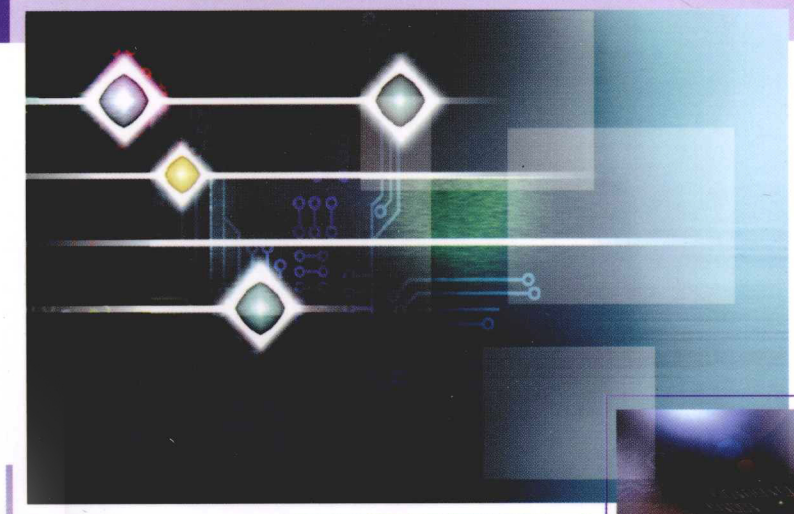
应用型特色

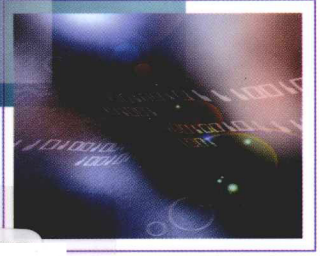
高等学校应用型特色规划教材

XINXILILUN YU BIANMA

信息理论与编码

姚善化　主　编
许恒迎　副主编





免费赠送电子课件

- 内容结构清晰明了，强调信息理论与编码方法在知识上的连贯性。
- 以通俗语言叙述代替高深繁琐的数学推导，便于读者理解。
- 突出基本概念，强调基本方法。

清华大学出版社

高等学校应用型特色规划教材

信息理论与编码

姚善化　主　编
许恒迎　副主编

清华大学出版社
北　京

内 容 简 介

本书以香农信息论为基础，分两大部分共 8 章向读者系统介绍信息理论与编码理论的基本思想：第一部分主要介绍了香农信息论的基本概念和性质，包括信息熵、信息率失真函数和信道容量，力求从基本概念上帮助读者理解和掌握信息理论的基本内容；第二部分以三个基本概念相对应的香农三大编码定理为基础，从满足通信系统的有效性、可靠性和安全性三项性能指标为出发点，详细介绍了无失真信源编码、限失真信源编码和信道编码的基本理论与方法。

本书内容结构清晰明了，以通俗语言叙述代替高深繁琐的数学推导，强调信息理论与编码方法在知识上的连贯性，以满足工科类本科专业学生的学习要求。

本书可作为普通高等院校电气信息类电子信息工程专业和通信专业的教材或教学参考书。

图书在版编目(CIP)数据

信息理论与编码/姚善化主编；许恒迎副主编. —北京：清华大学出版社，2011.1
(高等学校应用型特色规划教材)
ISBN 978-7-302-23896-6

Ⅰ. 信… Ⅱ. ①姚… ②许… Ⅲ. ①信息论 ②信道编码—编码理论 Ⅳ. ①TN911.2

中国版本图书馆 CIP 数据核字(2010)第 185430 号

责任编辑：李春明　郑期彤
装帧设计：杨玉兰
责任校对：王　晖
责任印制：李红英
出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　　编**：100084
社　总　机：010-62770175　　**邮　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质　量　反　馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京市清华园胶印厂
经　　销：全国新华书店
开　　本：185×260　**印　张**：13.75　**字　数**：327 千字
版　　次：2011 年 1 月第 1 版　　**印　次**：2011 年 1 月第1 次印刷
印　　数：1～4000
定　　价：25.00 元

产品编号：038733-01

前　言

自从美国科学家香农(C.E.Shannon)在 1948 年发表了一篇题为“通信的数学理论”的经典学术论文，一门涉及通信技术、自动控制、信号处理、计算机技术、网络技术等众多学科的研究领域——信息理论便诞生了。信息理论是概率论与通信技术相结合的边缘学科，研究领域从自然科学渗透到了政治、经济、军事甚至社会科学。现代科学技术的快速发展已使人们充分认识到，作为信息科学和技术基本理论的信息论在 21 世纪的信息时代起着非常重要的作用。

本书不仅系统和详细地介绍了经典信息论的基本概念、基本性质和基本方法，而且从通信系统基本目的出发，围绕优化和提高通信系统的三项性能指标，系统论述了解决实际问题的理论基础——香农三大编码定理的基本原理、方法和运用。全书分为两大部分共 8 章。第一部分的 4 章主要围绕信息熵、信息率失真函数和信道容量三个基本概念组织内容。该部分各章内容编排如下：第 1 章介绍了信息的基本概念与分类，以及研究信息论的目的和意义；第 2 章介绍了离散信源、离散序列信源和连续信源的数学模型、信源特征以及信息熵，主要包括信息熵和互信息量的概念、性质以及最大熵定理；第 3 章的内容包括离散信源和连续信源的信息率失真函数 $R(D)$的基本概念、性质和计算方法；第 4 章介绍了信道容量的定义，离散信道和连续信道的容量计算，其中包括几个特殊的离散信道容量计算，重点对香农信道容量公式作了详细的介绍。第二部分的 4 章根据第一部分提出的三个基本概念，引入了香农三大编码定理，详细介绍三大定理的主要内容和重要意义，以及信源编码和信道编码。该部分各章内容编排如下：第 5 章介绍了通信系统需要解决的三个问题，以及解决这三个问题所对应的香农编码定理的主要内容；第 6 章介绍了无失真信源编码定理即香农第一编码定理，包括等长编码、变长编码，主要介绍了最优无失真信源编码——霍夫曼编码方法；第 7 章主要内容包括限失真信源编码的保真度准则、限失真信源编码定理即香农第二编码定理，以及常见的语音和图像编码方案；第 8 章介绍了提高系统可靠性的信道编码即香农第三编码定理，包括信道编码概念和准则及几种常见的信道编码方案。

学习本课程需要学生具有高等数学、工程数学以及通信原理等基础知识的储备，作者力求以形象生动的语言对所学内容进行描述，以激发广大读者对本专业基础课的学习兴趣。本书可作为普通高等院校电气信息类电子信息工程专业和通信工程专业的教材，还可作为相关专业的教学参考书。

本书的第 1、5 章由长沙学院的谢明华老师编写，第 2 章由聊城大学东昌学院的郭宗林老师编写，第 3、4、7 章由安徽理工大学的姚善化老师编写，第 6 章由宁夏大学的李春树老师编写，第 8 章和附录部分由聊城大学的许恒迎老师编写，全书由安徽理工大学姚善化担任主编并负责统稿，聊城大学许恒迎担任副主编。另外，本书的编写得到了清华大学出版社的鼎力支持，在此深表感谢。

鉴于信息技术的发展日新月异，新的理论和研究成果层出不穷，本书无法做到全面覆盖，加之作者的学术水平和视野有限，错误和疏漏之处在所难免，敬请各位老师和同学批评指正。

编　者

目　录

第 1 章　概论 1

1.1　信息的概念 1
1.2　信息的分类 3
1.3　信息论的研究目的与意义 5

第 2 章　信源与信息熵 6

2.1　离散信源与信息熵 6
2.1.1　离散信源的数学模型与统计特征 6
2.1.2　自信息量和平均自信息量(熵) .. 7
2.1.3　熵函数的基本性质和最大熵定理 11
2.1.4　联合熵与条件熵 17
2.1.5　互信息 20
2.2　离散序列信源 28
2.2.1　离散序列信源的数学模型 28
2.2.2　离散序列信源的信息熵 29
2.2.3　马尔可夫信源 33
2.2.4　信源冗余度 37
2.3　连续信源 39
2.3.1　连续信源的熵 39
2.3.2　连续信源的最大熵及熵功率 ... 44
习题 46

第 3 章　信息率失真函数 50

3.1　失真测度 51
3.1.1　失真函数与平均失真度 51
3.1.2　信息率失真函数的定义 54
3.2　离散信源的信息率失真函数 55
3.2.1　信息率失真函数的性质 55
3.2.2　信息率失真函数的计算 58
3.3　连续信源的信息率失真函数 66
3.3.1　信息率失真函数的定义与性质 67
3.3.2　信息率失真函数的计算 68
习题 69

第 4 章　信道与信道容量 71

4.1　信道的分类与数学模型 71
4.1.1　信道的分类 71
4.1.2　信道的数学模型 72
4.1.3　信道容量的定义 77
4.2　信道容量的代价函数和信道冗余度 ... 79
4.2.1　信道容量的代价函数 79
4.2.2　信道冗余度 80
4.3　离散信道及其容量计算 81
4.3.1　单符号离散信道的信道容量 ... 81
4.3.2　多符号离散信道的信道容量 ... 89
4.4　连续信道及其容量 90
4.4.1　时间离散信道的信道容量 90
4.4.2　时间连续信道的信道容量 93
4.4.3　限频率、限时、限功率的 AWGN 信道容量 94
4.5　信道容量 C 与信息率失真函数 $R(D)$ 的区别 97
4.6　多用户信道 98
4.6.1　多址接入信道 99
4.6.2　广播信道 101
习题 102

第 5 章　香农三大定理 105

5.1　香农第一定理 105
5.2　香农第二定理 108
5.3　香农第三定理 110

第 6 章 无失真信源编码 112
6.1 信源编码概述 112
6.1.1 信源编码的一般模型 112
6.1.2 信源产生冗余的原因 113
6.2 无失真信源编码概述 115
6.2.1 编码的有关概念 115
6.2.2 几个简单的信源编码器 116
6.2.3 几种常见码 117
6.3 等长码与等长信源编码定理 119
6.3.1 无失真编码条件 119
6.3.2 信源序列渐近均分性 121
6.3.3 信源序列分组定理 122
6.3.4 渐近均分特性 124
6.3.5 等长编码定理 124
6.4 变长编码 126
6.4.1 码树 127
6.4.2 异前置码 127
6.4.3 克拉夫特-麦克米伦不等式 128
6.4.4 变长编码定理 129
6.5 最佳变长编码(霍夫曼编码) 132
6.5.1 二进制霍夫曼编码 133
6.5.2 多进制霍夫曼编码 138
习题 139
第 7 章 限失真信源编码定理 141
7.1 限失真信源编码概述 141
7.2 限失真信源编码逆定理 147
7.3 保真度准则下的码率压缩标准 149
7.3.1 语音压缩编码标准 149
7.3.2 图像压缩编码标准 154
7.4 几种常用的有损压缩编码技术 160
7.4.1 差分脉冲编码调制(DPCM) 160
7.4.2 预测编码 161
7.4.3 正交变换编码 165
7.4.4 小波变换编码 171
习题 173
第 8 章 信道编码 175
8.1 信道编码的基本概念 175
8.1.1 检错和纠错原理 175
8.1.2 检错和纠错能力的判断 176
8.1.3 信道编码的分类 177
8.1.4 常用的简单检错纠错码 178
8.2 错误概率 180
8.2.1 译码规则的概念 180
8.2.2 错误概率与译码规则 181
8.2.3 三种译码准则 182
8.2.4 错误概率与编码方法 187
8.3 线性分组码 190
8.3.1 线性分组码的基本概念 190
8.3.2 校验矩阵和生成矩阵 190
8.3.3 汉明码 193
8.3.4 线性分组码的译码 194
8.4 循环码 195
8.4.1 循环码的定义和多项式描述 195
8.4.2 码多项式的运算 196
8.4.3 循环码的生成多项式和生成矩阵 197
8.4.4 循环码的校验矩阵 198
8.4.5 循环码的编码和译码方法 199
8.5 卷积码 201
8.5.1 卷积码的概念 201
8.5.2 卷积码的编码 201
8.5.3 卷积码的图形表示 202
8.5.4 卷积码的译码 204
习题 206
附录 A 常用概率公式 208
附录 B 詹森不等式 209
参考文献 210

第1章 概 论

人们从开始意识到信息的存在到进入信息社会，对信息的认识随着社会文明程度的提高不断深入。在计算机技术的大力推动下，从20世纪末到21世纪初，信息技术获得了飞速的发展，现代电子信息产品几乎渗透了社会的各个领域，有力地推动了生产力的发展和社会信息化程度的提高。信息理论是概率论与通信技术相结合的边缘学科，研究领域从自然科学渗透到了政治、经济、军事甚至社会科学。现代科学技术的快速发展已使人们充分认识到，作为信息科学和技术基本理论的信息论在21世纪的信息时代起着非常重要的作用，它对我们的生产方式、生活方式、学习方式及思维方式产生了深远的影响。

1.1 信息的概念

人类从出生起就开始接收并向外界传递信息，那么，信息是什么？它都有哪些表现形式呢？在人们的日常生活中，信息无处不在，人与人之间的交流，包括一段谈话、一个眼神、一个微笑等都包含了信息，文字、图片、交通标志等也承载着信息；在自然界中，风向、温度、云朵等的状态包含了丰富的天气信息，植物的生长、动物的行为等也在向外界传递着信息；在现代通信系统中，大量的图像、声音、文字等信息通过光纤、无线通信等方式在网络中传播，人们可以通过手机、电脑等设备获得并发送信息。

从信息的表现形式可以看出，信息可以带来新的知识，例如听老师讲课可以使学生获得新的知识；信息以不同的形式存在，如一个新闻事件所包含的信息可以以电视、广播或报纸的形式向公众传播；信息仅仅在接收者能感受到的时候才有意义，例如警报器所发出的声音信息可以提醒人们警觉并采取相应的措施，但是对于没有听到或者没有注意到报警声音的人，则该信息完全没有意义。但是如果换一个角度来看信息，可能会得出不同的结论。信息并不总能带来新的知识，例如某同学在交谈中告诉你一个你已经很熟悉的知识，在该同学看来，信息已经发送出去了，而且你也准确无误地接收到了对方传递过来的信息，但是该信息的获得并没有增加你本身所具有的信息的总量，所以对于那些已经具有了的信息，后续接收到的同样信息对接收者没有任何意义。同样的信息在不同的时间或背景下对同一个接收者来说所能获得的有效信息的多少也有区别，例如某同学拿到成绩单的时候发现成绩提高了，如果他前段时间认真努力学习了，成绩提高在他的预料之中，那么成绩提高这件事情对他来讲就不算意外，或者说该信息早就在他的预料之中，成绩单给他带来的新的信息并不多；反之，通过努力之后，成绩反而下降了，也就是成绩与预期差距太大，那得知成绩的时候可能就会觉得非常不可思议，同时他接收到的新的信息就更多一些。类似的，不同的接收者接收到同样信息时所获得的有用信息往往也并不一样。另外，信息并不一定是正确的，它可以为真也可以为假。例如对通信造成严重干扰并导致产生误解的噪声信号也是信息的一种形式。

正是因为信息的抽象性与复杂性，所以信息如何定义、信息是否可以度量、如何度量、信息如何进行正确的传输、接收方接收到信息之后如何判断信息是否可靠、如何对信息进行存储等都是研究信息必须要考虑的问题。

信息是独立于物质和能量的客观世界的第三要素，它与物质和能量存在本质的区别，属于一种非实体的存在。它看不见摸不着，没有具体的形态、没有大小、没有重量，但是它却通过依附在如图片、声波、人类本身等物质形式上处处反映出信息对象及信息影响力的存在。它存在于我们周围，与物质与能量互相区别，又互相依赖。信息可能依附在一切存在的物质上，我们称信息的依附物质为信息的载体。所以信息本身并不会直接占用空间，直接占用时间和空间的是信息的载体。从广义上说，信息是人们对客观事物运动规律及其存在状态的映射与显示。哲学层次上来看，信息就是通过特定载体对存在者状态及变化过程的映射与显示，是存在者间接的非实体性存在。我们主要研究的是通信系统中的信息，属于狭义信息范畴，即可以承载在如电信号、光信号、无线电信号等具体信号上的信息。信息以信号的形式体现它的存在，信号是信息的载体，也可以看作是信息的外壳或外在表现形式。所以信号与信息的概念有本质的区别，它们属于不同的层次。另外，消息与信息也不完全相同，消息是用来描述事物的特征和状态的信息，属于信息概念中的语义信息。

信息作为一种客观存在，具有如下重要性质。

(1) 存在的普遍性。信息存在于自然界，也存在于人类社会，其本质是事物的运动变化。哪里有事物的运动和变化，哪里就会产生信息。

(2) 信息与载体的不可分性。信息来源于物质，但又不是物质本身，信息必须通过依附在物质载体上才能产生并传播，世界上不存在游离于物质载体之外的信息。

(3) 信息的时效性。这指的是信息从产生、发送、接收到进入利用的时间间隔及其效率。

(4) 信息的可扩充与可压缩性。由于信息存在的普遍性，在任何时刻、任何领域都会不断地产生信息，信息的总量将会不断地变大。同时，为了更好地利用不断增加的信息，需要对信息进行整理、浓缩，以使信息更便于存储、理解和学习。

(5) 信息载体的可替代性。传递信息可以使用不同的信息载体，例如想向他人传达某种信息可以通过电话、面对面交谈或者信件等不同的物质载体形式来实现。

(6) 信息的共享性。信息可以多人分享，例如老师讲课，听课的学生都能共享到老师传授的新的知识，都获得了信息。

(7) 信息的相对性。面对发出的相同信息，不同的接收方最终接收到的信息可能存在差异，或者虽然接收到相同的信息，但最终从该信息中所获得的有效信息量可能并不相同。

(8) 信息的可度量性。信息虽然是一种非实体性存在，但是信息的多少是可以度量的，我们用信息量来表示信息的多少。

(9) 信息的可传递性与可扩散性。信息可以通过物质载体进行空间和时间上的传输，通过传输，可以使信息迅速扩散开来。

(10) 有序性。信息可以用来消除认识主体对于事物运动状态和方式的不确定性，增加认识主体信息系统的有序性。

对信息的研究就是信息科学，信息论是信息科学的主要理论基础之一。1924 年，奈奎斯特(Harry Nyquist)解释了信号带宽和信息传输速率之间的关系；1928 年，哈特利(R.V.Hartley)

提出用对数度量信息；1936 年，阿姆斯特朗(E.H.Armstrong)提出增大带宽可以使抗干扰能力增强；1939 年，达德利(H.Dudley)提出通信所需带宽至少应该与待传送消息的带宽一样。

美国数学家、电子工程师、计算机科学家香农(C. E. Shannon)(见图 1-1)在 1941—1944 年对通信和密码进行了深入的研究，用概率论和数理统计的方法对通信系统进行了深入的剖析。他认为信息是不确定性的减少或消除，提出了通信系统传输的对象就是信息，并对信息给出了科学的度量方法，首次提出了信息熵的概念，还指出了通信系统的核心问题就是如何在有噪信道上稳定可靠地传输信息，以及通过有效的编码方法来保证通信系统的运行。香农在 1948 年 10 月的《贝尔系统电话杂志》上发表了他的经典论文“通信的数学理论”。这篇论文中总结了他的研究成果，从此便诞生了一门涉及通信技术、自动控制、信号处理、计算机技术、网络技术等众多学科的研究领域。随着人们对信息理论的研究日益广泛和深入，其基本思想和方法已经渗透到了许多学科，尤其是在人类社会进入信息时代的今天，信息理论还会发挥更大的作用。

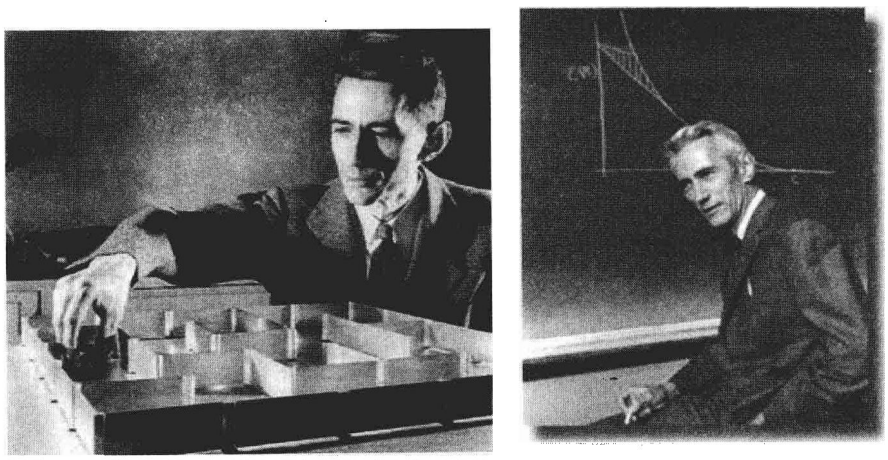


图 1-1 香农

1.2 信息的分类

信息是普遍存在，并不断变化和发展的概念，各个领域的信息总量在不断增加，为了使人们对信息的多元化、多层次有一个更清晰的认识，以更好地研究信息，需要对各种信息进行分类。

按照信息的内容可将信息分为自然信息和社会信息。自然信息是自然界客观存在的各种生命与非生命物质的特征、运行状态和方式以及互相之间内在联系的反映。社会信息是人类在社会实践中为生产、生活和社会发展产生、传播和利用的信息，包括政治、经济，军事、文化等人类社会特有的内容。

按照认识主体的认识层次可以将信息分为语法信息、语义信息和语用信息。语法信息是信息认识过程的第一个层次，指的是事物的状态和状态的改变方式，研究事物运动出现的各种可能状态和这些状态之间的联系。语义信息是信息认识过程的第二个层次，表示认

识主体感知的事物运动状态及方式的具体含义，它不仅反映事物运动变化的状态，而且还要揭示事物运动变化的意义。语用信息是信息认识过程的第三个层次，指事物运动状态及其状态变化的方式，以及其含义对观察者的效用或价值，研究的是信息的客观价值。信息的有用性取决于观察者对信息的需求状况，也就是由观察者的信息状态与信源发出的信息间的相关性所决定。语法信息、语义信息和语用信息三者不是彼此独立的，而是密切相关、互相依存。

在上述三个层次中，语法信息是最基本、最抽象的类型，它只表现事物的现象而不考虑信息的内涵。在通信系统中，语法信息主要研究信息发送端的性质，如正确表达信息所需要的符号、编码方式，信号在信道中的传输模式、速度等，是迄今为止在理论上研究最多的类型。语法信息还可以进一步分为有限和无限两种状态。另外，根据事物运动的状态可以将语法信息分为连续状态语法信息和离散状态语法信息，或者分为明晰语法信息和模糊语法信息；根据事物运动状态的改变方式还可以将语法信息分为概率信息、偶发信息、确定信息和模糊信息四种细分类型。香农信息论主要研究的就是语法信息中的概率信息。

语义信息又称意义信息，它不考虑使用者个人的主观因素。比如，说信息是“真实的”、“确切的”，是语义性的评判，而说信息是“有用的”、“有价值的”，则为语用性的断语。对语义信息的度量是以同构为基础的，即传递的符号系统是否与大脑内部语言同构，这种大脑内部语言又是否与外部世界同构。同构链上的转换即为外部世界信息意义的传递。如下面三个逻辑为真的语句：“有列火车要开出”，“一列火车今天从广州到北京”，“X次特快Y时从广州开往北京”。其中以第三句的定义最为精确，同构最为严格，因此语义信息量最大。由于语言的开放性，语义信息的度量变得极为困难。信源和信宿对语义的理解差异，以及语音、语调的变化，都是语义信息度量的障碍。通过信息加工可以增加信息的语义。如决策就是一种典型的信息加工过程。在决策过程中，信息被赋予新的语义。但是这种过程是不可逆的。

语用信息是最高层次的信息。它以语法、语义信息为基础，不仅要考虑状态和状态之间关系以及它们的含义，还要进一步考察这种关系及含义对于信息使用者的效用和价值，语用信息和语义信息一样，都能减少人们对所描述事物意义上的不确定性，但语义信息通常只回答“是什么意思”的问题，而语用信息则回答“为什么是这个意思”的问题。例如，一般人都知道熟透了的苹果会落到地上这个现象，这就是人们从这个现象获得的语义信息。而牛顿则不仅知道这个现象，而且还从这个现象中发现了导致苹果落地的原因，也就是说，他从这个现象中还获得了语用信息。由于人的知识水平和认识能力不同，因而不同的人会从同一语法信息中得到不同的语义信息和语用信息。语用信息比语义信息更依赖于接收者，而且与时间的关系更密切。一个过时的信息，将会变得毫无价值，从中得不到有用的信息。可见，对于语用信息，不仅包含了语义信息这种复杂因素，而且还要包含效用这种带主观意义的因素，这使信息理论更加符合实际，但也更加复杂。

信息分类还有许多种不同的准则和方法。如按信息的地位可以将信息分为客观信息和主观信息；按信息的作用可以将信息分为有用信息、无用信息和干扰信息；按信息的逻辑意义可以将信息分为真实信息、虚假信息和不确定信息；按信息的应用部门可以将信息分为工业信息、农业信息、政治信息、经济信息和军事信息等；按信息是否经过加工可以将信息分为原始信息与加工信息。

总之，只有针对不同类型的信息，研究它的特征，并建立具体的描述方法与度量方法，才能更有效地利用信息。

1.3 信息论的研究目的与意义

信息论研究的是事物运动状态及其改变方式的外在形式，即语法信息。它采用概率论与数理统计方法来研究通信和控制系统中普遍存在的信息传输过程的共同规律以及信息的获取、传输、存储、转化等问题，是当代信息科学的理论基础和核心。信息论研究的内容十分广泛，一般把信息论的研究范畴概括为以下三个层次。

(1) 狭义信息论。也叫香农信息论、经典信息论、概率信息论、语法信息论等。是一门应用数理统计方法来研究信息处理和信息传递的科学。它主要研究信息通信过程中普遍存在着的共同规律，以及如何提高各信息传输系统的有效性和可靠性的通信理论。

(2) 一般信息论。主要是研究通信的一般理论问题，除了包括以香农为代表的狭义信息论之外，还包括以维纳(Norbert Wiener)的研究成果为代表的噪声理论、信号滤波与预测、调制与信息处理等问题。香农研究的对象是从发送端到接收端之间的信息传输过程，是收发端联合最优化问题，其重点是编码，包括信息测度理论、信道容量理论和编码理论。而维纳的研究独立于香农，重点是在接收端。他将统计方法引入通信工程，从带直流电流或者至少可看作直流电流的电路出发来研究信息论，致力于研究信息在传输过程中被某些因素干扰时，如何在接收端去除干扰的影响，恢复该信息。

(3) 广义信息论。是指利用狭义信息论的观点来研究一切信息通信问题的理论，不仅包括狭义信息论和一般信息论的问题，而且还包括所有与信息相关的领域，如医学、生物学、心理学、语言学、神经心理学、遗传学等。或者说，所有有关信息的提取、识别、变换、传输、存储等信息处理的一般规律以及相关的技术手段的学科，都属于广义信息论的范畴。

研究信息论的目的就是要找到信息传输过程的共同规律，以提高信息传输的可靠性和有效性。可靠的或者说不失真的信息能够带来应用价值，任何接收方都希望接收到的信息是与发送方完全一致的、准确无误的信息。但是，信息的不可靠性广泛存在，从严格的角度来讲，要获得绝对可靠的信息是相当困难的。例如，用于产生并发送信息的人的大脑、感官或各种仪器设备都存在一定的不可靠性；信息的传递过程中，也将受到各种各样的干扰与破坏等。所以信息传输系统的可靠性取决于信息产生、传递、接收或存储等复杂过程中的可靠性。信息社会中，面对急剧增长的信息，如何更好地获取信息，传递信息，存储信息，提高信息传输系统的可靠性，将各种受影响或干扰的信息进行估计、修复和提取，显然是非常重要的。信息传输系统的有效性研究是指如何在尽可能短的时间，使用尽可能少的资源实现一定量信息的传送，或者在每一个传送符号内携带尽可能多的信息量。在信息传输系统中，提高可靠性和提高有效性往往是互相矛盾的，所以，若要获得信息传输系统的最佳性能，就需要统筹考虑。

本书仅限于介绍狭义信息论，即香农信息论，主要探讨通信系统的数学描述与定量分析，如信源、信道的分析及描述，信息的度量，系统的最优状态与优化理论(包括信道与通信系统之间的匹配以及通信系统的优化等)。

第 2 章　信源与信息熵

通信的根本目的是有效、可靠、安全地传递与交换信息，而信息是由信源产生的，香农信息论所研究的信源都具有随机性，根据随机性，信源可分为离散信源和连续信源，相应的输出为离散消息和连续消息。其中离散信源的样本空间和取值都是离散的，离散信源如果是单符号的，则其对应于一个离散型的随机变量；如果是多符号的，就对应于离散随机序列，称为离散序列信源或离散矢量信源。研究信源的核心问题就是要弄清楚信源输出消息的统计特征、消息所包含信息的度量、信源平均不确定性的度量，而信息熵则是信源消息平均携带信息量大小的度量，所以信息熵是本章重点讨论的内容。本章首先讨论了信源的数学模型和统计特征，给出了自信息量、平均自信息量和互信息量的表达式，然后讨论了信息熵的基本性质和最大熵定理，分析了信源的相关性和冗余度，最后介绍了连续信源和连续熵的基本性质。

2.1　离散信源与信息熵

2.1.1　离散信源的数学模型与统计特征

什么是信源？信源就是信息的来源，发源地。从广义上讲，信源可以是人、物、设备和其他任何事物。在通信系统中，最常见的信源就是语言、文字、图像、数据等，具体的输出形式可以是离散的，也可以是连续的。从信息论的角度来看，信源就是产生消息符号、消息序列和连续消息的源。信源发出消息，消息是承载信息的工具，而消息又具有不确定性，因此可以用不确定的随机变量、随机序列或随机过程来描述信源发出的消息。因此信源的特征是具有随机不确定性，从数学的角度上可以用概率来描述这种不确定性。

信源输出的消息若以一个个符号的形式出现，例如文字、字母、电报、计算机代码等，表示这些信源输出的消息数是有限的或是可数无穷的，则称这样的信源为离散信源；若每次输出的只是一个消息符号，只涉及一个随机事件，则称为离散单符号信源，其输出的消息在时间上和幅值上均是离散的。

最简单、最基本的信源就是离散单符号信源，它是组成实际信源的基本单元。实际生活中有很多这样的信源，例如投硬币、掷骰子、书信文字、计算机的代码、电报符号、阿拉伯数字符号码等。其样本空间和输出取值都是离散的，所谓离散就是集合的元素个数是至多可列的，即有限的或者可数的，所以可以用离散随机变量的样本空间和概率空间来描述其数学模型。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有序对$[X,P(X)]$称为信源的概率空间，有

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} X=x_1 & \cdots & X=x_i & \cdots & X=x_n \\ p(x_1) & \cdots & p(x_i) & \cdots & p(x_n) \end{bmatrix} \tag{2-1}$$

其中，X 表示信源输出的消息整体，x_i 表示某个消息，$p(x_i)$ 表示消息 x_i 出现的概率。n 是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。$p(x_i)$ 满足概率空间的非负性和完备性，即

$$0 \leqslant p(x_i) \leqslant 1, \sum_{i=1}^{n} p(x_i) = 1 \tag{2-2}$$

当信源给定，其相应的概率空间就已给定了；反之，如果概率空间给定，这就表示相应的信源给定。所以概率空间能表征离散信源的统计特性，因此有时也把这个概率空间称为信源空间。

这里需要特别注意的是字母的大小写不能混淆，大写 X 代表的是随机变量，指的是信源整体，小写 x 代表的是随机事件发生的某一结果或信源的某个元素。

例如，掷一枚均匀的骰子，落地后朝上一面的点数是一个离散型的随机变量 X，呈离散型均匀分布，基本的样本点为 1 点，2 点，3 点，4 点，5 点，6 点，样本空间为 {1，2，3，4，5，6}，每个样本点的概率相同。也可以将掷骰子看成一个信源，输出的消息是“朝上的是 1 点”、“朝上的是 2 点”、……、“朝上的是 6 点”等六个不同的消息，实验之前不知道是哪个消息出现，但必然是这六个消息中的一个，而且这六个消息是互斥的，即不能同时发生。这六个消息构成了互不相容的基本事件集合，用符号 x_i，$i=1,\cdots,6$ 来表示这些消息，得到这个信源的样本空间为符号集 $A:\{x_1,x_2,x_3,x_4,x_5,x_6\}$。因此，可以用一个离散型随机变量 X 来描述这个信源输出的消息。这个随机变量 X 的样本空间就是符号集 A；而 X 的概率分布就是各消息出现的先验概率，为

$$P\{X=x_1\}=P\{X=x_2\}=P\{X=x_3\}=P\{X=x_4\}=P\{X=x_5\}=P\{X=x_6\}=\frac{1}{6}$$

因此，这个信源的数学模型就为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1, & x_2, & x_3, & x_4, & x_5, & x_6 \\ \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6} \end{bmatrix}$$

并满足

$$\sum_{i=1}^{6} P(x_i) = 1$$

上式表示信源的概率空间必定是一个完备集，信源输出的消息只可能是符号集 $A:\{x_1,x_2,x_3,x_4,x_5,x_6\}$ 中任何一个，而且每次必定选取其中一个。所以这是一个离散单符号的信源。

2.1.2 自信息量和平均自信息量(熵)

要想定量地研究信息的传输问题，我们首先要解决信源能输出多少消息？每个消息的出现又携带了多少信息量？下面我们来讨论这个问题。

长期以来，人们就想对信息进行度量，就像对体积和重量一样。1928 年，信息论的奠

基者之一哈特利对具有 n^m 个组合的单个消息信源进行了研究，并给出最早的信息度量公式

$$I = \log n^m = m \log n \tag{2-3}$$

其中，n 为单个消息信源输出的消息符号的可取样值数，m 为消息符号的组合长度。

1948 年，美国数学家香农对上述公式做了进一步研究，保留了其对数度量信息的合理性(因为对数能直接反映信息的可加性)，并从概率的角度进行了分析。他发现，信源包含的信息与其不确定性密切相关，而根据概率论知识，概率可以作为衡量不确定性的一种指标。香农将特殊的等概率信源进一步推广至一般的不等概率的信源，并推断出，信源所发出的消息是不确定的随机事件，随机事件包含的信息度量应是其发生概率的函数。下面就从直观的概念来分析概率信源的信息度量。

我们从信源输出一个消息，也就是根据信源的概率分布在样本空间中随机抽取一个消息，输出的结果是不确定的；而输出之后，我们可以得到一个确定的结果，获得有关信源的一定量的信息，不确定性也随之消失。由此可知，信息是和不确定性相关联的。

究竟有什么关联呢？我们进一步讨论这一点，不妨假设比较极端的情况。比如有人告诉你一个消息："某人买彩票没中奖"，你会觉得很正常，并没有得到什么新信息，因为这个消息发生的概率几乎为 1，所以不确定性基本没有；但是如果有人告诉你"某人中了一亿元大奖"，你会很震惊，获得了大量的新信息，因为这个事件发生的概率非常小，所以不确定性非常大。在这种情况下，信息量和不确定性是等价的。

从上面的讨论我们发现，事件发生的不确定性与事件发生的概率有关。事件发生的概率越小，我们猜测它有没有发生的困难程度就越大，不确定性就越大。而事件发生的概率越大，我们猜测这事件发生的可能性就越大，不确定性就越小，对于发生概率等于 1 的必然事件，就不存在不确定性。因此，随机事件的自信息量 $I(x_i)$ 是该事件先验概率 $P(x_i)$ 的函数，并且 $I(x_i)$ 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就越大。

(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) 两个独立事件的联合信息量应等于它们各自的信息量之和，即统计独立信源的信息量等于它们各自的信息量之和。

可以证明，满足以上公理化条件的函数形式是对数形式，即

$$I(x_i) = \log \frac{1}{P(x_i)} \tag{2-4}$$

其中，$I(x_i)$ 表示事件 x_i 发生所含有的信息量，即自信息量。$P(x_i)$ 的定义域是[0,1]，且 $P(x_i) \geqslant 0$。

自信息 $I(x_i)$ 的含义具有双重性：当事件 x_i 没有发生时，它表示事件 x_i 发生的不确定性；当事件 x_i 发生以后，则表示事件 x_i 所含有或能提供的信息量。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，$I(x_i)$ 表示收信者接收到消息后所获得的信息量，即消除了 $I(x_i)$ 大小的不确定性，所获得这么大小的信息量。通俗地说，你手里有 100 块钱，没花时，说明你有 100 块的购买能力；若你全部花掉了，那么就是你花了 100 块钱或者提供了 100 块钱的购买能力。

在式(2-4)的定义中，对数的底没有加以说明，选择不同的底仅仅改变计量的尺度——单位。最常见的底为 2，e 和 10。如果以 2 为底，则所得的信息量单位称为比特(bit，binary unit 的缩写，即二进制数的缩写)；如果以 10 为底，则所得的信息量单位称为哈特(Hart，Hartley 的缩写，以纪念哈特利首先提出用对数来度量信息)；如果以 e 为底，则所得信息量单位称为奈特(nat，nature unit 的缩写)。

根据对数换底公式有

$$\log_a x = \frac{\log_b x}{\log_b a}$$

进一步可得

$$1 \text{ 奈特} = \log \mathrm{e} \approx 1.443 \text{ 比特}$$

$$1 \text{ 哈特} = \log_2 10 \approx 3.332 \text{ 比特}$$

在通信中一般都采用以 2 为底的对数，为了书写简洁，把底数“2”略去不写。

从自信息的定义可以看出，如果 $P(x_i)=\frac{1}{2}$ 则 $I(x_i)=1$ 比特。因此一个比特的信息量就是从两个等可能事件中任取一个时所含的信息量，比如投硬币问题的信息量就是 1 比特。

【例 2-1】 同时掷一对相同的骰子，每个骰子的六个面分别表示为“1”、“2”、“3”、“4”、“5”、“6”，假定每个骰子各面出现的概率均为 1/6，试问：

(1) 骰子中“2”和“6”同时出现这一事件的信息量是多少比特？

(2) 骰子中两个“3”同时出现这一事件的信息量是多少比特？

(3) 两个骰子中至少有一个“1”这一事件的信息量是多少比特？

(4) 两个骰子出现的两点之和为“4”这一事件的信息量是多少比特？

解：

(1) 设事件 A 表示“2”和“6”同时出现，由于“2”和“6”各自出现的概率为 1/6，则第一个骰子出现“2”、第二个骰子出现“6”这一事件的概率为 1/36；第一个骰子出现“6”、第二个骰子出现“2”这一事件的概率也为 1/36。

则事件 A 出现的概率为

$$P(A)=1/36+1/36=1/18$$

事件 A 出现后所提供的信息量为

$$I(A)=-\log 1/18=\log 18=\log 2+2\log 3=4.16 \text{(比特)}$$

(2) 设事件 B 表示两个“3”同时出现，由于两个“3”同时出现的概率为 $P(B)=1/36$，则事件 B 出现后所提供的信息量为

$$I(B)=-\log 1/36=\log 36=\log 2+\log 18=5.16 \text{(比特)}$$

(3) 设事件 C 表示两个骰子均不出现“1”，由于第一个骰子不出现“1”的概率为 5/6，同样第二个也为 5/6，则两个骰子均不出现“1”的联合概率 $P(C)=25/36$，那么两个骰子至少有一个“1”的事件 $\overline{C}$ 出现的概率为 $P(\overline{C})=11/36$。

事件 $\overline{C}$ 出现后提供的信息量为

$$I(\overline{C})=-\log 11/36=1.71 \text{(比特)}$$

也可以这样做：两个骰子同时为“1”的概率为 1/36；第一个骰子为“1”、第二个骰子不为“1”的概率为 1/6×5/6=5/36；第一个骰子不为“1”、第二个骰子为“1”的概率为 5/6×1/6=5/36，则两个骰子中至少有一个骰子出现“1”这一事件 $\overline{C}$ 的联合概率为

$$P(\overline{C})=1/36+5/36+5/36=11/36$$

事件 $\overline{C}$ 出现后所提供的信息量为

$$I(\overline{C})=-\log 11/36=1.71\text{(比特)}$$

(4) 设事件 D 表示两个骰子出现两点数的之和为“4”。由于第一个骰子为“1”、第二个骰子必出“3”的条件概率为 1/36；第一个骰子为“2”、第二个骰子必出“2”的条件概率为 1/36；第一个骰子为“3”、第二个骰子必出“1”的条件概率为 1/36，

则事件 D 出现的联合概率为

$$P(D)=3\times 1/36=1/12$$

事件 D 出现后所提供的信息量为

$$I(D)=-\log 1/12=\log 12=3.58\text{(比特)}$$

由于自信息 $I(x_i)$ 是消息 x_i 的函数，所以 $I(x_i)$ 也是一个随机变量，不能用来表征整个信源的不确定度。所以我们用平均自信息量来表征整个信源的不确定度。平均自信息量就是信源输出所有消息的自信息的数学期望，又称为信息熵、信源熵，简称熵。即

$$H(X)=E\left[\log\frac{1}{p(x_i)}\right]=-\sum_{i=1}^{n}p(x_i)\log p(x_i) \tag{2-5}$$

在统计物理学中，热熵是一个物理系统杂乱性(无序性)的度量。香农从物理学中把热熵的概念借用过来，用熵来描述信源的平均不确定性。由于在热力学中，对于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，信息熵正好相反，只会减少不会增加，所以有人称信息熵为负热熵。字母 H 是用来纪念波尔兹曼(L.E.Boltzmann)的，他第一个给出这种类型(气体统计力学)的定义，并指定用 H 来表示。

熵的单位也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/符号。通常以 2 为底时，信息熵写成 $H(X)$ 的形式，单位为比特/符号。

r 进制信息熵 $H_r(X)$ 与二进制信息熵 $H(X)$ 的关系是

$$H_r(X)=\frac{H(X)}{\log r} \tag{2-6}$$

信息熵 $H(X)$ 是对信源的平均不确定性的描述。它是从平均意义上来表征信源的总体信息测度的。对于某特定的信源(概率空间给定)，其信息熵是一个确定的数值。

信息熵具有如下三种物理含义。

第一，信息熵 $H(X)$ 是表示信源输出后，每个消息(或符号)所提供的平均信息量。

第二，信息熵 $H(X)$ 是表示信源输出前，信源的平均不确定性。

例如有两个信源，其概率空间分别为

$$\begin{bmatrix}X\\P(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2\\0.99 & 0.01\end{bmatrix},\quad \begin{bmatrix}Y\\P(y)\end{bmatrix}=\begin{bmatrix}y_1 & y_2\\0.5 & 0.5\end{bmatrix}$$

则信息熵分别为

$$H(X)=-0.99\log 0.99-0.01\log 0.01=0.08\text{（比特/符号）}$$

$$H(Y)=-0.5\log 0.5-0.5\log 0.5=1\text{(比特/符号)}$$

可见 $$H(Y)>H(X)$$

即信源Y比信源X的平均不确定性要大。通俗地说，第一类情况有A、B两种选择，A成功的可能性是99%，B是1%，当你选择时，你会毫不犹豫地选A；而另一类情况中，A和B成功的可能性都一样，当你选择时，你将犹豫不决。所以第一类情况的不确定性小，第二类情况的不确定性大。因此，信息熵正好反映了信源输出消息前，接收者对信源存在的平均不确定程度的大小。

第三，信息熵$H(X)$可表征变量X的随机性。如上例，变量Y取y_1和y_2是等可能的，所以其随机性大。而变量X取x_1比x_2的概率大很多，这时，变量的随机性就小。因此，$H(X)$反映了变量的随机性。信息熵正是描述信源输出每个离散消息所提供的平均信息量。

【例2-2】 随机变量X服从于几何分布，其概率分布为

$$p(x_i)=p(1-p)^{i-1},i=1,2,\cdots,\infty$$

求$H(X)$。

解：

$$H(X)=-\sum_{i=1}^{\infty}p(1-p)^{i-1}\log\left(p(1-p)^{i-1}\right)=-\sum_{i=1}^{\infty}p(1-p)^{i-1}\left(\log p+\log(1-p)^{i-1}\right)$$

$$=-\log p\sum_{i=1}^{\infty}p(1-p)^{i-1}-p\log(1-p)\sum_{i=1}^{\infty}(i-1)(1-p)^{i-1}=-\log p-p\frac{(1-p)}{p^2}\log(1-p)$$

$$=-\log p-\frac{(1-p)}{p}\log(1-p)=\frac{-p\log p-(1-p)\log(1-p)}{p}=\frac{H(p)}{p}\text{(比特/符号)}$$

由此可以看出，自信息量与信息熵的含义是不同的。

(1) 信息熵是表征信源本身统计特性的一个物理量，它表示信源的平均不确定性，是信源输出的每一个消息所能提供的平均信息量；自信息量表示的是每一个消息的信息量度。

(2) 信息熵是针对信源的，是信源输出的信息量，表示信源输出前的平均不确定性；自信息量是针对信宿的，是接收者在消除了信源不确定性后所获得的信息的度量。

(3) 若信道无干扰，接收者获得的信息量在数量上等于信源的熵，若有干扰时，则两者不相等。

2.1.3 熵函数的基本性质和最大熵定理

信息熵$H(X)$是随机变量X的概率分布$P(x)$的函数，函数关系由式(2-5)确定，所以又称为熵函数。当信源含有n个离散消息时，熵函数又可以写成概率矢量$\boldsymbol{P}=(p_1,p_2,\cdots,p_n)$的函数形式，记为$H(\boldsymbol{P})$，有

$$H(X)=-\sum_{i=1}^{n}p_i\log p_i=H(p_1,p_2,\cdots,p_n)=H(\boldsymbol{P}) \tag{2-7}$$

以后常用$H(X)$来表示以离散随机变量X描述的信源的信息熵。而用$H(\boldsymbol{P})$来表示概率矢量为$\boldsymbol{P}$的n个消息信源的信息熵。例如，若当$n=2$时，2个消息的熵函数可写成$H(\boldsymbol{P})$。

熵函数$H(\boldsymbol{P})$是一种特殊函数，其函数表达式为