中 $N(D+\delta)<d[x_i,f(x_i)]<Nd_{\max}$,

$$\overline{D}_N(C)=\sum_K p(C_K)\frac{1}{N}\cdot N(D+\delta)+\sum_K p(C_K)\frac{1}{N}\cdot Nd_{\max}\cdot\sum_{x_i\in\overline{U}}p(x_i)$$

$$= D + \delta + d_{\max} \sum_{K} p(C_K) \sum_{x_i \in \overline{U}} p(x_i) \tag{7.70}$$

$x_i \in \overline{U}$意味着 x_i 与 C_K 中每一个码字序列 $y_j(j=1,2,\cdots,M)$的失真均大于 $N(D+\delta)$.

设一个示性函数

$$\Delta(x_i, y_j) = \begin{cases} 1 & \text{若 } d[x_i, f(x_i)] \le N(D+\delta) \\ 0 & \text{若 } d[x_i, f(x_i)] > N(D+\delta). \end{cases} \tag{7.71}$$

则

$$\sum_{x_i \in \overline{U}} p(x_i) = \sum_{i=1}^{r^N} p(x_i)[1 - \Delta(x_i, y_1)][1 - \Delta(x_i, y_2)]\cdots[1 - \Delta(x_i, y_M)]$$

$$= \sum_{i=1}^{r^N} p(x_i) \prod_{j=1}^{M} [1 - \Delta(x_i, y_j)] \tag{7.72}$$

这样把有限制的求和号变成无限制的求和号 所以有

$$\overline{D}_N(C) \le D + \delta + d_{\max}\left\{ \sum_{i=1}^{r^N} p(x_i) \sum_{y_1 \in Y^N} \cdots \sum_{y_M \in Y^N} \prod_{j=1}^{M} p(y_j)[1 - \Delta(x_i, x_i)] \right\} \tag{7.73}$$

在数学上若某函数 $f(x)$定义在集合 L 上,则

$$\left[\sum_{x \in L} f(x)\right]^M = \sum_{x_1 \in L} \cdots \sum_{x_M \in L} f(x_1) f(x_2) \cdots f(x_M) \tag{7.74}$$

式(7.72)可写成

$$\overline{D}_N(C) \le D + \delta + d_{\max} \sum_{i=1}^{r^N} p(x_i)\left[\sum_{y_j \in Y^N} p(y_j)[1 - \Delta(x_i, y_j)]\right]^M$$

$$= D + \delta + d_{\max} \sum_{i=1}^{r^N} p(x_i)\left[1 - \sum_{y_j \in Y^N} p(y_j)\Delta(x_i, y_j)\right]^M \tag{7.75}$$

要对方括号内求和号的上限进行估计,为此再定义一个示性函数

$$D_O(x_i, y_j) = \begin{cases} 1 & \text{当 } d(x_i, y_j) \le N(D+\delta) \text{且 } I(x_i; y_j) \le N[R(D) + \delta] \\ 0 & \text{其他} \end{cases} \tag{7.76}$$

因为 $d(x_i, y_j) \le N(D+\delta)$,且 $I(x_i; y_j) \le N[R(D) + \delta]$的$(x_i, y_j)$只能满足 $d(x_i, y_j) \le N(D+\delta)$的集合的一部分,所以

$$\sum_{y_j \in Y^N} p(y_j) D_O(x_i, y_j) \le \sum_{y_j \in Y^N} p(y_j) \Delta(x_i, y_j) \tag{7.77}$$

$$\left[1 - \sum_{y_j \in Y^N} p(y_j) \Delta(x_i, y_j)\right]^M \le \left[1 - \sum_{y_j \in Y^N} p(y_j) D_O(x_i, y_j)\right]^M \tag{7.78}$$

当 $D_O(x_i, y_j) = 1$ 时, $I(x_i; y_j) \le N[R(D) + \delta]$,即

$$\ln \frac{p(y_j \mid x_i)}{p(y_j)} \leq N[R(D)+\delta] \tag{7.79}$$

$$p(y_j) \geq p(y_j \mid x_j)\exp\{-N[R(D)+\delta]\} \tag{7.80}$$

因此

$$\left[1-\sum_{y_j\in Y^N} p(y_j)\Delta(x_i,y_j)\right]^M \leq \left[1-\sum_{y_j\in Y^N} p(y_j)D_O(x_i,y_j)\right]^M$$

$$\leq \left[1-\mathrm{e}^{-N[R(D)+\delta]}\sum_{y_j\in Y^N} p(y_j\mid x_i)D_O(x_i,y_j)\right]^M \tag{7.81}$$

利用一个有趣的不等式：

当 $x\geq 0, y\leq 1, M>0$ 时，$(1-xy)^M \leq 1-x+\mathrm{e}^{-yM}$.令

$$x=\sum_{y_j\in Y^N} p(y_j\mid x_i)D_O(x_i,y_j)$$

$$y=\mathrm{e}^{-N[R(D)+\delta]}$$

符合 $x\geq 0, y\leq 1$ 的条件，所以有

$$\left[1-\sum_{y_j\in Y^N} p(y_j)\Delta(x_i,y_j)\right]^M \leq 1-\sum_{y_j\in Y^N} p(y_j\mid x_i)D_O(x_i,y_j)+\exp\{-\mathrm{e}^{-N[R(D)+\delta]}\cdot M\} \tag{7.82}$$

若取 $M=\mathrm{e}^{N[R(D)+2\delta]}$，当 $N\to\infty$ 时，$\exp\{-\mathrm{e}^{-N[R(D)+\delta]}\cdot M\}=\exp\{-\mathrm{e}^{N\delta}\}\to 0$

选择适当的 δ 可使 $\exp(-\mathrm{e}^{N\delta})\leq\dfrac{\delta}{d_{\max}}$，则

$$\left[1-\sum_{y_j\in Y^N} p(y_j)\Delta(x_i,y_j)\right]^M \leq 1-\sum_{y_j\in Y^N} p(y_j\mid x_i)D_O(x_i,y_j)+\exp(-\mathrm{e}^{-N\delta})$$

$$\leq 1-\sum_{y_j\in Y^N} p(y_j\mid x_i)D_O(x_i,y_j)+\frac{\delta}{d_{\max}} \tag{7.83}$$

$$\overline{D}_N(C) \leq D+\delta+d_{\max}\sum_{i=1}^{r^N} p(x_i)\left[1-\sum_{y_j\in Y^N} p(x_i)\Delta(x_i,y_j)\right]^M$$

$$\leq D+\delta+d_{\max}\left\{\sum_{i=1}^{r^N} p(x_i)-\sum_{i=1}^{r^N}\sum_{j=1}^{s^N} p(x_i)p(y_j\mid x_i)D_O(x_i,y_j)+\sum_{i=1}^{r^N} p(x_i)\frac{\delta}{d_{\max}}\right\}$$

$$= D+2\delta+d_{\max}\left[1-\sum_{i=1}^{r^N}\sum_{j=1}^{s^N} p(x_i,y_j)D_O(x_i,y_j)\right]$$

$$= D+2\delta+d_{\max}\left\{\sum_{i=1}^{r^N}\sum_{j=1}^{s^N} p(x_i,y_j)[1-D_O(x_i,y_j)]\right\} \tag{7.84}$$

上式花括号内是对所有满足 $d(x_i,y_j)>N(D+\delta)$ 或者 $I(x_i,y_j)>N[R(D)+\delta]$ 的

(x_i, y_j)求和,满足条件 $d(x_i, y_j) > N(D+\delta)$的$(x_i, y_j)$集称为 V,满足 $I(x_i, y_j) > N[R(D)+\delta]$的$(x_i, y_j)$集称为 W.

$$\overline{D}_N(C) \leq D + 2\delta + d_{\max}[P_r(V) + P_r(W)] \tag{7.85}$$

$$d(x_i, y_j) = \sum_{k=1}^{N} d(x_{i_k}, y_{j_k}) \tag{7.86}$$

式(7.86)是服从同一概率分布的 N 个彼此统计独立的随机变量之和,并且每个随机变量 $d(x_i, y_j)$的平均失真度 $\overline{D} \leq D$ 根据弱大数定理

$$\lim_{N\to\infty} P_r(V) = \lim_{N\to\infty} P_r[d(x_i, y_j) > N(D+\delta)] = 0 \tag{7.87}$$

同理,因为 $I(x_i, y_j) = \sum_k NI(x_{i_k}, y_{j_k})$,

每个随机变量是 $I(x_i, y_j)$的统计平均

$$I(X;Y) = E[I(x_i, y_j)] \leq R(D) \tag{7.88}$$

根据弱大数定理

$$\lim_{N\to\infty} P_r(W) = \lim_{N\to\infty} P_r\{I(x_i, y_j) > N[R(D)+\delta]\} = 0 \tag{7.89}$$

对于足够长的 N,可使

$$P_r(V) < \frac{\delta}{d_{\max}} \tag{7.90}$$

$$P_r(W) < \frac{\delta}{d_{\max}} \tag{7.91}$$

所以

$$\overline{D}_N(C) \leq D + 4\delta \tag{7.92}$$

令 $\varepsilon = 4\delta$,得

$$\overline{D}_N(C) \leq D + \varepsilon \tag{7.93}$$

这时码字个数

$$M = e^{N[R(D)+\frac{\varepsilon}{2}]} \leq e^{N[R(D)+\varepsilon]} \tag{7.94}$$

所以,证得在随机选择的信源编码集中,每一码含有 $M \leq e^{N[R(D)+\varepsilon]}$ 个码字,并且平均失真度 $\overline{D}_N(C) \leq D+\varepsilon$ 那么,在随机信源编码集中至少有一个码 C_O,它的平均失真度 $\overline{D}_N(C) \leq \overline{D}_N(C_O) \leq D+\varepsilon$.因此只要码长 N 足够长,一定存在一种信源编码,信息率为 $R(D)$,而码的平均失真度 $\overline{D} \leq D+\varepsilon$.

证毕

7.4.2 限失真信源编码逆定理

定理 7.2 不存在平均失真度 D 而信息传输率 $R' < R(D)$ 的任何信源编码.即对任

意码长为 N 的信源码 C, 若码字个数 $M < e^{NR(D)}$, 则 $\overline{D}_N(C) > D$.

逆定理说明, 如果编码后平均每个信源符号的信息传输率 $R' < R(D)$, 就不能在保真度准则下再现信源的信息.

证明

设存在一各信源码 $C_K = y_1, y_2, \cdots, y_M)$, $M < \mathrm{e}^{NR(D)}$, 它能使 $\overline{D}(C_K) \leq D$, 信源码 y_j 与信源序列 x_i 的变换关系仍采用

$$d[x_i, f(x_i)] = \min_{y_j \in C_K} d(x_i, y_j) \tag{7 95}$$

$$f(x_i) = y_j \tag{7 96}$$

把这种编码方法看成一个信道 $p_0(y_j \mid x_i)$

$$p_0(y_j \mid x_i) = \begin{cases} 1 & \text{当 } y_j \in C_K \text{ 且 } d(x_i, y_j) = d[(x_i, f(x_i)] \\ 0 & \text{其他} \end{cases} \tag{7 97}$$

又因为在这个信道中, 噪声熵 $H(Y \mid X) = 0$, 所以

$$I(X; Y) = H(Y) \leq \log_2 M \tag{7 98}$$

根据假设 $\overline{D}_N(C_K)$, 可得

$$\begin{aligned} \sum_{h=1}^{N} I(X_h; Y_h) &\geq \sum_{h=1}^{N} R(D_h) \\ &\geq N \sum_{h=1}^{N} \frac{1}{N} R(D_h) \\ &\geq NR\left(\frac{1}{N} \sum_{h=1}^{N} D_h\right) \\ &\geq NR(D) \end{aligned} \tag{7 99}$$

因此

$$R' = \frac{1}{N} \log_2 M \geq R(D) \tag{7 .100}$$

即

$$M \geq \mathrm{e}^{NR(D)} \tag{7 .101}$$

这个结果与本定理的假设矛盾. 所以不存在失真度小于 D 的编码其信息率小于 $R(D)$.

证毕

限失真信源编码定理也是一个极限存在定理, 不能像无失真信源编码定理那样从证明过程中引出概率匹配的编码方法. 一般只能从优化的思路去求最佳编码, 至今尚无合适的可实现的编码方法来接近 $R(D)$ 这个极限. 常用的限失真信源编码有量化编码、预测编码、变换编码.

7.5 熵压缩编码具体方法

前面讲解了熵压缩编码的理论基础——信息率失真函数和限失真编码定理,下面介绍几种常用的熵压缩编码算法.

7.5.1 标量量化

对连续无记忆信源进行熵压缩编码的常用方法是标量量化.设标量量化器的输出符号共有 M 个:$\{q_1, q_2, \cdots, q_M\}$,量化的过程为将取值为任意实数的信源输出符号 x 映射成 M 个可能的离散值中的一个.该映射是通过 $M-1$ 个门限 $T_1 < T_2 < \cdots < T_{M-1} < T_M$ 实现的.

$$q(x)=\begin{cases} q_1 & x \leq T_1 \\ q_k & T_{k-1} < x \leq T_k,\ k=2,3,\cdots,M-1 \\ q_M & x > T_{M-1} \end{cases} \tag{7.102}$$

标量量化器的输入符号序列为 $\alpha=(x_1, x_2, \cdots, x_N)$,其中 $x_n \in (-\infty, +\infty)$,其输出序列共有 M^N 种,记为 $\beta=(y_1, y_2, \cdots, y_N)$,其中 $y_n \in \{q_1, q_2, \cdots, q_M\}$.每个符号 q_k 出现的概率为

$$p(q_k)=\int_{T_{k-1}}^{T_k} p(x)\mathrm{d}x \tag{7.103}$$

由于此映射为多对一映射,所以通过该量化器传输的信息率 R_M 为

$$\begin{aligned} R_M &= \frac{1}{N} I(X^N; Y^N) \\ &= I(X; Y) \\ &= H(Y) - H(Y|X) \\ &= H(Y) \\ &= -\sum_{k=1}^{m} p(q_k)\log_2 p(q_k) \end{aligned} \tag{7.104}$$

而量化带来的平均失真 D_M 为

$$\begin{aligned} D_M &= E\left[\frac{1}{N}\sum_{n=1}^{N} d(X_n, Y_n)\right] \\ &= \frac{1}{N}\sum_{n=1}^{N} E[d(X_n, Y_n)] \\ &= E[d(X, Y)] \end{aligned}$$

$$= \sum_{k=1}^{M} \int_{T_{k-1}}^{T_k} d(x, q_k)\, p(x)\, \mathrm{d}x \tag{7.105}$$

一般来讲,标量量化器输出符号的概率分布不一定是均匀分布.量化器输出的最大可能比特速率 R_M^* 为

$$R_M^* = \log_2 M \tag{7.106}$$

信息传输速率 R_M、平均失真 D_M 和二元符号速率 R_M^* 是标量量化器的 3 个主要性能指标,选择不同的 T_k 和 q_k,量化器将有不同的 R_M, D_M 和 R_M^*.

7.5.2　矢量量化

在前面所述的标量量化中,我们总是对每个信源符号单独处理,这样的处理方法忽略了信源符号之间的相关性,因此信源的冗余度没有得到有效的压缩.通过将多个符号构成的序列作为一个整体进行量化的方法可以避免标量量化的固有缺点,这种方法称为矢量量化.

先将待编码的信源符号序列划分为一个个等长的分组,每个分组含有若干个符号,形成一个矢量.每个矢量与预先生成的矢量码书中的每个码字进行比较,求出相应的失真,然后用失真最小的码字的编号作为量化器的输出就实现了矢量量化.

由于在译码端有一个同样的码书,所以译码工作只是通过接收的码字序号在码书中搜索相应的码字,算法简单,容易实现.

矢量量化的关键在于矢量码书的构造,目前最流行的算法是由 Y.Linde、A.Buzo 和 R.M.Gray 共同提出的 LBG 算法:

(1) 采集用于构造码书的训练数据,为了得到性能较好的矢量码书,一般要求训练样本的数量 N 和码字的个数 L 满足: $N \geq 20L$;

(2) 构造初始码书,常用的构造方法有随机码书法、白噪声码书法等等;

(3) 按照初始码书对所有的训练样本进行矢量量化,得到分属不同码字的 L 个样本集合和相应的量化误差;

(4) 对每个样本集合进行聚类,根据聚类的结果修正初始码字,得到新的码书;

(5) 判断量化误差是否小于门限值、迭代次数是否超出规定值,若是,则训练结束,否则转第 3 步继续.

矢量量化中第二个重要的问题是量化误差的度量问题,即如何选择一个能准确反映量化前后失真大小的度量函数.这个问题与被编码的对象的具体特性有关,没有一个统一的答案.

矢量量化的第三个问题是搜索运算量问题.在量化的过程中,需要计算待量化矢量与所有码字的失真,然后选取失真最小的码字.当码书的规模较大时需要很大的计算量.可以通过为码书安排某种特定的结构(如二叉树)来有效地降低运算量.

矢量量化具有如下特点:

(1) 矢量维数 N 越大,量化失真越小;

(2) 码字个数 L 越大,量化失真越小,相应的存储空间开销和搜索时间开销也越大 .

7 .5 .3 变换编码

为了去除信源符号之间的相关性,通常将信源符号序列划分为较长的分组,然后对每个分组进行统一地编码 这仅仅是去除相关性的最简单的一种方法 .还可以通过某种数学变换来去除信源符号之间的相关性 .

将信源符号经过数学变换后再进行编码就称为变换编码 .常用的数学变换有傅里叶(Fourier)变换、小波(Wavelet)变换、沃尔什-哈达玛(Walsh-Hadamard)变换、K-L(Karhunen-Loeve)变换、余弦变换、正弦变换等等 .其中,K-L 变换去相关性最好,但算法复杂,实现困难;而离散余弦变换(DCT)的性能接近 K-L 变换,也容易实现,被选为众多图像压缩编码技术标准的基本算法 .

*7 .5 .4 预测编码

预测编码的概念最早是由 Peter Elias 在 1955 年提出的,它是目前得到广泛应用的一种实用的熵压缩树码 .

在预测编码的编码器和译码器中都存储有过去的符号值,并以此来预测或估计未来符号的值;编码器的输出不是信源符号本身,而是实际信源符号与预测值之间的差值;在译码端,译码器将接收到的这一差值与译码器的预测值相加,从而恢复信源符号 .预测编码的编译码过程如图 7 .4 和 7 .5 所示 .

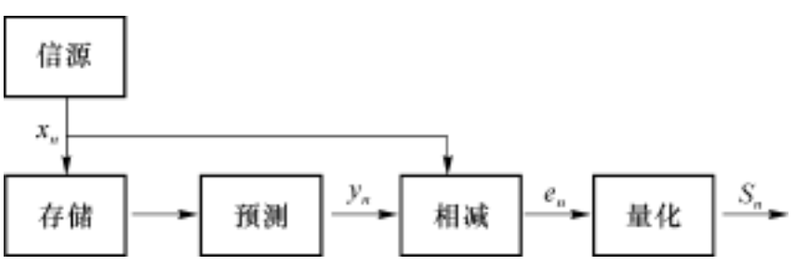


图 7 .4 预测编码器

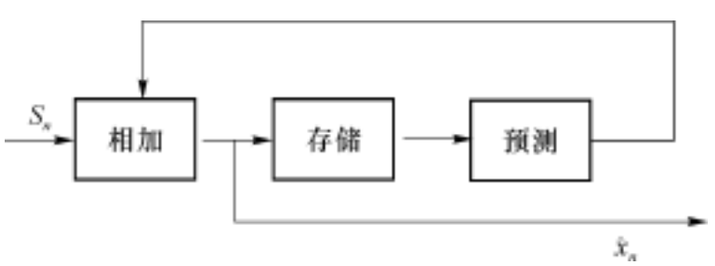


图 7 .5 预测译码器

根据预测器中预测值与信源符号过去时刻的值之间的函数关系,可将预测器分为线性预测器和非线性预测器,相应的预测编码称为线性预测编码和非线性预测编码 .

设信源输出的符号序列为$(\cdots, x_{-2}, x_{-1}, x_0, x_1, x_2, \cdots)$,预测器根据所存储的过去的符号值对时刻 n 的符号进行预测,得到的预测值 y_n 与实际值 x_n 之间的差值称为预测误差 e_n:

$$e_n = y_n - x_n \tag{7.107}$$

预测编码设计中的核心问题是如何选取预测函数以使预测误差 e_n 满足某种最佳条件 .常见的几种条件是最小均方误差、最小平均决定误差和最大零误差概率 .

不同准则下的最佳预测给出不同的预测值和不同的预测误差值,但是预测误差的分布与信源符号的一维分布具有相同的形状,其差别只是因为减去预测值所导致的分布的平移 .

习　题　7

7.1　试证明对离散信源, $R(D=0) = H(X)$ 的充要条件是失真矩阵 D 的每行中至少有一个 0,而在每列中至多有一个 0 .

7.2　利用 $R(D)$ 的性质,画出一般 $R(D)$ 的曲线并说明其物理意义?试问为什么 $R(D)$ 是非负且非增的?

7.3　一随机变量 X 的符号集为一个无限可数的集合 χ,假设失真测度定义为 Hamming 失真,

(1) 试证明 $R(D) \geq H(X) - \log_2(|\chi| - 1) - H(D)$　" $0 \leq D \leq D_{\max}$;

(2) 试证明当 X 在 χ 上为均匀分布的情况下,上问中的下限是严格的 .

7.4　令 $\chi = \{0,1\} = \hat{\chi}$,设失真矩阵为

$$d(x, \hat{x}) = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$$

对于一个等概分布的 Bernoulli 随机变量,求 $R(D) \in (0,1)$ 对应的定义域 $(D_{\min}, D_{\max})$.

7.5　令 X 为等概分布的 Bernoulli 随机变量,相应的失真测度定义为

$$\chi = \{0,1\},\ d(x, \hat{x}) = \begin{pmatrix} 0 & 1 & \infty \\ \infty & 1 & 0 \end{pmatrix},\ \hat{\chi} = \{0, e, 1\}$$

(1) 请找出使得 $R(D)$ 非平凡的 $(D_{\min}, D_{\max})$;

(2) 计算 $R(D)$.

7.6　考虑一个离散信源 $X \in \chi = \{1,2,\cdots, m\}$,其分布为 $p_1, p_2, \cdots, p_m$,失真测度定义为 $d(i,j)$,令 $R(D)$ 为此信源和相应失真测度对应的信息率失真函数 .通过 $d'(i,j) =$

$d(i,j)-w_i$生成一个新的失真测度定义,相应的信息率失真函数变为 $R'(D)$.

(1) 试证明 $R'(D)=R(D+w')$,其中 $w'=\sum p_i w_i$;

(2) 根据上述结论说明我们可以通过归一化失真使得:对任意的 $x\in\chi$,至少存在一个符号 $\hat{x}$使得 $d(x,\hat{x})=0$.

7 .7 三元信源的概率分布为{0 .4,0 .4,0 .2},失真函数为

$$d(i,j)=\begin{cases}0, \text{当 } i=j \text{ 时}\\ 1, \text{当 } i\neq j \text{ 时}\end{cases}$$

试求:(1) 信息率失真函数 $R(D)$;

(2) 若此信源用容量为 1 比符/ 符号和 0 .1 比特/ 符号的信道传输,请分别计算出最小误码率 P_E .

附 录 A

上机作业

A.1 信道容量的迭代计算

已知:信源符号个数 r、信宿符号个数 s、信道转移概率矩阵 P.

算法:

1. **初始化信源分布**: $p_i = \frac{1}{r}$, **循环变量** $k = 1$, **门限** Δ, $C^{(0)} = -\infty$

2.

$$c_{ij}^{(k)} = \frac{p_i^{(k)} p_{ji}}{\sum_{i=1}^{r} p_i^{(k)} p_{ji}}$$

3.

$$p_i^{(k+1)} = \frac{\exp\left[\sum_{j=1}^{s} p_{ji} \log_2 c_{ij}^{(k)}\right]}{\sum_{i=1}^{r} \exp\left[\sum_{j=1}^{s} p_{ji} \log_2 c_{ij}^{(k)}\right]}$$

4.

$$C^{(k+1)} = \log_2\left[\sum_{i=1}^{r}\exp\left(\sum_{j=1}^{s} p_{ji}\log_2 {}^{(k)}_{ij}\right)\right]$$

5.若

$$\frac{\left|C^{(k+1)} - C^{(k)}\right|}{C^{(k+1)}} > \Delta$$

则 $k = k+1$,转第 2 步.

6.输出 $P^* = \left[p_i^{(k+1)}\right]_r$ 和 $C^{(k+1)}$,终止.

要求:

1.允许使用的编程语言:C、C++、Basic、Pascal、Fortran、JAVA、Perl、TK/TCL

2.输入:任意的一个信道转移概率矩阵.信源符号个数、信宿符号个数和每个具体的转移概率在运行时从键盘输入.

3.输出:最佳信源分布 P^*,信道容量 C.

4.源程序格式整齐清晰,注释简单明了.

A.2 唯一可译码判决准则

已知:信源符号个数 r、码字集合 C

算法:

1.考查 C 中所有的码字,若 W_i 是 W_j 的前缀,则将相应的后缀作为一个尾随后缀码放入集合 F_0 中;

2.考查 C 和 F_i 两个集合,若 $W_i \in C$ 是 $W_j \in F_i$ 的前缀或 $W_i \in F_i$ 是 $W_j \in C$ 的前缀,则将相应的后缀作为尾随后缀码放入集合 F_{i+1} 中;

3. $F = \bigcup_i F_i$ 即为码 C 的尾随后缀集合;

4.若 F 中出现了 C 中的元素,则算法终止,返回假(C 不是唯一可译码);否则若 F 中没有出现新的元素,则返回真.

要求:

1.允许使用的编程语言:C、C++、Basic、Pascal、Fortran、JAVA、Perl、TK/TCL

2.输入:任意的一个码.码字个数和每个具体的码字在运行时从键盘输入.

3.输出:判决(是唯一可译码/不是唯一可译码).

4.源程序格式整齐清晰,注释简单明了.

A.3　Huffman 编码

已知:信源符号个数 q、信源符号 $s_0, \cdots, s_{q-1}$,信源概率分布 $p_0, \cdots, p_{q-1}$.

算法:

1. 如果 $q=2$ 则返回编码: $s_0 \to 0$, $s_1 \to 1$
2. 否则
   (a) 重新排序 $s_0, \cdots, s_{q-1}$ 和 $p_0, \cdots, p_{q-1}$
   (b) 创建一个符号 s',其概率为 $p' = p_{q-2} + p_{q-1}$
   (c) 递归调用本算法以得到 $s_0, \cdots, s_{q-3}, s'$ 的编码 $w_0, \cdots, w_{q-3}, w'$,它的概率分布为 $p_0, \cdots, p_{q-3}, p'$
   (d) 返回编码: $s_0 \to w_0, \cdots, s_{q-3} \to w_{q-3}, s_{q-2} \to w'0, s_{q-1} \to w'1$

要求:

1. 允许使用的编程语言:C、C++、Basic、Pascal、Fortran、JAVA、Perl、TK/ TCL
2. 输入:信源符号个数、每个信源符号的概率分布在运行时从键盘输入.
3. 输出:每个信源符号及其对应的码字.
4. 源程序格式整齐清晰,注释简单明了.

A.4　LZW 编码

已知:待压缩的数据文件.

算法:参见正文 5.5.3 节.

要求:

1. 允许使用的编程语言:C、C++、Basic、Pascal、Fortran、JAVA、Perl、TK/ TCL
2. 输入:任意的数据文件.
3. 输出:压缩后的数据文件.
4. 源程序格式整齐清晰,注释简单明了.

附 录 B

数学预备知识

B.1 概率论简单回顾

设 X 是一个具有有限个或无限可数个元素的样本空间上的离散随机变量,则 $X\sim p(x)$ 表示 $p(x)$ 是 X 的概率分布函数:

$$p(x)=p_X(x)=P_r\{X=x\},\ "\ x\in X \tag{B.1}$$

X 为 X 的取值空间,一般来讲,$X=\{x_1,x_2,\cdots,x_n\}$ 或 $X=\{x_1,x_2,\cdots,x_n,\cdots\}$,此时,我们简记 $p(x_i)$ 为 p_i .

两个或多个随机变量的联合概率分布函数也可以简洁地表示为如下形式:

$$p(x\ y)=p_{XY}(x,y)=P[X=x,Y=y] \tag{B.2}$$

$$p(x\ y\ z)=p_{XYZ}(x,y,z)=P[X=x,Y=y,Z=z] \tag{B.3}$$

条件概率也可以简写为

$$\begin{aligned}p(y|x)&=p_{Y|X}(y|x)\\&=P[Y=y|X=x]\\&=\frac{P[X=x,Y=y]}{P[X=x]}\end{aligned}$$

$$= \frac{p(x, y)}{p(x)} \tag{B.4}$$

条件概率定义隐含了下述规则:

$$p(x\ y) = p(x)\,p(y|x) = p(y)\,p(x|y) \tag{B.5}$$

$$p(x_1\ x_2 \cdots\ x_n) = p(x_1)\,p(x_2|x_1)\cdots p(x_n|x_{n-1}) \tag{B.6}$$

若 X 和 Y 两个随机变量相互独立,则:

$$p(x\ y) = p(x)\cdot p(y) \tag{B.7}$$

$$p(y|x) = p(y) \tag{B.8}$$

$$p(x|y) = p(x) \tag{B.9}$$

B.2　Jensen 不等式

定义 B.1　如果对于任意的 $x_1, x_2 \in (a, b)$以及 $0 \leq \lambda \leq 1$,函数 $f(x)$均满足

$$f[\lambda x_1 + (1-\lambda)\,x_2] \geq \lambda f(x_1) + (1-\lambda)\,f(x_2) \tag{B.10}$$

则称此函数为上凸函数.如果等号仅当 $\lambda = 0$ 或 $\lambda = 1$ 时取得,则称函数 f 为严格上凸函数.

定义 B.2　如果 f 是上凸函数,则 $-f$ 为下凸函数.

定理 B.1　如果函数 f 存在非负的二阶导数,则此函数为下凸函数.

定理 B.2(Jensen 不等式)　如果 f 为一个上凸函数, X 为一个随机变量,则

$$E[f(X)] \leq f[E(X)] \tag{B.11}$$

证明

用数学归纳法进行证明,归纳的对象是离散概率分布的样值数量.

对于只有两个样值的概率分布来讲,不等式变为

$$p_1 f(x_1) + p_2 f(x_2) \leq f(p_1 x_1 + p_2 x_2) \tag{B.12}$$

根据上凸函数的定义我们可以直接得到式(B.10).

假设对于有 $k-1$ 个样值的概率分布,不等式(B.10)成立,则可以将有 k 个样值的概率分布改写为

$$p_1' = p_1/(1-p_k),\ p_2' = p_2/(1-p_k),\ \cdots,\ p_{k-1}' = p_{k-1}/(1-p_k),\ p_k$$

因此有

$$\sum_{i=1}^{k} p_i f(x_i) = p_k f(x_k) + (1-p_k)\sum_{i=1}^{k-1} p_i' f(x_i) \tag{B.13}$$

$$\leq p_k f(x_k) + (1-p_k)\,f\left[\sum_{i=1}^{k-1} p_i' x_i\right] \tag{B.14}$$

$$\leq f\left[p_k x_k + (1 - p_k)\sum_{i=1}^{k-1} p_i' x_i\right] \tag{B.15}$$

$$= f\left[\sum_{i=1}^{k} p_i x_i\right] \tag{B.16}$$

证毕

上述证明能够简单地扩展到连续随机变量的情况.

B.3 马尔可夫链

定义 B.3 设$\{X_n, n\in N\}$为一随机序列,时间参数集 $N=\{1,2,\cdots\}$,其状态空间 $S=\{s_1, s_2, \cdots, s_J\}$,若对所有 $n\in N^+$,有

$$P(X_n = s_{i_n} \mid X_{n-1} = S_{i_{n-1}}, X_{n-2} = s_{i_{n-2}}, \cdots, X_1 = s_{i_1}) = P(X_n = s_{i_n} \mid X_{n-1} = s_{i_{n-1}}) \tag{B.17}$$

其中,$s_{i_1}, \cdots, s_{i_{n-2}}, s_{i_{n-1}}, s_{i_n} \in S$,则称$\{X_n, n\in N\}$为马尔可夫链.

系统在现在时刻$(n-1)$处于状态 $s_{i_{n-1}}$,那么将来时刻 n 的状态 s_{i_n} 与过去时刻 $n-2, n-3, \cdots, 1$ 的状态 $s_{i_{n-2}}, \cdots, s_{i_1}$ 无关,仅与现在时刻$(n-1)$的状态 $s_{i_{n-1}}$ 有关.

马尔可夫链通常用转移概率来描述其统计特性:

$$p_{ij}(m, n) = P\{X_n = s_j \mid X_m = s_i\} \tag{B.18}$$

$p_{ij}(m, n)$表示已知在时刻 m 系统处于状态 s_i 或者说 X_m 取值 s_i 的条件下经$(n-m)$步后在时刻 n 转移到状态 s_j 的概率.也可以把 $p_{ij}(m, n)$理解为已知在时刻 m 系统处于状态 s_i 的条件下,在时刻 n 系统处于状态 s_j 的条件概率,故转移概率实际上是一个条件概率,满足条件概率的性质:

(1) $p_{ij}(m, n)\geq 0 \qquad s_i, s_j \in S$

(2) $\sum_j p_{ij}(m, n) = 1$

当 $n-m=1$ 时,把 $p_{ij}(m, n)$记为 $p_{ij}(m)$, $m\geq 0$,称为基本转移概率,也称为一步转移概率:

$$p_{ij}^{(1)}(m) = p_{ij}(m) = P\{X_{m+1} = s_j \mid X_m = s_i\} \tag{B.19}$$

把 k 步转移概率写成:

$$p_{ij}^{(k)}(m) = P\{X_{m+k} = s_j \mid X_m = s_i\} \tag{B.20}$$

它表示在时刻 m, X_m 的状态为 s_i 的条件下,经过 k 步转移到达状态 s_j 的概率.由于系统在任一时刻可以处于状态空间 $S=\{s_1, s_2, \cdots, s_J\}$中的任一状态,因此,转移概率是一个矩阵:

$$P^{(k)}(m)=\{p_{ij}^{(k)}(m),s_i,s_j\in S\} \tag{B 21}$$

称为 k 步转移概率矩阵.一步转移概率矩阵为

$$P(m)=\{p_{ij}(m),s_i,s_j\in S\} \tag{B 22}$$

定义 B.4　如果在马尔可夫链中,$p_{ij}(m)=P\{X_{m+1}=s_j\mid X_m=s_i\}=p_{ij}$,即从状态 s_i 转移到状态 s_j 的概率与起始时刻 m 无关,则称这类马尔可夫链为时齐马尔可夫链或齐次马尔可夫链,也称为具有平稳转移概率的马尔可夫链.

这里的平稳仅仅是转移概率的平稳,还不是平稳过程.

由一步转移概率 p_{ij} 可以写出其一步转移概率矩阵:

$$P=\{p_{ij},s_i,s_j\in S\}$$

或

$$P=\begin{pmatrix}p_{11}&p_{12}&p_{13}&\cdots\\p_{21}&p_{22}&p_{23}&\cdots\\p_{31}&p_{32}&p_{33}&\cdots\\\cdots&\cdots&\cdots&\end{pmatrix} \tag{B 23}$$

矩阵中每一个元素都是非负的,并且每行之和均为 1.如果马尔可夫链中状态空间 $S=\{s_1,s_2,\cdots,s_J\}$ 是有限的,则称为有限状态的马尔可夫链.如果状态空间 $S=\{s_1,s_2,\cdots\}$ 是无穷集合,则称为可数无穷状态的马尔可夫链.

对于具有 $(m+r)$ 步转移概率的齐次马尔可夫链,存在下述 C-K 方程:

$$P^{(m+r)}=P^{(m)}P^{(r)} \tag{B 24}$$

或写成:

$$p_{ij}^{(m+r)}=\sum_k p_{ik}^{(m)}p_{kj}^{(r)}$$

对于齐次马氏链来说,一步转移完全决定了 k 步转移概率 $P^{(k)}=(P)^k$.

为了确定无条件概率(绝对概率) $P(X_k=s_j)$,还需引入初始概率,记:

$$p_{0i}=P(X_0=s_i) \tag{B 25}$$

$$P(X_k=s_j)=\sum_i P(X_k=s_j,X_0=s_i)=\sum_i p_{0i}p_{ij}^{(k)} \tag{B 26}$$

一般情况下,绝对概率是与初始分布 p_{0i} 有关的,但是当 $\lim\limits_{k\to\infty}p_{ij}^{(k)}$ 极限存在,且等于一个与起始状态 s_i 无关的被称为平稳分布的 W_j,即 $\lim\limits_{k\to\infty}p_{ij}^{(k)}=W_j$ 与 s_i 无关时,则不论起始状态是什么,此马氏链最终可以达到稳定,即所有变量 X_k 的概率分布均匀不变.这时

$$\lim_{k\to\infty}P(X_k=s_j)=\lim_{k\to\infty}\sum_i p_{0i}p_{ij}^{(k)}=\sum_i p_{0i}W_j=W_j \tag{B 27}$$

这时,马氏链达到了稳定的分布,稳定分布只与转移概率有关,而与初始分布无关.起始状态只使前面有限个变量的分布改变,如同电路中的暂态一样.

定义 B.5　若齐次马尔可夫链对一切 i,j 存在不依赖于 i 的极限:$\lim\limits_{k\to\infty}p_{ij}^{k}=W_j$ 并且 $W_j=\sum\limits_i W_i p_{ij}$,$\sum\limits_j W_j=1$,$W_j\geq 0$,则称其具有遍历性(各态历经性),$W_j$ 为平稳分布.

遍历性的直观意义是,不论起始状态是哪个状态 s_i,当转移步数 k 足够大时,转移到状态 s_j 的概率 $p_{ij}^{(k)}$ 都近似等于某个常数 W_j,反过来说,如果转移步数 k 充分大,就可以用常数 W_j 作为 k 步转移概率 $p_{ij}^{(k)}$ 的近似值 .

这意味着,马尔可夫信源在初始时刻可以处在任意状态,然后状态之间可以任意转移,经过足够长的时间之后,信源处在什么状态已与初始状态无关 这时,每种状态出现的概率已达到一种稳定分布,即平稳分布 .就像电路经过暂态后进入稳态一样,到这时,信源才是一个离散平稳信源 .

定理 B .3 W_j 是满足方程组 $WP = W$ 和 $\sum_j W_j = 1$, $W_j \geq 0$ 的唯一解 .

事实上,用 $\lim_{k\to\infty} p_{ij}^{(k)} = W_j$ 求稳态分布是比较困难的,如果能判断稳态分布存在,一般用解方程组 $W_j = \sum_i W_i p_{ij}$, $\sum_j W_j = 1$, $W_j \geq 0$ 来求 W_j .怎样判断马氏链的稳态分布存在呢 ?

定理 B .4 设 P 为某一马氏链的状态转移矩阵,则该马氏链稳态分布存在的充要条件是,存在一个正整数 N,使矩阵 P^N 中的所有元素均大于零 .

【例 B .1】

设有一马氏链,其状态转移矩阵为

$$P = \begin{pmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

问是否存在稳态分布 如果存在,求其稳态分布 .

解

为了验证它是否满足 3 .5 .2 的条件,计算矩阵

$$P^2 = \begin{pmatrix} 0 & 0 & * \\ * & * & * \\ * & * & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & * \\ * & * & * \\ * & * & 0 \end{pmatrix} = \begin{pmatrix} * & * & 0 \\ * & * & * \\ * & * & * \end{pmatrix}$$

$$P^3 = \begin{pmatrix} * & * & 0 \\ * & * & * \\ * & * & * \end{pmatrix}\begin{pmatrix} 0 & 0 & * \\ * & * & * \\ * & * & 0 \end{pmatrix} = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}$$

其中 * 表示非零元素 .因此,这个马氏链是遍历的,其稳态分布存在 .

由定理 B .3, W 满足方程组 $WP = W$,并且 $\sum_j W_j = 1$, $W_j \geq 0$,其中矢量 W 写成分量的形式 $W = (W_1 \quad W_2 \quad W_3)$ 带入 $WP = W$,得到

$$\begin{cases} \frac{1}{2}W_2 + \frac{1}{2}W_3 = W_1 \\ \frac{1}{3}W_2 + \frac{1}{2}W_3 = W_2 \\ W_1 + \frac{1}{6}W_2 = W_3 \end{cases}$$

并且 $W_1 + W_2 + W_3 = 1$,则可求得 $W_1 = \frac{1}{3}$, $W_2 = \frac{2}{7}$, $W_3 = \frac{8}{21}$.

定理 2 所给定的条件等价于存在一个状态 s_j 和正整数 N,使得从任意原始状态出发,经过 N 步转移之后,一定可以到达状态 s_j .

也就是说,只有在转移一定步数后各状态之间均可相通的条件下,当转移步数足够大时,各状态出现的概率才能稳定在某一极限值,存在状态的极限概率 所谓‘ 各态历经 ”,其含义之一就是各态相通,均可经历;其含义之二就是由各态历经过程产生的每个序列,都有相同的统计特性 如果 P 不含零元素,即任一状态经一步转移便可达其他状态,则稳态分布必然存在 .

时齐马氏链可以用状态转移图来表示 从状态转移图(香农线图)可以判断状态相通的情况 .图 B .1 是一个有着 6 个状态的时齐马氏链 .

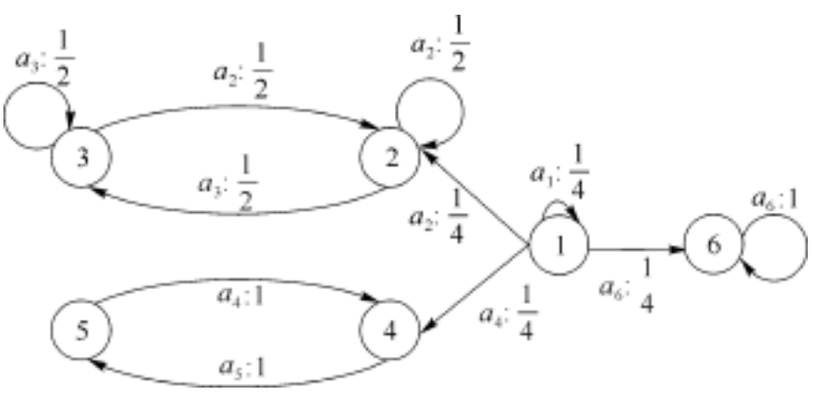


图 B .1　有 6 个状态的时齐马氏链

时齐马氏链的状态可以根据其性质进行如下的分类:

(1) 若一个状态经过若干步后总能到达某一个其他状态,但不能从其他状态返回,则称此状态为过渡态,如图 B .1 中的状态 1 .

(2) 经过有限步后迟早要返回的状态称为常返态,如图 B .1 中的状态 2、3、4、5、6 均为常返态 .

(3) 在常返态中,若仅当 n 能被某整数 $d(d>1)$ 整除时才有 $p_{ii}^{(n)} > 0$,则称此状态为周期态,如状态 4、5 .

(4) 在常返态中,若对于所有 $p_{ii}^{(n)} > 0$ n 值的最大公约数为 1,即 $d = 1$,则称此状态为非周期态,如状态 2、3 .

(5) 非周期的常返态称为遍历态 .

根据状态空间中的可达性和相通性,可以对状态空间进行分解 若状态空间中的某一子集中的任何一个状态都不能到达子集以外的任何状态,则称该子集为闭集,如{2,3,4,5},{2,3},{4,5},{6},闭集中除自身之外再没有其他闭子集的闭集称为不可约的,如{2,3},{4,5},{6} .

从不可约的非周期态出发,在转移一定步数后,各状态之间可相通,经过足够长时间后,就可以使各状态出现的概率稳定在某一极限值 .

(6) 既约、非周期、有限状态的马尔可夫链,其 n 步转移概率在 n 很大时趋于一个和初始状态无关的极限概率 W_j,它是满足方程组 $W_j = \sum_i W_i p_{ij}$, $\sum_j W_j = 1$, $W_j \geq 0$ 的唯一解 称 W_j 为马尔可夫链的平稳分布,即当时间足够长之后系统处于状态 s_j 的概率,此时马尔可夫链是平稳的或称为遍历的 .

B .4 信道容量定理的引理

引理 $\lim_{\theta\to 0}\frac{1}{\theta}\{I[\theta Q + (1-\theta)P] - I(P)\} = \sum_{i=1}^{r}(q_i - p_i)\frac{I(P)}{p_i}$

这里把 $I(X;Y)$写成概率矢量 P、Q 的函数 $I(P)$和 $I(Q)$的形式, P、Q 为不同的概率矢量:

$$P = (p_1, p_2, \cdots, p_r),\ Q = (q_1, q_2, \cdots, q_r)$$

证明

$$\begin{aligned}
&I[\theta Q + (1-\theta)P] - I(P)\\
&= I[p_1 + \theta(q_1 - p_1), p_2 + \theta(q_2 - p_2), p_3 + \theta(q_3 - p_3), \cdots, p_r + \theta(q_r - p_r)]\\
&\quad - I(p_1, p_2, \cdots, p_r)\\
&= I[p_1 + \theta(q_1 - p_1), p_2 + \theta(q_2 - p_2), p_3 + \theta(q_3 - p_3), \cdots, p_r + \theta(q_r - p_r)]\\
&\quad - [p_1, p_2 + \theta(q_2 - p_2), p_3 + \theta(q_3 - p_3), \cdots, p_r + \theta(q_r - p_r)]\\
&\quad + [p_1, p_2 + \theta(q_2 - p_2), p_3 + \theta(q_3 - p_3), \cdots, p_r + \theta(q_r - p_r)]\\
&\quad - [p_1, p_2, p_3 + \theta(q_3 - p_3), \cdots, p_r + \theta(q_r - p_r)]\\
&\quad + [p_1, p_2, p_3 + \theta(q_3 - p_3), \cdots, p_r + \theta(q_r - p_r)]\\
&\quad \cdots\\
&\quad + [p_1, p_2, p_3, \cdots, p_r + \theta(q_r - p_r)]\\
&\quad - I(p_1, p_2, \cdots, p_r)
\end{aligned}$$

上式为 r 对 r 个分量的多元函数的减式,每一对只有一个分量不同,其余分量相同。对每一对减式除以 θ 并求当 $\theta\to 0$ 时的极限,这可以看作多元函数对该分量求偏导,即当 $\theta\to 0$

时,有

$$\lim_{\theta\to 0}\frac{1}{\theta}I[p_1,p_2,\cdots,p_i+\theta(q_i-p_i),\cdots,p_r]-I[p_1,p_2,\cdots,p_i,\cdots,p_r]$$
$$=(q_i-p_i)\frac{I(P)}{p_i}$$

所以
$$\lim_{\theta\to 0}\frac{1}{\theta}\{I[\theta Q+(1-\theta)P]-I(P)\}$$
$$=(q_1-p_1)\frac{I(P)}{p_1}+(q_2-p_2)\frac{I(P)}{p_2}+\cdots+(q_r-p_r)\frac{I(P)}{p_r}$$
$$=\sum_{i=0}^{r}(q_i-p_i)\frac{I(P)}{p_i}$$

B.5　契比雪夫不等式

如果随机变量 X 是离散的,则其均方值由如下期望给出:

$$E(X^2)=\sum_{i=-\infty}^{+\infty}x_i^2p(x_i) \tag{B.28}$$

如果变量是连续的,则其均方值为

$$E(X^2)=\int_{-\infty}^{+\infty}x^2\,p(x)\mathrm{d}x \tag{B.29}$$

由于被积函数是恒正的,所以对于任何 $\varepsilon>0$ 均有

$$\int_{-\infty}^{+\infty}x^2\,p(x)\mathrm{d}x\geq\int_{|x|\geq\varepsilon}x^2\,p(x)\mathrm{d}x \tag{B.30}$$

式(B.30)又可写成

$$E(X^2)\geq\int_{|x|\geq\varepsilon}x^2\,p(x)\mathrm{d}x\geq\varepsilon^2\int_{|x|\geq\varepsilon}p(x)\mathrm{d}x \tag{B.31}$$

所以可

$$E(X^2)\geq\varepsilon^2P_r[|X|\geq\varepsilon] \tag{B.32}$$

或

$$P_r[|X|\varepsilon]\leq\frac{E(X^2)}{\varepsilon^2} \tag{B.33}$$

这就是著名的契比雪夫(Chebyshev)不等式.

B.6　大数定理

如果我们作大量的贝努利(Bernouli)实验(即实验成功或失败由一个随机变量 X 表

示，且各次实验之间是相互独立的)，那么成功次数与实验总次数的比值是多少？这一比值与每次实验成功的概率之间是什么关系？大数定理回答了这个问题．

我们考虑随机变量 X_i，它在第 i 次实验成功时取 1，失败时取 0．其均值的数学期望为

$$E\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = \frac{1}{n}\sum_i E(X_i) = \mu \tag{B 34}$$

其中，μ 是单次实验的期望值．

其均值的方差为

$$D\left(\frac{1}{n}\sum_i X_i\right) = E\left\{\frac{1}{n^2}\left[\sum_i (X_i - \mu)\right]^2\right\}$$

$$= \frac{1}{n^2}E\left[\sum_i (X_i - \mu)^2\right] \tag{B 35}$$

因为这些实验都是相同的实验，所以

$$D\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{n^2}\sum_i D(X_i) = \frac{1}{n}D(X_i) = \frac{\sigma^2}{n} \tag{B36}$$

这样，按契比雪夫不等式就得到

$$P_r\left[\left|\frac{1}{n}\sum_i (X_i - \mu)\right| > \varepsilon\right] \leq \frac{1}{\varepsilon^2}E\left\{\frac{1}{n^2}\left[\sum_i (X_i - \mu)\right]^2\right\} \leq \frac{\sigma^2}{n\varepsilon^2} \tag{B 37}$$

对相反的事件有

$$P_r\left[\left|\frac{1}{n}\sum_i (X_i - \mu)\right| < \varepsilon\right] \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \tag{B 38}$$

定理 B 5(大数定理)　若 $X_1, X_2, \cdots, X_n$ 是 n 个独立同分布的随机变量，均值为 μ，方差为 σ^2．则对任意 $\varepsilon > 0$ 和 $\delta > 0$，一定存在一个整数 n_0 使得当 $n > n_0$ 时

$$\mu - \varepsilon \leq \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \leq \mu + \varepsilon \tag{B 39}$$

的概率大于 $1 - \delta$．

这个定理说明平均值依概率收敛于期望值 μ．

B .7　渐进等同分割性和 ε 典型序列

当随机试验的次数很大时，事件发生的频率具有稳定性．比如反复进行抛掷硬币的随机实验，出现正面或反面的次数是不定的，但是随着试验次数的增加，出现正面或反面的频率逐渐将稳定于 1/ 2，这就是随机事件的统计规律性．

对于独立同分布的随机变量 X_1，X_2，…，X_N，只要 N 足够大，其算术平均值

$\frac{1}{N}\sum_{i=1}^{N} X_i$ 接近其数学期望值 $E(X)$,即

$$\lim_{N\to\infty} P_r\left[\left\{\left|\frac{1}{N}\sum_{i=1}^{N} X_i - E(X)\right| < \varepsilon\right\}\right] = 1 \tag{B.40}$$

也就是说其算术平均值依概率收敛于数学期望.当 N 很大时,其算术平均值将几乎变成一个常数 $E(X)$,这就是大数定理.

把$\frac{1}{N}\sum_{i=1}^{N} X_i$ 看成一个随机变量,$E\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = E(X)$,$D\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{\sigma^2}{N\varepsilon^2}$,根据契比雪夫不等式,对于独立同分布的随机变量 X_1, X_2, …, X_N,和任意 $\varepsilon > 0$,有不等式

$$P_r\left[\left|\frac{1}{N}\sum_{i=1}^{N} X_i - E(X)\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{N\varepsilon^2} \tag{B.41}$$

和

$$P_r\left[\left|\frac{1}{N}\sum_{i=1}^{N} X_i - E(X)\right| \leq \varepsilon\right] \geq 1 - \frac{\sigma^2}{N\varepsilon^2} \tag{B.42}$$

成立,其中 σ^2 为随机变量 X_1, X_2, …, X_N的方差.

考虑一个离散无记忆信源

$$\begin{bmatrix} S \\ P(S) \end{bmatrix} = \begin{bmatrix} s_1 & \cdots & s_i & \cdots & s_q \\ p(s_1) & \cdots & p(s_i) & \cdots & p(s_q) \end{bmatrix}$$

的 N 次扩展信源

$$\begin{bmatrix} S \\ P(S) \end{bmatrix} = \begin{bmatrix} s_1 & \cdots & s_j & \cdots & s_q \\ p(s_1) & \cdots & p(s_j) & \cdots & p(s_q) \end{bmatrix}$$

这里,$S = S_1 S_2 \cdots S_N$ 是 N 维随机矢量,而 $s_j = s_{j_1} s_{j_2} \cdots s_{j_N}$,其中 s_{j_1}, s_{j_2}, …, $s_{j_N} \in \{s_1, \cdots, s_i, \cdots, s_q\}$.

因为是离散无记忆信源的扩展信源,所以有

$$p(s_j) = p(s_{j_1})\, p(s_{j_2}) \cdots p(s_{j_N}) = \prod_{k=1}^{N} p(s_{j_k}) \tag{B.43}$$

$$\begin{aligned} I(s_j) &= -\log_2 p(s_j) \\ &= -\log_2\left[\prod_{k=1}^{N} p(s_{j_k})\right] \\ &= -\sum_{k=1}^{N} \log_2 p(s_{j_k}) \\ &= \sum_{k=1}^{N} p(s_{j_k}) \end{aligned} \tag{B.44}$$

$I(s_j)$是一个随机变量,其数学期望就是 S 的熵.

$$E[\ I(s_j)\ ] = H(S) = \sum_{k=1}^{N} E[\ I(s_{j_k})\ ] = NH(S)$$

$$D[\ I(s_j)\ ] = ND[\ I(s_i)\ ]$$

$$D[\ I(s_i)\ ] < \infty$$

所以当 q 为有限值时，$D[\ I(s_j)\ ] \leq \infty$.

由于相互统计独立的随机变量的函数也是相互统计独立的随机变量，所以由 S_1，S_2，…，S_N 是相互统计独立且服从同一概率分布的随机变量，可以推出其自信息量 $I(s_{j_k})$，$k=1,2,\cdots,N$ 也是相互统计独立且服从同一分布的随机变量 .

$$\frac{I(s_j)}{N} = \frac{1}{N}\sum_{k=1}^{N} I(s_{j_k}) \tag{B.45}$$

$$E\left[\frac{I(s_j)}{N}\right] = \frac{1}{N}H(S) = \frac{1}{N}\sum_{k=1}^{N} E[\ I(s_{j_k})\ ] = H(S) \tag{B.46}$$

所以，$\dfrac{I(s_j)}{N}$ 依概率收敛于 $H(S)$（大数定理），这称为渐进等同分割性（AEP）.

离散无记忆信源的 N 次扩展信源，N 维随机矢量中每一维随机变量相互独立，当序列长度 N 变得很大时，由于统计规律性，N 个随机变量的算术平均，将变成一个常数（随机变量的数学期望），也就是 N 维随机矢量中平均每一维随机变量的自信息非常接近每一维随机变量的自信息量的数学期望，而 $D\left[\dfrac{I(s_j)}{N}\right] = D[\ I(s_{j_k})\ ]/\ N$.

根据契比雪夫不等式，有以下不等式成立：

$$P_r\left\{\left|\frac{I(s_j)}{N} - H(S)\right| \geq \varepsilon\right\} \leq \frac{D[\ I(s_i)\ ]}{N\varepsilon^2} \tag{B.47}$$

和

$$P_r\left\{\left|\frac{I(s_j)}{N} - H(S)\right| \leq \varepsilon\right\} \geq 1 - \frac{D[\ I(s_i)\ ]}{N\varepsilon^2} \tag{B.48}$$

令 $\dfrac{D[\ I(s_i)\ ]}{N\varepsilon^2} = \delta(N,\varepsilon)$，可知 $\lim\limits_{N\to\infty}\delta(N,\varepsilon) = 0$.

这样，我们可以把扩展信源输出的 N 长信源符号序列集合，分成两个子集 G_ε 和 $\overline{G_\varepsilon}$：

$$G_\varepsilon = \left\{ s_j : \left|\frac{I(s_j)}{N} - H(S)\right| < \varepsilon \right\} \tag{B.49}$$

$$\overline{G_\varepsilon} = \left\{ s_j : \left|\frac{I(s_j)}{N} - H(S)\right| \geq \varepsilon \right\} \tag{B.50}$$

且 $P_r\{G_\varepsilon\} + P_r\{\overline{G_\varepsilon}\} = 1$ .G_ε 称为 ε 典型序列集，它表示 N 长序列中平均每一维随机变量的自信息非常接近每一维随机变量的自信息的数学期望的一类序列的集合 .而 $\overline{G_\varepsilon}$ 表示 N

长序列中不在 G_ε 集中的序列的集合,称为非 ε 典型序列集.它们的差别在于 $I(s_j)/N$ 与 $H(S)$ 的差是否小于正数 ε.

下面推导 ε 典型序列集的一些性质:

1. G_ε 和 $\overline{G_\varepsilon}$ 的概率

$$1 \geq P_r\{G_\varepsilon\} \geq 1 - \delta(N, \varepsilon) \tag{B 51}$$

$$0 \leq P_r\{\overline{G_\varepsilon}\} \leq \delta(N, \varepsilon) \tag{B 52}$$

2. G_ε 和 $\overline{G_\varepsilon}$ 中序列出现概率的范围

根据 ε 典型序列集的定义, G_ε 中序列 $I(s_j)/N$ 与 $H(S)$ 的差小于正数 ε.

$$-\varepsilon \leq \frac{I(s_j)}{N} - H(S) \leq \varepsilon \tag{B 53}$$

$$N[H(S) - \varepsilon] \leq I(s_j) \leq N[H(S) + \varepsilon] \tag{B 54}$$

而 $I(s_j) = -\log_2 p(s_j)$,所以

$$\text{最大值}\frac{2^{-N[H(S)-\varepsilon]}}{} \geq p(s_j) \geq \frac{2^{-N[H(S)+\varepsilon]}}{}\text{最小值} \tag{B 55}$$

3. G_ε 和 $\overline{G_\varepsilon}$ 中序列的个数

设 G_ε 中序列数为 M_G,

$$1 \geq P_r\{G_\varepsilon\} \geq M_G \frac{2^{-N[H(S)+\varepsilon]}}{}\text{最小值} \tag{B 56}$$

$$1 - \delta(N, \varepsilon) \leq P_r\{G_\varepsilon\} \leq M_G \frac{2^{-N[H(S)-\varepsilon]}}{}\text{最大值} \tag{B 57}$$

因此,有

$$\begin{aligned}
[1 - \delta(N, \varepsilon)]2^{N[H(S)-\varepsilon]} &= \frac{1 - \delta(N, \varepsilon)}{2^{-N[H(S)-\varepsilon]}} \\
&\leq \frac{P_r\{G_\varepsilon\}}{2^{-N[H(S)-\varepsilon]}} \\
&\leq M_G \\
&\leq \frac{P_r\{G_\varepsilon\}}{2^{-N[H(S)+\varepsilon]}} \\
&\leq \frac{1}{2^{-N[H(S)+\varepsilon]}} \\
&= 2^{N[H(S)+\varepsilon]}
\end{aligned} \tag{B 58}$$

即

$$[1 - \delta(N, \varepsilon)]2^{N[H(S)-\varepsilon]} \leq M_G \leq 2^{N[H(S)+\varepsilon]} \tag{B 59}$$

因此, N 次扩展信源中信源序列可分为两大类,一类是 ε 典型序列,是经常出现的信源序列.当 $N \to \infty$ 时,这类序列出现的概率趋于 1,并且,所有的 ε 典型序列接近等概分布

$p(s_j)\approx 2^{-N[H(S)]}$.另一类是低概率的非 ε 典型序列,是不经常出现的信源序列.当 $N\to\infty$时,这类序列出现的概率趋于零.

信源的这种划分性质就是渐近等同分割性.

G_ε 中序列虽然是高概率序列,但是,G_ε 中序列数占序列总数的比例却很小:

$$\zeta=\frac{M_G}{q^N}\leq\frac{2^{N[H(S)+\varepsilon]}}{q^N}=2^{-N[\log_2 q-H(S)-\varepsilon]} \tag{B.60}$$

B.8 n 维欧式空间

n 维空间的意思仅仅是说有 n 个独立的变量 x_1, x_2, …, x_n.欧几里得(Euclidean)空间是指在这一空间中我们采用毕达哥拉斯(Pythagorean)距离,即

$$x_1^2+x_2^2+\cdots+x_n^2=r^2 \tag{B.61}$$

同时,式(B.17)还定义了半径为 r 的 n 维球.

n 维球的体积与半径 r 之间的关系是与 r^n 成正比,体积的公式为

$$V_n(r)=C_n r^n \tag{B.62}$$

其中,C_n 是一个依赖于 n 的常数,例如

$$C_1=2\quad C_2=\pi\quad C_3=\frac{4\pi}{3}\quad\cdots$$

利用 Gamma 函数进行推导,我们可以得到 C_n 的表达式

$$C_n=\frac{\pi^{n/2}}{\Gamma(n/2+1)} \tag{B.63}$$

不难证明

$$C_n=\frac{2\pi}{n}C_{n-2} \tag{B.64}$$

并计算得到表 B.1.

表 B.1 n 维单位球的体积

n	C_n	n	C_n
1	$2=2.00000\cdots$	6	$\frac{\pi^3}{6}=5.16771\cdots$
2	$\pi=3.14159\cdots$	7	$\frac{16\pi^3}{105}=4.72477\cdots$
3	$\frac{4\pi}{3}=4.18879\cdots$	8	$\frac{\pi^4}{24}=4.05871\cdots$
4	$\frac{\pi^2}{2}=4.93480\cdots$	…	…
5	$\frac{8\pi^2}{15}=5.26379\cdots$	$2k$	$\frac{\pi^k}{k!}=\to 0$

从这张表可以看出,单位球的体积,或者说式(B.55)中 r^n 的系数 C_n,在 $n=5$ 时达到最大值,然后在 $n\to\infty$ 时很快地趋向于零.

现在我们考虑 n 维球中与球面的距离小于 ε 的一个球壳的体积,有

$$\frac{\text{壳体积}}{\text{球体积}}=\frac{C_n r^n - C_n(r-\varepsilon)^n}{C_n r^n}=1-\left(1-\frac{\varepsilon}{r}\right)^n \tag{B.65}$$

这一比值当 $n\to\infty$ 时趋向于 1.即对于一个很高维数的空间来说,球体的体积几乎全部集中在表面,而在其内部只有很小的体积.

参 考 文 献

1 傅祖芸 . 信息论——基础理论与应用 . 北京: 电子工业出版社,2001

2 周荫清 . 信息理论基础 . 北京: 北京航空航天大学出版社,2002

3 T M Cover, J A Thomas . Elements of Information Theory .New York: John Wiley & Sons . Inc .,1991

4 [美] R W 汉明著 . 编码和信息理论 . 朱雪龙译 .北京: 科学出版社,1984

5 朱雪龙 . 应用信息论基础 . 北京: 清华大学出版社,2001

6 姜丹 . 信息论与编码 . 合肥: 中国科技大学出版社,2001

7 陈运 . 信息论与编码 . 北京: 电子工业出版社,2002

8 曹雪虹 . 信息论与编码 . 北京: 北京邮电大学出版社,2001

9 吴伟陵 . 信息处理与编码 . 北京: 人民邮电出版社,2003