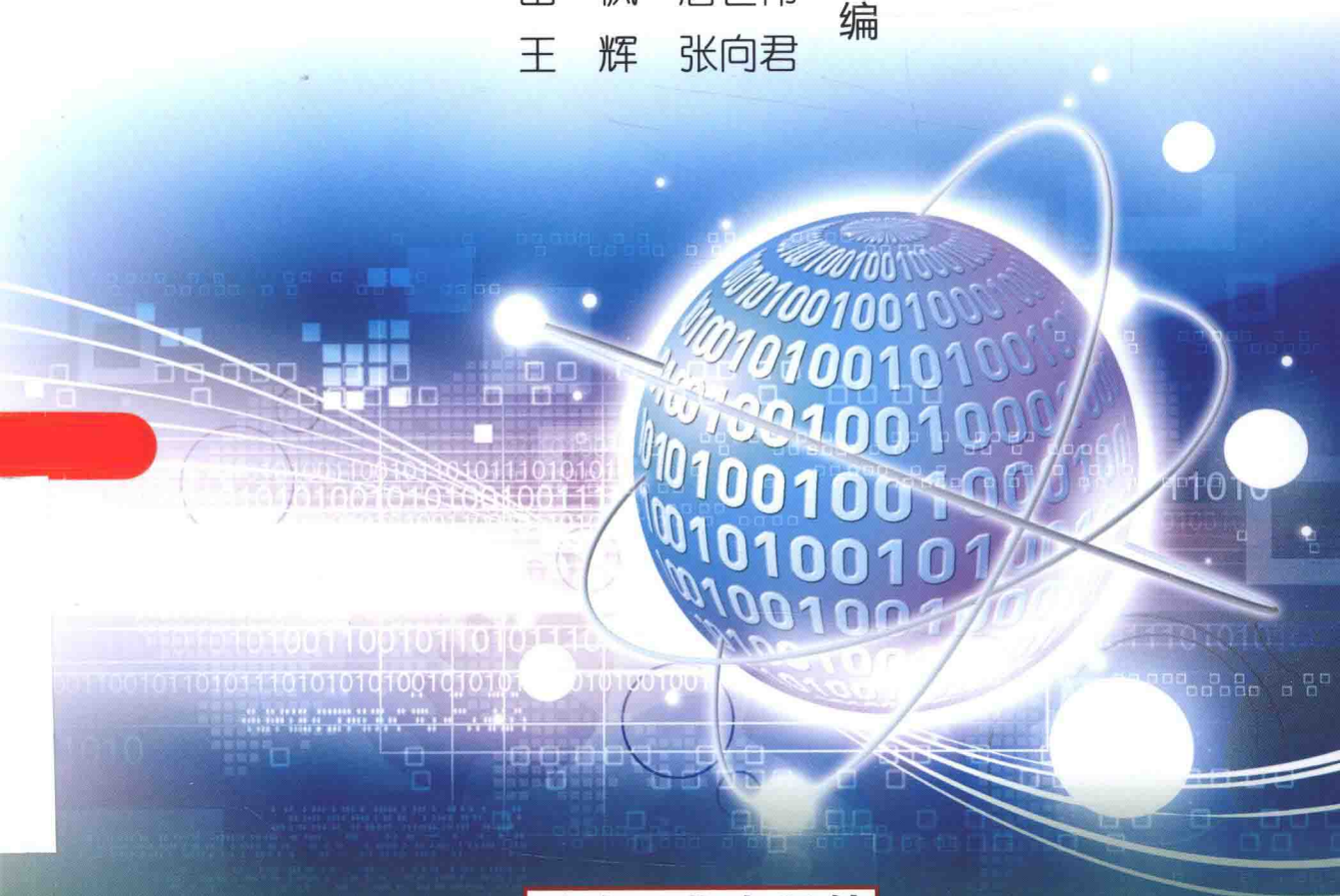
高等学校教材

信息论与编码

田　枫　唐世伟
王　辉　张向君　编



石油工业出版社
Petroleum Industry Press

高等学校教材

信息论与编码

田　枫　唐世伟
王　辉　张向君　编

石油工业出版社

内 容 提 要

本书系统介绍了由香农理论发展而来的信息论的基本理论以及编码的理论、实现原理。具体内容包括：信息的基本概念和度量方法，无失真信源编码理论和实用无损编码方法，限失真信源编码理论和有损压缩编码方法，信道容量与信道编码理论和编码方法，网络信息安全和密码学理论等。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推演过程和证明。

本书适合作为高等院校电子信息、通信工程、信息与计算科学等本科专业的教材，也可作为相关专业研究人员和工程人员的参考书。

图书在版编目(CIP)数据

信息论与编码/田枫等编．
北京：石油工业出版社，2015.8
（高等学校教材）
ISBN 978－7－5183－0834－7

Ⅰ．信…
Ⅱ．田…
Ⅲ．①信息论－高等学校－教材
②信源编码－高等学校－教材
Ⅳ．TN911.2

中国版本图书馆 CIP 数据核字(2015)第 180810 号

出版发行：石油工业出版社
（北京市朝阳区安华里 2 区 1 号　100011）
网　址：www.petropub.com
编辑部：(010)64256990　图书营销中心：(010)64523633
经　销：全国新华书店
排　版：北京乘设伟业科技有限公司
印　刷：北京晨旭印刷厂

2015 年 8 月第 1 版　2015 年 8 月第 1 次印刷
787×1092 毫米　开本：1/16　印张：15.25
字数：382 千字

定价：32.00 元
（如出现印装质量问题，我社图书营销中心负责调换）

前　言

信息论与编码是研究信息的存储、传输和处理规律的学科。它在方法论层面研究如何提高信息系统的可靠性、有效性、保密性和认证性，以达到信息系统的最优化。随着现代信息技术水平的迅猛发展和提高，信息论与编码理论也在不断发展。类似于无失真和限失真信源编码（即数据有损压缩和无损压缩原理）、信道编码原理（即纠错码理论）、保密编码原理（即数据加密与解密理论）等内容，在现代信息处理与通信等工程实践中都得到了广泛的应用。

伴随着信息技术的发展，信息论与编码理论的研究在不断扩大和深化，从狭义信息论发展到广义信息论，并迅速地渗透到其他相关学科领域，如无线电技术、自动控制、人工智能、信号处理、计算机技术、生命科学、材料科学、心理学、密码学、质量管理、市场营销、信息经济、美学等，从而形成了一门具有划时代意义的新兴学科。

本书系统介绍了信息论与编码理论，既有实际应用背景，又有清晰的数学思想。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、循序渐进。

全书共分 7 章，第 1 章为绪论，介绍信息的基本概念和定义，信息论的起源、发展和研究内容；第 2 章为信源与信源熵，介绍信源熵的概念、性质、定理等基础内容；第 3 章为无失真信源编码，介绍了信源的定长和变长编码定理，以及实用的无失真信源编码，即无损压缩编码方法；第 4 章为限失真信源编码，介绍了信息率失真理论和有损压缩编码；第 5 章为信道及信道容量，介绍了单符号离散信道、多符号离散信道和多用户信道的信道模型及信道容量的计算；第 6 章为信道编码，介绍了信道编码的基本概念、信道编码定理和检错码编码方法；第 7 章为网络信息安全与密码学，介绍了密码学的基本概念和加密方法。

本书由田枫、唐世伟、王辉、张向君共同编撰完成。具体编写分工如下：第 1 章、第 5 章、第 6 章由唐世伟、张向君编写；第 2 章、第 7 章由王辉编写；第 3 章、第 4 章由田枫编写；模拟试题由本书编者共同编写。全书的整理和审稿工作由田枫、唐世伟负责。

在编写过程中，参阅了相关同仁的前期工作，均列于参考书目中，在此谨向作者表示深切谢意。有关书中的不妥和错误之处，殷切希望广大读者予以批评指正。

编者

2015 年 6 月

目　录

第1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、接收和存储的一门学科。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深远的影响。

1.1 信息的基本概念

信息科学、材料科学和能源科学一起被称为当代文明的“三大支柱”。一位美国科学家说：“没有物质的世界是虚无的世界，没有能源的世界是死寂的世界，没有信息的世界是混乱的世界。”由此可见信息的重要性。随着社会信息化进程的加速，人们对信息的依赖程度会越来越高。

1.1.1 信息的不同定义

信息是信息论的最基本和最重要的概念，它既是信息论的出发点，也是信息论的归宿。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。

那么信息究竟是什么呢？

信息自古就有，但是古代社会文明程度很低，信息传递手段落后，获取信息困难，人们没有意识到信息的存在。随着人类社会的不断进步，人们才意识到信息的存在。对信息的认识随着社会文明程度的提高不断扩大和深入。然而，信息学科毕竟还是一门年轻的学科，人们对信息还没有一个全面的、系统的、准确的、一致的认识。从不同的学科、不同的角度、不同的方面、不同的层次、不同的深度，对信息有不同的认识。

信息的概念十分广泛，不同的定义在百种以上。例如，“信息是事物之间的差异”、“信息是事物联系的普遍形式”、“信息是物质和能量在时间和空间中分布的不均匀性”、“信息是物质的普遍属性”、“信息是收信者事先所不知道的报道”、“信息是用以消除随机不确定性的东西”、“信息是负熵”、“信息是作用于人类感觉器官的东西”、“信息是通信传输的内容”、“信息是加工知识的原材料”、“信息是控制的指令”、“信息就是数据”、“信息就是情报”、“信息就是知识”……数学家认为“信息是使概率分布发生改变的东西”，哲学家认为“信息是物质成分的意识成分按完全特殊的方式融合起来的产物”……

1928年，美国数学家哈特莱(Hartley)在《贝尔系统电话杂志》上发表了一篇题为《信息传输》的论文，把信息理解为选择通信符号的方式，并用选择的自由度来计量这种信息的大小。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。例如，从符号表中选择了这样一些符号：“I am well.”他就发出了“我平安”的信息；如果选择了“I am sick.”这些符号，他就发出了“我病了”的信息。发信者选择的自由度越大，所能发出的信息量也就越

大。此外，哈特莱还注意到，选择的具体物理内容是无关紧要的，重要的是选择的方式。也就是说，不管符号代表的意义是什么，只要符号表的符号数目一定，“字”的长度一定，那么，发信者所能发出的信息的数量就被限定了。所以他认为“信息是选择的自由度”。

事隔 20 年，另一位美国数学家香农(C. E. Shannon)在《贝尔系统电话杂志》发表了题为《通信的数学理论》的长篇论文。这篇论文以概率论为工具，深刻阐述了通信工程的一系列基本理论问题，给出了计算信源信息量和信道容量的方法和一般公式，得到了一组表征信息传递重要关系的编码定理，从而创立了信息论。但是香农并没有给出信息的确切定义，他认为“信息就是一种消息”。

后来，随着认识的进一步深化，人们把信息理解为广义通信的内容。美国数学家、控制论的主要奠基人维纳(Winner)在 1950 年出版的《控制论与社会》一书中写道，“人通过感觉器官感知周围世界”、“我们支配环境的命令就是给环境的一种信息”，因此，“信息就是我们在适应外部世界，并把这种适应反作用于外部世界的过程中，同外部世界进行交换的内容的名称”，“接收信息和使用信息的过程，就是我们适应外界环境的偶然性的过程，也是我们在这个环境中有效地生活的过程”。在这里，维纳把人与外部环境交换信息的过程看作是一种广义的通信的过程，认为“信息是人与外界相互作用的过程中所交换的内容的名称”。

这些定义都或多或少地从某种程度上描述了信息的一些特征，但是都不够全面、系统和准确。例如，消息、信号、数据、情报和信息都是在通信系统中传送的东西，但是这些概念之间有着原则的区别。消息是信息的外壳，信息则是消息的内核。同样多的消息，所包含的信息量可能差异很大。信号也不等同于信息，它只是信息的载体，信息是信号所载荷的内容。同样的信息可以用多种不同的信号来载荷。至于数据，它只是记录信息的一种形式，而且不是唯一的形式，因此不能把它等同于信息本身。情报一词在日语中的确就是信息，但是在汉语中，情报只是一类专门的信息，是信息的一个子集。

(1)维纳的定义：信息是人与外界相互作用的过程中所交换的内容的名称。信息既不是物质又不是能量，信息就是信息。

维纳对信息的认识也不够准确。因为在人与外界相互作用的过程中，参与内容交换的还有物质和能量，而不仅仅是信息。后来维纳自己也认识到“信息既不是物质又不是能量，信息就是信息”。这句话起初被人批评为唯心主义，也有人笑话维纳“说了等于没说”。但是人们后来才意识到，正是维纳揭示了信息的特性，即信息是独立于物质和能量之外存在于客观世界的第三要素。

(2)本体论层次的信息定义：事物的信息是该事物运动的状态和状态改变的方式。

上述定义虽然各不相同，实质内容并无太大的差异。主要差异在于侧重面不同、详略不同、抽象的程度不同和概括的层次高低不同。根据不同的条件区分不同的层次，可以给信息下不同的定义。最高的层次是最普遍的层次，也是无约束条件的层次，定义事物的信息是该事物运动的状态和状态改变的方式。我们把它叫作本体论层次。在这个层次下定义的信息是最广义的信息，使用范围也最广。每引入一个约束条件，定义的层次就降低一点，使用的范围就变窄一点。

(3)认识论层次的信息定义：信息是认识主体(生物或机器)所感知的或所表述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。

例如，引入一个最有实际意义的约束条件：认识主体，即站在认识主体的立场上定义信息。这样，本体论层次的信息定义就转化为认识论层次的信息定义。其中认识主体所感知的东西是外部世界向认识主体输入的信息，而认识主体所表述的东西则是其向外部世界输出的信息。

本体论层次和认识论层次的信息定义之间有着本质上的联系，即两者关心的都是“事物的运动状态及其变化方式”。但是两者之间又有原则的区别：前者从“事物”本身的角度出发，就“事”论事；后者是从“主体”的角度出发，就“主体”论事。

虽然认识论比本体论的层次要低一些，所定义信息的使用范围也要窄一些，但是信息概念的内涵比本体论层次要丰富得多。因为认识主体具有感觉能力、理解能力和目的性，能够感觉到事物运动状态及其变化方式的外在形式、内在含义，并能够判断其效用价值。对认识主体来说，这三者之间是相互依存、不可分割的关系。因此，在认识论层次上研究信息的时候，“事物的运动状态及其变化方式”就不再像本体论层次上那样简单了，它必须同时考虑到形式、含义和效用三个方面的因素。

事实上，认识主体只有在感知了事物运动状态及其变化的形式，理解了它的含义，判明了它的效用之后，才算真正掌握了这个事物的认识论层次信息，才能做出正确的决策。一般把同时考虑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的认识论层次信息称为“全信息”，而把仅仅考虑其中形式因素的部分称为“语法信息”，把考虑其中含义因素的部分称为“语义信息”，把考虑其中效用因素的部分称为“语用信息”。换句话说，认识论层次的信息是同时考虑语法信息、语义信息和语用信息的全信息。

香农信息论仅考虑了事物运动状态及其变化方式的外在形式，实际上研究的是语法信息。从这个角度出发，可以对信息下这样的定义：信息是对事物运动状态和变化方式的表征，它存在于任何事物之中，可以被认识主体（生物或机器）获取和利用。从数学观点出发研究香农信息论，可以认为信息是对消息统计特性的一种定量描述。

信息存在于自然界，也存在于人类社会，其本质是运动和变化。可以说哪里有事物的运动和变化，哪里就会产生信息。信息必须依附于一定的物质形式存在，这种运载信息的物质，称为信息载体。

人类交换信息的形式丰富多彩，使用的信息载体非常广泛。概括起来，有语言、文字和电磁波。语言是信息的最早载体；文字使信息保存得更持久、传播范围更大；电磁波则使载荷信息的容量和速度大为提高。

1.1.2 信息的特征

信息本身既看不见，又摸不着，没有气味、没有颜色、没有形状、没有大小、没有重量……总之，它是非常抽象的东西。但它又处处存在，呼之塞耳，示之濡目。它既区别于物质和能量，又与物质和能量有相互依赖的关系。综合起来，信息有以下主要特征：

(1)信息来源于物质，又不是物质本身；它从物质的运动中产生出来，又可以脱离源物质而相对独立地存在。

(2)信息来源于精神世界，但又不局限于精神领域。

(3)信息与能量息息相关，但又与能量有本质的区别。

(4)信息具有知识的本性，但又比知识的内涵更广泛。

(5)信息可以被认识主体获取和利用。

1.1.3 信息的性质

根据上述特征和信息的基本定义，可以导出信息的一些重要性质：

(1)存在的普遍性。信息的本质是事物的运动和变化，只要有事物的存在，就会有事物的运动和变化，就会产生信息。绝对静止的事物是没有的，因此，信息普遍存在。

(2)有序性。信息可以用来消除系统的不确定性，增加系统的有序性。认识论层次的信息是认识主体所感知和表述的事物运动的状态和方式。获得了信息，就可以消除认识主体对于事物运动状态和方式的不确定性。信息的这一性质对人类有特别重要的价值，要使一个系统从无序变为有序，必须从外界获取信息。

(3)相对性。对于同一个事物，不同的观察者所能获得的信息量可能不同。

(4)可度量性。信息虽然很抽象，但它是可以度量的。信息的多少用信息量表示。

(5)可扩充性。信息并非一成不变，随着时间的推移，大部分信息将得到不断扩充。例如，人类对于宇宙的认识就是不断扩充的。人们对信息的认识也在不断扩充。香农创立信息论之前，很少有人意识到信息的客观存在，如今人们对信息的研究已经非常广泛和深入。

(6)可存储、传输与携带性。信息依附于信息载体而存在，而任何物质都可以成为信息的载体。既然物质可以存储、传输和携带，所以信息可通过信息载体以多种形式存储、传输和携带。

(7)可压缩性。人们得到信息之后，并非原封不动拿来应用，往往要进行加工、整理、概括、归纳，使信息更加精练、可靠，从而浓缩。信息论研究的主要问题之一就是信息的压缩。

(8)可替代性。信息能替代劳动力、资本、物质材料甚至时间，正确、及时、有效地利用信息，可创造更多的物质财富，开发或节约更多的能量，节省更多的时间，收到巨大的经济效益。

(9)可扩散性。信息可以在短时间内较大范围地扩散开来。如广播、电视信息，顷刻之间即传遍全球。

(10)可共享性。信息与实物不同，可以大家共享。甲传递一件东西给乙，乙得到，甲便失去。但信息持有者传递一条信息给另一个人的时候，他自己所拥有的信息并不会丧失。正像教师把知识传授给学生一样，学生掌握了知识，但教师的知识并不会因此减少。信息的这种特性对人类具有特别重要的意义。可以说没有信息的共享性就没有人类社会的发展和进步。

(11)时效性。信息以事实的存在为前提。它不是一成不变的死东西，可以随着事实的不断扩大而增值，也会随着事实的过去而衰老，从而失去本身的价值，是有“寿命”的。

信息在信息化程度越来越高的社会中将起到越来越重要的作用，是比物质和能量更为宝贵的资源，全面掌握信息的概念，正确、及时、有效地利用信息，能够为人类创造更多的财富。

1.1.4 信息的分类

通过对信息概念、特征和性质的讨论，使我们对信息有了定性的认识。但要全面、准确地掌握信息的概念，必须对信息有定量的认识。这就要求首先能够确切地描述信息，即对信息进行分类。

信息分类有许多不同的准则和方法。

(1)按照信息的性质，可以分成语法信息、语义信息和语用信息。

(2)按照观察的过程，可以分成实在信息、先验信息和实得信息。

(3)按照信息的地位,可以分成客观信息和主观信息。

(4)按照信息的作用,可以分成有用信息、无用信息和干扰信息。

(5)按照信息的逻辑意义,可以分成真实信息、虚假信息和不定信息。

(6)按照信息的传递方向,可以分成前馈信息和反馈信息。

(7)按照信息的生成领域,可以分成宇宙信息、自然信息、社会信息和思维信息。

(8)按照信息的应用部门,可以分成工业信息、农业信息、军事信息、政治信息、科技信息、文化信息、经济信息、市场信息和管理信息等。

(9)按照信息的来源,可以分成语声信息、图像信息、文字信息、数据信息、计算信息等。

(10)按照信息载体的性质,可以分成语声信息、图像信息、文字信息、电磁信息、光学信息和生物信息等。

(11)按照携带信息的信号性质,还可以分成连续信息、离散信息和半连续信息等。

研究信息的目的,就是要准确地把握信息的本质和特点,以便更有效地利用信息。因此,在众多的分类原则和方法中,最重要的就是按照信息性质的分类。在语法信息、语义信息和语用信息三个基本类型中,最基本也是最抽象的类型是语法信息。它是迄今为止在理论上研究得最多的类型。

语法信息考虑的是事物运动状态和变化方式的外在形式。根据事物运动状态和方式在形式上的不同,语法信息还可以进一步分成有限状态和无限状态;其次,事物运动状态可能是连续的,也可能是离散的,于是,又可以分成连续状态语法信息和离散状态语法信息;再者,事物运动状态还可能是明晰的或者是模糊的,这样,又可以分成状态明晰的语法信息和状态模糊的语法信息。

当然,按照事物运动的方式,还可以把信息进一步细分为概率信息、偶发信息、确定信息和模糊信息。香农信息论主要讨论的是语法信息中的概率信息,本书也以概率信息为主要研究对象。

上述按照信息性质的分类可以用图 1.1.1 直观地表示。

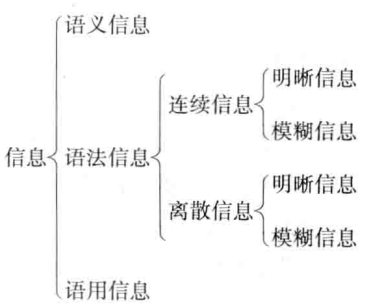


图 1.1.1　不同性质的信息分类

1.1.5　信息科学

信息科学是研究信息的概念、相关理论和应用的科学,是一门新兴的边缘学科。

信息科学具有以下特点:第一,多学科性,信息科学与许多基础科学和应用技术有关,互相渗透,如数学、逻辑学、心理学、语言文字学、生物学、控制论、计算机科学、通信技术、仿生学和人工智能技术;第二,产业化性,信息科学服务于国民经济和社会生活的各个方面,从而形成一个新兴产业——信息产业。

信息科学的研究范围有：

(1)信息源：自然信息源(物理、化学、天体、地理、生物)；社会信息源(管理、金融、商业)；知识信息源(古今中外)等。

(2)信息载体：第一载体(语言)；第二载体(文字)和第三载体(电磁波)。

(3)信息的采集与转换：传感器、雷达、视、听、触、力、声、光等。

(4)信息的传输：光、电磁波、神经、意念。

(5)信息的存储与处理：计算机、视听系统。

1.2 信息论的起源、发展及研究内容

1.2.1 信息论的起源

信息论理论基础的建立，一般来说开始于1948年美国数学家香农在《贝尔系统电话杂志》发表了题为《通信的数学理论》的长篇论文。这篇论文以概率论为工具，深刻阐述了通信工程的一系列基本理论问题，给出了计算信源信息量和信道容量的方法和一般公式，得到了一组表征信息传递重要关系的编码定理，从而创立了信息论。

信息论自诞生到现在不过60多年，在人类科学史上是相当短暂的。但它的发展对学术界及人类社会的影响是广泛而深刻的。信息作为一种资源，如何开发、利用、共享，是人们普遍关心的问题。

信息论是研究信息的传输、存储和处理的学科，亦称它为“通信的数学理论”。它主要研究在通信系统设计中如何实现信息传输的有效性和可靠性。因此，信息论与通信技术、统计数学、信号处理等密切相关。

1.2.2 信息论的发展

在人类历史的长河中，信息传输和传播手段经历了五次重大变革，正是在不断的变化中，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为提高。第三次变革是印刷术的发明。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。它开始了人类的电信时代，通信理论和技术迅速发展。第五次变革是计算机技术与通信技术的结合，促进了网络通信的发展。

1924年，奈奎斯特(Nyquist)解释了信号带宽和信息速率之间的关系。20世纪30年代，新的调制方式，如调频、调相、单边带调制、脉冲编码调制和增量调制的出现，使人们对信息能量、带宽和干扰的关系有了进一步的认识。1936年，阿姆斯特朗(Armstrong)指出增大带宽可以使抗干扰能力加强，并根据这一思想提出了宽频移的频率调制方法。1939年，达得利(Dudley)发明了带通声码器，指出通信所需带宽至少同待传送消息的带宽应该一样。声码器是最早的语言数据压缩系统。这一时期还诞生了无线电广播和电视广播。通信技术的进步使人们考虑到更深入的问题：究竟如何定量地研究通信系统中的信息，怎样才能更有效和更可靠地传递信息，现有的各种通信体制如何改进，等等。

1928年，哈特莱首先提出了用对数度量信息的概念。哈特莱的工作给香农很大的启示，他在1941—1944年对通信和密码进行深入研究，用概率论和数理统计的方法系统地讨论了通信的根本问题，得出了几个重要而带有普遍意义的结论：(1)阐明通信系统传递的对象就是信息，并对信息给予科学的定量描述，提出了信息熵的概念；(2)指出通信系统的中心问题是在噪声下如何有效而可靠地传递信息，以及实现这一目标的方法是编码等。这些成果在1948年以《通信的数学理论》(A mathematical theory of communication)为题公开发表，标志着信息论的正式诞生。与此同时，维纳在研究火控系统和人体神经系统时，提出了在干扰作用下的信息最佳滤波理论，成为信息论的一个重要分支。

20世纪50年代，信息论在学术界引起了巨大反响。1951年，美国无线电工程师协会(IRE)成立了信息论组，并于1955年正式出版了《信息论汇刊》。这一时期，包括香农本人在内的一些科学家做了大量工作，发表了许多重要文章，将香农的科学论断进一步推广，同时信道编码理论有了较大的发展。信源编码的研究落后于信道编码。1959年，香农发表了《保真度准则下的离散信源编码定理》(Coding theorems for discrete source at the fidelity criterion)，系统地提出了信息率失真理论(rate－distortion theory)，为信源压缩编码的研究奠定了理论基础。

20世纪60年代，信道编码技术有了较大发展，成为信息论的又一重要分支。它把代数方法引入到纠错码的研究，使分组码技术达到了高峰，找到了可纠正多个错误的码，并提出了可实现的译码方法。其次是卷积码和概率译码有了重大突破，提出了序列译码和维特比(Viterbi)译码方法。

1961年，香农的重要论文《双路通信信道》开拓了多用户信息理论的研究。

宽带综合业务数字网(B－ISDN，Broad－Integrated Service Digital Network)的出现，给人们提供了除电话服务以外的多种服务，使人类社会逐渐进入了信息化时代。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。20世纪70年代以后，多用户信息论成为中心研究课题之一。

后来，随着通信规模的不断扩大，人们逐渐意识到信息安全是通信系统正常运行的必要条件。于是，把密码学也归类为信息论的分支。如今信息安全已是网络通信和电子商务系统中不可缺少的重要环节。

人们对信息的认识越来越深入，先后提出了加权熵、动态熵等概念，建立在模糊数学基础之上的模糊信息的研究也取得了一定的进展。信息论不仅在通信、广播、电视、雷达、导航、计算机、自动控制、电子对抗等电子学领域得到了直接应用，还广泛地渗透到诸如医学、生物学、心理学、神经生理学等自然科学的各个方面，甚至渗透到语言学、美学等领域。

从20世纪60年代开始，一些社会学家在研究社会问题和社会现象时，先后提出了后工业社会和信息社会的概念，信息论开始向经济学和社会科学领域渗透。1977年，美国经济学家马克·波拉特发表了长达九卷的《信息经济》报告，用信息论的基本概念研究经济现象和社会现象，将信息论的研究从自然科学领域正式移植到经济学和社会科学领域。

1.2.3 信息论研究的主要内容

信息论的研究对象是广义通信系统。不仅是电子的、光学的信号传递系统，对于任何系统，只要能够抽象成通信系统模型，都可以用信息论研究，如神经传导系统、市场销售系统等。关于信息论的研究内容，一般有以下三种解释。

1. 信息论基础

信息论基础亦称香农信息论或狭义信息论。它主要研究信息的测度、信道容量、信息率失真函数，以及与这三个概念相对应的香农三定理以及信源和信道编码。

2. 一般信息论

一般信息论主要是研究信息传输和处理问题。除了香农基本理论之外，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。后一部分内容以美国科学家维纳为代表。虽然维纳和香农等人都是运用概率和统计数学的方法研究准确或近似再现消息的问题，都是通信系统的最优化问题。但他们之间有一个重要的区别，维纳研究的重点是在接收端。研究消息在传输过程中受到干扰时，在接收端如何把消息从干扰中提取出来。在此基础上，建立了最佳过滤理论(维纳滤波器)、统计检测与估计理论、噪声理论等。香农研究的对象是从信源到信宿的全过程，是收、发端联合最优化问题，重点是编码。香农定理指出，只要在传输前后对消息进行适当的编码和译码，就能保证在有干扰的情况下，最佳地传送消息，并准确或近似地再现消息。为此，发展了信息测度理论、信道容量理论和编码理论等。

3. 广义信息论

广义信息论是一门综合性的新兴学科，至今并没有严格的定义。概括说来，凡是能够用广义通信系统模型描述的过程或系统，都能用信息基本理论来研究。它不仅包括一般信息论的所有研究内容，还包括如医学、生物学、心理学、遗传学、神经生理学、语言学、语义学，甚至社会学和经济管理中有关信息的问题。反过来，所有研究信息的识别、控制、提取、变换、传输、处理、存储、显示、价值、作用以及信息量的大小的一般规律以及实现这些原理的技术手段的工程学科，也都属于广义信息论的范畴。

总之，人们研究信息论的目的是为了高效、可靠、安全并且随心所欲地交换和利用各种各样的信息。

1.2.4 信息论的应用

信息论从它诞生那时起就吸引了众多领域学者的注意，他们竞相应用信息论的概念和方法去理解和解决本领域中的问题。例如，信息论在生物学、医学、经济、管理、图书情报等领域都有不同程度的应用，这使信息论成为一门新兴的横断科学。在这里，简要介绍一下信息论在生物学、医学、管理科学、经济学中的应用。

1. 信息论在生物学中的应用

生命体本身是一个复杂的信息传递、存储、处理、加工和控制的系统。理论上说，信息论应该和生物学有着密切关系。近几十年来，由于生物学的发展非常迅速，人们对生命现象的研究，已经从整体深入到细胞、亚细胞、分子水平和量子水平上，以揭示生命现象的本质。尤其是在遗传信息方面的研究取得了重大进展和成效，从此确立了信息理论在生物学研究方面的重要作用和地位。

特别是20世纪90年代以来，伴随着分子结构测定技术的突破和各种基因组测序计划的展开，生物学数据大量出现，如何分析这些数据，从中获得生物结构、功能的相关信息成为困扰生物学家的一个难题。生物信息学就是在此背景下发展起来的综合运用生物学、数学、统计学、物理学、化学、信息科学以及计算机科学等诸多学科的理论和方法的崭新的交叉学科。

目前，国际上公认的生物信息学的研究内容大致包括以下几个方面：

(1)生物信息的收集、储存、管理和提供。

(2)基因组序列信息的提取和分析。

(3)功能基因组相关信息分析。

(4)生物大分子结构模拟和药物设计。

(5)生物信息分析的技术与方法研究。

(6)应用与发展研究。

2. 信息论在医学中的应用

医学是研究人的生命活动的本质、疾病发生发展的规律、诊断和防治疾病、恢复和保护人的身体健康的学科。信息论在医学上的应用，大大促进了医学的现代化。

从信息论的观点看，有机体是不断接收与输出信息的，以维持正常的生命活动。有机体中，信息熵标志着系统组织结构复杂的有序状态，由于新陈代谢的作用，有机体内部有序结构不断遭到破坏，这时熵增加，反之机体不断从外界接收信息——负熵，在机体内合成高度的有序结构，使熵降低。因此运用信息理论来分析生命系统，可以把生命系统看作是接收信息和传递信息的调节控制系统。

3. 信息论在管理科学中的应用

在现代化管理中，信息论已成为与系统论、控制论等相并列的现代科学的主要方法论之一。信息价值、信息量、信息反馈、信息时效性和真实性，信息处理和传递，以及信息论与信息科学是现代化管理的运动命脉。实际上，现代化管理与信息已融为一体，并形成一种特殊形态的信息运动形式，即管理系统信息流。

在整个管理世界里，管理信息依据不同的分类方法，可以分为各种不同的类别，而在这繁多的种类中，总的可分为两大形式：管理自然信息和管理社会信息。管理自然信息指的是：管理系统以时间、效益形式呈现的自身形态、结构、运动过程与主体(主要是管理者)的同样以时间、效益形式呈现的自身形态、结构、运动过程相互作用而在人脑中留下的与该管理系统同态的响应。管理社会信息指的是：一切经过管理者利用语言、文字、符号、图像等加工过的管理自然信息。管理方面的知识、情报、指令、告示、法律等全都属于管理社会信息。

对于任何管理者来说，他将随时都会同时面临着这两种信息，并深刻地影响着自己的管理活动。由此可见，信息论在企业管理中拥有重要的应用价值和应用前景。

4. 信息论在经济学中的应用

信息论在经济学领域中有着广泛的渗透：一方面，可以用经济学的观点来研究信息的一般问题，特别是信息的价值问题；另一方面，又可以用信息科学的观点和方法来重新认识和探讨经济活动的规律。

目前，在经济学领域活跃着一门新的学科：信息经济学(Economics of Information)。截至目前，信息经济学可概括为五大领域：不完全信息经济学、信息转换经济学、信息的经济研究、信息经济的研究、信息经济的社会学研究。

1.3 编码理论概述

1.3.1 编码理论的基本概念

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，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，通常称它为通信系统模型，如图 1.3.1 所示。

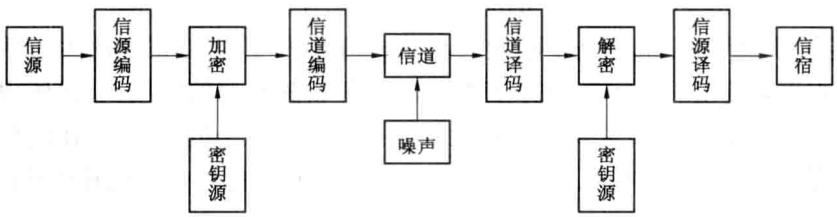


图 1.3.1 通信系统模型

图 1.3.1 所示的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。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。信息传输或通信的目的是要把收方不知道的信息及时、可靠、完整、安全而又经济地传送给指定的收方。该模型按功能可分为信源、编码器、信道、译码器、信宿五部分。

(1)信源是产生消息和消息序列的源，它可以是人、生物、机器或其他事物，它是事物各种运动状态或存在状态的集合。信源发出的消息有语音、图像、文字等，人的大脑思维活动也是一种信源。信源的输出是消息，消息是具体的，但它不是信息本身。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)编码器可分信源编码器、信道编码器、保密编码器三种。信源编码是对信源输出的消息进行适当的变换和处理，把信息变换成信号，目的是为了提高信息传输的效率，使传输更为经济、有效，还要去掉一些与被传信息无关的多余度；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；保密编码是为了保证信息的安全性。在信息传输或处理过程中，除了指定的接收者外，还有非指定的或非授权的用户，他们通过各种技术手段企图窃取机密信息。因此，为了保证被传送信息的安全和隐私，必须对信源的输出进行加密或隐藏，同时还要求信息传递过程中保证信息不被伪造和篡改。通信系统中的传输媒质有电缆、明线、光纤和无线电波的传播空间等。信号通过这些媒质时，是很不安全的，存在着各种天然和人为干扰使被传信号产生错误。除此以外，非指定用户或敌人还会通过各种方法(如搭线、电磁波接收、声音接收等)对所传输的信号进行侦听(称被动攻击)。更有甚者，有些非法入侵者主动对系统进行骚扰，采用删除、更改、增添、重放、伪造等手段，向系统注入信号或破坏被传的信号，以达到欺骗别人，有利于自己的目的，这种攻击称为主动攻击。因此，保护系统中所传消息的真实性、完整性，是一个更为困难的问题，也是密码系统所必须完成的另一个更为艰巨的任务。由于在传

输信息的媒质中总是存在着各种人为或天然的干扰和噪声，因此，为了提高整个通信系统传输信息的可靠性，就需要对加密器输出的信息进行一次纠错编码，人为地增加一些多余信息，使其具有自动检错或纠错功能。这种功能由图 1.3.1 中信道编码器完成。当然，对于各种实际的通信系统，还应包括换能、调制、发射等各种变换处理。

(3)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。信道除了传送信号以外，还有存储信号的作用。在信道中还存在噪声和干扰，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。由于干扰或噪声往往具有随机性，所以信道的特性也可以用概率空间来描述。

(4)译码就是把编码器输出的编码信号进行反变换，一般认为这种变换是可逆的。译码器也可分成信源译码器、信道译码器及保密译码器三种。

(5)信宿是消息传送的对象，即接收消息的人或机器。图 1.3.1 给出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，而卫星通信则是多个输入、多个输出的多向传输的通信。

1.3.2 编码理论的发展

1948 年，香农在《通信的数学理论》的论文中，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。香农理论的核心是：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性的信息传输，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。

从无失真信源编码定理出发，1948 年，香农在论文中提出并给出了简单的编码方法(香农编码)，1952 年，费诺(Fano)提出了一种费诺码，同年哈夫曼(D. A. Huffman)构造了一种哈夫曼编码方法，并证明了它是最佳码。哈夫曼码是有限长度的块码中的最好的码，亦即代码总长度最短的码。1956 年，麦克米伦(B. McMillan)首先证明了唯一可译变长码的克拉夫特(Kraft)不等式。

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，使其更为简化、易于实现。

1977 年由齐弗(J. Ziv)和兰佩尔(A. Lempel)提出了 LZ 算法，它是适用于通用信源的编码算法之一。1978 年他们又提出了改进算法，而且齐弗也证明此方法可达到信源的熵值。1990 年，贝尔(T. C. Bell)等在 LZ 算法基础上又做了一系列改进，现在 LZ 码已广泛应用于文