



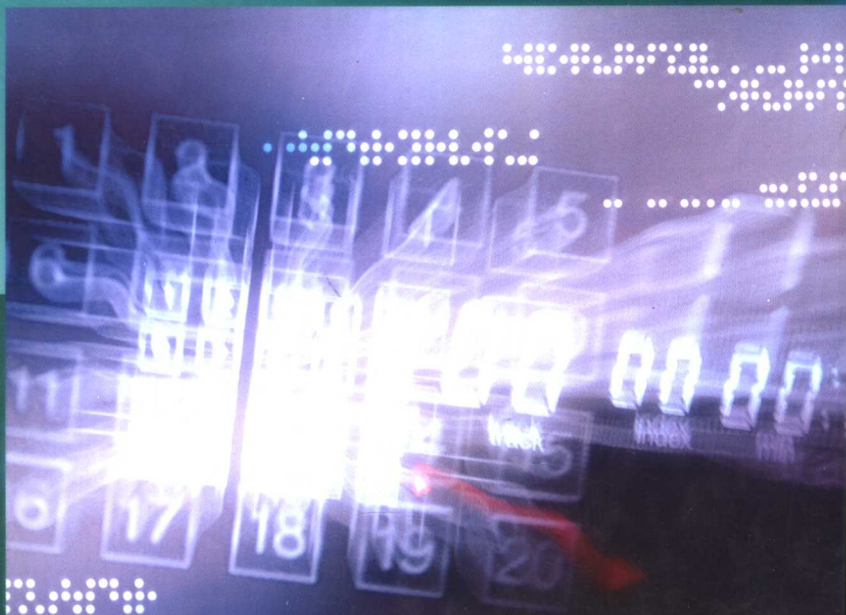
● 普通高等学校信息与计算科学专业系列丛书



教育科学“十五”国家规划课题研究成果

信息论 基础与应用

沈世镒 吴忠华 编著



高等教育出版社
HIGHER EDUCATION PRESS

0110
日地同寺 学校信息与计算科学专业系列丛书

教育科学“十五”国家规划课题研究成果

信息论基础与应用

沈世镒 吴忠华 编著



高等教育出版社
HIGHER EDUCATION PRESS

图书在版编目(CIP)数据

信息论基础与应用/沈世镒,吴忠华编著. —北京:
高等教育出版社, 2004.7(2005重印)
ISBN 7-04-014374-7

I.信... II.①沈...②吴... III.信息论-高等学
校-教材 IV.G201

中国版本图书馆CIP数据核字(2004)第059374号

策划编辑 李蕊 责任编辑 赵天夫 封面设计 王凌波
责任绘图 郝林 责任印制 陈伟光

出版发行 高等教育出版社
社 址 北京市西城区德外大街4号
邮政编码 100011
总 机 010-58581000

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>

经 销 北京蓝色畅想图书发行有限公司
印 刷 北京外文印刷厂

开 本 787×960 1/16
印 张 19
字 数 340 000

版 次 2004年7月第1版
印 次 2005年1月第2次印刷
定 价 25.50元(配光盘)

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究
物料号 14374-00

总 序

根据教育部 1998 年颁布的普通高等院校专业目录,“信息与计算科学”专业被列为数学类下的一个新专业(它覆盖原有的计算数学及其应用软件、信息科学与运筹控制等专业)。这一新专业的设置很好地适应了新世纪以信息技术为核心的全球经济发展格局下的数学人才培养与专业发展的需要。然而,作为一个新专业,对其专业内涵、专业规范、教学内容与课程体系等有一个自然的认识与探索过程。教育部数学与统计学教学指导委员会数学类专业教学指导分委员会(下称教指委)经过过去两年艰苦细致的工作,对这些问题现在已有了比较明确的指导意见,发表了《关于信息与计算科学专业办学现状与专业建设相关问题的调查报告》及《信息与计算科学专业教学规范》(讨论稿)(见《大学数学》第 19 卷 1 期(2003))。为此,全国高等学校教学研究中心在承担全国教育科学“十五”国家级规划课题——“21 世纪中国高等教育人才培养体系的创新与实践”研究工作的基础上,根据教指委所颁布的新的教学规范,组织国内各高校的专家教授,进行其子项目课题“21 世纪中国高等学校信息与计算科学专业教学内容与课程体系的创新与实践”的研究与探索。为推动本专业的教材建设,该项目课题小组与高等教育出版社联合成立了“信息与计算科学专业系列教材编委会”,邀请有多年教学和科研经验的教师编写系列教材,由高等教育出版社独家出版,并冠以教育科学“十五”国家规划课题研究成果。

按照新的《信息与计算科学专业教学规范》(讨论稿),信息与计算科学专业是以信息技术和计算技术的数学基础为研究对象的理科学类专业。其目标是培养学生具有良好的数学基础和数学思维能力,掌握信息与计算科学基础理论、方法与技能,受到科学研究的训练,能解决信息技术和科学与工程计算中的实际问题的高级专门人才。毕业生能在科技、教育、信息产业、经济与金融等部门从事研究、教学、应用开发和管理工 作,能继续攻读研究生学位。根据这一专业目标定位和落实“强基础、宽口径、重实际、有侧重、创特色”的办学指导思想,我们认为,本专业在数学基础、计算机基础、专业基础方面应该得到加强,而各学校在这三个基础方面可大体一致,但专业课(含选修课)允许各校自主选择、体现各自特点。考虑到已有大量比较成熟的数学基础与计算机基础课程教材,本次教材编写主要侧重于专业基础课与专业课(含选修课)方面。

信息与计算科学,就其范畴与研究内容而言,是数学、计算机科学和信息

工程等学科的交叉,已远远超出数学学科的范畴。但作为数学学科下的一个理科专业,信息与计算科学专业则主要研究信息技术的核心基础与运用现代计算工具高效求解科学与工程问题的数学理论与方法(或更简明地说,研究定向于信息技术与计算技术的数学基础),这一专业定位明显地与计算机科学与信息工程专业构成区别。基于这一定位,信息与计算科学专业可包括信息科学与科学计算(计算数学)两个大的方向。科学计算方向在我国已有长期的办学经验,通常被划分为偏微分方程数值解、最优化理论与方法、数值逼近与数值代数、计算基础等学科子方向。然而,对于信息科学,它到底应该怎样划分学科子方向?应该怎样设置专业与专业基础课?所有这些都仍是正在探索的问题。

任何技术都可以认为是延伸与扩展人的某种功能的方式与方法,所以信息技术可以认为是扩展人的信息器官功能的技术。人的信息器官主要包括感觉器官、传导器官(传导神经网络)、思维器官和效应器官四大类型,其功能则主要是信息获取、信息传输、信息处理和信息应用(控制),因而感测技术、通信技术、智能技术与控制技术通常被认为是最基本的信息技术(常称之为信息技术的四基元),其它信息技术可认为是这四种基本技术的高阶逻辑综合或分解衍生。所以可以把信息科学理解为是“有关信息获取、信息传输、信息处理与信息控制基础的科学”。从这个意义上,我们认为:信息处理(包括图像处理、信号分析等)、信息编码与信息安全、计算智能(人工智能、模式识别等)、自动控制等可构成信息科学的主要学科子方向。这一认识也是教指委设置信息与计算科学专业信息科学方向课程的基本依据。

本系列教材正是基于以上认识,为落实新的《信息与计算科学专业教学规范》(讨论稿)而组织编写的。我们相信,该系列教材的出版对缓解本专业教材的紧缺局面,对推动信息与计算科学专业的快速与健康发展会有裨益。

从长远的角度看,为适应不同类型院校和不同层次要求的课程需求,本系列教材编委会还将不断组织教材的修订和编写新的教材,从而使本专业的教学用书做到逐步充实、完善和多样化。我们诚恳希望采用本系列教材的教师、同学们及广大读者对书中存在的问题及时指正并提出修改意见和建议。

信息与计算科学专业系列教材编委会

2003年8月31日

前 言

自 2000 年数学学科成立信息与计算科学专业以来,各高等院校招收该专业本科生的数量逐年增加,如何为该专业本科生提供一本关于信息论基本知识的教材已是一个迫切的问题.在信息论教材的编写过程中,我们也作过不少尝试,但是如何提供一本适用范围较广、内容较为丰富、重点突出的教材是一个困难的问题.应高等教育出版社的邀请,也为解决该教材的迫切需要问题,我们决定尝试编写本教材,希望能达到我们预期的目的.

本书的写作的基本思路有以下几点.其一是香农信息论仍是本教材的核心内容,我们仍把信源与信道编码定理作为信息论的基本理论与主要内容,这些内容不仅是近代信息处理的理论基础,具有深刻而又严格的数学描述与证明,而且在通信领域得到重要应用.通过这些内容的学习,可使我们看到数学理论在通信技术与通信工程中的重要作用,也可以看到数学理论与这些技术与工程问题的相结合过程.

其二是我们希望尽量增加本教材内容的范围,除了介绍信息与编码理论在通信技术领域的应用之外,我们还增加以下两部分内容,即增加对信息科学、信息技术与信息度量的一般问题的介绍与探讨.另外我们还介绍了信息论在纠错码、密码、卷积码、汉字编码与图形码中的应用,以及在金融、统计与智能计算等领域中的应用,这样就可使数学类专业的学生在更广泛的意义上学习和理解信息论,这也是本教材与以往国内其他信息与编码理论教材的一个主要区别.

其三是我们增加了信息处理的实验内容,我们选择了若干与信息处理过程密切相关的几个理论问题,给出它们的实验课题.由于信息与编码理论是一门理论与实践并重的学科,香农信息编码理论不仅在理论上得到广泛而又系统的发展,而且在近代通信技术与信息处理中得到实现与应用.因此,这些成果都可在教材与教学中得到体现,我们希望在教材中进行尝试,以此来提高学生在学习信息论的兴趣,加深对基础理论内容的理解.这也是本教材与以往其他信息与编码理论教材的区别.

本书内容由四部分组成,第一部分介绍信息论与信息科学的基本知识,其中包括信息科学概论,信息论的产生、发展与应用,信息与信息度量的内容要点等.

第二部分介绍香农信息论的基本内容，其中包括通信系统的基本要素与模型，信源、信道编码理论，有失真与多用户编码定理等。这部分内容是香农信息论的基础与核心。

第三部分介绍信息的应用，其中包括编码与密码学的要点，数据压缩编码与调制码理论及其他几种实用型的信息编码方法，如汉字编码与条码技术等。另外，我们还介绍了信息论在金融、统计与智能计算等领域中的应用。

第四部分我们给出了有关信息处理的几中典型问题的实验内容，如无失真信源编码问题，有、无失真数据压缩编码问题，卷积码模拟系统，图形码等。我们为这些实验提供国内外标准数据文本及实验基本要求方案。

本书在介绍以上内容时，叙述力求通俗、简练，较长的定理证明我们将通过指定参考书给出，各章附有习题与注解说明。

本书供高等院校信息科学专业或相关专业的本科生作为教材使用，亦可作为相关领域中的研究生、教学与科研人员和工程技术人员的参考用书。

目 录

第一部分 信息科学概论	1
第一章 信息科学与信息技术概论	3
§1.1 信息的定义与特征	3
1.1.1 信息的定义	3
1.1.2 信息的特征	4
1.1.3 信息的类型与表达方式	6
§1.2 信息科学、信息技术与信息产业	7
1.2.1 信息科学的研究内容、目的与意义	7
1.2.2 信息技术与信息产业	11
§1.3 信息论的产生、发展与应用	13
1.3.1 信息论的早期酝酿	14
1.3.2 香农 (Shannon) 信息论的建立与发展	16
1.3.3 信息论与信息科学的发展趋势与展望	23
§1.4 本书的内容简介与符号说明	26
1.4.1 内容简介	26
1.4.2 有关问题的说明	28
§1.5 习题	29
第二章 信息的度量问题	31
§2.1 信息度量问题概述	31
2.1.1 信息的可度量性	31
2.1.2 信息度量的原则	31
§2.2 香农熵 —— 随机变量的不确定性度量	32
2.2.1 随机变量和它的不确定性	32
2.2.2 不确定性的特征表示与香农熵的产生	33
2.2.3 香农熵的简单性质与例子	36
2.2.4 联合熵与条件熵	36
§2.3 熵的基本性质	39
2.3.1 对数函数的基本不等式与熵函数的最大值	40

2.3.2	熵函数的可加性	41
2.3.3	信息论中的重要不等式	42
§2.4	互熵与互信息	45
2.4.1	互熵的定义与性质	45
2.4.2	互信息的定义与性质	46
2.4.3	条件互信息的定义与性质	47
§2.5	连续型随机变量的信息量	49
2.5.1	关于连续型随机变量信息量的推广问题	49
2.5.2	连续型随机变量的熵与互信息	50
2.5.3	若干连续型随机变量熵的计算公式	51
2.5.4	其他连续型随机变量的信息量	54
§2.6	最大熵原理	55
2.6.1	最大熵原理的一般方法	55
2.6.2	最大熵分布的求解问题	56
§2.7	习题	58
 第二部分 信息论基础		 61
 第三章 通信系统概论		 63
§3.1	通信系统的基本模型与要素	63
3.1.1	通信系统的基本要素	63
3.1.2	通信系统中信息的传递过程	64
3.1.3	通信系统的概率统计模型	65
§3.2	单字母表通信系统的概率统计模型	65
3.2.1	信源与信道的概率统计模型	65
3.2.2	通信系统的数学模型	67
3.2.3	由通信系统决定的随机变量	68
§3.3	通信系统的序列模型与编码问题	69
3.3.1	通信系统的序列模型	69
3.3.2	通信系统序列中的编码问题	71
3.3.3	关于编码理论的基本要点	72
§3.4	习题	73

第四章 信源编码问题	75
§4.1 信源编码的基本问题	75
4.1.1 信源编码的基本要求与分类	75
4.1.2 即时码与前缀码	78
§4.2 变长码的编码问题	80
4.2.1 变长码的平均码长与编码问题	80
4.2.2 Kraft 不等式	81
4.2.3 变长码的平均码长估计问题	84
§4.3 哈夫曼 (Huffman) 码与算术码	86
4.3.1 哈夫曼码的构造运算	86
4.3.2 算术码	90
§4.4 通用码简介	91
4.4.1 通用码的一般概念	91
4.4.2 LZW 码的编码运算	92
4.4.3 LZW 码的译码运算	94
4.4.4 Yang-Kieffer 码简介	95
4.4.5 通用码的一般理论	97
§4.5 信源定长码的编码定理	98
4.5.1 信源序列的一般编码问题	98
4.5.2 信源序列的一般编码定理	100
4.5.3 无记忆信源序列的编码定理	100
§4.6 习题	102
第五章 信道编码定理	105
§5.1 信道编码问题	105
5.1.1 通信中编码的数量与质量关系问题	105
5.1.2 信道序列的编码问题	107
§5.2 离散无记忆信道	112
5.2.1 信道容量	112
5.2.2 几种典型的无记忆信道	113
5.2.3 无记忆信道序列的信道容量	115
§5.3 信道序列的正、反编码定理	117

5.3.1	一般信道序列的正、反编码定理	118
5.3.2	无记忆信道的正、反编码定理	122
§5.4	信道容量的计算问题	124
5.4.1	拉格朗日乘法	124
5.4.2	信道容量的递推算法	128
§5.5	可加高斯信道	132
§5.6	习题	135
第六章	有失真的信源编码定理	139
§6.1	有失真信源编码问题	139
6.1.1	有失真信源编码问题的意义	139
6.1.2	有失真信源的率失真函数	140
6.1.3	有失真信源的编码问题	144
§6.2	有失真信源编码定理	146
6.2.1	一般信源序列的正编码定理	146
6.2.2	一般信源序列的反编码定理	148
6.2.3	无记忆信源序列的正、反编码定理	150
§6.3	率失真函数的计算问题	154
6.3.1	计算率失真函数的一般方法	154
6.3.2	几种特殊信源的率失真函数的计算	156
§6.4	习题	159
第三部分	信息论应用	161
第七章	编码与密码	163
§7.1	代数码理论的基本特点与应用	163
7.1.1	代数码的基本特点	163
7.1.2	线性码	169
7.1.3	线性码的译码问题	175
7.1.4	汉明码	177
§7.2	信息论与密码学	179
7.2.1	密码学的发展概况	180

7.2.2	密码体制的基本要素与模型	181
7.2.3	对密码体制的随机分析	183
7.2.4	几种典型的密码体制	186
§7.3	习题	189
第八章	数据压缩与卷积码	193
§8.1	数据压缩理论的应用	193
8.1.1	数据压缩问题概述	193
8.1.2	静态图像数据压缩的基本运算	194
8.1.3	数据压缩技术分析的主要指标	198
8.1.4	数据压缩中的其他问题	199
§8.2	卷积码	200
8.2.1	卷积码的构造	200
8.2.2	卷积码的树与格子结构	204
8.2.3	卷积码的译码算法	208
§8.3	习题	211
第九章	信息论在统计与金融领域中的应用	213
§9.1	信息论在统计中的应用	213
9.1.1	概况	213
9.1.2	EM 算法	213
9.1.3	互熵与 Fisher 矩阵	218
§9.2	组合投资决策 (Portfolio) 中的应用	221
9.2.1	组合投资决策问题	221
9.2.2	最优组合投资决策的计算	222
§9.3	习题	225
第十章	几种实用的编码问题	227
§10.1	汉字编码问题	227
10.1.1	汉字的分类问题	227
10.1.2	汉字的编码类型	228
10.1.3	汉字的形、音结构	228
10.1.4	汉字编码的应用	228

§10.2 图形码概述	229
第四部分 信息处理实验	231
第十一章 无失真与有失真信源编码实验	233
§11.1 无失真的数据压缩	233
11.1.1 无失真的数据压缩的标准文本	233
11.1.2 哈夫曼编码的实现	236
11.1.3 LZW 数据压缩算法	243
§11.2 有失真的数据压缩	248
11.2.1 数字图像及其失真度量	248
11.2.2 JPEG 数据压缩算法的主要步骤	250
第十二章 其他编码问题	263
§12.1 卷积码与 Viterbi 译码	263
§12.2 图形码	268
12.2.1 条形码	268
12.2.2 二维条形码	278
参考文献	287

第一部分

信息科学概论

第一章 信息科学与信息技术概论

信息、信息科学、信息技术与信息产业是当今信息社会中最常见的名词，它在社会发展中的重大意义已在许多著作中给以论述，我们先对它们的含义、内容与特点作概要介绍。

§1.1 信息的定义与特征

信息是一个十分通俗而又广泛的名词，当今信息社会中，常把它作为人类认识世界的向导与智慧的源泉，也是社会与社会生产力发展的动力与资源。但到底什么是信息？什么是信息科学？这又是一个不太好回答的问题，从各种不同的学科领域出发，对它的重要定义与解释有 39 种之多。因此，对“信息”这个名词的含义我们只能从这些定义与解释中作综合理解。

1.1.1 信息的定义

关于信息，从不同的学科及不同的角度有不同的解释定义，我们大体把它归结为三大类。

第一种类型就是从日常生活知识来理解，这在世界著名的百科全书或词典中都有说明，在那里信息被看作新闻、消息与知识，如在英国牛津字典中，信息是谈论的事情、新闻和知识；在韦氏字典中，信息是在观察研究过程中获得的数据情报、新闻和知识；在日本广辞苑中，信息是所观察事物的知识等，这种定义较符合在日常生活中对信息的理解，与每个人都有密切关系，因此有十分广泛的意义。

第二种类型是有关科学家的论述，他们的出发点是在对信息的处理中所产生的理解与定义，因此较偏重于组成信息的形式、结构等内容等角度来定义，把信息作为事物的联系、变化与差异。如在香农 (C. E. Shannon) 理论中，把信息与信息的度量作为事物的“不肯定性”与“不肯定性”的度量；艾什比把信息定义为“差异度”，维纳 (N. Wiener) 把信息定义为一种“组织性”的度量，并把它作为控制系统在进行调节活动时的相互作用、相互交换的内容。近期有一种趋势就是把熵的概念统一成事物“复杂度”的度量，把信息作为运动变化与交换中结构形态的差异性度量。由此可见，这类定义是把信息看作事物结构形态的“复杂度”、“差异度”与“不肯定度”。

第三种类型是有关哲学家的论述，几种典型的说法如：“信息是一切物质的属性”，“信息就是信息，不是物质也不是能量”。这种定义从哲学的角度来讨论并确定信息在现实世界中的地位。

从这些简单的摘录中，我们可大体了解信息的含义。对信息的定义过程实际上就是信息科学的发展过程，这个过程还远未终结，随着科学技术的进步与发展，对信息作出更为全面、深刻的定义一定还会出现。为避免偏离本书所要讨论的主题，我们对这些命题不展开讨论。

1.1.2 信息的特征

了解信息的特征不仅有助于我们对信息这个概念的理解，也有助于我们今后对信息处理的理解。信息有以下主要特征：

1. 信息必须依附于某种载体，这种载体包含确定的物质内容，并通过一定的结构形式体现。如语言文字信息，它所依附的物质载体是纸张、油墨，仅仅纸张、油墨并不能构成信息，油墨在纸张上形成文字后才能构成信息，不同的文字就是油墨在纸张上的不同表现形式。

在相同的物质载体条件下，构成信息的结构形式是千变万化的，同样 26 个英文字母，由它们的组合可以产生各种各样的文学、图画、指令等内容，这种组合变化与物质所提供的条件并不按照比例关系增长，而是按指数的比例，甚至按指数的指数比例增长，这种不按比例的增长关系为信息爆炸的可能性提供依据，当物质与信息的增长关系进入按指数的比例变化时代就是信息的爆炸时代。

2. 信息需在交换、交流、分析与比较中发挥作用。因此，在信息处理中有信息的生成体、载体与受体的区别，如在文字的信息交换中有作家、出版物与读者的区别，通电话时有打电话、听电话与传电话的区别，这些都是信息的生成体、载体与受体。因此在信息论中，把信息处理分为信源、信道与信宿三大要素，对此我们在下文中还会详细讨论。

信息在交换、交流中不仅有生成体、载体与受体的区别，而且在不同的生成体与受体中信息所发挥的作用与价值也会不相同的。同样一篇文章，对不同的读者可能会产生不同的理解与影响。另外，有的信息虽然文字不多，但对受体来讲会有起到关键性的重要作用，有的信息虽然长篇大论，但对受体来讲也许毫无意义。由此说明信息在不同的生成体与受体中的作用可能是不同的。

研究信息不同生成体、载体与受体之间信息的相互作用关系，对它们作