
# 信息论、编码与密码学

田丽华　编著

西安电子科技大学出版社
http://www.xduph.com

21 世纪高等学校电子信息类规划教材

# 信息论、编码与密码学

田丽华 编著

西安电子科技大学出版社

2008

## 内容简介

本书系统地介绍了信息理论、信源的压缩编码、信道的纠错编码、加密编码学及组合编码等内容的基本原理及应用，同时简单介绍了学习本书需要的数论及近代代数的相关知识。

本书主要内容有：信源及信息度量；信道及信道容量；信源压缩编码原理及编码方法；信道纠错编码的基本原理和编码方法；密码编码的基本原理和编码方法；消息认证的相关知识；组合编码原理及编码方法。

本书力求物理概念清晰、通俗易懂、由浅入深、循序渐进、重点突出，对基本概念和基本原理的阐述清晰明了，实用性强。本书可作为电子信息类、信息工程类、计算机等专业本科生和研究生的教材或参考书，也可供从事电子、信息、通信、计算机、自动化等专业的科技人员参考。

**图书在版编目(CIP)数据**

**信息论、编码与密码学**/田丽华编著. —西安：西安电子科技大学出版社，2008.4
21世纪高等学校电子信息类规划教材
ISBN 978-7-5606-1967-5

Ⅰ.信… Ⅱ.田… 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密码-理论-高等学校-教材 Ⅳ.TN911.2 TN918.1

**中国版本图书馆CIP数据核字(2007)第204882号**

策　　划　云立实
责任编辑　薛　媛　云立实
出版发行　西安电子科技大学出版社(西安市太白南路2号)
电　　话　(029)88242885　88201467　　邮　编　710071
http://www.xduph.com　　E-mail:xdupfxb@pub.xaonline.com
经　　销　新华书店
印刷单位　陕西华沐印刷科技有限责任公司
版　　次　2008年4月第1版　2008年4月第1次印刷
开　　本　787毫米×1092毫米　1/16　印张25.375
字　　数　595千字
印　　数　1～4000册
定　　价　36.00元
ISBN 978-7-5606-1967-5/TP·1016
**XDUP 2259001-1**

# 前　言

信息理论与编码既是一门工程科学，又是一门应用科学，同时也是一门不断发展的学科。信源压缩编码、信道纠错编码及密码编码学是信息论的三大核心内容。本书试图用有限的篇幅将信息理论、信源压缩编码、信道纠错编码、密码编码学以及组合编码等重要的基本原理及方法有机地结合起来讲述，力图使本书的内容具有知识性、研究性、实用性、先进性和综合性；章节的安排结构严谨、合理而系统；物理概念清晰、通俗易懂、由浅入深、循序渐进、示例丰富，便于读者学习；突出了其在信息传输系统中的应用，有助于读者了解产生理论和解决问题的实际背景，也提高了工科学生的学习兴趣。

本书所要求的概率论、随机过程、线性代数等数学基础是初等的，同时对学习本书所要求的数论、离散数学及近代代数中的基本知识，做了简单的介绍，供读者学习时参考。

全书共五篇，分 15 章。书中系统地介绍了信息理论、信源的压缩编码、信道的纠错编码、保密编码学以及组合编码等知识的基本原理及应用。同时，简单介绍了学习本书需要的数论及近代代数的相关知识。除了第 14 章及 10.5 节以外的绝大多数内容均适合于本科教学。其中，第 6 章可以不讲，仅供学生学习相关内容时参考。书中有些加宽、加深的内容，对本科生讲授时，可作适当取舍，或只讲授基本内容，如 BCH 码、Goppa 码、秩距离码、AES 算法、IDEA算法、EIGamal 算法、椭圆曲线加密算法、Turbo 码及 TCM 码等。

本书作者在编写本书的过程中，得益于对国内外不少信息编码方面优秀教材及专著的学习。同时，还参阅了许多文献、资料。在此，作者对这些著作的作者深表谢意。

本书在作者多年教学经验和研究实践的基础上编写而成，书中的所有图形均由蔡东杰副教授设计并绘制。

作者要感谢吉林大学提供的良好的教学及科研环境；感谢信息系的领导及同事提出的宝贵意见及帮助；感谢西安电子科技大学出版社的信任与支持及其工作人员的辛勤劳动；感谢博士导师王珂教授提出的许多宝贵意见；作者也要感谢教过的几届学生，他们的反馈为本书的改进提供了很好的帮助；感谢作者家人的支持与理解。最后，要感谢所有对本书编写过程中给予过热情帮助的前辈、同行及学生们：王新梅、王树勋、康键、云立实、杨晓萍、王国鸿、张巍、周文慧等。

限于作者的水平，书中不妥和谬误之处难免，欢迎读者将发现的错误、遗漏以及其他的建议发到 tlh85@sina.com。

作　者

2007 年 4 月 25 日

# 目　录

## 第三篇 信道纠错编码

## 第四篇 加密编码学

## 第五篇 组 合 编 码

# 第1章 绪 论

美国数学家香农(C. E. Shannon)在1948年发表了著名的论文“通信的数学理论”，开创了一门在现代科学技术中具有重大意义的崭新的学科——信息论。顾名思义，信息论是关于信息的理论，应有自己明确的研究对象和适用范围，但从信息论诞生的那时起人们就对它有不同的理解。

信息作为技术术语广泛使用，是在计算机特别是微处理器得到广泛应用之后。在计算机发展的早期，计算机处理的对象仍沿用过去的名词，如数据、记录、报表、文字等等。但随着计算机的不断发展，无论在计算机学术界或工业界都产生了一种明显的倾向，即希望有一个名称能把所有这些处理对象统统包含在内。信息这一名称恰好符合这一要求，因为只有这样一个含糊的术语才能对多种多样且在不断涌现的对象有一个统一的、全面的、不需时时改变的表达。信息作为一个可以严格定义的科学名词，首先出现在统计数学中，随后又出现在通信技术中。统计信息是一个抽象而明确的概念，它与作为技术术语用的信息仍有很大的区别。

## 1.1 信息传输系统

### 1.1.1 信息传输的目标

研究通信系统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、有效性、保密性和认证性，以达到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、不失真地再现在接收端。而有效性高，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。但提高可靠性和提高有效性常常会发生矛盾，这就需要统筹兼顾。例如，为了兼顾有效性(考虑经济效果)，有时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一定的失真，或者说允许近似地再现原来的消息。信息传输的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。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。

有效性、可靠性、保密性、认证性和经济性构成了现代通信系统对信息传输的全面要求。

### 1.1.2 信息传输系统模型

各种现代通信系统如电报、电话、无线电、电视、广播、因特网、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，如图 1－1 所示。通常称它为信息传输系统模型。

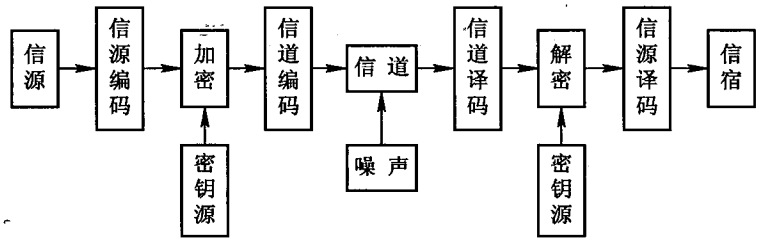


图 1－1　信息传输系统模型

图 1－1 信息传输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，人体、动物的神经网络系统、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**

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、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，信源输出的消息是随机的、不确定的，但又有一定的规律性。信源输出的消息有多种形式，可以是离散的或连续的、平稳的或非平稳的、无记忆的或有记忆的。

**2. 编码器**

编码器可分为信源编码器、信道编码器和保密编码器三种。信源编码是对信源输出的消息进行适当的变换和处理，目的是为了提高信息传输的效率。为了使传输更为经济有效，还要去掉一些与被传信息无关的消息。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。保密编码的目的是保证信息的安全性。由于传输信息的媒质如电波、天线等总是存在有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，就需要对保密编码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。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**

信道是信息传输和存储的媒介，是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，明线、电缆、波导、光纤、无线电波传播空间等，都是信道。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，往往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. 译码器**

译码器是编码器的反变换。一般认为这种变换是可逆的。译码器也可分成信源译码器、信道译码器和保密译码器三种。

**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 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，另外信息传输方向也可以双向进行，例如广播通信是一个输入、多个输出的单向传输的通信；因特网是多输入、多输出的多向传输的通信；卫星通信网也是多输入、多输出的多向传输的通信。

## 1.2 信息传输系统的发展与现状

### 1.2.1 信息论的形成与发展

从历史上看信息论的形成是两部分人共同努力的结果，一部分是通信工程方面的学者，另一部分是统计数学家。这两部分人虽然研究的是同一领域的问题，但他们感兴趣的方面和侧重点是有差异的。这种情况从信息论产生时起一直保持到现在，今天从事信息论研究的工作者仍然由这两部分人组成。

信息论的基本概念最初是从古典的统计理论与通信工程中提出的。但自从信息论产生以后，它的一些基本概念与方法就在一般的信号与信息处理中获得应用，并在应用过程中逐步丰富和发展了信息论的内容。

信息论最主要的内容是以通信技术基础理论的形式逐步形成和发展起来的。这一点有它内在的原因。一方面广义信息的含义极其复杂，而通信本身只涉及信息的表现形式或者说只对消息的表现形式感兴趣，而这是广义信息最简单最基础的方面。因此，可以认为正是对这最简单的方面的研究，形成了信息论。另一方面，当通信技术得到广泛发展和应用以致形成通信网以后，人们自然要问：通信传送的究竟是什么？而信息论正是对这一问题的全面和系统的问答。然而这个问题并不会自然地导致信息论的诞生，因为通信关心的是信息的表现形式，而这种形式在通信的传输过程中可能经过多次的变换。只要通信设备还能够把发送端输入的形式足够精确地在接收端输出处再现，人们是不会进一步追根究底的。只有当人们无法实现“准确再现”时，理论上的追根究底才有了动力，并导致信息论的诞生。

从历史上看，信息作为一个科学名词最早出现在统计数学中。1922 年 R. V. L. Hartley 发表了信息量的定义。1925 年统计数学家 R. A. Fisher 从古典统计理论的角度定义了一种信息量，这种信息量现在一般被称为 Fisher 信息量，Fisher 信息量在估计问题中迄今仍有着重要的价值。

1949 年香农把他在“通信的数学理论”一文中发展起来的概念用于保密系统，发表了

"保密系统的通信理论"一文。这一论文提出了完全保密性等重要概念，从而奠定了密码学的理论基础。1957 年 E. T. Jaynes 发表了"信息论与统计力学"，该文提出的最大熵原理不但对统计力学有重要意义，而且在随后的几十年中对信号处理产生了很大影响，成为信号处理的一个重要方法。20 世纪 60 年代人们又开始在模式识别与分类中应用信息论，其中较突出的代表是 S. Watanabe，他最早用熵解释模式分类过程。今天，随着信号与信息处理的深入发展，人们已经越来越深刻地认识到信号与信息处理的中心问题是信息。在非线性非高斯信号处理问题以及在信号分类识别问题和信号重建复原等问题中，信息论的方法应该取代诸如最小二乘误差等准则和方法，信息论应该成为信号与信息处理的一个理论基础。

### 1.2.2 信源压缩编码的发展

1948 年，香农在"通信的数学理论"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，得出了几个重要而具有普遍意义的结论。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。

当已知信源符号的概率特性时，可计算它的信息熵，用它表示每个信源符号所载有的信息量。编码定理不但证明了必存在一种编码方法，使代码的平均长度可任意接近但不能低于信息熵，而且还阐明达到这一目标的途径，就是使概率与码长匹配。信源编码定理出现后，编码方法就趋向于合理化。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，同年霍夫曼(D. A. Huffman)构造了一种霍夫曼编码方法，并证明了它是最佳码。霍夫曼码是有限长度的块码中最好的码，亦即代码总长度最短的码。1949 年 L. G. Kraft 提出了克拉夫特(Kraft)不等式，克拉夫特不等式指出了即时码的码长必须满足的条件，后来，B. McMillan 在 1956 年证明惟一可译码也满足此不等式，1961 年 J. karuSh 简化了 B. McMillan 的证明方法。

霍夫曼码在实际中已有所应用，但它仍存在一些块码及变长码所具有的缺点。例如概率特性必须精确地测定，它若略有变化，还需更换码表，以及对于二元信源，常需多个符号合起来编码，才能取得好的效果等。因此，在实用中常需作一些改进，同时也就有研究非块码的必要性。算术码就是一种非块码，它是从整个序列的概率的匹配来进行编码的。其实此概念也是香农首先提出的，后经许多学者改进，已逐渐进入实用阶段。1968 年前后，P. Elias 发展了香农—费诺码，提出了算术编码的初步思路。而 J. Rissanen 在 1976 年给出并发展了算术编码。1982 年他和 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。

对概率特性未知或不确知的信源进行有效的编码，上述方法已无能为力。对有些信源，要确知信源的统计特性相当困难，尤其对高阶条件概率是非常困难的，甚至有时信源

的概率特性根本无法测定，或是否存在也不知道。地震波信号就是如此，因为无法取得大量实验数据。当信源序列是非平稳时，其概率特性随时间而变更，要测定这种信源的概率特性也近乎不可能。因此，总希望能有一种编码方法，通用于各类概率特性的信源。通用编码就是在信源统计特性未知时，对信源进行编码，且使编码效率很高的一种码。

20 世纪 70 年代末，以色列学者 A. lempel 和 J. ziv 提出一种语法解析码，习惯上简称 LZ 码。1977 年他们首先提出这种基于字典的方法，1978 年他们又提出了改进算法，分别称为 LZ77 和 LZ78。1984 年 T. A. welch 以 LZ 编码中的 LZ78 算法为基础修改成一种实用的算法，后定名为 LZW 算法。LZW 算法保留了 LZ78 算法的自适应性能，压缩效果也大致相同，其显著特点是逻辑性强，易于硬件实现，且价格低廉，运算速度快。LZW 算法已经作为一种通用压缩方法，广泛应用于二元数据的压缩。

通用编码中最困难的问题是准则问题。这与概率匹配问题不同，此时已不能确定最佳的标准。当概率特性已知时，信源编码定理给出了极限值，达到这个界的就是最佳码。当概率特性未知时，就无法确定这个上界。一般认为它的概率特性是存在的，只是未能测量不确知而已。这样就可与该概率特性下的极限熵相比较，来确定某种通用编码是否渐近最佳。由此可见，通用编码不但在实用上，而且在理论上都需要进一步探讨。

前面介绍的无失真信源编码适用于离散信源或数字信号，不适用于连续信源或模拟信号，如语音、图像等信号的数字处理。因为连续信源的每个样值所能载荷的信息量是无限大的，而数字信号的值则为有限，对连续信源不引入失真是不可能的。并且，连续信号所对应的信宿一般是人，当失真在某一限度以下时是不易被感觉到的，因此是容许的。连续信源编成代码后就无法无失真地恢复原来的连续值，此时只能根据率失真理论进行限失真编码。限失真信源编码的研究较信道编码和无失真信源编码落后约十年左右。1948 年香农在其论文中已体现出了关于率失真函数的思想。1959 年他发表了“保真度准则下的离散信源编码定理”，首先提出了率失真函数及率失真信源编码定理。率失真信源编码理论是信源编码的核心问题，是频带压缩及数据压缩的理论基础，直到今天它仍是信息论研究的课题。

从率失真函数 $R(D)$ 出发的限失真编码定理虽给出了最佳编码的存在性，也就是在保证平均失真小于允许失真 $D$ 的情况下，最佳码的码率可以压缩到略大于 $R(D)$，但未能给出像概率匹配那样的具体编码途径。限失真编码实际上就是最佳量化问题。最佳标量量化常不能达到率失真函数所规定的值，因此就提出了矢量量化，就是多个信源符号合成一个矢量并对它进行编码。从理论上说，在某些条件下，用矢量量化来编码可达到上述的 $R(D)$ 值，但在实现上还非常困难，有待进一步的研究。1955 年，P. Elias 提出了预测编码方法，经过发展，现在已经成为美国军用通信语言压缩的标准算法。预测编码是利用前几个符号来预测后一个符号的值，预测值与实际值之差，即预测误差作为待编码的符号，这些符号间的相关性就大为减弱，这样可提高压缩比。变换编码是样值空间的变换，例如从时域变到频域，在某情况下，可减弱相关性，取得良好的压缩比。预测编码和变换编码已在实际中有所应用。从理论上说，怎样把有记忆信源转换成无记忆序列，尚无理想的方法，更没有较简单且能实际应用的方法。

以上简述了根据香农信源编码定理发展起来的各种信源编码方法，也就是从概率论形成的信息出发，去掉冗余而达到压缩码率的目的。