

高等院校信息科学系列教材

信息论与编码理论

沈世镒 陈鲁生 编著



科学出版社

高等院校信息科学系列教材

信息论与编码理论

沈世镒 陈鲁生 编著

科学出版社

2002

内 容 简 介

本书主要介绍了信息论与编码理论的基本内容,其特点是具有较严谨的数学描述与推导,同时注意到信息论的实用背景,其中许多典型问题在通信工程中已得到实际应用.全书共12章,主要内容包括:信息的度量和信源、信道编码问题与编码定理,编码理论中用到的基本代数知识,编码理论的基本概念和基本问题,线性码、汉明码以及循环码、BCH码、Reed-Solomon码及其他的一些重要的线性码,以及信息与编码理论的应用问题.本书每章末均附有习题,其中有些习题是对正文内容的补充,以供学生复习巩固书中所学内容.

本书可作为高等院校信息科学专业及相关专业的本科生教材,也可供相关领域中的研究生、教学与科研人员和工程技术人员参考.

图书在版编目(CIP)数据

信息论与编码理论/沈世镛,陈鲁生编著. —北京:科学出版社,2002
(高等院校信息科学系列教材)
ISBN 7-03-010496-X

I. 信… II. ①沈…②陈… III. ①信息论-教材②编码理论-教材
IV. ①G201②O157.4

中国版本图书馆CIP数据核字(2002)第037504号

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>

新蕾印刷厂 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)

2002年7月第一次印刷 印张:14

印数:1—5 000 字数:270 000

定价:19.00元

(如有印装质量问题,我社负责调换〈路通〉)

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年教材编写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.在此衷心希望采用该系列用书的教师、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

前 言

信息与信息科学是两个常见的名词,它们的内容十分广泛,尤其在当今的信息社会中,有着十分重要的意义与丰富的含义.哲学家把信息与物质、能量并列,作为构成世界的三大要素之一,并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产与生活中的比重越来越大.但是到底什么是信息?什么是信息科学?信息如何度量等问题一直是人们所关心的.随着科学的发展对以上问题的认识正在逐步深入.信息论与编码理论可以看成是信息科学产生的一个起点,有它的特殊意义,我们可从它所解决的一类编码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,如它的来源、意义与作用等问题,它是以后编码理论的基础,有了信息的度量才使信息论的研究进入定量化.信息论与编码理论是本书的基本内容,其中信息论一般指香农信息论,它是通信中的可行性问题,而不涉及它们的具体实现问题,因此在香农信息论中的编码理论又称为香农编码理论.本书中的编码理论是指代数编码理论,它们主要利用代数结构实现纠错编码与检错编码,有关的编码结构与译码算法都可在计算机上快速实现.

信息论与编码理论是一门理论与应用关系十分密切的学科.首先它的应用性很强,从它的产生背景、发展与应用内容来看,都可以看到它与电子、通信、计算机技术的发展密切相关,并得到一系列的重要应用.尤其是与近代网络通信、数据加密与安全技术、多媒体技术密不可分.在本书中我们回顾它的产生过程、发展历史、理论与应用的概况,由此我们可以看到它的全貌与在近代信息技术与产业中的重大作用.

信息论与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,再到定理与它的证明都是用严格的数学语言与逻辑推导完成,因此又可将该学科看成是一门数学学科分支.有如此严格的数学理论又有如此巨大的应用背景与成就就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.

本书是高等院校信息科学专业及相关专业本科生的教材,它是在我们多年讲授该课的基础上编写而成的.由于信息论和编码理论涉及的内容很多,在此我们只介绍它的基本内容.

全书共12章.第2章至第4章为香农信息论部分,其中第2章重点介绍信息度量的来源、意义、性质与作用.

第3、4章分别介绍信源、信道编码问题与编码定理.由香农信息论可把通信问题与普通的运输问题相比较.前者是转送信息,后者是转送货物,转送货物有货源、运输通道,而货源与运输通道又有货源的体积(或吨位)与通道的容量,当通道的容

量大于货源的体积(或吨位)时就能实现货物的正常运输. 而信源、信道编码理论则是解决信息的传输问题, 它可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)与信道容量, 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汉明码以及循环码, 第10、11章介绍 BCH 码和 Reed-Solomon 码及其他的一些重要的线性码. 这部分的主要内容是用代数结构来构造纠错与检错码与它们的译码算法. 因此这部分内容与香农信息论不同, 它不刻意追求实现香农信息论中的最佳通信目标, 而强调它们的可实现性(在通信中快速完成编、译码计算, 达到与通信实时、同步完成). 由此可见, 香农信息论与代数编码理论虽都是通信理论的组成部分, 但它们的具体目标与实现方法手段并不相同.

本书的第1章与第12章分别简单介绍了信息与编码理论的概况与应用. 其中第1章介绍了信息论的发展历史、意义及信息论与编码理论的产生、发展与应用过程, 还介绍了通信问题的基本要素, 及对这些要素的数学模型表述与记号. 第12章简单介绍了信息与编码的几个应用问题, 其中包括有失真编码、问题与数据压缩问题、卷积码理论及其应用、汉字与图形的信息编码问题等. 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章信源编码问题的继续, 在第3章的信源编码理论中只研究无失真的信源编码问题. 卷积码理论是代数码理论的又一重要方面, 同样在通信工程中有许多应用. 而汉字与图形的信息编码是两种应用范围很广的编码技术, 学习信息论时应当了解. 第12章中的各应用问题我们只做简单介绍, 深入学习与研究还需参考其他的著作或文献. 通过学习, 我们希望使读者可以对信息论与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、通信与计算机专业有关学科方向本科生、研究生的参考教材及这些专业的教学、研究与工程技术人员的参考书.

在本书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工作; 符方伟教授对本书内容编排方面提出了宝贵的建议, 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.

作者

2002年5月

目 录

第 1 章 概论	1
1.1 信息论的发展概况	1
1.2 信息与编码理论的主要内容	6
1.3 本书内容简介与文献介绍.....	13
习题 1	16

第一部分 信息论

第 2 章 信息量	18
2.1 熵.....	18
2.2 联合熵和条件熵.....	23
2.3 熵的基本性质.....	27
2.4 互熵与互信息.....	31
2.5 凸函数及其应用.....	33
2.6 连续型随机变量的信息量.....	36
2.7 最大熵原理.....	40
习题 2	44
第 3 章 信源编码	48
3.1 信源编码问题.....	48
3.2 前缀码与即时码.....	52
3.3 信源变长码的编码定理.....	56
3.4 霍夫曼(Huffman)信源编码算法	59
3.5 霍夫曼信源编码性能分析.....	63
3.6 信源定长码的编码定理.....	68
习题 3	72
第 4 章 信道编码定理	74
4.1 信道编码问题.....	74
4.2 离散无记忆信道.....	76
4.3 无记忆信道的信道容量.....	81
4.4 信道容量的计算.....	86
4.5 信道的编码与译码问题.....	94
4.6 信道的正、反编码定理	97

4.7 可加高斯(Gaussian)信道	103
习题 4	106

第二部分 编码理论

第 5 章 抽象代数的基本知识	108
5.1 群	108
5.2 环和域	111
5.3 理想和商环	113
5.4 域上的多项式	114
5.5 有限域	119
5.6 域上的线性代数	122
习题 5	126
第 6 章 编码理论的基本知识	128
6.1 代数码理论的基本特点	128
6.2 码的基本定义与纠错、检错能力	128
6.3 编码理论的基本问题	133
习题 6	139
第 7 章 线性码	141
7.1 线性码的定义	141
7.2 线性码的对偶码	142
7.3 线性码的译码方法	145
7.4 线性码的重量分布	148
习题 7	151
第 8 章 汉明码	154
8.1 汉明码的定义	154
8.2 汉明码的性质	155
8.3 汉明码的译码方法	156
8.4 极长码	158
习题 8	160
第 9 章 循环码	161
9.1 循环码的定义及其基本性质	161
9.2 循环码的校验矩阵及其对偶码	165
9.3 循环码的编码方法	168
9.4 循环码的检错性能	169
习题 9	170

第 10 章 BCH 码和 Reed-Solomon 码	172
10.1 BCH 码及其基本性质	172
10.2 Reed-Solomon 码及其基本性质	175
10.3 BCH 码和 Reed-Solomon 码的译码方法	176
10.4 Reed-Solomon 码和最大距离可分码的重量分布	179
习题 10	180
第 11 章 一些重要的线性码	182
11.1 Golay 码	182
11.2 Reed-Muller 码	183
11.3 平方剩余码	186
11.4 Goppa 码	187
习题 11	188
第 12 章 若干应用问题	190
12.1 有失真的数据压缩问题	190
12.2 卷积码理论及其应用	199
12.3 几种实用的编码问题	206
习题 12	210
主要参考文献	212

第 1 章 概 论

1.1 信息论的发展概况

在本章中我们首先介绍信息论与编码理论的基本情况,其中包括信息论与编码理论的产生、发展、理论的形成过程、重要的应用领域与问题,同时还要介绍信息度量的基本要求与通信系统的基本要素及它们的数学模型.通过这些介绍可使读者对信息论与信息编码理论的全貌有一个基本了解.

1.1.1 信息论的产生与发展

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、发展与通信、计算机技术的产生、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的几个阶段与特征,我们先对此作一个简单的回顾.

1.1.2 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的要求.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,人们往往从物理与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,如不同通信手段(如有线、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.但在物理、机电技术改进的同时,人们发现数学理论与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、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加大问题,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关系,如牺牲通信的数量可以达到提高质量的目的等.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,因此不能作定量化的说明.

到 20 世纪 20 年代,奈奎斯特(H. Nyquist)与哈特莱(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、对数函数等有关等.这些思想为以后香农(C. H. Shannon)信息论的建立打下基础.

3. 香农熵的其他产生途径

至 20 世纪 40 年代,控制论的奠基人维纳(N. Wiener)和美国统计学家费希尔(E. Fisher)与香农同时提出信息的一种度量——熵的定义形式.由此说明,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建立信息的度量.

4. 纠错与检错码的产生

与此同时,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.

1.1.3 香农信息论的建立与发展

信息论的产生以 1948 年香农的奠基性论文为起始,至今已有 50 多年的历史了.在这 50 多年中,电子、通信与计算机技术、产业与市场经历了空前的大规模的发展,信息技术的产品进入千家万户,成为工作、学习与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、思想先导与技术关键性的作用.现代的快速通信、多媒体与网络技术无不受益于信息与编码理论及它们的相关学科(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,对此我们在下面还要详细说明.

自 1948 年香农理论产生以后,它的发展大体经历了理论的确立与发展、理论的应用与近代发展阶段,为了使读者了解它的情况,下面我们对这些阶段的主要内容与特点做简单介绍.

1. 香农信息论的确立期

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

这一阶段工作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的.另外,一系列专著的完成也是香农信息论确立的标志.当时(70年代前后)的重要著作有B. McMillan, A. Feinstein(1954), Robert G. Gallager(1968), J. Wolfowitz(1978)等人的论著.这些著作基本上完成了对香农理论的阐明,在理论上解答了通信问题中提出的这些问题.

在这一时期,前苏联学者,如柯莫各洛夫(А. Н. Колмогоров),欣钦(А. Я. Хинчин),宾斯基(М. С. Пинскер),达布罗辛(Р. Л. Добрушин)等人有重要的工作,如:柯莫各洛夫和宾斯基关于信息量的研究,欣钦关于有限记忆信道的编码理论研究,达布罗辛的一般通信系统的研究理论等,这些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,沈世镒在香农第一、第二定理的关系与有限记忆信道编码定理等问题做了工作.

与香农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、域与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), Robert M. Fano(1961)等人的.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. 香农信息论的发展

20世纪70到80年代,信息论处于理论发展期.由于香农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有允许误差下的信源编码理论,该理论到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

由香农提出的,到 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符合.

1.1.4 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方向发展.其重要的发展方向有如下几种:

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,如加密运算中的完全性、剩余度等指标与信息量密切相关.

2. 算法信息论与分形数学

由于香农熵、柯莫各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、计算机科学与分形理论找到了它们的汇合点.人们发现香农熵、柯莫各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形理论,在本质上有共同之处.它们的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融合的新学科.

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,如用交互信息来取代统计中的相关系数,更能反映二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工作如下:

(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 信息矩阵是 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.日本的 S. Amari 教授、香港中文大学徐雷教授对此做了大量工作,并有专著问世.

(2) 信息计算与组合投资决策关系密切,T. Cover 教授把组合投资决策问题化成一个信息论的问题,并在最优决策的计算中给出了一个渐近递推算算法,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与构造试验设计表。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,并给出了这两类误差的下降速度。

1.1.5 信息论在 IT 领域中的应用

自从香农信息论与编码理论产生以来,随着电子、通信与计算机的发展,信息与编码理论的研究成果不断得到应用,典型的事例如下。

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大量应用,当时的通信技术正在从低速向高速发展,通信手段正向微波、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、R-S 码等为克服误差干扰发挥了重要作用,成为通信工程不可缺少的一个组成部分。另外,卷积码理论当时也有重要发展,如卷积码的 Viterbi 译码算法产生了新一代卫星的通信技术,成为当时信息与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 年代信息编码理论应用的两项重要成果是:调制解调码理论及数据压缩理论在多媒体技术领域中的应用。由 G. Ungerboeck 等人在 1982 年利用格子码与软判决理论,对高斯信道给出了调制解调码的结构与编、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,使调制解调码通信速度从原来的 1200bit/s 逐步增加到 30000bit/s。数据传输速度提高了 25 倍。从而使现有的网络通信成为实用性的技术。

3. 数据压缩理论与技术

数据压缩理论分无失真压缩与有失真压缩理论两大部分,它们有各自的理论基础与应用范围。无失真压缩理论与技术和数据存储密切相关,目前在计算机数据与文件的存储中得到广泛应用,成为计算机软件技术中不可缺少的一个组成部分。它又是无失真压缩理论与技术的基础。

有失真压缩理论在 20 世纪 70 年代就已提出并得到发展 T. Berger 的专著出版是该理论成熟的标志。但数据压缩技术的大量应用是在 90 年代,由于多媒体技术的需要,在综合无失真数据压缩技术、有失真压缩理论与信号处理技术的基础上,形成实用化的有失真压缩技术,该技术在不太影响人的视觉、听觉效果的前提下大大压缩了通信与存储中的数据量,如动态、彩色图像的数据压缩率达到了

50 : 1—60 : 1, 有了这样的压缩率, 多媒体技术才真正具有实用意义. 90 年代, 多媒体技术已成为 IT 产业中的一个重要组成部分, 其信息处理技术已实现标准化, 如 JPEG (Joint Photographic Experts Group), MPEG 是静态与动态图像数据压缩的技术标准.

信息与编码理论在近代工程技术中的应用还有许多方面, 如在存储介质中的信息处理问题等, 对此我们就不一一列举了.

由以上的讨论我们可以见到, 信息与编码的理论发展在工程技术中应用的重要意义与基本情况, 随着 21 世纪的来临, 信息与信息处理的理论与技术问题会大量涌出.

目前可以看到的有网络技术、量子领域、生物技术中的信息处理问题, 其中有的问题已经提出, 有的问题则刚刚提出, 这正是信息与编码理论今后发展的新起点.

4. 编码技术在其他领域中的应用

在信息与编码的应用问题中, 除了在电子、计算机与通信工程中有大量应用之外, 值得注意的应用问题还有很多, 如汉字信息处理问题与条码、磁码技术等.

汉字信息处理包括汉字信息处理的基础与应用, 其中基础研究包括: 汉字的音、形结构, 如拼音规则、部件与笔画结构; 关于汉字使用的频率统计; 汉字的编码方式等规定有的已成为国家标准或部门标准, 因此有关汉字信息处理的各种技术都要按此标准或规定执行.

汉字信息处理的应用十分广泛, 如汉字输入法、汉字印刷技术、汉字语音与图像识别技术等, 现有的汉字输入法有上百种, 它们利用汉字的音、形结构与编码结合形成各种不同类型的输入法, 目前汉字输入法的发展趋势是向联想型、智能化方向发展. 汉字印刷技术的关键是对汉字图像的设计与压缩技术, 数千汉字有数十种字体和数十种大小型号, 如果没有合理的数据压缩技术就不可能有高性能的汉字印刷技术. 汉字语音与图像识别技术有着十分广泛的应用, 如语音、图像自动记录、识别、输入等技术, 都与信息和编码理论有关, 也涉及到其他领域, 如信号处理、计算机软件与光、电技术等.

关于条码、磁码与 IC 卡技术就是把各种不同类型的信息用条码、磁码与 IC 卡进行存储, 它们的特点是可以利用光、电、磁的扫描技术实现快速阅读, 因此在物流管理、商品管理、电子货币等领域有大量应用. 值得注意的是在条码、磁码的结构中普遍采用了纠错码技术与数据加密技术.

1.2 信息与编码理论的主要内容

信息与编码理论的主要内容有信息的度量问题、通信系统与编码问题、信源与信道编码、代数码理论等, 我们对此做概要介绍.

1.2.1 信息的度量问题

什么是信息? 信息能否度量? 怎样度量? 这些是信息论所必须回答的问题. 在回答这个问题之前必须解决两个前提, 其一就是如何引进一种新的概念或新的度量, 从科学发展的历史来看, 这种新概念或新的度量的引进必须符合两个条件, 第一个条件是一种新概念或度量的引进必须大体合理, 这种新概念所反映的问题与我们以往所了解的特征大体相符, 第二个条件是这种新的概念最后可以解决什么问题. 这两个条件缺一不可, 它们相互呼应与补充. 因此我们既要在这种度量的引进来理解它的含义, 也要从它最终所解决的问题来加深并确定它的本质含义. 香农熵的引进正是从这两个条件出发, 它首先把信息的度量看作事物不肯定性的度量来引进, 同时在编码理论中发挥了重大作用, 只有在信息度量确定后, 通信的编码理论才得到了量化的描述, 正因为香农在 1948 年所发表的论文中同时解决了这两个问题, 所以人们才会把这篇论文作为信息论的奠基性论文.

需解决的第二个问题是: 既然信息是一个具有广泛含义的概念, 那么它的度量与物理量一样, 不可能通过一种量来确定它的全部度量, 在物理量中有速度、加速度、作用力、电量、电磁强度等不同的量, 它们从各种不同的角度来刻画这些物体的运动变化的定量特征. 因此在信息量中也有同样的情形, 在信息论中我们将介绍它们的多种度量形式, 对这种不同类型的度量我们不仅要从它们的引进定义来理解它们的意义, 而且也要从它们能解决的问题上进一步加深对它们的理解.

香农熵是香农信息论中信息的度量基础, 它与概率分布相联系, 以“不肯定性”作为它度量的基础, 对此我们在下面还要详细介绍. 由香农熵可引进联合熵、条件熵、互信息、信道容量、率失真函数与互熵等, 我们都把它们作为信息度量的各种不同形式. 这里需要特别指出的一点是香农熵虽以概率分布的不肯定性作为度量基础, 但是随着信息科学的不断发展, 人们对香农熵的理解也在日益加深与扩大, 在上面我们已经指出, 香农熵将作为事物的复杂度的普遍意义来理解, 并随着信息科学的发展, 新的信息度量与新的学科分支一定还会产生.

1.2.2 通信系统的基本模型与要素

一个通信系统由以下基本要素组成.

1. 信源: 产生信息的来源

如一个电话通信系统, 那么打电话的用户与内容就是这个系统的信源. 由信源产生的信息一般称为“消息”.

2. 信道: 传递信息的通道

如这个电话通信系统是一个有线电话系统, 那么连接两地的电话线与有关的

发送、接收设备就是信道。

由信道传送信息的载体称为信号。消息与信号虽都是信息的载体,但它们的表现形式不同。消息是指一般人们所了解的语言、文字等信息载体,而信号则是较为专业或特殊的信息符号。信道的传送信号又分输入、输出信号,它们分别是信道两端的发送与接收信号。信道的基本特征除了传送信号之外,还有干扰存在,它使信号在传送过程中产生误差。

3. 编码:由消息变为信号的运算

在通信过程中,消息一般都不能直接传送,而是要通过一定形式的电磁信号来传送。由消息变成信号的运算称为编码。

4. 译码:由信号还原成消息的运算称为译码

当信号被接收后还需要还原成消息,由接收信号还原成消息的运算称为译码运算或译码。

5. 通信系统与用户

一个通信系统由信源、信道、编码和译码四要素组成。通信系统的使用者为用户,用户分信源消息的发送者与接收者两种。

使输入信号产生变异或差错的因素为干扰,干扰的来源为干扰源。干扰也是通信系统的组成部分,它可以多种数学形式表示。

由此我们把通信系统归结为信源、信道、翻码、译码四要素。编码与译码合称编码,在国外文献中对编码、译码的用词有明显的区别,它们分别用 Incoding, Decoding 表示,在国内的有些信息论文献中把编码称为翻码。

1.2.3 通信系统中信息的传递过程

通信的传递过程可归结如下。首先由信源发出的消息,在信息论中把这种消息又称为原始消息。由编码将原始消息变为信号,并进入信道成为信道的输入信号(简称输入信号,或入口信号)。输入信号经信道的传送,到达通信的另一端,形成输出信号(或出口信号)。输出信号经译码运算把输出信号变为消息,这种消息是原始消息的还原,所以又称为还原消息。还原消息最终由接收者接收,实现通信的目标。

在信道中,信号的产生、发射、传送与接收实际上是一个很复杂的过程,如果输入是脉冲信号,那么它还要变成具有不同振幅的电磁波才能发射。因此,信道终端最初接收到的信号是电磁振荡信号,再由电磁振荡信号变为信道输出的脉冲信号,由译码运算把输出信号还原成消息。

由脉冲信号变为电磁振荡信号,再由电磁振荡信号变为信道输出的脉冲信号,在通信工程中称为调制解调,它涉及信号处理与电磁波理论,在本书中不做讨论。