

■ 高 等 学 校 教 材

信息理论与编码

XINXI LILUN YU BIANMA

吕 锋 王 虹 刘皓春 苏 扬 编著



人民邮电出版社
POSTS & TELECOM PRESS

高等学校教材

信息理论与编码

吕锋 王虹 刘皓春 苏扬 编著

人民邮电出版社

图书在版编目(CIP)数据

信息理论与编码/吕锋等编著. —北京:人民邮电出版社,2004.2

高等学校教材

ISBN 7-115-12067-6

I. 信... II. 吕... III. ①信息论—高等学校—教材②信源编码—编码理论—高等学校—教材③信道编码—编码理论—高等学校—教材 IV. TN911.2

中国版本图书馆 CIP 数据核字(2004)第 005715 号

内 容 提 要

本书系统地讲述了信息论的基础理论。全书分 8 章,分别讨论了信息的度量、信源无失真编码、信道编码、信息率失真函数、网络信息论以及信息安全的理论与方法。

本书注重基本概念,论述力求简明,可作为高等院校通信类、信息类、电子工程类及相关专业的教材,也可供有关科研人员参考。

高等学校教材

信息理论与编码

◆ 编 著 吕 锋 王 虹 刘皓春 苏 扬
策划编辑 滑 玉

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
读者热线 010-67194042
北京汉魂图文设计有限公司制作
北京朝阳展望印刷厂印刷
新华书店总店北京发行所经销

◆ 开本: 787×1092 1/16
印张: 15.5 2004 年 2 月第 1 版
印数: 1-5 000 册 2004 年 2 月北京第 1 次印刷

ISBN 7-115-12067-6/TP · 3828

定价: 21.00 元

本书如有印装质量问题,请与本社联系 电话:(010)67129223

编者的话

信息是事物运动的状态和方式，是关于事物运动的千差万别的状态和方式的知识。人类社会的生存、发展都离不开信息的获取、传递、处理、控制和利用。

信息论是整个信息科学发展的起源和基石。21 世纪是高度信息化时代，信息论不仅在通信领域中发挥越来越重要的作用，也是解决通信领域中相关问题的有力工具；而且由于信息理论所蕴涵的独特的、有效的、新颖的解决问题的思路和方法，已渗透到其它相关的自然科学、社会科学领域，与计算机技术、电子技术、控制技术、网络技术、人工智能、生物工程、医学工程、管理科学等学科密切结合，显示了它的勃勃生机。因此，在现代科学技术高度发展的过程中，学习和掌握信息科学显得尤为重要。

信息论是一门利用概率论、随机过程和数理统计等数学方法来研究信息的存储、度量、编码、传输、处理中一般规律的重要学科，是数学知识与通信技术相结合的边缘学科。自从香农 1948 年发表奠定信息论基础的“通信的数学理论”一文以来，信息科学有了很大的发展并已经延伸到许多领域中。当今，它主要研究如何提高信息系统的可靠性、有效性、保密性和认证性，从而获取最优信息系统。

本书系统地介绍香农信息论的基本内容及其应用。全书注重基本概念、基本理论和基本分析方法的论述，并结合实例给出详细的数学推演过程和证明，力求概念清晰、结构严密、内容由浅入深、章节循序渐进。

全书共分八章。其中前三章是全书的理论基础，主要介绍信息论的基本理论知识。第一章阐述了信息的概念、信息论的研究内容以及应用范围等，期望为读者展示一个信息论的轮廓；第二章详细地讨论了信息的度量方法，侧重于信息的数学建模；第三章讨论信道，描述和分析了各种不同类型信道的模型和特性；第四章讨论信源无失真编码，目的在于提高信源的信息含量效率；第五章讨论信道编码，借此可解决传送的可靠性问题；第六章介绍信源有失真编码方法，对连续信源只能进行有失真编码；第七章介绍了网络信息论的一些基本理论和新成果，论述多用户通信系统的信道容量、信道编码定理、实现编码定理的码的结构问题等理论、技术。第八章简要地探讨了信息安全、密码学，阐述用密码技术如何保证电子信息的有效性、保密性、完整性。

信息论是一门既具有严密的逻辑演绎推理系统，又具有丰富生动的时代气息的科学理论。全书按照理论联系实际的原则，基于信息论中的重要结论，分析方法，解决思路的论述，并力求反映近年来国内外信息理论的新发展。

本书的第一、二章由吕锋教授编写，第三、五章和第六章的 1、2 节由刘皓春副教授编写，第七、八章和第六章的 3、4 节由王虹副教授编写，第四章由苏扬副教授编写。全书由吕锋教授统稿。

本书可作为高等院校通信、信息类专业和相关专业本科生教材或教学参考书，也可作为科研人员、工程技术人员的参考书。

限于水平，书中难免出现错误和缺憾，殷切期望广大读者批评指正。

编著者

目 录

第一章 绪论	1
1.1 信息的基本概念	1
1.1.1 信息概念的复杂性	1
1.1.2 信息的定义	3
1.2 信息论的研究目的和内容	4
1.2.1 信息传输基本模型	4
1.2.2 信息论研究的内容	6
1.2.3 目前信息论的主要研究成果	7
1.3 信息论的发展历程与应用概述	9
1.3.1 信息论发展简史	9
1.3.2 信息论的应用	12
习题	15
第二章 信息的度量	16
2.1 信源模型	16
2.2 信息的描述	17
2.3 不确定性与信息	18
2.3.1 自信息量	19
2.3.2 联合自信息量	19
2.3.3 条件自信息量	20
2.3.4 自信息量的性质和相互关系	21
2.3.5 互信息量及其性质	22
2.4 离散随机变量的(统计)平均不确定性度量——离散熵	25
2.4.1 离散熵	25
2.4.2 离散熵的性质	27
2.5 联合熵和条件熵	29
2.5.1 联合熵	30
2.5.2 条件熵	30
2.5.3 各类熵之间的关系	30
2.6 平均互信息量及其性质	31
2.7 离散无记忆信源的扩展	32
2.8 离散有记忆信源的熵	34
2.9 马尔可夫信源的信息熵	34
2.9.1 马尔可夫链	35

2.9.2 马尔可夫信源	36
2.9.3 马尔可夫信源的信息熵	38
2.10 离散信源的信息(速)率和信息含量效率	40
2.11 连续随机变量的熵和平均互信息量	40
2.11.1 连续随机变量的熵	41
2.11.2 连续随机变量的联合熵、条件熵以及平均互信息量	43
2.11.3 微分熵的极大化问题	44
2.11.4 连续信源的熵功率	47
本章主要概念	47
习题	49
第三章 信道模型和信道容量	53
3.1 信道模型与信道分类	53
3.2 离散无记忆信道的数学模型	54
3.3 概率的计算问题	55
3.4 信道的疑义度、散布度和平均互信息	57
3.4.1 信道的疑义度	57
3.4.2 信道的散布度	59
3.4.3 信道的平均互信息	60
3.5 信道容量	63
3.5.1 信道容量的定义	63
3.5.2 离散无噪信道的信道容量	64
3.5.3 离散对称信道	67
3.5.4 一般 DMC 达到信道容量的充要条件	72
3.5.5 信道容量的迭代算法	75
3.6 扩展信道及其信道容量	76
3.6.1 扩展信道的数学模型	77
3.6.2 扩展信道的平均互信息量和信道容量	78
3.7 信道的组合	80
3.7.1 串联信道	80
3.7.2 独立并联信道	81
3.8 信源与信道的匹配	82
3.9 连续信道及其信道容量	82
3.9.1 连续信道的数学模型	83
3.9.2 加性高斯噪声信道的信道容量	84
3.9.3 一般加性噪声信道的信道容量的界	86
3.10 波形信道及其信道容量	88
本章主要概念	91
习题	94

第四章 离散无记忆信源无失真编码	97
4.1 信源编码概论	98
4.2 码的惟一可译性	100
4.2.1 常见码及其惟一可译性	100
4.2.2 码树和 Kraft 不等式	101
4.3 定长编码定理和定长编码方法	103
4.4 变长编码定理	106
4.5 变长编码方法	107
4.5.1 霍夫曼编码	107
4.5.2 费诺编码	113
4.5.3 香农编码	114
4.6 几种实用的无失真信源编码	115
4.6.1 游程编码	115
4.6.2 算术编码	119
4.6.3 基于字典的编码	122
本章主要概念	124
习题	125
第五章 有噪信道编码	128
5.1 译码规则与错误概率	128
5.2 两种典型的译码规则	130
5.3 平均差错率与信道编码	133
5.3.1 “简单重复”编码	133
5.3.2 对符号串编码	135
5.4 汉明距离	137
5.5 有噪信道编码定理	140
5.5.1 联合典型序列	140
5.5.2 有噪信道编码定理的证明	142
5.6 Fano 不等式和有噪信道编码逆定理	144
5.7 线性分组码	146
5.7.1 线性分组码的生成矩阵和校验矩阵	146
5.7.2 汉明距离和码的纠、检错能力	148
5.7.3 线性码的伴随式与伴随式译码	150
本章主要概念	151
习题	153
第六章 限失真信源编码	154
6.1 失真测度	154

6.2 信息率失真函数及其性质	156
6.2.1 信息率失真函数的定义	156
6.2.2 信息率失真函数的性质	157
6.3 限失真信源编码定理	159
6.4 信息率失真函数的计算	160
6.4.1 离散信源信息率失真函数的参量表示计算方法	160
6.4.2 离散信源信息率失真函数的迭代计算方法	167
本章主要概念	170
习题	171
第七章 网络信息论基础	173
7.1 概论	173
7.2 网络信道的分类	174
7.3 典型信源编码模型	176
7.4 多随机变量联合典型序列	177
7.5 相关信源编码	180
7.6 多址接入信道	183
7.7 高斯多址接入信道	187
7.8 广播信道	190
7.9 中继信道	194
7.10 具有边信息的信源编码和数据压缩	195
本章主要概念	197
习题	199
第八章 信息安全与密码学基础	201
8.1 信息安全概述	201
8.2 网络模型与安全服务功能	202
8.2.1 开放系统互联 OSI 模型	202
8.2.2 安全分层原则	203
8.2.3 安全服务功能	203
8.2.4 网络安全对策	205
8.3 密码学基础知识	207
8.3.1 基本术语	207
8.3.2 代替密码	209
8.4 密码算法的数学背景	212
8.4.1 信息论	212
8.4.2 复杂性理论	214
8.4.3 数论基础	215
8.5 数据加密标准(DES)	218

8.5.1 数据加密标准的开发	218
8.5.2 DES 算法概要	218
8.5.3 初始置换	219
8.5.4 密码运算函数 $f(R, K)$	219
8.5.5 密钥置换	219
8.5.6 扩展置换	221
8.5.7 S 盒替代	221
8.5.8 P 盒置换	222
8.5.9 逆初始置换	223
8.5.10 DES 的安全性	223
8.5.11 DES 的硬件实现	225
8.6 公开密钥算法	225
8.6.1 公开密钥密码体制	225
8.6.2 背包公钥密码	225
8.6.3 RSA 公钥加密	227
8.6.4 数字签名	230
本章主要概念	232
习题	234
参考书目	236

第一章 绪 论

信息论是关于信息的本质和传输规律的科学理论，是研究信息的度量、发送、传递、交换、接收和储存的一门新兴学科。它不仅是现代信息科学大厦的一块重要基石，而且还广泛地渗透到生物学、医学、管理学、经济学等其他各个领域，对社会科学和自然科学的发展都有着深远的影响。

本章首先阐述信息的基本概念，然后讨论信息论的研究目的和研究内容，最后对信息论的发展历程和应用做一个简单的介绍。

1.1 信息的基本概念

任何一门科学都有它自己的基本概念，理解和掌握这些基本概念是学习这门科学的基础。传统科学的基本概念是物质和能量，而信息论的最基本和最重要的概念就是信息。因此信息既是信息论的出发点，也是它的归宿。具体地说，信息论的出发点是认识信息的本质和它的运动规律；它的归宿则是利用信息来达到某种具体的目的。

1.1.1 信息概念的复杂性

当今社会，“信息”一词，在各种场合都被广泛采用，但如同数学中的“集合”一词一样，要给它下一个严格的定义却异常之难。即使是信息论的奠基人香农（C. E. Shannon），在其著名论文“通信的数学理论”中，也没有给信息下一个明确的定义，留下了一个悬案。香农论文发表之后，由于其方法新颖，引来许多专家学者对信息进行深入研究，研究中碰到的首要问题就是要给“信息”一词下一个明确的定义。很多学者都给“信息”下过定义，流行的说法不下百种，而且对此还展开了一些重要的哲学争论，到现在为止还没有一个定论。钟义信先生在《信息科学原理》一书中引述了三十多种比较典型和有代表性的说法，这些说法要么出发点不同，要么所站角度不一样，有些甚至带有较明显的学科倾向，但都在一定层面上对信息概念做了描述。作者从中选取一部分，按个人理解略作归类分析，供读者参考。

1. 用人们熟知的、与信息有某种联系的概念来定义信息

(1) 信息是消息。

(2) 信息是数据。

此类定义只接触到与信息相关的一些表象，未触及到信息的本质。在日常交流中，信息常以消息的形式出现，消息中可能含有信息，但不是信息本身。数据则是记录信息的一种形式，并非是唯一形式。

2. 用某些学科的专门术语和名词来定义信息

- (1) 信息是集合之间的变异度。
- (2) 信息是一种场。
- (3) 信息是信号。

“集合”和“场”分别是数学和物理学中引入的概念，姑且不谈其自身在所属学科或领域存在的争议，就概念本身就很专业、难懂，拿来定义信息，叫人更加不好理解。信号是电专业的专用名词，是表现信息的一种形式，或传输信息的一种载体，不是信息本身。

3. 从广义的角度、或者说从哲学意义上来定义信息

- (1) 信息就是信息，既不是物质也不是能量。
- (2) 信息是事物之间的差异。
- (3) 信息是事物相互作用的表现形式。
- (4) 信息是事物联系的普遍形式。
- (5) 信息是物质的普遍属性。

这类定义很多，它们的共性是力争从一般意义上来定义信息，哲学味道很浓，但难以从定义出发引出信息的度量方法。此类定义容易引起哲学论争。当然，论争的过程实际上也是认识深入的过程，工科学生了解这些定义，对开阔自己的视野是非常有益的，其是非曲直就靠各人自己理解了。

4. 从控制论和系统论的角度来定义信息

- (1) 信息是与控制系统相联系的一种功能现象。
- (2) 信息是控制的指令。
- (3) 信息是系统组织程度的度量。
- (4) 信息是有序性的度量。
- (5) 信息是负熵。

前两种是从控制论的角度下的定义，这两说法不能说无道理，因为控制系统必须有信息与之相关联，才能实施有效的控制，而控制指令中必然含有信息，但根据这两种说法，我们还是不知道信息到底是什么。后三种是从系统论角度下的定义，三种说法意思差不多，因为系统的有组织程度与有序性意思相近，而熵代表无组织程度，负熵即为有组织程度。这三种定义与香农对信息的理解较为接近，香农虽未明说，但从其论文可看出，香农把信息理解为消除确定性的东西。

5. 从随机不确定性的角度来定义信息

- (1) 信息是收信者事先不知道的报道。
- (2) 信息是用以消除不确定性的东西。

下此类定义者通常对通信有所了解，有些本来就是通信领域的专家学者。他们在定义信息时考虑到了通信的实际情况，即收信者在收到信息之前是心存疑问或不确定性的，收到信息之后，不确定性会减少或消除，因此说，信息是用以消除不确定性的东西。例如，假设巴

西足球队和阿根廷足球队将进行一场重要的国际比赛，你在未得到赛事的任何消息之前，你可能会根据两队以往的战绩判断，两队势均力敌、胜负参半，即比赛结果存在不确定性，而且这时不确定性是最大的。若从电视新闻中得到消息是“上半场巴西队 3:0 领先”，这时你会重新判断比赛结果，认为巴西队取胜的可能性大大增加，这说明比赛结果的不确定性大大减少了，有部分不确定性消除了。若新闻中说：“巴西队 3:0 取胜”，则比赛结果的不确定性完全消除了。无论是不确定性部分消除还是完全消除，都是因为收到了信息的缘故。因此将信息定义为消除不确定性的东西是有道理的。

1.1.2 信息的定义

由于信息概念的复杂性，众说纷纭，叫人无所适从。定义信息尽管困难，但却是不可回避的问题，为此，本书采用钟义信先生提出的信息定义体系，即按适用面的不同来分层次地定义信息。层次越高，适用面越宽。适用面不同意味着给定义附加的约束条件不同，约束条件越多，适用面越窄。

钟先生将最高层次的信息称为**本体论层次信息**，定义为：**某事物的（本体论层次）信息，就是事物运动的状态和方式，也就是事物内部结构和外部联系的状态和方式。**

信息是事物运动的状态与方式，具体地讲，就是事物内部结构和外部的联系和运动的状态与方式。在此，“事物”泛指一切可能的研究对象，包括外部世界的物质客体，也包括主观世界的精神现象；“运动”泛指一切意义上的变化，包括机械运动、物理运动、化学运动、生物运动、思维运动和社会运动等；“运动方式”是指事物运动在时间上所呈现的过程和规律；“运动状态”则是事物运动在空间上所展示的性状与态势。

本体论层次的信息是最广义的信息，是无条件的信息，与是否有（认识或观察）主体无关。世上一切事物始终处于运动之中，有些是已知的，有些是未知的，因此都在产生信息。

本体论层次的信息定义适用面最广，但使用起来却不太方便，它与我们后面将要给出的信息度量方式难以产生直接的联系。鉴于此，可给定义附加一些约束条件，降低定义的层次，当然适用面会相应变窄。最具实际意义的约束条件是加入人类这个认识或观察的主体，从认识论的角度来定义信息。此层次的信息称为**认识论层次信息**，定义为：**某主体关于某事物的（认识论层次）信息，是指该主体所感知的相应事物的运动状态及其变化方式，包括状态及其变化方式的形式、含义和效用。**

认识论层次下的信息同时考虑了事物运动状态及其变化方式的外在形式、内在含义和效用价值三个因素，因此，较之本体论层次信息，尽管适用面变窄，但内涵却丰富得多。之所以如此，原因在于加入了人这个认识主体，人有感觉能力、理解能力以及判断能力，因此，对事物的运动状态及其变化方式，可感觉其外在形式、理解其内在含义以及判断其效用价值。

由于认识论层次信息内涵丰富，笼统地加以研究实属困难，为此，钟先生借用语言学中的术语，将认识论层次信息进行细分。将涉及形式因素的信息部分称为语法信息，涉及含义因素的信息部分称为语义信息，涉及效用因素的信息部分称为语用信息。

语法信息：语法信息是事物运动状态和状态改变的方式的本身。所以它不涉及这些状态的含义和效用、是最抽象最基本的层次。它只研究事物运动中各种可能出现的状态，以及状态之间的关系。香农的信息定义正是属于这个层次，它是从概率统计角度来研究事物运动中各种可能出现的状态以及状态间的关系，因此是概率性的语法信息。它能较好地解决通信工

程这样一类信息传递的问题。

语义信息：语义信息是主体所感知或表述的事物运动状态和方式的具体含义。这是研究各种状态和实体间的关系，即研究信息的具体含义。

语用信息：语用信息是事物运动状态和方式及其含义对观察者（主体）的效用。这是研究事物运动状态和方式与主体的关系，即研究信息的主观表示部分。

1.2 信息论的研究目的和内容

从上一节中我们讨论的信息定义及其分类中，知道信息可分为语法信息、语义信息和语用信息。信息是信息论研究的主要内容。根据研究内容范围的大小，可对信息论进行分类。

狭义信息论，也称经典信息论。它主要研究信息的测度、信道容量以及信源和信道编码理论等问题。这部分内容是信息论的基础理论，又称香农基本理论。

一般信息论，主要是研究信息传输和处理问题。除了香农理论以外，还包括噪声理论、信号滤波和预测、统计检测与估计理论、调制理论、信息处理理论以及保密理论等。后一部分内容是以美国科学家维纳（N. Wiener）为代表。

可以看出，狭义信息论和一般信息论研究的都是关于语法信息的问题。

广义信息论，它不仅包括上述两方面的内容，而且包括所有与信息有关的自然和社会领域。如模式识别、计算机翻译、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，即广义信息论不但研究语法信息，还将研究领域延伸到语义信息和语用信息的范畴。

在本书中，我们讨论的范围限于一般信息论之内，因此除特别说明外，我们所说的信息论均指一般信息论。

1.2.1 信息传输基本模型

信息论所处理的问题可用图 1.1 所示的信息传输基本模型加以说明。

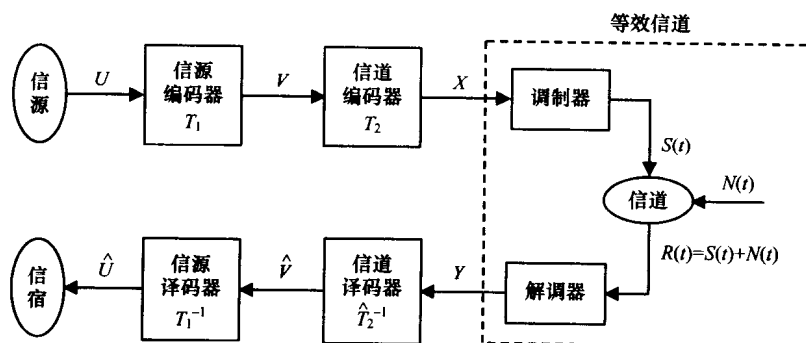


图 1.1 信息传输基本模型

图 1.1 所示的实际上是一个点到点的单向信息传输系统模型，实际的通信系统未必如此简单，但用图 1.1 所示的模型可解释各种通信系统中的一些共性问题，对这些共性问题进行总结分析，会得到一些重要的基本概念。

通常,实际的信息传输系统中,事先给定的是图中椭圆框出的部分,即发出信息的信源,接收信息的信宿和传递信息的物理媒质信道,其余中间环节都是由人来设计的。信息传输性能的好坏,很大程度取决于这些中间环节设计的优劣。

1. 信源、信宿和信道

信源是发送消息的源,根据其输出的性质,有离散信源和模拟信源之分。离散信源输出离散的符号或数字消息序列,如电报机输出在时间上离散的符号序列;模拟信源输出连续波形信号,如麦克风输出连续语音信号。

信源是信息论的主要研究对象之一,但在信息论中并不探讨信源的内部结构和物理机理,而把注意力放在信源的输出上,重点讨论信源输出的描述方法及性质。在认识主体看来,信源的输出都是随机的(具有不确定性),因此,可将信源输出的消息视为某个随机实验的输出或某个随机变量的取值,因此,可用随机数学方法予以处理。另外,从等效的观点来看,图 1.1 中每一个环节的输出,都可视为一个等效信源的输出。信源的数学模型、不确定性测度以及信息度量,将在第二章介绍。

信宿取的是信息归宿之意,亦即收信者或用户,是信息传送的终点或目的地。语义信息和语用信息就是由信宿来理解和判断的。

传输信息的物理媒质通常称为物理信道,如空气、双绞线、同轴电缆、光纤等。物理信道的输入信号是 $S(t)$, 输出信号是 $R(t)$ 。 $R(t)$ 通常是 $S(t)$ 的不完全复现。之所以不完全,是因为存在随机干扰信号,即噪声 $N(t)$, 对于加性噪声,有 $R(t) = S(t) + N(t)$ 。

各种物理信道都有其固有的通过频带。为了使载荷信息的信号频谱结构与信道的通过频带相匹配,在信号送入物理信道之前,必须对信号进行调制,即进行信号频谱迁移,这就是调制器的作用;当信号传送到信道输出端时,对信号进行解调,将信号复原。调制与解调技术,是“通信原理”课程重点讨论的内容之一。在本课程中,我们不专门讨论调制与解调技术,而是将两者与物理信道合并到一起,作为一个等效信道来处理。

其实,图 1.1 中任一输入至任一输出之间的通道,都可看作是一个等效信道。信息论中研究的信道都是等效信道,所关心的是:在噪声干扰下,信道输入至输出之间的状态转移关系。研究信道时,噪声是我们关注的重点。噪声的来源很复杂,主要有以下几种情况:(1) 电路中由于元器件发热而产生的热噪声;(2) 电子和光子设备中的发射噪声;(3) 来源于地球、太阳以及其他宇宙体的电磁辐射。实际上,图 1.1 中每个环节都存在噪声干扰,我们将全部噪声集中等效成一个加在信道上的噪声 $N(t)$, 这样做主要是为了分析方便。

第三章将讨论信道的数学模型,以及信道容量的概念和计算方法。

2. 信源编码器与译码器

前面说过,信息传输系统中,通常给定的是信源、物理信道以及信宿,其余环节都是为保证有效通信而人为加入的。日常生活中进行信息传输时,有时不需要这些中间环节。如两人当面讲话,甲说乙听,甲是信源,乙是信宿,空气则是信道;甲发出的声波,直接通过空气传到乙,中间环节全无。若两人相隔很远,甲费尽全力喊话,由于声波在空气中传播声波会逐渐衰减,所以乙还是听不见。这种情况下,必须借用别的通信手段,其中电通信是较好的选择之一。

要进行电通信, 首先要把实际信源发出的非电信号, 如声音、图像、文字等, 转换成电信号, 这个过程称为换能。换能的方法和技术, 是“检测与转换”研究的内容, 我们这里研究的信源, 都是经过换能之后的等效信源, 即图 1.1 中信源的输出已经是电信号。

以离散的情况为例, 信源发出一个离散符号序列 $\bar{u} = u_1 u_2 \cdots u_N$ 。该序列携带一定量的信息, 这些信息分散在各个符号 u_i 之中。从信息传输的角度看, 总是希望信息传输的效率尽量高, 即希望以最小的代价 (如最短的时间、最小的能量等) 传递尽可能多的信息。如果传送一个序列符号所耗时间是固定的, 那就希望各个符号所携带的信息尽量多, 理想情况下, 希望各个符号携带的信息同样多, 并达到最大。但实际信源未必如此。一般, 信源发出的符号序列中, 各符号携带信息的多少相差很大, 即信息分布不均匀, 因此有必要对这个符号序列加以变换, 使得变换之后的序列信息分布均匀化, 这种变换称为信源编码。信源编码器所输出序列, 其信息分布大致均匀, 且接近最大。因此, 编码之后的序列较“紧凑”, 而编码之前的序列较“松散” (有信息的冗余), 这种由“松散”变为“紧凑”的过程也称数据压缩。总之, 信源编码的实质就是为了去掉信源中的信息冗余。

对于自然性质较好的离散信源, 如我们今后要重点研究的离散无记忆信源, 可以做到无失真编码。离散信源无失真编码的理论与方法, 是第四章讨论的主题。

有些信源, 不可能做到无失真编码。例如, 为了进行数字通信, 必须对模拟信源的输出进行采样, 将其变为离散序列。这样, 量化误差就不可避免了, 即存在编码失真。允许一定失真的编码, 称为限失真编码, 其理论将在第六章讨论。

信源译码是信源编码的逆过程, 如果把信源编码视为变换或映射 T_1 , 信源译码通常就是 T_1 的简单求逆, 即 T_1^{-1} 。

3. 信道编码器与译码器

信道编码也可以看作是一种变换 T_2 , 主要作用是提高信息传送的可靠性。因为有噪声干扰, 等效离散信道在传送某个信息位 (或序列) 时, 总有出错的可能。比如说, 信道的输入为“0”, 但在输出端收到的可能是“1”。为了减小这种传送出错的可能性, 最简单的办法是将这个“0”重复传送多次, 如重复传送 3 次, 即先将“0”变成“000”, 再送入信道传送。把“0”变成“000”是由信道编码器来完成的。“000”中的第一个“0”是载荷信息的, 称为信息位; 后两位是为提高传送可靠性而加入的, 不载荷信息, 称为 (信息) 冗余位。信道编码通常是在信息序列中有目的地加入冗余, 从而使其变“长”, 这与信源编码的做法刚好相反。

由于噪声干扰, 传送“000”或“111”时某些位可能出错, 信道可能输出的是“000, 001, ..., 111”, 要将其恢复成“0”或“1”, 需要进行信道译码——变换 T_2^{-1} 。显然, T_2^{-1} 不是 T_2 的简单反变换。信道译码 (变换 T_2^{-1}) 规则要根据信道的噪声特性而定, 通常不是一一变换, 而是多一变换。此问题比信源译码复杂得多, 需专门讨论。

信道编码与译码的有关问题, 将在第五章讨论。

1.2.2 信息论研究的内容

归纳起来, 信息论研究的内容, 大致包括以下几个方面。

1. 通信的统计理论研究

主要研究利用统计数学工具分析信息和信息传输的统计规律。其具体内容有:

- (1) 信息的测度。
- (2) 信息速率与熵。
- (3) 信道传输能力——信道容量。

2. 信源的统计特性

主要包括：

- (1) 文字（如汉字）、字母（如英文）的统计特性。
- (2) 语音的参数分析和统计特性。
- (3) 图片及活动图像（如电视）的统计特性。
- (4) 其他信源的统计特性。

3. 编码理论与技术的研究

主要包括：

(1) 有效性编码：用来提高信息传输效率，它主要是针对信源的统计特性进行编码，所以也称为信源编码。

(2) 抗干扰编码：用来提高信息传输的可靠性，它主要是针对信道统计特性进行编码，所以也称为信道编码。

4. 提高信息传输效率的研究

主要包括：

- (1) 功率的节约。
- (2) 频带的压缩。
- (3) 传输时间的缩短，即快速传输问题。

5. 抗干扰理论与技术的研究

主要包括：

- (1) 各种调制制式的抗干扰性。
- (2) 理想接收机的实现。

6. 噪声中信号检测理论与技术的研究

主要包括：

- (1) 信号检测的最佳准则。
- (2) 信号最佳检测的实现。

1.2.3 目前信息论的主要研究成果

1. 语音信号压缩

语音信号一直是通信网中传输的主要对象。自从通信网数字化以来，降低语音信号的编码速率就成为通信中的一个重要问题。根据信息理论的分析，语音信号所需的编码速率可以

远远低于仅按奈奎斯特采样定理和量化噪声分析所决定的编码速率。几十年来的研究工作已在这方面取得巨大的进展：长途电话网标准的语音编码速率已从 1972 年原 CCITT G. 711 标准中的 64kbit/s，降低到 1995 年原 CCITT G. 723.1 标准中的 6.3kbit/s。在移动通信中 1989 年欧洲 GSM 标准中的语音编码速率为 13.2kbit/s，1994 年在为半码速 GSM 研究的 VSELP 编码算法中，码速率为 5.6 kbit/s，IS-96 是美国高通（Qualcomm）公司为 CDMA 移动通信研制的一种 CELP 编码，具有 4 种码速率。对语音音质且要求较低的军用通信，美国 NSA 标准的速率在 1975 年时已达到 2.4kbit/s。目前，在实验室中已实现 600bit/s 的低速率语音编码，特别是按音素识别与合成原理构造的声码器其速率可低于 100bit/s，已接近信息论指出的极限。

2. 图像信号压缩

图像信号的信息量特别巨大，这对图像信号的传输及存储都带来极大的不便。经过多年的研究，到 20 世纪 80 年代，图像信号压缩逐步进入建立标准的阶段。1989 年原 CCITT 提出电视电话/会议电视的压缩标准 H. 261，其压缩比达到 25 : 1 到 48 : 1 左右。1991 年原 CCITT 与 ISO 联合提出的“多灰度静止图像压缩编码”标准 JPEG，其压缩比为 24 : 1。在运动图像方面，运动图像专家组（MPEG）继成功定义了 MPEG-1 和 MPEG-2 之后，于 1993 年 7 月开始制订全新的 MPEG-4 标准，并分别于 1999 年初和 2000 年初正式公布了版本 1 和版本 2。到 2001 年 10 月，MPEG-4 已定义了 19 个视觉类（Visual Profile），其中新定义的简单演播室类（Simple Studio Profile）和核心演播室类（Core Studio Profile）使 MPEG-4 对 MPEG-2 类别保留了一些形式上的兼容，其码率可高达 2Gbit/s。随着 MPEG-4 标准的不断扩展，它不但能支持码率低于 64kbit/s 的多媒体通信，也能支持广播级的视频应用。

3. 降低信息传输所需的功率

在远距离无线通信，特别是深空通信中如何降低信息传输所需的功率至关重要。因为在这种情况下发送设备的功率和天线的尺寸都已成为设备生产和使用中的一个困难问题。幸运的是，在这个领域，信息论获得了一批令人信服的成果。从 20 世纪 60 年代后期起，NASA 发射的所有深空探测器无一例外的在其通信设备中采取了信道编码措施，因为根据信息理论的分析，采用低码率的信息编码可以降低传送单位比特所需的能量 E_b 与噪声功率谱密度 N_0 之比。现在利用不太复杂的信道编码就可以使同样误码率下所需的 E_b/N_0 比不采用信道编码时低 6dB 左右。其中一些好的方案（如用 RS 码作为外码、卷积码作为内码的方案）可以使误码率在 10^{-5} 的情况下所需的 E_b/N_0 降到 0.2dB，比不用信道编码时所需的 10.5dB 降低了 10 dB 多。

4. 计算机网中的可靠性数据传输

随着计算机技术的发展，计算机设备的布局变得愈来愈分散，各种终端及外围设备离主机也越来越远，这就产生了计算机网。近年来，由于计算机网还与分布式计算机系统相联系，因而变得更为重要。在用各种电缆连接而成的计算机网中电噪声和各种外界的电磁干扰是必须考虑的，因为它使传输的信息发生差错。一般情况下，局域网中的差错率在 10^{-8} 左