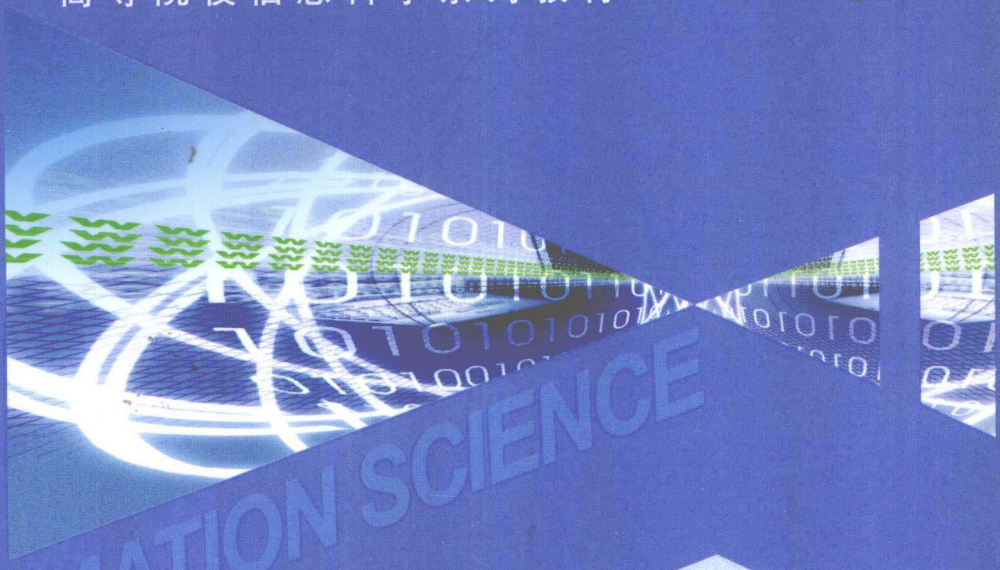
普通高等教育“十一五”国家级规划教材

高等院校信息科学系列教材



信息论与编码理论

(第二版)

科学出版社
www.sciencep.com

普通高等教育“十一五”国家级规划教材

高等院校信息科学系列教材

信息论与编码理论

(第二版)

沈世镒　陈鲁生　编著

科学出版社

北京

内 容 简 介

本书主要介绍信息论和编码理论的基本内容，其特点是简明扼要，可读性强，既具有较严谨的数学描述与推导，又注意到信息论的实用背景，其中许多典型问题已在通信工程中得到实际应用.

全书共分 12 章. 主要内容包括：信息的度量、信源编码、信道编码定理、编码理论中用到的基本抽象代数知识、编码理论的基本概念和基本问题、线性码、Hamming 码、循环码、BCH 码、Reed-Solomon 码、Golay 码、Reed-Muller 码、平方剩余码、Goppa 码以及信息论和编码理论的应用. 本书每章末均附有习题，其中部分习题是对正文内容的补充.

本书可作为高等院校信息科学专业、计算机科学专业、通信专业以及相关专业的本科生教材，也可供相关领域的研究生、教学与科研人员，以及工程技术人员参考.

图书在版编目(CIP)数据

信息论与编码理论/沈世镒，陈鲁生编著. —2版. —北京：科学出版社，2010

(普通高等教育“十一五”国家级规划教材·高等院校信息科学系列教材)

ISBN 978-7-03-029158-5

Ⅰ. ①信… Ⅱ. ①沈… ②陈… 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. ①TN911.2

中国版本图书馆 CIP 数据核字(2010) 第 193478 号

责任编辑：鞠丽娜／责任校对：柏连海

责任印制：吕春珉／封面设计：三函设计

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

铭浩彩色印装有限公司 印刷

科学出版社发行　　各地新华书店经销

*

2002 年 7 月第　一　版　开本：B5(720×1000)

2010 年 10 月第　二　版　印张：16

2010 年 10 月第八次印刷　字数：315 000

印数：21 001—24 000

定价：27.00 元

(如有印装质量问题，我社负责调换〈环伟〉)

销售部电话 010-62134988　　编辑部电话 010-62138978-8002(HI08)

序　言

1998 年教育部进行高校专业调整时设立了“信息与计算科学”专业. 该专业的设立, 受到很多高等院校的热烈响应, 据不完全统计, 几年来已有约 280 所院校招收了该专业的本科生, 其中大部分院校计划开设信息科学方面的系列课程.

为了配合高等院校在学科专业设置上的改革与深化, 来自几十所高等院校的有关专业的部分领导和教师, 于 1999 年、2000 年召开了第一、二届“信息科学专业发展与学术研讨会”, 与会者热烈讨论并探讨了许多与信息学科的学科发展和建设的基本问题. 会议一致认为教材建设是目前最为紧迫的任务, 因此成立了教材编审协调组来组织该系列教材的编写.

2001 年教材编审协调组召集了有多位经验丰富的教师和出版社参加的教材建设会议. 会议明确了教材建设是一项长期的工作, 并决定首先编写和出版这套教材来满足近期急需. 为了保证教材的质量, 会议对每本教材的要求、内容和大纲进行了具体研讨, 并请具有多年教学经验的重点院校教授担任各教材的负责人.

为了贴近教学的实际, 每部教材都配有习题或思考题, 同时对内容也做了结构化安排, 以便教师能根据实际情况部分选讲. 本套教学用书不仅适用于教学, 也可供相关读者参考.

在教材编写和出版过程中, 作者对内容的取舍、章节的安排、结构的设计以及表达方式等方面多方听取意见, 并进行了反复修改. 在感谢作者们辛勤劳作的同时, 编委会还特别感谢科学出版社的鞠丽娜编辑, 她不辞辛劳, 在统筹印刷出版、督促进度、征求意见、组织审校等方面做了大量工作. 这套教材能在保证质量的前提下, 及时与读者见面, 和她的努力是分不开的.

从长远的教学角度考虑, 为了适应不同类型院校、不同要求的课程需要, 教材编审协调组将不断组织教材的修订、编写 (译), 从而使我国信息科学教学用书做到逐步充实、完善、提高和多样化. 在此衷心希望采用本系列用书的教师、学生和读者对书中存在的问题及时提出修改意见和建议.

高等院校信息科学系列教材编委会

2002 年 3 月

第二版前言

本书作为高等院校信息科学系列教材之一已出版 8 年, 在此期间已重印 7 次, 受到国内多所高校师生的欢迎. 本书于 2006 年被列入普通高等教育 "十一五" 国家级规划教材. 借此机会, 我们根据多年来讲授信息论和编码理论的教学实践以及有关反馈信息, 对第一版内容做了一些修订.

随着计算机和通信网络的迅速发展, 信息论和编码理论无论在理论上还是在应用上都得到了长足的发展, 其内容变得越来越丰富. 本书的写作目的是为高等院校信息科学专业或相关专业的本科生提供一本关于信息论和编码理论的教材. 信息论和编码理论的内容非常丰富, 要想在一个学期里深入地讲授是困难的. 因此, 作为一本面向普通高等院校本科生的基础教材, 本书只介绍其基本内容.

多年来, 我们一直在南开大学数学科学学院为信息科学专业的本科生讲授信息论和编码理论. 本书可以作为一学期的信息论和编码理论的教材使用, 也可以作为两门课程的教材使用, 第一部分和第四部分作为信息论的教材, 第二部分和第三部分作为编码理论的教材.

对于本次修订, 我们在内容组织上进行了修改, 以便更适合于作为教材使用, 同时对全书的一些符号以及文字表达进行了修订.

本书第一部分和第四部分由沈世镒负责编写, 第二部分和第三部分由陈鲁生负责编写. 陈鲁生对全书进行了修订.

尽管本书做了修订, 但仍然难免会有不妥之处. 敬请读者批评指正.

陈鲁生

2010 年 6 月

第一版前言

信息和信息科学是两个常见的名词, 其内容十分广泛, 尤其是在当今的信息社会中, 有着十分重要的意义和丰富的含义. 哲学家将信息与物质和能量并列, 作为构成世界的三大要素之一. 信息和信息科学成为推动当今社会文明与发展的主要因素, 信息技术的产业与产品在社会生活中的比重也越来越大. 但是, 到底什么是信息? 什么是信息科学? 信息如何度量? 这些问题一直是人们所关心的. 随着科学的发展, 对上述问题的认识正在逐步深入. 信息论和编码理论可以看成是信息科学产生的一个起点, 有其特殊意义. 我们可以从它所解决的问题去了解它在信息科学中的作用.

在本书中, 我们首先介绍信息的度量问题. 它是信息论的基础. 有了信息的度量, 才使信息论的研究进入定量化. 信息论和编码理论是本书的基本内容, 其中信息论一般指 Shannon 信息论, 它研究的是通信的可行性问题, 而不涉及具体的实现问题. 因此, Shannon 信息论中的编码理论又称为 Shannon 编码理论. 本书中的编码理论是指代数编码理论, 它们主要是利用代数结构来实现信道的纠错编码和检错编码, 相应的编码算法和译码算法都可在计算机上快速实现.

信息论和编码理论是一门理论与应用关系十分密切的学科. 它的应用性很强. 从它的产生背景和发展历史以及应用内容, 都可以看到它与电子、通信、计算机技术的密切相关性, 尤其是与现代网络通信、信息安全技术、多媒体技术密不可分. 在本书中, 我们回顾了它的产生过程、发展历史、理论与应用的概况. 由此我们可以看到它的全貌以及在现代信息技术与产业中的重大作用.

信息论和编码理论又是一门十分严谨的理论学科. 从它的概念定义到问题模型, 再到定理及其证明, 都是用严格的数学语言与逻辑推导完成的. 因此, 又可将该学科看成是数学学科的一个分支. 有如此严格的数学理论, 又有如此巨大的应用背景和成就, 这是本学科的一个重要特点. 我们希望读者在学习本书时注意这个特点.

本书的目的是为高等院校信息科学专业或相关专业的本科生提供一本关于信息论和编码理论的教材. 它是在我们多年讲授的基础上编写而成的. 信息论和编码理论涉及的内容很多, 在本书中我们只介绍它的基本内容.

全书共分 12 章. 第 2 章至第 4 章为 Shannon 信息论部分, 其中第 2 章重点介绍信息度量的来源、意义、性质以及作用.

第 3 章和第 4 章分别介绍信源、信道编码问题和编码定理. 对于 Shannon 信息论, 可将通信问题与普通的运输问题做类比. 前者是传送信息, 后者是传送货物.

传送货物有货源和运输通道, 而货源和运输通道又有货源的体积 (或吨位) 和运输通道的容量. 当运输通道的容量大于货源的体积 (或吨位) 时, 就能实现货物的正常运输. 信源和信道编码理论则是解决信息的传输问题. 可以通过信息的度量来确定信源的信息量 (可形象地称为信号体积) 和信道容量. 当信道容量大于信号体积时, 就能实现信息的正确传输.

第 5 章至第 11 章为代数编码理论部分. 其中第 5 章主要介绍编码理论中用到的基本代数知识, 第 6 章介绍编码理论的基本概念和基本问题, 第 7 章至第 9 章分别讨论线性码和 Hamming 码以及循环码, 第 10 章和第 11 章分别介绍 BCH 码和 Reed-Solomon 码以及其他一些重要的线性码. 这部分的内容是用代数结构来构造纠错码和检错码, 并讨论它们的译码算法. 因此, 这部分内容与 Shannon 信息论不同, 它不刻意追求实现 Shannon 信息论中的最佳通信目标, 而是强调它们的可实现性 (在通信中快速完成编码和译码计算, 达到与通信实时同步完成). 由此可见, Shannon 信息论和代数编码理论虽都是通信理论的组成部分, 但它们的具体目标和实现方法并不完全相同.

本书的第 1 章和第 12 章分别介绍信息论和编码理论的概况与应用. 其中第 1 章介绍信息论和编码理论的发展历史和意义, 还介绍通信问题的基本要素, 以及对这些要素的数学模型表述与记号. 第 12 章简单介绍信息论和编码理论的几个应用问题, 其中包括有失真编码问题和数据压缩问题, 卷积码理论及其应用, 汉字和图形信息的编码问题等. 有失真编码问题是信源编码领域的一个重要发展与应用方向, 它是本书第 3 章中信源编码问题的继续. 第 3 章的信源编码理论只研究无失真的信源编码问题. 卷积码理论是代数编码理论的一个重要方面, 在通信工程中有许多应用. 汉字和图形信息的编码是两种应用范围很广的编码技术, 学习信息论时应当有所了解. 第 12 章中的应用问题我们只做简单介绍, 深入学习与研究还需参阅其他的著作或文献. 通过学习, 我们希望使读者可以对信息论和编码理论的全貌与应用背景有更多的了解, 而不局限于几个具体的编码定理之中.

本书可作为数学类信息与计算科学专业本科生的基础教材, 也可作为电子、通信、计算机等有关学科方向的本科生、研究生的参考教材以及这些专业的教学、研究与工程技术人员的参考书.

在本书的编写过程中, 岳庭海博士和常祖领博士对本书的初稿完成做了大量的工作. 符方伟教授对本书内容的编排也提出了宝贵的建议. 在此一并致谢.

由于时间仓促, 书中难免有错误和不当之处, 敬请读者批评指正.

沈世镒

2002 年 5 月

目　录

第一部分　信息论

第四部分　信息论和编码理论的应用

第1章 引 言

信息和信息科学是两个常见的名词, 它们的内容十分广泛, 尤其是在当今的信息社会中, 有着十分重要的意义与丰富的含义. 哲学家将信息与物质、能量相并列, 作为构成世界的三大要素之一. 信息已成为推动当今社会文明与发展的主要因素. 信息技术的产业和产品在社会生产和生活中的比重越来越大.

在本章中, 我们将介绍信息论和编码理论的基本情况, 其中包括信息论和编码理论的发展过程以及重要的应用领域, 同时还将介绍信息度量的基本要求和通信系统的数学模型. 通过这些介绍, 可使读者对信息论和编码理论的全貌有一个基本的了解.

1.1 信息论的发展概况

信息的概念是一个普遍的概念. 从细胞分裂到植物、动物的遗传与进化, 从语言文字到动作表情都有信息表达与传递. 但是信息论的产生和发展与通信、计算机技术的产生和发展密切相关. 回顾它的历史, 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. 我们对此做一个简单的回顾.

1.1.1 信息论的早期酝酿

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标. 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. 在 20 世纪初, 信息论开始进入早期酝酿期. 为了提高通信的质量与效率, 人们往往从物理和数学两个不同的方面来考虑. 在物理方面, 主要研究与改进通信的物理手段与条件, 如通信手段 (有线、无线通信) 的采用, 发射与接收设备的改进, 波段的选择与信噪比的提高等. 在物理和机电技术改进的同时, 人们发现数学理论和工具的使用变得十分重要, 许多问题甚至是根本原理性的. 对通信中的许多问题, 如果没有数学的描述就无法说明. 从 19 世纪到 20 世纪 40 年代, 信息论的一些基本问题开始形成.

1. 早期编码问题

在有线和无线电通信产生的同时, 编码技术随之产生. 早期的编码有 Morse 码和 Bodo 码等. 它们将文字通过点、划、空等信号给以表达. 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, Morse 码和 Bodo 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, 再用电码发射.

2. 通信的有效性与可靠性

随着通信距离的加大, 出现了信号强度的衰减与噪声干扰的问题. 因此, 如何克服噪声干扰问题就成为通信技术中的一个迫切问题. 为解决这些问题, 人们就对通信中的各种因素进行分析. 结果发现, 频带的加宽可以提高通信效率, 而且它的作用较信号强度更为重要. 又发现在通信技术中, 通信的数量与质量存在相互制约的关系. 如果牺牲通信的数量, 则可以达到提高质量的目的. 这种概念虽符合人们的日常生活经验 (说话的多次重复可让对方听得更清楚), 但在理论上应如何解释还是无法说明, 因为那时对信息没有度量化, 所以不能做定量化的说明.

到了 20 世纪 20 年代, H. Nyquist 和 L. Hartley 提出了解决以上问题的一系列的讨论, 如信息传递的速率与带宽成正比, 信息的度量与信号的概率分布、对数函数有关等. 这些思想为以后 Shannon 信息论的建立打下了基础.

3. Shannon 熵的其他产生途径

至 20 世纪 40 年代, 控制论的奠基人 N. Wiener 和美国统计学家 E. Fisher 以及 C. E. Shannon 几乎同时提出了对信息的一种度量, 就是熵的定义形式. 至此, 信息论的前期酝酿已经成熟. 信息的度量有广泛的背景, 从不同的学科方向都可以导致它的确立.

4. 纠错与检错码的产生

在信息论产生的过程中, 纠错与检错码的概念也在逐步形成. 人们发现, 由点、划、空等信号所构成的码可形成一定的结构, 由这些结构可产生抗干扰能力. 抗干扰的概念实际上就是一种码的纠错能力.

1.1.2 Shannon 信息论的建立与发展

信息论的产生以 1948 年 C. E. Shannon 发表的 "通信的数学理论" 这一奠基性论文为起始, 至今已有 60 多年的历史了. 在这 60 多年中, 电子、通信以及计算机技术、产业和市场都经历了空前大规模的发展, 信息技术的产品进入了千家万户, 成为工作、学习以及生活中不可缺少的组成部分. 可以毫不夸张地说, 信息论在这场空前的技术革命的许多问题中起到了理论基础、思想先导以及技术关键的作用. 现代的快速通信、多媒体和网络技术无不受益于信息论和编码理论以及它们的相关学科 (如信号处理等) 的发展. 这些发展推动了信息技术的革命, 也丰富了信息论的内容. 对此我们下面还要进行详细说明.

自 1948 年 Shannon 理论产生以后, 其发展大体上经历了理论的确立与发展、理论的应用与近代发展等阶段. 为了使读者了解它的情况, 我们对这些阶段的主要

内容和特点做一简单介绍.

1. Shannon 信息论的确立期

我们把自 1948 年到 20 世纪 60 年代这一时期称为理论的确立期, 其主要特点是对 Shannon 理论的研究与说明, 其中包括对通信系统的数学模型和基本问题的说明与讨论. 其中的主要问题包括对信息量、Shannon 熵的来源、意义与作用的讨论; 关于通信基本问题的讨论; 信源和信道编码问题的模型、本质问题与意义的讨论; 信源和信道编码的编码定理及其证明; 信源和信道编码的实现与应用等问题.

这一时期完成的主要标志是对以上问题实现了严格的数学描述与论证. 从信息的度量到通信模型, 从编码问题到主要编码定理的证明都是在严格的数学定义与证明中完成. 另外, 一系列专著的完成也是 Shannon 信息论确立的标志. 当时 (20 世纪 70 年代前后) 的重要著作有 B. McMillan, A. Feinstein(1958), Robert G. Gallager(1968), 以及 J. Wolfowitz(1978) 等人的著作. 这些著作基本上完成了对 Shannon 理论的阐明, 从理论上解答了通信中提出的一些问题.

在这一时期, 前苏联学者, 如 A. N. Kolomogolov, A. Y. Xinqing, M. S. Pinsker, 以及 R. L. Dobrushin 等人都有重要的工作与著作, 其中重要的工作有 A. N. Kolomogolov 和 M. S. Pinsker 关于信息量的研究, A. Y. Xinqing 关于有限记忆信道的编码理论研究, R. L. Dobrushin 的一般通信系统的研究理论等, 这些工作都是信息论早期的基础性工作.

国内学者, 如王寿仁、江泽培、胡国定、蔡长年、章照止、沈世镒等人也有重要的工作与著作, 他们为信息论的早期发展与国内的引进做出了重要贡献. 胡国定先生关于信息量的集合化理论, 信源与信道编码的基本定理等成果已成为信息论研究的重要基础. 章照止在互信息公理化与卷积码分析, 沈世镒在 Shannon 第一和第二定理的关系与有限记忆信道编码定理等问题上做了工作.

在 Shannon 信息论发展变化的同时, 代数编码理论也得到迅速发展. 人们发现, 利用群、环、域以及线性子空间理论可将码赋予一定的代数结构. 这种结构可使通信信号具有纠错与检错的能力, 并称这种码为纠错与检错码. 在这一时期, 纠错与检错码已开始进入实用化的通信技术领域. 代数码的重要经典著作有 W. W. Peterson(1961) 和 Robert M. Fano(1961) 等人的著作. 另外, 由 Wozencraft-Reiffen(1961) 提出的序贯译码理论为以后的卷积码和它的译码算法奠定了基础.

国内万哲先等人研究与发展的代数码与移位寄存器理论也有专著问世, 他们在代数码与移位寄存器理论研究中做了许多工作.

2. Shannon 信息论的发展

自 20 世纪 70 到 80 年代, 信息论处于理论发展期. 由于 Shannon 理论的阐明

与通信技术的发展, 信息论的研究范围日益扩大. 这一时期发展的主要内容在 "率失真理论" 与 "多用户信息论" 方面. 率失真理论实际上是一种在允许误差下的信源编码理论. 该理论到 20 世纪 80、90 年代成为数据压缩技术的理论基础. 多用户信息论的最早思路是由 Shannon 提出的, 到 20 世纪 70、80 年代得到迅速发展, 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相符合.

1.1.3 信息论的近期发展

信息论近期发展的主要特点是向多学科结合的方向发展, 其重要的发展方向有以下几种.

1. 信息论与密码学

信息论与密码理论和技术相结合是信息论的一个发展方向. 通信中的安全和保密问题是通信编码问题的又一种表示形式. 由 Shannon 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, 如加密运算的完善性、剩余度等指标都与信息量密切相关.

2. 算法信息论与分形数学

由于 Shannon 熵、Kolomogolov 复杂度, 以及 Hausdorff 维数的等价性在理论上已得到证明, 从而使信息论与计算机科学以及分形理论找到了它们的汇合点. 人们发现 Shannon 熵和 Kolomogolov 复杂度以及 Hausdorff 维数都是某种对事物复杂性的度量, 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、计算机程序复杂度以及分形理论, 在本质上有共同之处. 它们相结合后产生了一些新的学科方向, 如算法信息论就是信息论与计算复杂性理论相结合的新学科方向.

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, 信息的度量为非线性问题的研究提供了工具, 如用交互信息来取代统计中的相关系数更能反映两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的出现 (如金融、股票数据、生物数据等), 使许多计算问题成为可能. 因此信息量的作用在统计中必将发挥更大的作用. 下面列出一些信息论与统计理论相结合的典型工作.

(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, 如在

由概率分布族所组成的微分流形中, Fisher 信息矩阵是 Kullback-Laiber 熵的偏微分. 由此关系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 算法等都与此有关. 日本的 S. Amari 教授、香港中文大学的徐雷教授等都对此做了大量的工作, 并有专著问世.

(2) 信息计算与组合投资决策关系密切. T. Cover 教授将组合投资决策问题化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和收敛性.

(3) 编码理论在与试验设计和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和方法, 这些码在一定意义下具有正交性. 因此, 这些码可直接用来设计和构造试验设计表. 另外, 信息编码定理在证明假设检验中的两类误差的指数下降性和估计下降速度的过程中发挥了重要作用.

1.1.4　信息论在信息技术领域中的应用

自从 Shannon 信息论和编码理论产生以来, 随着电子和通信以及计算机的发展, 信息论和编码理论的研究成果不断得到应用.

1. 编码技术在快速通信领域中的应用

20 世纪 70、80 年代的编码理论在快速通信技术中得到了大量应用. 当时的通信技术正在从低速向高速过渡, 通信手段正向微波和卫星等方向发展. 因此, 误差干扰问题就突现出来. 利用纠错码可大大降低通信中的差错率. 当时的代数码, 如 BCH 码、Reed-Solomon 码等为克服误差干扰发挥了重要作用, 成为通信工程中不可缺少的一个组成部分. 另外, 卷积码理论当时也有重要的发展, 如卷积码的 Viterbi 译码算法改造了新一代卫星的通信技术, 成为当时信息论和编码理论在工程技术应用中的一个光辉典范.

2. 在调制解调码技术中的应用

20 世纪 80、90 年代调制解调码理论和数据压缩理论在多媒体技术领域中的应用是信息论和编码理论应用的两项重要成果. G. Ungerbock 等人在 1982 年利用格子码和软判决理论对 Gaussian 信道给出了调制解调码的结构和编码译码算法. 调制解调码的出现, 从根本上改变了数据通信的状况, 使调制解调码的通信速度从原来的 1200bit/s 逐步提高到 30000bit/s. 数据传输速度提高了近 30 倍, 从而使现有的网络通信成为实用性的技术.

3. 数据压缩理论和技术

数据压缩理论分为无失真压缩和有失真压缩两大部分. 它们有各自的理论基础和应用范围. 无失真压缩理论和技术与数据存储密切相关, 目前在计算机数据和