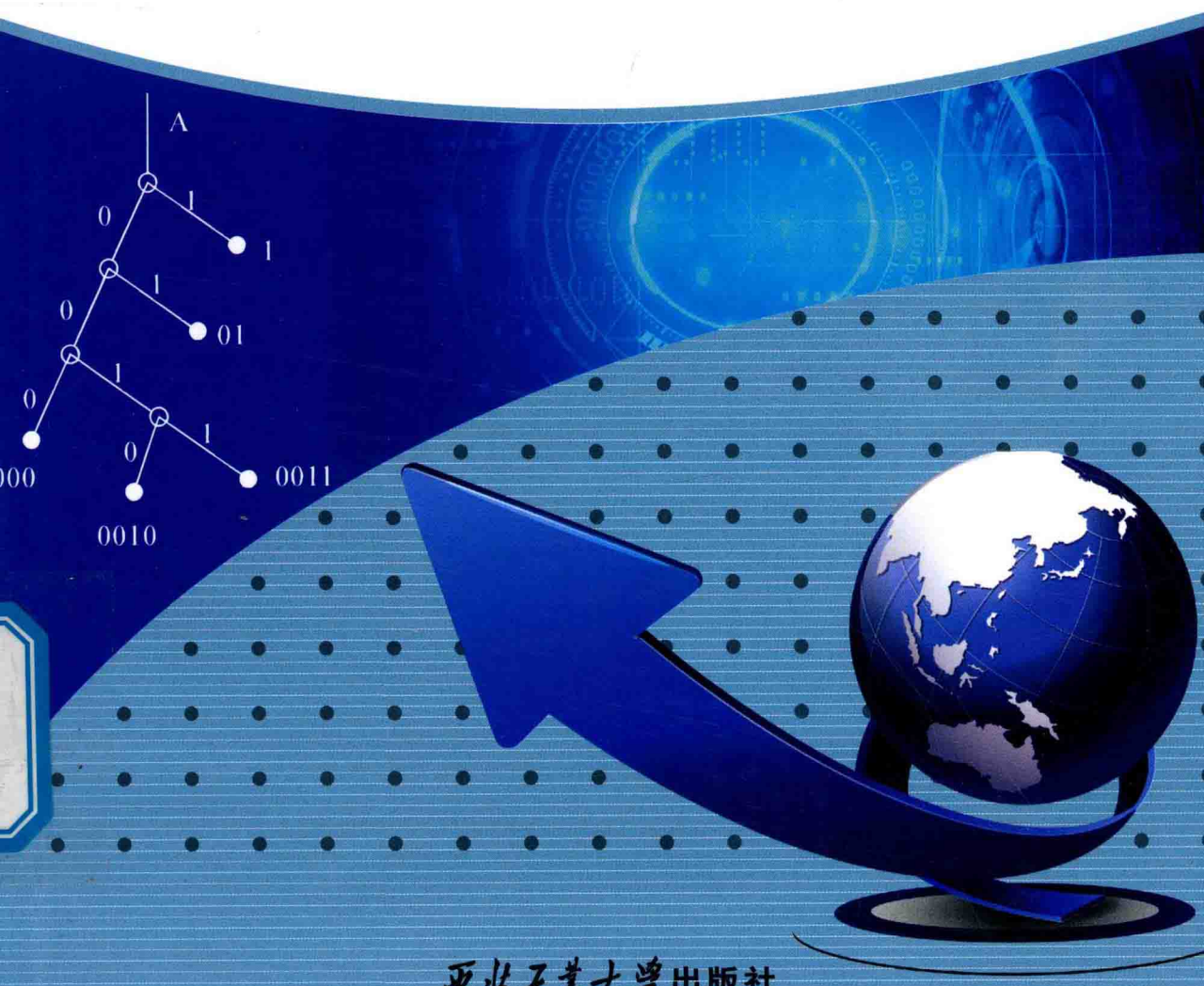
高等学校“十三五”规划教材

信息论与编码理论

李　敏　邢宇航　王利涛◎编著



西北工业大学出版社

高等学校“十三五”规划教材

XINXILUN YU BIANMA LILUN

信息论与编码理论

李敏　邢宇航　王利涛　编著

西北工业大学出版社
西　安

【内容简介】 本书系统阐述了经典信息论和编码的基本原理及应用，主要内容分为基本概念和物理量的计算、信源编码、信道编码、安全编码和实际应用五篇共9章，在突出实际应用的同时，着重介绍信息与编码理论的基本概念和基本方法，体现信息传输追求的有效性、可靠性和安全性。本书内容丰富，文字通俗，深入浅出，力求理论联系实际。

本书可作为高等学校电子信息类专业的研究生及本科生教材，也可供信息科学及系统工程专业领域教学、科研人员参考。

图书在版编目(CIP)数据

信息论与编码理论/李敏，邢宇航，王利涛编著．—西安：西北工业大学出版社，2018.5

ISBN 978-7-5612-6007-4

Ⅰ.①信… Ⅱ.①李… ②邢… ③王… Ⅲ.①信息论—高等学校—教材 ②信源编码—高等学校—教材 Ⅳ.①TN911.2

中国版本图书馆CIP数据核字(2018)第100531号

策划编辑：杨 军
责任编辑：张 友

出版发行：西北工业大学出版社
通信地址：西安市友谊西路127号 邮编：710072
电 话：(029)88493844 88491757
网 址：www.nwpup.com
印 刷 者：陕西金德佳印务有限公司
开 本：787 mm×1 092 mm 1/16
印 张：11.5
字 数：276千字
版 次：2018年5月第1版 2018年5月第1次印刷
定 价：39.00元

前　言

2000年10月6日，著名信息论与编码学者Dr. Richard Blahut在C. E. Shannon(香农)塑像落成典礼上致辞："……两三百年之后，当人们回过头来看我们这个时代的时候，他们可能不会记得谁曾是美国的总统，他们也不会记得谁曾是影星或摇滚歌星，但是仍然会知晓Shannon的名字，学校里仍然会讲授信息论……"

1948年，香农发表了划时代文章《通信的数学理论》，宣告了一门崭新学科——信息论的诞生。信息论为计算机和远程通信奠定了坚实的理论基础，是20世纪产生的对人类最伟大的贡献之一，它解决了通信的有效性、可靠性和安全性问题。自信息论诞生以来，信息理论和编码技术已广泛应用于我们的生活中，信息论的研究领域从自然科学扩展到经济、管理科学，甚至人文社会科学，从狭义信息论发展到如今的广义信息论，成为涉及面极广的信息科学。

信息论也称为香农信息论，主要研究信息理论与编码技术，是用概率论与随机过程的方法研究通信系统传输有效性和可靠性的理论，是现代通信与信息处理技术的理论基础，也是通信与电子信息类专业的重要基础课程。信息论以信息熵为基本概念，以香农三个编码定理——无失真信源编码理论、有噪信道编码理论和限失真信源编码定理为核心内容，研究通信系统中信息的度量、信源的压缩编码，以及信息通过信道有效、可靠和安全传输的问题。

本书分为基本概念和物理量的计算、信源编码、信道编码、安全编码和信息编码理论应用等五篇，共9章。第1章绪论，主要介绍信息的定义，信息论的发展历史、现状和趋势，信息论学科的研究对象和应用，信息系统传输模型及功能；第2章数学及编码基础知识，主要围绕信息论的数学基础，阐述信息论中涉及的概率论和编码理论的基础知识和重要结论，以方便后续对信息理论知识的学习和掌握；第3章信息度量与信息熵，介绍信源的分类和数学模型，信息量和信息熵的概念、性质、定理及计算方法等；第4章信息率失真理论与信息率失真函数，主要介绍允许压缩信源输出的信息率，信息率与允许失真之间的关系，信息率失真函数的概念、性质及计算等；第5章信源编码，主要介绍信源编码的基本概念、常见的三类信源编码技术(统计编码、变换编码和预测编码)、无失真信源编码定理及限失真信源编码定理；第6章信道与信道容量，主要讨论信道的分类和数学模型，信道容量的定义、计算方法和有噪信道编码定理等；第7章信道编码，介绍抗干扰信道编码的基本原理，常见的信道编码方法等；第8章安全编码，介绍实现信息传输

安全性要求的技术，密码技术的产生和发展，加密的基本原理，现代密码体制的分类，信息熵测度密码学的基本概念等；第9章信息编码理论实际应用，主要结合信息论和编码技术在多媒体数据压缩、计算机网络通信和移动通信等领域中的实际应用，开阔视野、拓展内容。

本书是在笔者多年从事教学科研实践的基础上编写而成的，系统地介绍了信息技术领域的基础理论，综合利用概率论、信息度量、信源编码、信道编码和安全编码等技术解决信息传输处理问题。具体编写分工如下：李敏负责内容与结构安排组织，并编写了第四、六、七章；邢宇航编写了第一至三章；王利涛编写了第五、八、九章。

在编写本书的过程中，曾参阅了相关文献，在此谨向其作者深表谢意。

由于水平所限，书中疏漏之处在所难免，欢迎广大读者批评指正。

作　者

2018年2月

目　录

基本概念及物理量的计算篇

第 1 章　绪论 …… 1

1.1　信息与信息论 …… 1
1.2　信息系统传输模型 …… 7
1.3　信息论的研究内容 …… 9
习题 1 …… 9

第 2 章　数学及编码基础知识 …… 10

2.1　概率论基础知识 …… 10
2.2　编码基础知识 …… 12
习题 2 …… 15

第 3 章　信息度量与信息熵 …… 16

3.1　信源的分类 …… 16
3.2　离散单符号信源信息度量 …… 17
3.3　离散多符号序列信源信息度量 …… 38
3.4　连续信源信息度量 …… 39
习题 3 …… 41

信源编码篇

第 4 章　信息率失真理论与信息率失真函数 …… 45

4.1　引言 …… 45
4.2　$R(D)$函数的性质 …… 48
4.3　离散信源 $R(D)$函数 …… 49
4.4　连续信源 $R(D)$函数 …… 56
习题 4 …… 59

第 5 章　信源编码 …… 61

5.1　信源编码概念及常见格式 …… 61
5.2　定长码及其编码定理 …… 64

5.3 变长码及其编码定理 …… 68
5.4 限失真编码定理 …… 73
5.5 信源编码技术 …… 75
习题 5 …… 87

信道编码篇

第 6 章 信道与信道容量 …… 89

6.1 信道的分类与数学模型 …… 89
6.2 离散单符号信道及其容量 …… 93
6.3 离散多符号信道及其容量 …… 97
6.4 连续信道及其容量 …… 98
6.5 多用户信道的信道容量 …… 102
习题 6 …… 109

第 7 章 信道编码 …… 111

7.1 信道编码概述 …… 111
7.2 线性码 …… 127
7.3 循环码 …… 135
习题 7 …… 144

安全编码篇

第 8 章 安全编码 …… 148

8.1 密码技术的起源与发展 …… 148
8.2 密码编码基础 …… 149
8.3 现代密码体制 …… 153
8.4 密码体制的安全性测度 …… 161
习题 8 …… 164

信息编码理论应用篇

第 9 章 信息编码理论实际应用 …… 165

9.1 信息编码理论在多媒体技术中的应用 …… 165
9.2 信息编码理论在计算机网络通信中的应用 …… 169
9.3 信息编码理论在数字移动通信中的应用 …… 172
习题 9 …… 175

参考文献 …… 176

基本概念及物理量的计算篇

第1章 绪 论

虽然信息论自诞生到现在只有60多年，但它的发展对学术界及人类社会的影响是广泛和深刻的。在人类历史的长河中，信息传输和传播手段经历了六次重大变革，在不断的变化中，人们逐渐认识到信息的存在及重要作用。

第一次变革是语言的产生。人们用语言准确地传递感情和意图，使语言成为传递信息的重要工具。第二次变革是文字的产生，不久又发明了纸张，人类开始用书信的方式交换信息，使信息传递的准确性大大提高。第三次变革是印刷术的发展。它使信息能大量存储和大量流通，并显著扩大了信息的传递范围。第四次变革是电报、电话的发明。开始了人类电信时代。通信理论与技术迅速发展。第五次变革是计算机技术与通信技术相结合，促进了网络通信的发展。信息理论的研究得到进一步的发展，多用户理论的研究取得了突破性的进展。至此，香农的单用户信息论已推广到多用户信息论。第六次变革是大数据技术的出现。促进了信息收集、处理、存储和管理等全寿命周期的信息技术发展。

信息论是由通信技术与概率论、随机过程和数理统计相结合而逐步发展起来的一门科学。香农(C. E. Shannon)在1948年发表了著名的论文《通信的数学理论》，为信息论奠定了理论基础。随着信息理论的迅猛发展和信息概念不断深化，信息论所涉及的内容早已超越了狭义的通信工程范畴，进入了信息科学这一更广、更新的领域。在人类文明的早期，人们就已经知道利用信息与信息传递等手段来实现某些目的，如古代的烽火台，就是用烽烟来传递外敌入侵的信息。但是，大量信息的运用还是在有线、无线电通信产生以后。

本章首先引出信息的概念，进而讨论信息论学科的研究对象、目的和内容，并简述其发展历史、现状和趋势。介绍信息传输编码理论的相关内容，其中包括编码理论的产生、发展，理论的形成过程、重要的应用领域等问题。同时介绍信息系统传输模型及功能。

1.1 信息与信息论

1.1.1 信息的定义

什么是信息呢？信息是信息论中最基本、最重要的概念，它是一个既抽象又复杂的概念。这一概念是在人类社会互通情报的实践过程中产生的。在现代信息理论形成之前的漫长时期

中，信息被看作是通信的消息的同义词，没有赋予它严格的科学定义。

最早对信息进行科学定义的，是哈特莱(R. V. Hartley)。他认为，发信者所发出的信息，就是他在通信符号表中选择符号的具体方式。哈特莱的这种理解存在着严重的局限性。首先，他所定义的信息不涉及信息的价值和具体内容，只考虑选择的方法。其次，但没有考虑各种可能选择方法的统计特性。

1948年，控制论的创始人之一，美国科学家维纳(N. Wiener)指出："信息是信息，不是物质，也不是能量"。这就是说，信息就是信息自己，它不是其他什么东西的替代物，它是与"物质""能量"同等重要的基本概念。后来，维纳提出："信息是人们适应外部世界并且使这种适应反作用于外部世界的过程中，同外部世界进行互相交换内容的名称。"又说："接收信息和使用信息的过程，就是我们适应外部世界环境的偶然性变化的过程，也是我们在这个环境中有效地生活的过程。要有效地生活，就必须有足够的信息。"的确，信息对人类的生存是很重要的；但是，信息不仅仅与人类有关，不仅仅是人与外部世界交换的内容。人们在与外部世界相互作用过程中，还进行着物质与能量的交换。这样，就又把信息与物质、能量混同起来。所以，维纳关于信息的定义是不确切的。

关于信息的定义，有人认为"信息就是差异"。这种说法的典型代表意大利学者朗梅(G. Longe)提出："信息是反映事物的形式、关系和差别的东西。信息是包含于客体间的差别中，而不是在客体本身中。在通信中差别关系是重要的。"也就是说，他定义信息是客体之间的相互差异。的确，宇宙内到处存在着差异，差异的存在使人们存在着"疑问"和"不确定性"。从这个角度看，差异的确是信息。但是，并不能说没有差异就没有信息。所以，这样定义的信息也是不全面、不确切的。

而香农在1948年发表的著名论文《通信的数学理论》中，从研究通信系统传输的实质出发，对信息作了科学的定义，并进行了定性和定量的描述。

在各种通信系统中，其传输的形式是消息。但消息传递过程的一个最基本、最普通却又不十分引人注意的特点是，收信者在收到消息以前是不知道消息的具体内容的。在收到消息以前，收信者无法判断发送者将会发来描述何种事物运动状态的具体消息，更无法判断是描述这种状态还是那种状态。再者，即使收到消息，由于干扰的存在，他不能断定所得到的消息是否正确和可靠。总之，收信者存在着"不知""不确定"或"疑问"。通过消息的传递，收信者知道了消息的具体内容，原先的"不知""不确定"和"疑问"消除或部分消除了。因此，对收信者来说，消息的传递过程是一个从不知到知的过程，或是从知之甚少到知之甚多的过程，或是从不确定到部分确定或全部确定的过程。

例如，在电报通信中，收报人在收到报文"弟高中"后，才能确定是他家人告诉他弟弟的高考情况。其次，报文"弟高中"是弟弟考学结果的一种描述。收信者在看到报文以前，他不能确定弟弟考学结果如何，也存在"不确定性"。只要报文是清楚的，在传递过程中没有差错，那么，他收到报文以后，他原来所有的"不确定性"都没有了，他就获得了所有的信息。如果在传递过程中存在着干扰，使报文完全模糊不清，收信者收到报文以后，报文所具有的不确定性一点也没有减少，他就没有获得任何信息。如果干扰使报文发生部分差错，使收信者对报文的不确定性减少了一些，但没有全部消除，他就获得了一部分信息。所以，通信过程是一种消除不确定性的过程。在不确定性消除后，就获得了信息。原先的不确定性消除得越多，获得的信息就越多。如果原先的不确定性全部消除了，就获得了全部的信息；若消除了部分不确定性，就获得

了部分信息;若原先不确定性没有任何消除,就没有获得任何信息。由此可见,信息是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。

1.1.2 信息论发展简介

信息论是从实践中经过抽象、概括、提高而逐步形成的,是在长期的通信工程实践和理论研究的基础上发展起来的。通信系统是人类社会的神经系统,即使在原始社会也存在着最简单的通信工具和通信系统。日常生活、工农业生产、科学研究以及战争等等,一切都离不开信息传递和流动。

例如,1832 年莫尔斯电报系统中高效率编码方法对后来香农的编码理论是有启发的。1885 年凯尔文(Kelvin)曾经研究过一条电缆的极限传输信息率问题。1922 年卡逊(J. R. Carson)对调幅信号的频谱结构进行了研究,并明确了边带的概念。1924 年奈奎斯特(H. Nyquist)和屈夫缪勒(K. Kupfmuller)分别独立地指出,如果以一个确定的速度来传输电报信号,就需要一定的带宽。这证明了信号传输速率与信道带宽成正比。1928 年哈特莱(R. V. Hartley)发展了奈奎斯特的工作,并提出把消息考虑为代码或单语的序列。他提出"定义信息量 $H=N\log s$",即定义信息量等于可能消息数的对数。其缺点是没有统计特性的概念。他的工作对后来香农的思想是有很大影响的。1936 年阿姆斯特朗(E. H. Armstrong)提出增加信号带宽可以使抑制噪声干扰的能力增强,并给出了宽频移的调频方式,使调频实用化,出现了调频通信装置。1939 年达德利(H. Dudley)发明了声码器。当时他提出的概念:通信所需要的带宽至少应与所传送的消息的带宽相同。达德利和莫尔斯都是研究信源编码的先驱者。

20 世纪 40 年代初期,维纳发表了《平稳时间序列的外推、内插与平滑及其工程应用》的论文。他把随机过程和数理统计的观点引入通信和控制系统中来,揭示了信息传输和处理过程的统计本质。他还利用早在 30 年代初他本人提出的"广义谐波分析理论"对信息系统中的随机过程进行谱分析。这就使得通信系统的理论研究面貌焕然一新,有了质的飞跃。

1948 年香农在贝尔系统技术杂志上发表了两篇有关"通信的数学理论"的论文。在这两篇论文中,他用概率测度和数理统计的方法系统地讨论了通信的基本问题,得出了几个重要而带有普遍意义的结论,并由此奠定了现代信息论的基础。香农理论的核心:揭示了在通信系统中采用适当的编码后能够实现高效率和高可靠地传输信息,并得出了信源编码定理和信道编码定理。从数学观点看,这些定理是最优编码的存在定理。但从工程观点看,这些定理不是结构性的,不能从定理的结果直接得出实现最优编码的具体途径。然而,它们给出了编码的性能极限,在理论上阐明了通信系统中各种因素的相互关系,为人们寻找最佳通信系统提供了重要的理论依据。

从 1948 年开始,信息论的出现引起了数学家的兴趣,他们将香农已得到的数学结论作了进一步的严格论证和推广,使这一理论具有更为坚实的数学基础。例如,1952 年费诺(R. M. Fano)给出并证明了费诺不等式,还给出了关于香农信道编码逆定理证明。1957 年沃尔夫维兹(J. Wolfowitz)采用了类似典型序列方法证明了信道编码强逆定理。1961 年费诺又描述了分组码中码率、码长和错误概率的关系,并提供了香农信道编码定理的充要性证明。1965 年格拉格尔(R. G. Gallager)发展了费诺的证明结论并提供了一种简明的证明方法。而科弗尔

(T. M. Cover)于1975年采用典型序列方法来证明。1972年阿莫托(S. Arimoto)和布莱哈特(R. Blahut)分别发展了信道容量的迭代算法。

香农在1948年的论文中首先分析和研究了高斯信道。1964年霍尔辛格(J. L. Holsinger)开展了有色高斯噪声信道容量的研究。1969年平斯克尔(M. S. Pinsker)提出了具有反馈的非白噪声高斯信道容量问题。科弗尔(T. M. Cover)于1989年对平斯克尔的结论作出了简洁的证明。

香农在1948年的论文中提出了无失真信源编码定理,也给出了简单的编码方法,即香农编码。麦克米伦(B. McMillan)于1956年首先证明了唯一可译变长码的克拉夫特(Kraft)不等式。关于无失真信源的编码方法,1952年费诺(Fano)提出了费诺编码方法。同年,霍夫曼(D. A. Huffman)首先构造了一种霍夫曼编码方法,并证明了它是最佳码。20世纪70年代后期开始,人们对与实际应用有关的信源编码问题产生了兴趣。于1968年前后,埃利斯(P. Elias)发展了香农—费诺码,提出了算术编码的初步思路。而里斯桑内(J. Rissanen)在1976年给出和发展了算术编码。1982年他和兰登(G. G. langdon)一起将算术编码系统化,并省去了乘法运算,更为简化,易于实现。通用信源编码算法——字典编码LZ码是于1977年由齐弗(J. Ziv)和兰佩尔(A. lempel)提出的。1978年他们俩又提出了改进算法,而且齐弗证明此方法可达到信源的熵值。1990年贝尔(T. C. Bell)等在LZ算法基础上又作了一系列变化和改进。

在研究香农信源编码定理的同时,另外一部分科学家从事寻找最佳编码(纠错码)的研究工作。早在1950年,汉明码出现后,人们把代数方法引入到纠错码的研究,形成了代数编码理论。由此找到了大量性能好的纠错码,并提出了可实现的编译码方法。但代数编码的渐近性能较差,不能实现香农信道编码定理所指出的结果。因此,于1960年左右提出了卷积码的概率译码,并逐步形成了一系列概率译码理论。尤其,以维特比(Viterbi)译码为代表的译码方法被美国卫星通信系统所采用,使香农理论成为真正具有实用意义的科学理论。限失真信源编码的研究与信道编码和无失真信源编码相比,落后约10年左右。香农在1948年的论文中已体现出了关于率失真函数的思想。一直到1959年他发表了《保真度准则下的离散信源编码定理》,首次提出了率失真函数及率失真信源编码定理。从此,发展成为信息率失真编码理论。1971年伯格尔著作的《信息率失真理论》一书是一本较全面地论述有关率失真理论的专著。率失真信源编码理论是信源编码的核心问题,是频带压缩、数据压缩的理论基础。一直到今天,它仍是信息论的研究课题。有关数据压缩、多媒体数据压缩又是另一独立的分支——数据压缩理论与技术。

香农1961年的论文《双路通信信道》开拓了网络信息论的研究。1970年以来,随着卫星通信、计算机通信网的迅速发展,网络信息理论的研究异常活跃,成为当前信息论的中心研究课题之一。1971年艾斯惠特(R. Ahlswede)和1972年廖(H. Liao)找出了多元接入信道的信道容量区。接着,1973年沃尔夫(K. Wolf)和斯莱平(D. Slepian)将它推广到具有公共信息的多元。

1.1.3 信息论的应用

信息是一个普遍的概念,信息论及编码技术的产生、应用与通信、计算机技术的产生、发展

密切相关。回顾信息论的历史，大体可以分为早期酝酿、理论建立、理论发展、理论应用与近代发展等几个阶段。

1. 早期编码问题

在有线、无线电通信产生的同时，编码技术随之产生，早期的编码有莫尔斯(Morse)码和波多(Bodo)码等，它们把文字通过点、划、空等信号来表达，这些码虽很原始，但它们实现了从文字到通信信号的转变。因此莫尔斯码和波多码是最早的编码方式。中文通信一直采用电报码方式，先将汉字变成数字，再用电码发出。

在20世纪七八十年代，由香农提出的信息理论成为这一时期信息论研究的一个主流课题，各种不同类型的多用户信源、信道模型被提出，许多相关的编码定理被证明。这些模型与当时的微波与卫星通信模型密切相关，当时的微波转播、通信卫星与广播卫星模型正与这些模型符合。

2. 信息论在其他学科的应用

信息论近期发展的主要特点是向多学科交叉方向发展，其重要的发展方向有以下几种：

(1)信息论与密码学。通信中的安全与保密问题是通信编码问题的又一种表示形式，由香农提出的保密系统模型仍然是近代密码学的基本模型。其中的许多度量性指标，如加密运算中的安全性、剩余度等指标与信息量密切相关。

(2)算法信息论与分开理论。由于香农熵、柯莫格洛夫复杂度与豪斯道夫(Hausdorff)维数的等价性在理论上已得到证明，从而使信息论、计算机科学与分开理论找到了它们的汇合点。人们发现香农熵、柯莫格洛夫复杂度与豪斯道夫维数都是某种事物复杂性的度量，它们在一定的条件下可以相互等价转化。由这三种度量分别产生了信息论、计算机程序复杂度与分开理论，在本质上有共同之外，它们结合后所产生的新兴学科方向具有跨学科的特点，如算法信息论就是信息论与计算复杂性理论的新学科。

3. 信息论在统计与智能计算中的应用

信息论与统计理论的结合已有许多突出的成果出现。其主要特点是统计理论正在从线性问题转向非线性问题，信息的度量可以作为研究非线性问题的工具，如用互信息来取代统计中的相关系数，更能反映随机变量的相互依赖程度。信息量的统计计算较为复杂，因此在统计中一直没有得到大量的应用，但由于近期大批海量数据(如金融、股票数据、生物数据等)的出现，使许多计算问题成为可能，因此信息论在统计中必将发挥更大的作用。信息论与统计理论结合的典型应用如下：

(1)智能计算中的信息统计问题。信息量与统计量存在许多本质的联系，在概率分布族所组成的微分流形中，Fisher信息距阵是Kullback-laiber熵的偏微分，由此关系而引出的信息几何理论是智能计算的基础，一些重要的智能计算方法，如EM算法、ACI算法、Ying-Yang算法都与此有关。

(2)信息计算与组合抽奖决策关系密切，T. Cover教授把组合抽奖决策问题提取成一个信息论的问题，在最优决策的计算中给出了一个渐近递推算法，并利用互熵关系证明了该算法的单调性与收敛性。

(3)编码理论在与试验设计、假设检验理论的结合中发挥了重要作用。在信息编码理论中有许多码的构造理论与方法，这些码在一定意义下具有正交性，因此这些码可直接设计和构造

试验设计表。另外，利用信息编码定理可以证明在假设检验中两类误差的指数下降性，并给出这两类误差的下降速度。

人类从产生那天起，就生活在信息的海洋之中。人类社会的生存和发展，每时每刻都离不开接收信息、传递信息、处理信息和利用信息。自古以来，人们就对信息的表达、存储、传送和处理等问题进行了许多研究。近百年来，随着生产和科学技术的发展，信息的处理、传输、存储、提取和利用的方式及手段达到了更新更高的水平。近代电子计算机的迅速发展和广泛应用，尤其是个人微型计算机的普及，大大提高了人们处理信息、存储信息及控制和管理信息的能力。20 世纪后半叶，计算机技术、微电子技术、传感技术、激光技术、卫星通信和移动通信技术、航空航天技术、广播电视技术、多媒体技术、新能源技术和新材料技术等新技术的发展和应用，尤其近年来以计算机为主体的互联网技术的兴起和发展，它们相互结合、相互促进，以空前的威力推动着人类经济和社会高速发展。正是这些现代新科学新技术汇成了一股强大的时代潮流，将人类社会推入到高度化的信息时代。在当今社会中，人们在各种生产、科学研究和社会活动中，无处不涉及信息的交换和利用。迅速获取信息，正确处理信息，充分利用信息，就能促进科学技术和国民经济的飞跃发展。可见，信息的重要性是不言而喻的。

1.1.4 信息与情报等概念的区别和联系

在日常生活中，信息常常被认为就是“情报”“知识”“消息”“信号”等。的确，信息与它们之间是有着密切联系的。但是，信息的含义更深刻、更广泛，它是不能等同于情报、知识、消息和信号的。

信息不能等同于情报。情报往往是军事学、文献学方面的习惯用词。如“对敌方情况的报告”，“文献资料中对于最新情况的报道或者进行资料整理的成果”等称为情报。在情报学中，它们对于“情报”是这样定义的：“情报是人们对于某个特定对象所见、所闻、所理解而产生的知识”。可见，情报的含义要比“信息”窄很多。情报只是一类特定的信息，不是信息的全体。

信息不能等同于知识。知识是人们根据某种目的，从自然界收集得来的数据中，整理、概括、提取得到有价值的、人们所需的信息。知识是一种具有普遍和概括性质的高层次的信息。例如，获得大量的遥感图片数据，根据不同目的，处理后可以得到不同的知识（地质知识、地形知识、水源知识等等）。由此可知，知识是以实践为基础，通过抽象思维，对客观事物规律性的概括。知识信息只是人类社会中客观存在的部分信息。所以知识是信息，但不等于信息的全体。

信息不能等同于消息。人们也常常错误地把信息等同于消息，认为得到了消息，就是得到了信息。例如，当人们得到一封电报，接到一个电话，收听了广播或看了电视等以后，就说得到了“信息”。的确，人们从接收到的电报、电话、广播和电视的消息中能获得各种信息，信息与消息有着密切的联系。但是，信息与消息并不是一件事，不能等同。

在电报、电话、广播、电视（也包括雷达、导航、遥测）等通信系统中传输的是各种各样的消息。这些被传送的消息有着各种不同的形式，例如：文字、符号、数据、语言、音符、图片、活动图像等等。所有这些不同形式的消息都是能被人们感觉器官所感知的，人们通过通信，接收到消息后，得到的是关于描述某事物状态的具体内容。语言、报文、图像等消息都是对客观物质世界的各种不同运动状态或存在状态的表述。当然，消息也可用来表述人们头脑里的思维活动。

因此，用文字、符号、数据、语言、音符、图片、图像等能够被人们感觉器官所感知的形式，把客观物质运动和主观思维活动的状态表达出来就称为消息。

可见，消息中包含信息，是信息的载体。得到消息，从而获得信息，同一则信息可用不同的消息形式来载荷。而一则消息也可载荷不同的信息，它可能包含非常丰富的信息，也可能只包含很少的信息。因此，信息与消息是既有区别又有联系的。

信息不同于消息，也不同于信号。在各种实际通信系统中，往往为了克服时间或空间的限制而进行通信，必须对消息进行加工处理。把消息变换成适合信道传输的物理量，这种物理量称为信号(如电信号、光信号、声信号、生物信号等)。信号携带着消息，它是消息的运载工具。信号携带信息，但不是信息本身。同样，同一信息可用不同的信号来表示。同一信号也可表示不同的信息。所以，信息、消息和信号是既有区别又有联系的三个不同的概念。

1.2 信息系统传输模型

从信息概念的讨论中，可以看到：各种通信系统如电报、电话、电视、广播、遥测、遥控、雷达和导航等，虽然它们的形式和用途各不相同，但本质是相同的，都是信息的传输系统。为了便于研究信息传输和处理的共同规律，我们将各种通信系统中具有共同特性的部分抽取出来，概括成一个统一的理论模型，如图1-1所示。通常称它为通信系统模型。

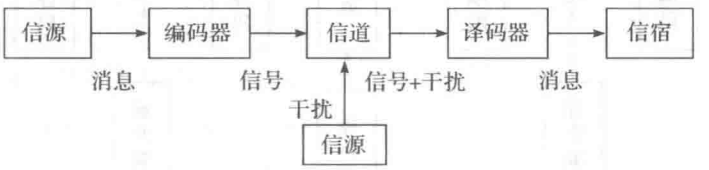


图1-1 通信系统模型

这个通信系统模型也适用于其他的信息流通系统，如生物有机体的遗传系统、神经系统、视觉系统等，甚至人类社会的管理系统都可概括成这个模型。

信息论研究的对象正是这种统一的通信系统模型。人们通过系统中消息的传输和处理来研究信息传输和处理的共同规律。

这个模型主要分成下述五部分：

(1)信息源(简称信源)。顾名思义，信源是产生消息和消息序列的源。它可以是人、生物、器或其他事物。它是事物各种运动状态或存在状态的集合。信源的输出是消息，消息是具体的，但它不是信息本身。消息携带着信息，消息是信息的表达者。另外，信源可能出现的状态(即信源输出的消息)是随机的、不确定的，但又有一定的规律性。

(2)编码器。编码是把消息变换成信号的措施，而译码就是编码的反变换。编码器输出的是适合信道传输的信号，信号携带着消息，它是消息的载荷者。编码器可分为两种，即信源编码器和信道编码器。信源编码是对信源输出的消息进行适当的变换和处理，目的是为了提高信息传输的效率。而信道编码是为了提高信息传输的可靠性而对消息进行的变换和处理。当然，对于各种实际的通信系统，编码器还应包括换能、调制、发射等各种变换处理。

(3)信道。信道是指通信系统把载荷消息的信号从甲地传输到乙地的媒介。在狭义的通

信系统中实际信道有明线、电缆、波导、光纤、无线电波传播空间等，这些都是属于传输电磁波能量的信道。当然，对广义的通信系统来说，信道还可以是其他的传输媒介。

(4)译码器。译码就是把信道输出的编码信号(已叠加了干扰)进行反变换。一般认为这种变换是可逆的。译码器也可分成信源译码器和信道译码器。

(5)信宿。信宿是消息传送的对象，即接收消息的人或机器。

图 1-1 所示的模型只适用于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。更一般的情况是，信源和信宿各有若干个，即信道有多个输入和多个输出，另外信息传输也可以双向进行。例如广播通信是一个输入、多个输出的单向传输的通信，而卫星通信网则是多个输入、多个输出和多向传输的通信。因此，图 1-1 所示的通信系统模型是最基本的。

近年来，以计算机为核心的大规模信息网络，尤其是互联网的建立和发展，对信息传输的质量要求更高了。不但要求快速、有效、可靠地传递信息，而且要求信息传递过程中保证信息的安全保密，不被伪造和窜改。因此，在编码器这一环节中还需加入加密编码。相应地，在译码器中加入解密译码。

为此，我们把图 1-1 所示的通信系统模型中编(译)码器分成信源编(译)码、信道编(译)码和加密(解密)编(译)码三个子部分。这样，信息传输系统的基本模型如图 1-2 所示。

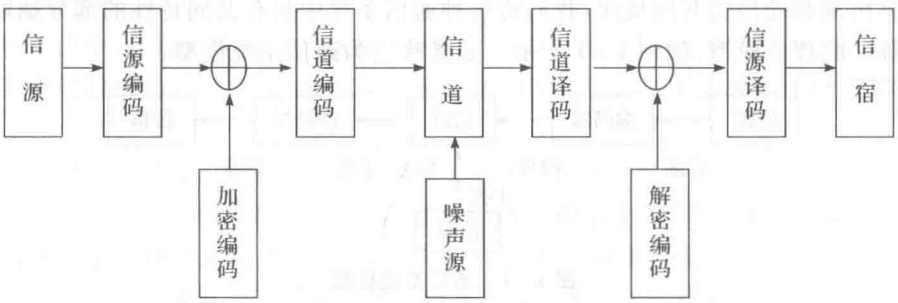


图 1-2　信息传输系统模型

研究这样一个概括性很强的通信系统，其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、有效性、保密性和认证性，使信息传输系统最优化。

所谓可靠性高，就是要使信源发出的消息经过信道传输以后，尽可能准确地、不失真地再现于接收端。而所谓有效性高，就是经济效果好，即用尽可能短的时间和尽可能少的设备来传送一定数量的信息。

所谓保密性，就是隐蔽和保护通信系统中传送的消息，使它只能被授权接收者获取，而不能被未授权者接收和理解。所谓认证性是指接收者能正确判断所接收的消息的正确性，验证消息的完整性，而不是伪造的和被窜改的。有效性、可靠性、保密性和认证性四者才构成现代通信系统对信息传输的全面要求。

信息传输系统模型不是不变的，它根据信息传输的要求而定。当研究信息传输有效性时，可只考虑信源与信宿之间的信源编(译)码，将其他部分都看成一无干扰信道。当研究信息传输可靠性时，可将信源、信源编码和加密编码都等效成一个信源，而将信宿、信源解码和解密译码都等效成一信宿。当考虑信息传输的保密性和认证性时，可将信源和信源编码等效成一信源，将信道编码、信道、噪声源和信道译码等效成一无干扰信道，而将信源译码和信宿等效于信宿。

1.3 信息论的研究内容

信息论的研究对象是广义通信系统。任何系统，只要能够抽象成通信系统模型，都可以用信息论研究。一般有信息论基础、一般信息论和广义信息论之分。信息论与编码是一门应用概率论、随机过程、数理统计和近世代数的方法，来研究广义的信息传输、提取和处理系统中一般规律的学科。它的主要目的是提高信息系统的可靠性、有效性、保密性和认证性，以便达到系统最优化。它的主要内容(或分支)包括香农理论、编码理论、维纳理论、检测和估计理论、信号设计和处理理论、调制理论、随机噪声理论和密码学理论等。

由于信息论与编码研究的内容极为广泛，而各分支又有一定的相对独立性，因此本书仅论述信息论的基础理论即香农信息理论。

习 题 1

1.1 请给出最简单通信系统的物理模型并说明各基本单元的主要功能。

1.2 通信系统要解决的根本问题是什么？

1.3 消息的定义是什么？有什么特征？

1.4 信息的定义是什么？有什么特征？

1.5 信息传输系统的基本模型是什么？

1.6 信息论与编码研究的主要内容是什么？

第 2 章　数学及编码基础知识

研究信息论，实际涉及很多方面的知识。通过香农的信息理论可以知道，建立信息论的理论基础，必须以数学作为支撑，尤其是概率论的知识。本章主要围绕信息论的数学基础，介绍信息论中涉及的概率论和编码理论的基础知识和重要结论，以方便后续信息理论知识的学习和掌握。

2.1　概率论基础知识

2.1.1　基本概念

基本事件：随机试验的每一个可能的结果(样本点)。

样本空间：基本事件的集合。

复杂事件：多个基本事件所组成的事件。

随机事件：无论基本事件还是复杂事件，它们在试验中发生与否，都带有随机性。

事件域：基本事件和复杂事件是样本空间的子集，事件域是所有子集的全体。

概率空间三要素：样本空间、事件域(集合)、概率。

事件 A 的概率：A 中样本点数与样本空间中样本点之比。

先验概率：根据以往的统计规律得到的概率。

1. 随机试验

随机试验是一个概率论的基本概念。在概率论中把符合下述三个特点的试验叫做随机试验：

(1)每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；

(2)进行一次试验前无法确定哪一个结果会出现；

(3)可以在同一条件下重复进行试验。

例 2.1　掷骰子。基本事件：骰子朝上面的点数。

(1)以下几种情况中，求样本空间的大小。

掷一个骰子：样本空间大小为 6；

掷两个骰子：样本空间大小为 11。

(2)以下几种情况中，求骰子朝上面的点数>5 的概率。

掷一个骰子：骰子朝上面的点数>5 的概率为 1/6；

掷两个骰子：骰子朝上面的点数>5 的概率为 26/36。