



清华大学
电子与信息技术系列教材

研究生教学用书

教育部研究生工作办公室推荐

应用信息论基础

Fundamentals of Applied Information Theory

朱雪龙 编著



清华大学出版社

<http://www.tup.tsinghua.edu.cn>

清华大学电子与信息技术系列教材

应用信息论基础

朱雪龙 编著

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书从基本概念、基本方法和基本应用三个方面较系统全面地介绍了信息理论。本书取材既重视基础理论,又面向实际应用;既讲述成熟的原理,又适当介绍发展中的前沿课题。全书共分9章。第1章和第2章分别为概述和基本概念,第3章至第6章分别介绍信源的冗余度压缩编码、熵压缩编码和信道的容量与信道编码。第7章讨论最大熵与最小鉴别信息原理及其应用。第8章讨论非统计意义下的信息理论,内容包括组合信息、算法信息与通用编码。最后在第9章中简要介绍了通信网中的信源编码与信道容量问题。各章附有习题。

本书可作为高等学校与科研院所信息类专业研究生教材或教学参考书使用,也可供有关科技人员在学习专业基础理论时参考。

图书在版编目(CIP)数据

应用信息论基础/朱雪龙编著. --北京:清华大学出版社,2000

清华大学电子与信息技术系列教材

ISBN 7-302-04154-7

I. 应… II. 朱… III. 信息学-高等学校-教材 IV. G201

中国版本图书馆 CIP 数据核字(2000)第 78583 号

出版者:清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者:北京市清华园胶印厂

发行者:新华书店总店北京发行所

开 本:787×960 1/16 印张:22 字数:394 千字

版 次:2001 年 3 月第 1 版 2001 年 5 月第 2 次印刷

书 号:ISBN 7-302-04154-7/TN·116

印 数:3001~6000

定 价:26.00 元

前 言

自香农(C. E. Shannon)提出信息理论以来,在大学中开设信息理论课已有近半个世纪的历史。尽管课程的中心内容都离不开香农提出的基本概念、方法和定理,但课程设置的对象、目的、课程内容的组织、取舍以至讲授的方式方法都因校、因时、因人而异,存在众多显著的差别。

信息理论涉及两个学科是造成上述差别的主要原因。在数学界,信息理论被看成是概率论的分支,是遍历性理论的分支,一个涉及不变变换理论的分支。所以在数学系,这一课程是为统计数学专业的学生开设的。而在工学界,信息理论被看成是通信理论的一部分,是通信的统计理论,是数字通信的基础理论。因此在电气或电子工程系内它是为通信专业的学生开设的。了解同一名称的课程在不同系开设时的差异不但是有益的,而且是必要的,这可使授课者或学习者都能把他们的精力有效地放在他们关心的问题 and 关心的方向上。这一本书是从工程应用的角度编写的,所以我们在信息论前加上应用两字,以明确本书的性质和特点。

其次,正如汉明(R. W. Hamming)所说,从逻辑上讲是编码理论导致信息理论,而信息理论又给出编码的性能极限。因此信息理论常和编码理论放在一起讲授,甚至把课程称为信息与编码理论。但目前这样做时其中的编码理论往往只指基于近世代数的信道编码构造理论。由于迄今在信源编码方面尚没有一个统一的构造性理论,因此目前还难于开设一个在内容上组织均衡的本来意义上的信息与编码理论。基于这一考虑,本书对信道编码的构造理论仅作简单的介绍,也不把编码理论四字纳入教材名称,以明确本书完全以信息理论为中心。

香农提出的信息理论是一种基于统计意义上的信息理论。这一理论对通信技术的发展产生了持久和深刻的影响,但它对信息技术的其他一些方面,如人工智能等,则很少有理论指导作用。所以自香农以来,人们对更广泛意义下的信息的研究一直没有停止。迄今为止,较为成熟的研究成果有:E. T. Jaynes 在 1957 年提出的最大熵原理的理论,S. K. Kullback 在 1959 年首次提出后又为 J. S. Shore 等人在 1980 年后发展了的鉴别信息及最小鉴别信息原理的理论,A. N. Kolmogorov 在 1965 年提出的关于信息量度定义的三种方法——概率

法,组合法,计算法,A. N. Kolmogorov 在 1968 年阐明并为 J. Chaitin 在 1987 年系统发展了的关于算法信息的理论。这些成果大大丰富了信息理论的概念、方法和应用范围。首先,它把信息的统计定义进一步推广并对非统计意义的信息给出了一种量度。其次,信息量度的意义已不再限于信源编码和信道编码。信息的量度已系统地发展成为信息处理的一种准则,这一准则在信息技术领域正逐渐取代代表功率的最小均方误差准则。可以认为信息理论已从通信的数学理论发展成为信号与信息处理的基础理论。基于这一考虑,本书对上述列举的研究成果也作了系统的介绍。

本书共分 9 章,第 1 章介绍信息论和信息论方法。通过这一章读者可以明了在学习信息理论后什么是他们可以得到的,什么是不能得到的,以及为了他自己的学习目的,他应该如何学习。第 2 章介绍信息论的基本概念:熵、互信息和鉴别信息。这是学习本书的基础。第 3 章至第 6 章介绍香农的信源编码和信道编码理论,同时对实用的信源编码方法(如矢量量化、线性预测编码、变换编码)和信道编码方法(如线性码等)作简单的介绍。主要目的是说明信息理论如何应用于实际以及理论与实际之间存在什么差距。第 7 章介绍最大熵与最小鉴别信息原理。这一章具体体现了熵与鉴别信息作为准则在信息处理中的作用。非统计意义上的信息理论在第 8 章中介绍。这一章同时给出了通用编码方法的代表——Lempel-Ziv 编码方法及其性能界。第 9 章介绍网络中的信源编码与信道容量问题。这些内容主要针对以电路交换为基础的通信网,但不适用于以分组交换为基础的通信网。在以分组交换为基础的通信网中信息传输的载体——信号,除了在信号幅度上受到干扰外还在信号传输的延迟时间上受到干扰,这两者都对信号携带的信息量起到限制作用。它们的分析方法与这一章介绍的方法差异很大,限于篇幅,这部分内容在书中未作介绍。

作者希望通过这 9 章的内容能为读者提供有关信息理论的比较全面和比较系统的知识。在全书各章的叙述中作者都尽可能地使基本概念和基本方法的描述清晰易懂,为此省略了某些数学细节。少数不涉及基本概念且较深入的内容在目录中用 * 号标出。全书除第 1,2,8 章外其余各章都有一些内容涉及信号处理和通信技术,这对了解信息理论的应用是有益的。为此学习本书时除了需要有概率论、随机过程的必要基础知识外,还需具有信号处理和通信技术方面的基础知识。

我国已故院士常迥教授,信息科学界的前辈周炯槃院士,吴佑寿院士以及知名学者仇佩亮、章照止、钟义信、周荫清、孟庆生、王新梅、贾世楼、金振玉、方军、

姜丹等都在信息理论方面有很好的著作或教材。国外学者如 T. M. Cover, T. Berger, R. E. Blahut, R. G. Gallager 等也都有很好的专著或教材。作者在编写此书时得益于早期对他们著作的学习,在此顺表谢意。

作者感谢清华大学电子工程系对作者的信任,使作者有机会多年担任此课程的讲授。本书正是在这一基础上写成的。

作者还感谢艾红梅、邓北星及郁杨在整理出版本书中的帮助。

本书不当之处,敬请来函赐教。

朱雪龙

2000 年 7 月

于清华大学电子工程系

目 录

第 1 章 信息论与信息论方法概述	1
1.1 信息、信息科学和信息论	1
1.2 信息论方法的应用及其取得的成果	3
1.3 信息论的形成与发展	7
1.3.1 通信技术的理论基础	7
1.3.2 统计数学的一个分支	11
1.3.3 信号与信息处理的一般理论基础	12
第 2 章 信息论的基本概念	14
2.1 离散熵	14
2.1.1 熵的定义	14
2.1.2 熵函数的性质	19
2.1.3 熵函数形式的唯一性	23
2.1.4 联合熵与条件熵	26
2.2 离散互信息	29
2.2.1 互信息的定义	29
2.2.2 多个随机变量下的互信息	31
2.2.3 互信息函数的性质	35
2.3 连续随机变量下的熵与互信息	38
2.3.1 连续随机变量下的微分熵	38
2.3.2 随机变量函数的微分熵	40
2.3.3 连续随机变量下的互信息	42
2.4 鉴别信息	44
2.4.1 鉴别信息的定义	44
2.4.2 Kullback 与香农两种信息量度之间的关系	47
2.4.3 鉴别信息的性质	51
2.4.4 鉴别信息函数形式的唯一性	57
2.5 对信息论基本概念的若干评注	70

习题	71
第 3 章 信源的熵率、冗余度与冗余度压缩编码	74
3.1 信源、信源模型与信源编码	74
3.1.1 信源	74
3.1.2 信源模型	75
3.1.3 信源编码	77
3.2 离散稳恒信源的熵率与冗余度	78
3.3 离散无记忆信源的渐近等同分割性与信源的定长编码定理	82
3.3.1 渐近等同分割性	82
3.3.2 定长编码定理	84
3.4 离散无记忆信源的变长编码	86
3.4.1 前缀码与 Kraft 定理	87
3.4.2 唯一可译码定理	89
3.5 变长编码的平均码长与最优编码	90
3.5.1 变长编码的平均码长	90
3.5.2 最优编码	92
3.6 离散无记忆信源的变长树码	95
3.6.1 算术码	95
3.6.2 算术码的存在性	97
3.7 离散马尔可夫信源的熵率	100
3.7.1 马尔可夫链的基本概念	100
3.7.2 离散马尔可夫信源熵率的计算	102
3.8 离散马尔可夫信源的编码定理与最优编码	105
习题	108
第 4 章 信道、信道容量与信道的有效利用	110
4.1 信道、信道模型与信道分类	110
4.1.1 信道	110
4.1.2 信道模型与信道分类	111
4.2 离散无记忆信道及其信道容量	112
4.3 离散无记忆信道容量的计算	116
4.3.1 信道容量解的充要条件	116
4.3.2 某些简单情况下信道容量的解	119

4.3.3	一般情况下信道容量的解	122
4.3.4	信道容量的迭代解法	124
4.4	级联信道和并联信道的信道容量	125
4.4.1	级联信道	125
4.4.2	并联信道	128
4.5	信道达到充分利用时输入输出字母概率分布的唯一性	131
4.6	连续信道的信道容量	135
4.6.1	无记忆加性噪声信道的信道容量费用函数	136
4.6.2	无记忆加性高斯噪声信道的信道容量费用函数	137
4.6.3	一般无记忆加性噪声信道的信道容量费用函数的界	141
4.6.4	无记忆加性高斯噪声信道的级联和并联	142
4.7	模拟信道的信道容量	145
4.7.1	模拟随机信号的正交展开	145
4.7.2	模拟信道下的信道容量费用函数及其计算	148
4.8	限带加性白色高斯噪声信道的极限性能及其与传输要求的匹配	156
4.8.1	限带加性白色高斯噪声信道的性能及其极限	157
4.8.2	信道与信息传输要求的匹配	160
4.9	限带模拟信道的数字化	163
4.10	蜂窝移动通信条件下信道的有效利用	165
4.10.1	蜂窝移动通信与系统的频谱利用效率	165
4.10.2	不同接入方法下蜂窝移动通信系统的频谱利用效率及其比较	167
	习题	171
第5章	信道编码	173
5.1	信道编码概述	173
5.2	信道译码准则	176
5.3	联合典型序列与联合渐近等同分割定理	179
5.4	信道编码定理	181
5.5	信道编译码方法的最初范例——汉明码	183
5.6	分组码之一:线性码	185
5.6.1	线性码的定义、编码与生成多项式	185

5.6.2	线性码的伴随式与伴随式译码	187
5.7	分组码之二:循环码	190
5.7.1	循环码的定义	191
5.7.2	循环码的编码与生成多项式	191
5.7.3	循环码的伴随式与译码原理	193
5.8	树码、网格码与卷积码	195
5.9	关于信道编码理论的若干评注	199
	习题	200
第6章	信源的信息速率失真函数与熵压缩编码	203
6.1	熵压缩编码和信源的信息速率失真函数	204
6.1.1	熵压缩编码	204
6.1.2	离散无记忆信源的熵压缩分组编码及信源的信息速率失真函数	204
6.2	信息速率失真函数的性质	206
6.3	离散无记忆信源下的信息速率失真函数的计算	211
6.3.1	信息速率失真函数解的充要条件及参数方程	211
6.3.2	求解信息速率失真函数的迭代算法	217
6.3.3	信息速率失真函数解的唯一性问题	220
6.3.4	乘积信源的信息速率失真函数	222
6.4	连续无记忆信源的信息速率失真函数	224
6.4.1	连续无记忆信源信息速率失真函数的定义及其解的充要条件	224
6.4.2	差值失真量度下连续无记忆信源信息速率失真函数的计算	226
6.5	标量量化	233
6.5.1	均匀量化	235
6.5.2	Lloyd-Max 算法	235
6.6	有记忆连续信源与模拟信源的信息速率失真函数	236
6.6.1	有记忆连续信源的信息速率失真函数的定义	236
6.6.2	模拟信源的信息速率失真函数的定义	237
6.6.3	高斯有记忆连续信源的信息速率失真函数	237
6.6.4	高斯模拟信源的信息速率失真函数	240

6.7	变换编码——实用的熵压缩分组编码	242
6.8	预测编码——实用的熵压缩树码	248
6.8.1	最小均方误差预测器	249
6.8.2	最小平均绝对误差预测器	249
6.8.3	最大零误差概率预测器	249
	习题	251
第7章	最大熵原理与最小鉴别信息原理	254
7.1	非适定问题与最大熵和最小鉴别信息原理	254
7.1.1	非适定问题的提出	254
7.1.2	最大熵原理与最小鉴别信息原理	255
7.2	最大熵原理的合理性	258
7.3	最小鉴别信息原理与最大熵原理的公理化推导	261
7.3.1	最小鉴别信息原理的推导	261
7.3.2	最大熵原理的推导	269
7.4	最小失真意义下的最大熵原理与最小鉴别信息原理	269
7.4.1	方向正交与投影失真	270
7.4.2	投影失真的一般形式	272
7.4.3	最小失真准则与熵准则	274
7.5	最大熵与最小鉴别信息原理的应用及其解的分布	276
7.5.1	最大熵与最小鉴别信息原理的应用	276
7.5.2	最大熵分布与最小鉴别信息分布	281
	习题	282
第8章	组合信息、算法信息与通用编码	284
8.1	信源统计特性不确定时的信源编码问题	284
8.1.1	统计特性失配时统计编码的性能	285
8.1.2	自适应统计编码	285
8.2	基于组合的信息量度与通用编码	287
8.2.1	基于组合的信息量度	287
8.2.2	通用编码	288
8.2.3	Fitingof 通用编码的性能	289
8.3	算法信息量	292
8.3.1	单一事件或数值下的信息量度问题	292

8.3.2 Kolmogorov 算法熵	293
8.4 二元字符序列的算法熵	294
8.5 算法熵的不可计算性	296
8.6 有限状态压缩编码器	298
8.7 Lempel-Ziv 编码	302
8.8 LZ 编码压缩比与香农熵	305
习题	306
第 9 章 通信网中的信源编码与信道容量	309
9.1 概述	309
9.2 反馈信道	314
9.3 多个随机变量下的联合典型序列	316
9.4 多源接入信道	320
9.4.1 多源接入信道的容量	320
9.4.2 相关信源输入下的多源接入信道	324
9.5 高斯多源接入信道	324
9.5.1 高斯多源接入信道的容量域	324
9.5.2 对高斯多源接入信道容量域的讨论	326
9.5.3 多源接入信道容量域与多址方法的关系	327
9.6 分布信源编码	329
习题	333
参考文献	335
索引	336

第 1 章 信息论与信息论方法概述

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。随着信息和信息科学对现代社会生活各方面影响的不断加大和深化,人们对信息论的意义的认识 and 价值的估计也不断变化。在这一章中,我们将简要地从工程技术或技术科学的角度讨论一下什么是信息,什么是信息科学和信息论,并对信息论形成、发展的过程及其已取得的应用成果作简要的回顾。我们的目的主要是通过历史发展的实际过程来说明信息论的研究对象、方法和适用的范围。

1.1 信息、信息科学和信息论

什么是信息?信息一词在我国由来已久。据辞海记载,我国南唐诗人李中有诗云“梦断美人沉信息,目穿长路依楼台”,可见信息泛指音讯和消息。在近代,信息一词又被用作英语中 information 的译名,information 在英语中来自词根 inform(通知),乃指被告知的事实或知识。information 在牛津英文字典里给出的解释是“某人被通知或告知的内容、情报、消息”。在这样的解释中,信息一词显然不是作为科学名词或技术术语来定义的,因此无法对信息一词作更深入的推敲。实际上在不同的字典中对信息一词还有不同的解释,更不用说工作在不同领域的人们对信息有不同的理解了。

尽管信息一词的含义模糊和难于捉摸,但人人都感觉到它的存在。每时每刻我们都在通过对周围世界的观察去获取它,并且通过一定的方法把它传送给别人、进行交换或把它存储起来留作以后使用。这种目前尚难明确定义的信息我们暂时可以把它称为广义理解的信息。对这种广义的信息自然是值得研究的,实际上语言学家已经对其作了认真的研究,如 Zellig Harris 的著作《语言和信息的理论》就是这方面的代表。但迄今为止尚未有得到普遍接受的有关广义信息的定义。

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以后的事。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得到一个统一的、全面的、不需时时改变的表达。作为一个技术术语的信息其意义当然要比前面广义信息的含义具体得多,但仍然是比较笼统和含混不清的。

信息作为一个可以用严格的数学公式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。无论是在统计数学中还是在通信技术中定义的信息都是一种统计意义上的信息,我们可以把它简称为统计信息。统计信息是非常明确的,同时其适用范围要比广义信息狭隘得多。我们在本书中讨论的信息论正是关于这种统计信息的理论。

统计信息是一个抽象的然而明确的概念,它与作为技术术语用的信息仍有很大的区别。后者比较笼统,没有确切定义但却比较具体。有一种意见认为作为技术术语用的信息实际上是指一切符号、记号、信号等表达信息所用的形式或载体,这种意见实际上把信息的形式或载体和它的具体内容区分开来。计算机所能处理的(特别是通信所能传送的)都是信息的载体或表达形式。计算机可把信息的一种形式转换成另一种形式,如把英语文本翻译成法语文本,把数据库中的数据整理成所需形式的报表,或把气象数据进行处理后给出某一地区的气温等,而通信则把信息的具体载体或形式从甲地传送到乙地。这种看法虽然未得到广泛的承认,但基本上反映了目前的实际情况。

上述讨论归纳起来可以说:广义信息是把信息的形式、内容等全都包含在内的最广泛意义上的信息。作为技术术语的信息主要指信息的具体表达形式,虽然信息的形式总是与信息的内容有一定的联系,且不可能存在没有内容的形式,但作为技术术语的信息的确不考虑信息的内容。而统计信息则是一种有明确定义的科学名词,它与内容无关,而且不随信息具体表达形式的变化(如把文字翻译成二进制码)而变化,因而也独立于形式。它反映了信息表达形式中统计方面的性质,是一个统计学上的抽象概念。

现在我们转过来讨论信息科学和信息理论,但为此我们必须先谈谈信息技术。什么是信息技术?对此我们又无法给予明确的回答。但在实际中我们知道它是泛指计算机所涉及的各种各样的技术。有意思的是这一名词的笼统和不明确反倒成为一个优点,使它能对计算机涉及的种类繁多且在不断发展的技术得到一个统一的、全面的、高层次的表述。信息技术的出现当然使人们联想到信息

科学,因为所谓技术乃是对科学原理加以应用的具体实践。我们知道能源技术和材料技术都是有相应的科学原理作为基础的,但信息技术迄今为止还没有上述意义的信息科学作为基础,或者确切地说系统的信息科学迄今还没有完全成熟。信息科学作为一个名词来看,最早出现在图书馆学中,主要研究图书文献的检索。在计算机出现以后,信息科学被赋予新的含义,但在不同国家中它的含义不尽相同。在日本信息科学的含义和美国的计算机科学的含义相似,主要研究科学计算的理论和方法。而在美国信息科学原先主要指科学计算以外,如商业、服务业、管理统计部门等所需要的涉及大量数据但计算比较简单的数据处理问题。20世纪80年代以来,信息科学的含义不断扩大,不但逐渐把计算机科学的内容统一包含在内,而且有把信息技术涉及的所有科学理论统统包含在内的趋势。20世纪80年代初美国信息科学学会就曾多次举行学术讨论会研究这种意义上的信息科学,即作为信息技术理论基础的信息科学。因此,尽管人们还在不同意义上使用信息科学这一名词,但从发展和长远的观点来看,把信息科学理解成能与信息技术相对应的基础科学可能是合适的。在这样的理解下信息科学与信息理论的关系又如何呢?在美国信息科学学会所组织的一次学术讨论会上这一问题成为讨论的中心问题之一。有一部分人的意见认为统计信息理论不能解决诸如人工智能中如何进行启发式搜索和学习等问题,因此在未来的信息科学中没有信息理论的位置。但多数人的意见认为信息理论尽管有其局限性和不足,但仍然是未来信息科学所不可缺少的一块基石。近年来信息论的发展和信息论方法的逐渐广泛应用表明后一种看法是比较全面和客观的看法,我们有理由相信今后的发展还会证明这一点。

1.2 信息论方法的应用及其取得的成果

信息论从它诞生的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。50年过去了,这期间虽有失败的教训但也不乏成功的例子,下面我们就列举若干成功的范例。

(1) 语音信号压缩

语音信号一直是通信网中传输的主要对象。自从通信网数字化以来,压缩语音信号的编码速率就成为通信中的一个重要问题。根据信息理论的分析,语音信号所需的编码速率可以远远低于仅按 Nyquist 采样定理和量化噪声分析所决定的编码速率。几十年来的工作已在这方面取得巨大的进展:长途电话网标

准的语音编码速率已从 1972 年 CCITT G. 711 标准中的 64 kbit/s 降低到 1992 年 CCITT 标准中的 16 kbit/s。在移动通信中 1988 年欧洲 GSM 标准中的语音编码速率为 13.2 kbit/s, 而 1989 年美国 CTIA 标准中的速率为 7.95 kbit/s。对语音质量要求较低的军用通信, 美国 NSA 标准的速率在 1975 年时已达到 2.4 kbit/s。目前, 在实验室中已实现 600 bit/s 的低速率语音编码, 特别是按音素识别与合成原理构造的声码器其速率可低于 100 bit/s, 已接近信息论指出的极限。

(2) 图像信号压缩

图像信号的信息量特别巨大, 这对图像信号的传输及存储都带来极大的不便。经过多年的研究, 到 20 世纪 80 年代时图像压缩逐步进入建立标准的阶段。1989 年 CCITT 提出电视电话/会议电视的压缩标准 H. 261, 其压缩比达到 25:1 到 48:1 左右。1991 年 CCITT 与 ISO 联合提出的“多灰度静止图像压缩编码”标准 JPEG, 其压缩比为 24:1。对常规电视信号的压缩在 1991 年提出的国际标准 MPEG-1 中其平均压缩比可达 50:1。这些巨大的进展为图像信号进入通信领域以及多媒体计算机的发展创造了条件。此外, 随着全数字高清晰度电视技术的发展, 相应的压缩编码研究也已取得进展, 这就是 1993 年提出的 MPEG-2。

(3) 计算机文件的压缩

由于数据库的广泛应用, 存储计算机文件所需的存储量问题日益突出。在过去的二十多年中对计算机文件的压缩已发展了至少二十余种不同的算法。1988 年澳大利亚举行的压缩算法对抗赛对各种算法进行了一次大检阅。竞赛所用的文件内容多种多样, 如各种高级语言的源程序、目标码、论文、小说、科学数据、图书目录等等。压缩工作全在当时比较典型的微机上进行。测试结果表明, 其中较好的算法都能使各种文件压缩后所需的存储量只为原文件的 25% 到 50%, 其平均值约为 30% 左右。压缩所需的时间和存储器开销都不大。目前, 各种压缩算法已在计算机中得到广泛的应用。

(4) 模拟话路中数据传输速率的提高

20 世纪 50 年代初计算机开始在美国联网, 当时模拟话路是传输计算机数据的几乎唯一可用的信道。最早的调制器其速率只有 300 bit/s, 而信息论指出, 标称带宽为 4 kHz, 信噪比为 25 dB 的话路信道的极限速率应在 25 kbit/s 左右。所以在以后的三十多年中就开始了提高速率的长期的、现在看来是极其成功的工作。1967 年速率为 4 800 bit/s, 1971 年为 9 600 bit/s, 1980 年开始进

入 14.4 kbit/s, 1985 年时利用多维网格编码调制的调制器 Codex 2680 使速率达到 19.2 kbit/s, 非常接近于理论极限。信息论在模拟话路数据传输中获得的巨大成功, 其意义远远超出模拟话路本身。实际上由于模拟话路中所用的调制器可用 VLSI 电路实现, 价格低廉, 因而实际上成了信息论方法的试验场。各种在话路调制器中获得成功的调制方法都很快被推广应用到载波的高次群信道及数字微波、数字卫星信道中, 他们都同样获得了成功。

(5) 降低信息传输所需的功率

在远距离无线通信, 特别是深空通信中如何降低信息传输所需的功率至关重要, 因为在这种情况下发送设备的功率和天线的尺寸都已成为设备生产和使用中的一个困难问题。幸运的是正是在这个领域信息论获得了它第一批令人信服的成果。从 20 世纪 60 年代后期起, NASA 发射的所有深空探测器无一例外地在其通信设备中采取了信道编码措施, 因为根据信息理论的分析, 采用低码率的信道编码可以降低传送单位比特所需的能量 E_b 与噪声功率谱密度 N_0 之比。现在利用不太复杂的信道编码就可以使同样误码率下所需的 E_b/N_0 比不采用信道编码时低 6 dB 左右。其中一些好的方案(如用 RS 码作为外码、卷积码作为内码的方案)可以使误码率在 10^{-6} 的情况下所需的 E_b/N_0 降到 0.2 dB, 比不用信道编码时所需的 10.5 dB 降低了近 10 dB。

(6) 计算机网中数据传输可靠性的保证

随着计算机技术的发展, 计算机设备的布局变得愈来愈分散, 各种终端及外围设备离主机也越来越远, 这就产生了计算机网。近年来, 计算机网还与分布式计算机系统相联系而变得更为重要。在用各种电缆联接而成的计算机网中电噪声和各种外界的电磁干扰是不能不考虑的, 因为它使传输的信息发生差错。一般情况下局域网中的差错率在 10^{-8} 左右, 广域网中的差错率在 $10^{-3} \sim 10^{-5}$ 。这样高的差错率在实际应用中是无法接受的。

目前普遍采用的解决办法是带自动重发请求的差错检测码。差错检测的方法从最简单的奇偶检验到比较复杂的循环冗余检验都被采用, 但较大的网一般都用循环冗余检验。这种方法已被各种网络通信协议采用并成为标准。例如 ISO 制定的高级数据链路协议 HDLC 就采用 CCITT V. 41 的 CRC 码进行循环冗余检验。HDLC 在全世界已被广泛采用, 这一标准有很广的应用领域, 许多协议都是从它派生出来的。

(7) 计算机中的容错问题

计算机的存储器是计算机的重要组成部分。随着计算机技术的发展无论是