



21 世纪高等院校计算机网络与通信教材

# 信息论与编码理论

北京希望电子出版社 总策划  
曲 炜 主 编  
王结南 汪建立 副主编  
朱诗兵 刘作学 李迎春 编 著



 科学出版社  
[www.sciencep.com](http://www.sciencep.com)



21 世纪高等院校计算机网络与通信教材

# 信息论与编码理论

北京希望电子出版社 总策划  
曲 炜 主 编  
王结南 汪建立 副主编  
朱诗兵 刘作学 李迎春 编 著



 科学出版社  
[www.sciencep.com](http://www.sciencep.com)

## 内容简介

为了适应计算机科学与技术学科的发展和现代计算机教学的需要,作者在多年研究生、本科生和大专生的计算机网络教学、实践的基础上,以香农信息论为基础,编写了本教材。

在引言中,简单地介绍了信息论与编码理论的基本概念;第一部分讲解了香农信息论与编码理论的主要内容,如信息的统计度量、信源的信息熵、信道的信道容量、平均失真度与信息率失真函数以及信源编码和信道编码定理;第二部分阐述了一些基于香农编码理论的信道编码方法,如线性码、循环码、卷积码、Turbo 码等信道纠错编码。

本教材内容丰富翔实,重点突出,对基本概念和基础理论的阐述清晰明了,有实际应用价值,可作为通信工程及信息类专业的高年级本科生教材,亦可作为其他专业学生和有关科技人员的参考用书。

需要本书或需要得到技术支持的读者,请与北京中关村 083 信箱(邮编 100080)发行部联系,电话:010-82702660 010-82702658 010-62978181 转 103 或 238 传真:010-82702698 E-mail: tbd@bhp.com.cn

## 图书在版编目(CIP)数据

信息论与编码理论/曲炜主编,朱诗兵等编著. —北京:科学出版社, 2005.9

21 世纪高等院校计算机网络与通信教材

ISBN 7-03-015805-9

I. 信... II. ①曲...②朱... 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. TN911.2

中国版本图书馆 CIP 数据核字(2005)第 071011 号

责任编辑:曾华 / 责任校对:马君  
责任印刷:媛明 / 封面设计:梁运丽

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>

北京市媛明印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2005 年 9 第 一 版 开本: 787×1092 1/16  
2005 年 9 第一次印刷 印张: 18 3/4  
印数: 1-3 000 册 字数: 426816

定价: 26.00 元

# 21 世纪高等院校计算机网络与通信教材

## 编委会

主任 曲 炜

副主任 陆卫民 卢 昱 赵洪利 李新明

委员 (以姓氏笔画为序)

马彦恒 万定生 王擎天 王成友 王向阳

朱诗兵 刘作学 吴善培 何新华 何忠龙

张 文 周 辉 郑明红 罗建华 杨喜权

赵立军 姚秀芳 徐建华 徐远超 郭德纯

梁计春 韩素华 葛洪华 樊秀梅 穆道生

## 序

目前,中国固定和移动两大网络的规模都已位居世界第2位,上网用户2004年总数达9400万,中国的信息通信制造业也得到很大的发展。今后5年中国信息产业预计将仍会以高于20%的速度增长。中国将加快建设新一代信息通信网络,全面振兴信息通信产品制造业和软件业,建立能够支撑信息通信业发展的技术、生产体系。在向数字化、集成化、网络化转变的过程中,简单服务要向个性化服务发展,低带宽要向高带宽发展,电路交换要向分组交换发展。无线网络、网络多媒体、多媒体计算、人机自然语音通信是网络与通信专业重点建设的四大方向。

面对潜力巨大的中国市场,我国大学的相关专业需要培养具有知识创新能力的高素质人才,在通信高新技术的研究上争创国际先进水平,为我国在信息领域达到国际一流的目标作出贡献。

科技的发展使得教育要跟上时代发展的步伐,但是目前市面上还没有一套系统、完整的关于计算机网络与通信方面的教材。现有的教材有些偏重理论,有些则偏重实用,不太适合于课堂教学。而对于学习网络与通信的学生来说,不仅要懂得原理,还必须学会技术,这样才能符合“培养人才、创造知识、转化成果、服务社会”的教学宗旨,在人才培养、科学研究和技术应用等方面有所成就,为我国通信与信息领域的发展做出贡献。

为了获得与国际接轨的教学内容,达到提高整体教学水平的目的,北京希望电子出版社组织国内各大高校相关专业的教授、专家、学者,共同编选本套丛书。本套丛书强化学生实践能力和创新意识的培养,定位准确、内容创新、结构合理。在选材上主要采用了成熟的理论,并通过在目前研究现状的跟踪,补充了最新的研究成果;充分考虑了内容组织的系统性和完整性,从学生的认知规律出发,力求做到简明和便于教学的特色;以培养学生分析问题和解决问题的能力为目标,着重基本概念、基本原理和基本分析方法的论述。本套丛书特别突出了各项技术的实用性,可作为计算机网络和通信专业或相近专业本科生、研究生的教科书,同时,还可以作为从事网络系统开发的科研人员和相关行业技术人员、管理人员有用的参考资料。

在撰写过程中参阅了大量的参考书、论文和资料,这里谨向所有的作者致以崇高的敬意!

我们欢迎更多的优秀教师参与到教材建设中来,真诚希望广大教师、学生与读者朋友在使用本丛书过程中提出宝贵的意见和建议。若有投稿或建议,请发至本丛书出版者电子邮件: [textbook@bhp.com.cn](mailto:textbook@bhp.com.cn)

21世纪高等院校计算机网络与通信教材编委会

# 前 言

自从香农 1948 年发表奠定信息论理论基础的“通信的数学理论”一文以来，信息理论有了很大的发展并已经延伸到众多领域中。随着信息论的发展，它不仅在通信、计算机以及自动控制等电子学领域中得到直接的应用，而且还广泛地渗透到社会学、生物学、医学、生理学、语言学和经济学等各领域，向多学科结合方向发展。人们已经认识到，在现代科学技术高度发展的过程中，学习和掌握信息理论日益成为人们的一种需要。

本教材系统地介绍了信息论与编码理论的基本内容及其应用。全书分两大部分，共 10 章，其中第 1 章详细地介绍了信源的数学模型和各类信源的信息度量、信息熵及其性质；第 2 章介绍了信道的数学模型和分类，信道传输的平均互信息，信道容量的概念及其各种计算方法；第 3 章介绍了离散信源的无失真编码，主要包括离散无记忆信源的等长和变长编码定理以及典型的变长码的编码方法；第 4 章主要介绍了有噪信道的编码问题，包括译码规则、编码方法等信息在信道传输的影响以及在有噪信道中实现信息可靠传输的有噪信道编码定理；第 5 章主要介绍了信息率失真理论的基本内容，侧重讨论离散无记忆信源，包括信源的失真测度、信息率失真函数及其计算、限失真信源编码定理；第 6 章介绍了分组码的基础知识及校验矩阵和生成矩阵的概念，讨论了分组码纠错能力、译码规则和原理；第 7 章简要介绍了相关的代数知识，特别是有限域和多项式的理论，详细讲述了循环码的概念、编译码方法；第 8 章简要介绍了卷积码的基本概念、定义和它们的各种表示方法，主要讲解了维特比译码算法，分析了维特比译码的性能；第 9 章主要介绍 Turbo 编码器，译码器和迭代算法以及交织器的设计；第 10 章立足实际，分别介绍了分组码卷积码以及 Turbo 码的软硬件实现方法实例，本章内容对纠错编码的实际应用具有指导作用。

全书注重基本概念、基本理论和基本分析方法的阐述，力求物理概念清晰，数学结构简明、完整。在内容编排上力求由浅入深、循序渐进，结合实例讲解，以最易接受的方式介绍信息论与编码理论的基本内容及其应用。

本教材主要由朱诗兵、刘作学、李迎春、高娟、曲炜、王洁南和汪建立编写。在本书的编写过程中，得到了赵洪利教授、王擎天教授及其他多位老师的指导帮助，罗春燕、庞海生、徐珍妮、王明娟、陈海小、包永平、周剑峰等多人参与了编写、文字校对等工作，在此一并表示诚挚的感谢。限于编者视野及学术水平，书中难免还存在一些缺点和错误，殷切希望广大读者批评指正。

编 者

# 目 录

引言.....	1	0.3 信息论对编码的指导意义.....	3
0.1 信息的概念.....	1	0.4 信息论与编码理论的形成及发展.....	4
0.2 信息论的研究对象.....	2		

## 第一部分 信息论

<b>第 1 章 信源模型及信息的度量</b> .....	7	2.1.4 单符号离散信道的数学模型.....	54
1.1 信源模型及分类.....	7	2.2 信道传输的平均互信息.....	56
1.1.1 单符号的离散信源.....	7	2.2.1 损失熵和噪声熵.....	57
1.1.2 单符号的连续信源.....	8	2.2.2 平均互信息.....	58
1.1.3 多符号的离散信源.....	9	2.2.3 平均条件互信息.....	58
1.1.4 多符号的连续信源.....	10	2.2.4 平均互信息的特性.....	62
1.2 不确定性与信息量.....	10	2.3 离散信道的信道容量.....	68
1.2.1 非平均不确定性.....	11	2.3.1 信道容量的定义.....	68
1.2.2 非平均信息量.....	11	2.3.2 简单离散信道的信道容量.....	69
1.3 平均信息量.....	16	2.3.3 对称离散信道的信道容量.....	71
1.3.1 信息熵.....	16	2.3.4 一般离散信道的信道容量.....	75
1.3.2 条件熵.....	18	2.4 连续/波形信道的信道容量.....	83
1.3.3 联合熵.....	19	2.4.1 连续信道的信道容量.....	83
1.3.4 平均互信息量.....	20	2.4.2 波形信道的信道容量.....	84
1.3.5 各类熵之间的关系.....	21	2.5 信道的组合.....	85
1.3.6 信息熵的基本性质.....	24	2.5.1 串联信道及其信道容量.....	85
1.4 消息序列的熵.....	28	2.5.2 并联信道及其信道容量.....	89
1.4.1 消息序列信息量的一般表 达式.....	28	2.6 信源与信道的匹配.....	91
1.4.2 离散无记忆的扩展信源.....	30	2.7 习题.....	92
1.4.3 离散平稳信源.....	31	<b>第 3 章 无失真信源编码</b> .....	95
1.5 连续信源和波形信源的信息度量.....	39	3.1 编码器.....	95
1.5.1 连续信源的差熵.....	39	3.2 离散无记忆信源的等长编码.....	99
1.5.2 波形信源的差熵.....	44	3.3 离散无记忆信源的变长编码.....	102
1.6 信源的相关性和剩余度.....	45	3.3.1 码的分类和主要编码方法.....	102
1.7 习题.....	47	3.3.2 克拉夫特不等式和麦克米伦 不等式.....	103
<b>第 2 章 信道模型及信道容量</b> .....	50	3.3.3 惟一可译码判断准则.....	108
2.1 信道的数学模型和分类.....	50	3.4 变长编码定理.....	109
2.1.1 信道的分类.....	50	3.4.1 码的平均长度.....	109
2.1.2 离散信道的数学模型.....	51	3.4.2 变长无失真信源编码定理.....	112
2.1.3 连续/波形信道的数学模型.....	52		

3.5	变长编码方法.....	116
3.5.1	香农编码方法.....	116
3.5.2	费诺编码.....	118
3.5.3	霍夫曼编码.....	120
3.6	习题.....	126
第4章	有噪信道编码.....	129
4.1	错误概率和译码规则.....	129
4.2	两种典型的译码规则.....	131
4.3	错误概率与编码方法.....	135
4.4	汉明距离.....	141
4.5	有噪信道编码定理.....	144
4.6	习题.....	146
第5章	限失真信源编码.....	148
5.1	失真测度.....	148

5.1.1	失真度.....	149
5.1.2	平均失真度.....	152
5.2	信息率失真函数及其性质.....	153
5.2.1	$D$ 允许信道(试验信道).....	153
5.2.2	信息率失真函数的定义.....	153
5.2.3	信息率失真函数的性质.....	154
5.3	信息率失真函数的参量表述及其 计算.....	161
5.3.1	利用信源的对称性来计算 信息率失真函数.....	161
5.3.2	信息率失真函数的参量表述.....	163
5.4	限失真信源编码定理.....	170
5.5	习题.....	174

## 第二部分 编码理论

第6章	线性分组码.....	178
6.1	分组码的基本概念.....	178
6.1.1	引言.....	178
6.1.2	分组码的基本概念.....	178
6.1.3	分组码相关基础知识.....	180
6.2	线性分组码的生成矩阵和校验矩阵.....	183
6.2.1	生成矩阵.....	183
6.2.2	校验矩阵.....	185
6.2.3	汉明码.....	186
6.3	线性分组码的纠错能力.....	187
6.4	线性分组码的译码.....	190
6.4.1	相关定义和定理.....	190
6.4.2	标准阵列译码法与伴随式 译码法.....	191
6.5	由一个已知码构造新码的方法.....	193
6.5.1	扩展码.....	193
6.5.2	删除码.....	194
6.5.3	凿孔码.....	194
6.5.4	增广码(增信删余码).....	194
6.5.5	增余删信码.....	194
6.5.6	缩短码.....	195
6.5.7	延长码(增信码)和RM码.....	195
6.6	交织编码简介.....	196

6.6.1	引言.....	196
6.6.2	交织码的基本原理.....	196
6.6.3	分组交织器的基本性质.....	197
6.6.4	交织编码的应用实例.....	198
6.7	线性码的重量分布与译码错误概率 计算.....	199
6.7.1	线性码的重量分布.....	199
6.7.2	线性码译码错误概率计算.....	200
6.7.3	误码率计算.....	201
6.8	习题.....	202
第7章	循环码.....	204
7.1	循环码代数的基本知识.....	204
7.1.1	有限域的定义.....	204
7.1.2	域上的多项式.....	205
7.1.3	$GF(2^m)$ 的构成.....	206
7.2	循环码的定义及多项式表示.....	207
7.2.1	循环码的基本概念.....	207
7.2.2	码的多项式描述.....	208
7.3	循环码的矩阵描述.....	211
7.4	系统循环码的编码方法及实现电路.....	213
7.4.1	循环码的编码方法.....	213
7.4.2	多项式运算电路.....	214
7.4.3	系统循环码编码的实现电路.....	217



7.5	系统循环码的译码方法及实现电路	220	9.1.1	递归系统卷积码 (RSC 码)	253
7.5.1	接收多项式伴随式的计算	220	9.1.2	删余器	255
7.5.2	循环码的译码算法	223	9.2	Turbo 码的译码器	256
7.6	3 种重要的循环码	225	9.3	译码算法	257
7.6.1	BCH 码	225	9.3.1	MAP 算法	259
7.6.2	RS 码	227	9.3.2	MAP 算法在 Turbo 译码中 的应用	261
7.6.3	缩短循环码	228	9.3.3	SOVA 算法	262
7.7	习题	229	9.3.4	性能的仿真	264
<b>第 8 章</b>	<b>卷积码</b>	<b>231</b>	9.4	Turbo 码中的交织器	265
8.1	卷积码的基本概念	231	9.5	Turbo 码研究的方向	267
8.2	卷积码的描述	234	9.6	习题	268
8.2.1	卷积码的矩阵和多项式描述	234	<b>第 10 章</b>	<b>差错控制编码的实现</b>	<b>269</b>
8.2.2	卷积码的树图描述	237	10.1	分组码的 Systemview 实现	269
8.2.3	卷积码的状态图描述	238	10.2	分组码在超短波通信系统中的 实现	272
8.2.4	卷积码的篱笆图描述	239	10.2.1	BCH 编码与译码	273
8.3	卷积码的维特比译码算法	239	10.2.2	交织编码	276
8.3.1	最大似然译码	240	10.2.3	纠错码性能分析	278
8.3.2	维特比译码的基本原理	241	10.3	卷积码的 Systemview 实现	279
8.3.3	软判决维特比译码	244	10.4	卷积码的 DSP 实现	281
8.4	维特比译码的性能	246	10.5	Turbo 码的 Matlab 实现	285
8.4.1	二进制对称信道下的维特比 译码的性能	247	10.5.1	发送端的实现	286
8.4.2	高斯白噪声信道中维特比 译码器输出的误码率	247	10.5.2	信道的模拟实现	287
8.5	维特比译码的实用	249	10.5.3	接收端的实现	287
8.6	习题	251	<b>参考文献</b>	<b>288</b>	
<b>第 9 章</b>	<b>Turbo 编译码与迭代译码算法</b>	<b>252</b>			
9.1	Turbo 码编码器	252			

# 引 言

美国科学家香农 (C.E.Shannon) 于 1948 年发表了著名的论文《通信的数学理论》(A Mathematical Theory of Communication), 用概率测度和数理统计的方法系统地研究了通信的基本问题, 给出了信息的度量表示, 得出了带有普遍意义的重要结论, 由此奠定了信息论的理论基础, 为信息的表达、存储、传送和处理的具体实现提供了理论依据。

在香农信息论的指导下, 为提高通信系统信息传输的有效性和可靠性, 人们在信源编码和信道编码两个领域进行了卓有成效的研究, 取得了丰硕的成果。近几十年来, 随着信息概念的不断深化和信息理论的迅猛发展, 信息论所涉及的内容早已超出了狭义的通信工程范畴, 它已渗透到许多学科, 日益得到众多领域中的科学工作者的重视。

在引言里主要阐述与“信息论与编码理论”有关的一些基本概念, 即信息的概念, 信息论所研究的对象, 信息论对信源编码、信道编码研究的指导意义以及信息论与编码理论的发展历程。

## 0.1 信息的概念

信息是信息论中最基本、最重要的概念。那么什么是“信息”呢?

“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。

在日常生活中, 人们也常常错误地把信息等同于消息, 认为得到了消息, 就是得到了信息。在通信系统中传输的是各种各样的消息, 而这些被传送的消息有着各种不同的形式。

那么什么是消息呢? 用文字、符号、数据、语言、音符、图片、图像等能够被人们感觉器官所感知的形式, 把客观物质运动和主观思维活动的状态表达出来就成为消息。

从通信的观点出发, 构成消息的各种形式要具有两个条件, 一是能够被通信双方所理解; 二是可以传递。各种通信系统中, 其传输的形式是消息。但消息传递过程中的一个最基本、最普通却又不十分引人注意的特点是:

(1) 收信者在收到消息以前, 是不知道消息的具体内容的。也就是说, 在收到消息以前, 收信者无法判断发送者将会发来描述何种事物运动状态的具体消息。

(2) 收信者在收到消息以前, 无法判断是描述这种状态还是那种状态。

(3) 收信者在收到消息后, 由于干扰的存在, 不能断定所得到的消息是否正确和可靠。

总之, 收信者存在着“不知”、“不确定”或“疑问”。通过消息的传递, 收信者知道了消息的具体内容, 原先的“不知”、“不确定”和“疑问”消除或部分消除了。因此, 对收信者来说, 消息的传递过程是一个从不知到知的过程; 或是从知之甚少到知之甚多的过程; 或是从不确定到部分确定或全部确定的过程。如果一个通信系统不具备这样的特点, 那就失去了其存在的意义。

可见, 通信过程是一种消除不确定性的过程。不确定性的消除, 就获得了信息。原先的不确定性消除得越多, 获得的信息就越多。若原先的不确定性全部消除了, 就获得了全部的信息; 若消除了部分不确定性, 就获得了部分信息; 若原先不确定性没有任何消除,

就没有获得任何信息。

那么如何定义信息呢？信息是事物运动状态或存在方式的不确定性的描述。这就是香农信息的定义。

可见，在通信系统中形式上传输的是消息，但实质上传输的是信息。消息中包含信息，是信息的载体，得到消息，从而获得信息。同一则信息可以由不同形式的消息来载荷；而一则消息也可载荷不同的信息，它可能包含非常丰富的信息，也可能只包含很少的信息。可见，信息与消息是既有区别又有联系的。

既然信息不同于消息，当然也不同于信号。

在各种实际通信系统中，往往为了克服时间或空间上的限制而进行通信，必须对消息进行各种加工处理。把消息变换成适合信道传输的物理量，这种物理量称为信号（如电信号、光信号、声信号、生物信号等）。

信号是一个物理量，可测量，可描述，可显示，它携带着消息，是消息的运载工具。而信息与信号也有着本质的区别，信号是载荷信息的实体，信号仅仅是外壳，信息则是内核。可见，信息、消息和信号是既有区别又有联系的3个不同的概念。

## 0.2 信息论的研究对象

从关于信息概念的讨论中，可以看到：虽然各种通信系统的形式和用途各不相同，但本质是相同的，即都是信息的传输系统。为了便于研究信息传输和处理共同规律，可以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，即通信系统模型，如图0.1所示。

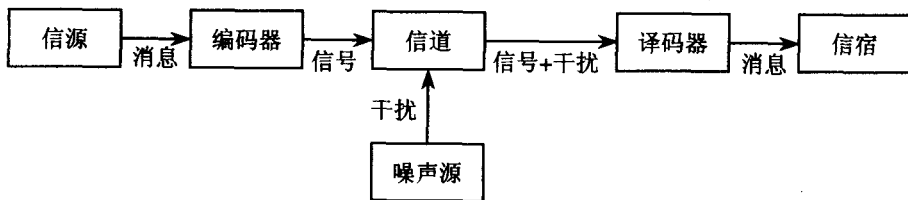


图 0.1 通信系统模型

信息论研究的对象正是这种统一的通信系统模型。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共同规律。

这个模型主要包括如下5个部分。

### 1. 信源

信源是产生消息或消息序列的源。信源的输出是消息，消息是具体的，但它不是信息本身。信源输出的消息有多种形式，可以是离散的或连续的，也可以是时间序列，它们分别可用离散型/连续型随机变量、随机矢量及随机过程等数学模型表示。

### 2. 编码器

编码是把消息变换成信号的措施，是对消息符号进行编码处理的过程，而译码就是编码的反变换。编码器输出的是适合信道传输的信号。信号携带着消息，它是消息的载荷者。

编码器可分为 2 种, 即信源编码器和信道编码器。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。而信道编码是为了提高信息传输的可靠性而对信号进行的变换和处理。香农信息论分别用几个重要的定理给出了编码的理论性能 极限。

### 3. 信道

信道是指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实际信道有明线、电缆、波导、光纤、无线电波传播空间等, 这些都是属于传输电磁波能量的信道。当然, 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, 还有存储信号的作用。如书写通信方式就是一例。

信息的传输不可避免的会引入噪声和干扰, 为了分析方便, 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。而噪声源的统计特性又是划分信道的依据。

### 4. 译码器

译码就是把信道输出的编码信号(已叠加了干扰)进行反变换, 以尽可能准确地恢复原始的信源符号。译码器也可分成信源译码器和信道译码器。

### 5. 信宿

信宿是消息传送的对象, 即接收消息的人或机器。

图 0.1 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。在实际的情况下, 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是双向进行的。例如, 广播通信是单个输入、多个输出的单向传输的通信; 而 Internet 通信则是多个输入, 多个输出的多向传输的通信。要研究这些通信系统, 只需对两端单向通信系统模型作些适当修正, 就可引出多用户通信系统模型。因此, 图 0.1 所示的是最基本的通信系统模型, 是信息论所要研究的对象。

## 0.3 信息论对编码的指导意义

信息传输的可靠性是所有通信系统努力追求的首要目标。要实现高可靠性的传输, 可采取诸如增大发射功率、增加信道带宽、提高天线增益等传统方法, 但这些方法往往难度比较大, 有些场合甚至无法实现。而香农信息论指出, 进行适当的信道编码后, 同样可以提高信道的传输可靠性, 这种编码即是信道编码。信道编码是在著名的信道编码定理指导下发展起来的, 几十年来已取得了丰硕的成果。

信息传输的有效性是通信系统追求的另一个重要目标。有效性是指在一定的时间内如何传输尽可能多的信息量, 或者在每一个传送符号内携带尽可能多的信息量, 这就需要对接源进行高效率的压缩编码, 尽量去除信源中的多余度。

实际信源, 如图像、语音、文本数据等都存在着大量的多余度(剩余度), 信源多余度有多种形式, 如统计多余度、结构多余度、视觉多余度、时间多余度、空间多余度等, 不

同类型的多余度需采用不同的方法消除。

香农信息描述的是事物的不确定性，因此，相应地香农信息论讨论的多余度是统计多余度。这种多余度包括信源前后符号间相关性带来的多余度和信源符号分布不均匀导致的多余度。统计多余度在各种信源中是普遍存在的，如何在无失真或限定失真的条件下对信源进行高效压缩是香农信息论研究的重点，香农第一定理和香农第三定理分别从理论上给出了无失真信源编码与限失真信源编码的压缩极限，对于压缩编码的研究具有重要的理论指导意义。香农信息论对信源统计多余度的透彻分析为各种具体压缩编码方法的研究提供了明确的思路，如变换编码、预测编码、统计编码等均是行之有效的信源压缩编码方法，并且在目前的视音频压缩国际标准中得到了广泛的采用。

值得注意的是，香农信息论仅讨论了统计多余度的去除，而未涉及到其他类型的多余度，事实上对它们的深入研究，同样可以对提高压缩编码效率有所作为，而这正是目前人们对小波变换、分形编码、模型编码等新压缩方法研究兴趣甚浓的原因。

在以后的学习中，可以看到，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，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。

## 0.4 信息论与编码理论的形成及发展

信息论从诞生到今天，已有 50 多年的历史，现已成为一门独立的理论科学。而编码理论与技术研究也有 40 多年历史了，并从刚开始时作为信息论的一个组成部分逐步发展成为比较完善的独立体系。回顾它们的发展历史，可以清楚地看到理论是如何在实践中经过抽象、概括、提高而逐步形成和发展的。

信息论与编码理论是在长期的通信工程实践和理论研究的基础上发展起来的，其形成的历史可以上溯到 19 世纪 30 年代。

1832 年，莫尔斯电报系统中高效率编码方法对后来香农的编码理论是有启发的。

1885 年，凯尔文 (L.Kelvin) 曾经研究过一条电缆的极限传信率问题。

1917 年，坎贝尔 (G.A.Campbell) 申请了第一个关于滤波器的专利，为频分复用信道提供了条件。

1922 年，卡逊 (J.R.Carson) 对振幅调制信号的频谱结构进行了研究，开始明确上下边带的概念。

1924 年，奈奎斯特 (H.Nyquist) 开始分析电报信号传输中脉冲速率与信道带宽的关系，建立了限带信号的采样定理。

1928 年，哈特莱 (R.V.Hartley) 发展了奈奎斯特的工作，第一次从通信的观点出发对信息量作了定义，提出把消息考虑为代码或单语的序列。在  $s$  个代码中选  $N$  个码即构成  $s^N$  个可能的消息，提出“定义信息量  $H = N \log s$ ”，即定义信息量等于可能消息数的对数。其缺点是没有统计特性的概念。哈特莱的工作对后来香农的思想是有很大影响的。

1930 年，维纳 (N.Wiener) 开始把 Fourier 分析方法全面引入到随机信号的研究中来。

1936 年，兰登 (V.D.Landon) 发表了他第一篇有关噪声的论文。

1936年,阿姆斯特朗(E.H.Armstrong)提出频率调制,指出增加信号带宽可以使抑制噪声干扰的能力增强,使调频实用化,出现了调频通信装置。

1939年,达德利(H.Dudley)发明了声码器。提出了通信所需要的带宽至少应与所传送的消息的带宽相同。达德利和莫尔斯都是研究信源编码的先驱者。

1939年,瑞弗(H.Reeve)提出了具有强干扰能力的脉冲调制。

20世纪40年代初期,维纳在研究防空火炮的控制问题时,发表了“平稳时间序列的外推、内插与平滑及其工程应用”的论文。他把随机过程和数理统计的观点引入到通信和控制系统中来,揭示了信息传输和处理过程的统计本质。他还利用自己提出的“广义谐波分析理论”对信息系统中的随机过程进行了谱分析。这使得通信系统的理论研究上了一个新台阶。

1948年,香农在贝尔系统技术杂志上发表了两篇有关“通信的数学理论”的文章。在这两篇论文中,他利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,得出了几个重要而带有普遍意义的结论,并由此奠定了现代信息论的基础。

香农的论文“通信的数学理论”发表后,在通信技术界,科学工作者的主要精力转到信源编码和信道编码的具体构造方法上,这方面取得了稳步的发展。

### 1. 无失真信源编码

在香农编码方法提出后,许多科学家对无失真信源编码进行了大量的研究。

1952年,费诺(Fano)提出了一种费诺编码方法。同年,霍夫曼(D.A.Huffman)提出了Huffman编码方法,并证明它是一种最佳码。

1963年,埃利斯(P.Elias)提出了算术编码方法。

1965年,柯尔莫哥洛夫(A.N.Kolmogorov)提出了通用编码方法。

这些编码方法经过改进都先后实用化。例如,Huffman编码用于传真图像的压缩标准,算术编码用于二值图像的压缩标准JBIG,通用编码用于计算机文件的压缩等。

### 2. 有失真信源编码

量化这一最古老的方法经过发展现在已经成为语音和图像压缩的最重要的手段。例如,北美移动通信标准IS-95中语音压缩的标准算法就是矢量量化算法。

1955年,埃利斯(P.Elias)提出了预测编码方法。经过发展,现已成为美国军用通信语音压缩的标准算法。

1959年,香农发表了“保真度准则下的离散信源编码定理”,以后发展成为“信息率失真理论”。这一理论是信源编码的核心问题,是频带压缩、数据压缩的理论基础。直到今天它仍是信息论研究的课题。

1969年,T.S.Huang提出了分组交换与量化方法。经过发展,现在已在电视图像压缩的各种标准如H.261、JPEG、MPEG中得到应用。

### 3. 面向数字信道的信道编码

另外一部分科学家从事寻找最佳编码(纠错码)的研究工作,并已经形成一门独立的分支——纠错码理论。

20世纪40年代末,M.J.E.Golay和汉明提出分组编码技术,把代数方法引入到纠错码

的研究，形成了代数编码理论，由此找到了大量可纠正多个错误的性能优异的码，而且提出了可实现的编译码方法。分组码中的不少码，如汉明码、Golay 码、Fire 码、BCH 码等都在通信、计算机技术中获得广泛应用。

但是代数编码的渐近性能很差，不能实现香农信道编码定理所指出的结果。因此，于 1960 年前后，提出了卷积码和概率译码，并逐步形成了一系列概率译码理论。尤其，以维特比（Viterbi）译码为代表的译码方法被美国卫星通信系统所采用，使香农理论成为真正具有实用意义的科学理论。1993 年提出的 Turbo 码在性能上已非常接近理论极限。

#### 4. 面向模拟信道的信道编码

1974 年，J.L.Massey 提出将编码与调制统一考虑的概念。1982 年，这一想法在 G.Ungerboeck 等人的研究下终于取得突破，这就是网格编码调制。网格编码调制在实际应用中发生的相位含糊问题在 1984 年被 L.E.Wei 所解决，这一方法随即被 CCITT（现为 ITU-T）所采纳成为一种标准。现在，网格编码调制正在向卫星通信、磁纪录等领域扩展其应用范围。

从上面所阐述的信息论与编码理论的形成及发展历程来看，信息论是从最初形成时仅提供性能极限和进行概念方法性指导，发展到今天具体指导通信系统的结构组织和部件的设计，这种趋势势必还会进行下去，而信息论也将在与通信理论、通信系统设计的理论日益融合的过程中得到进一步的发展。

# 第一部分 信息论

## 第 1 章 信源模型及信息的度量

为了提高信息传输的有效性，本章将对信息的来源——信源进行分析研究，重点讨论信源的统计特性和数学模型，以及各类信源的信息度量——熵及其性质，从而引入信息论的一些基本概念和重要结论。这部分内容是香农信息论的基础。

### 1.1 信源模型及分类

信源是信息的来源，但信息是较抽象的东西，所以要通过信息的表达者——消息来研究信源，这里对信源的内部结构，为什么产生和怎样产生各种不同的消息都不作研究，而只研究信源的输出，以及信源输出各种可能消息的不确定性。

如何对信源输出的消息进行描述呢？先理解下面 3 个基本概念：

#### 1. 样本空间

把某事物各种可能出现的不同状态，即所有可能选择的消息的集合，称为样本空间。每个可能选择的消息是这个样本空间的元素。

#### 2. 概率测度

对于离散消息的集合，概率测度就是对每一个可能选择的消息指定一个概率（非负的，且总和为 1）。

#### 3. 概率空间

一个样本空间和它的概率测度统称为一个概率空间。

正如引言中所述，在通信系统中受信者在未收到消息以前，对信源发出什么消息是不确定的，是随机的，所以可用随机变量或随机矢量来描述信源输出的消息。或者说，用一个样本空间及其概率测度——概率空间来描述信源。

不同的信源输出的消息不同，可以根据消息的不同随机性质来对各类信源进行分类，具体如下。

#### 1.1.1 单符号的离散信源

有些信源可能输出的消息数是有限的或可数的，而且每次只输出其中一个消息。例如，扔一颗均匀的骰子，分析研究其下落后，朝上一面的点数。每次实验结果必然是 1 点、2 点、3 点、4 点、5 点、6 点中的一个面朝上。其输出消息是“朝上一面是 1 点”、“朝上一面是 2 点”……“朝上一面是 6 点”等 6 个不同的消息。每次实验中，出现哪一种消息是随机的，但必定是出现这 6 个消息中的某一个消息。这 6 种不同的消息构成两两互不相容



的基本事件集合, 用符号  $a_i, i=1, \dots, 6$  来表示这些消息。另外由大量实验表明, 各消息都是等概率出现的, 即都等于  $1/6$ 。因此, 可以用随机变量  $X$  来描述这个信源输出的消息。这样, 这个信源抽象后得到的数学模型为:

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6} \end{bmatrix}$$

并且各事件的出现概率满足:

$$\sum_{i=1}^6 p(a_i) = 1$$

上式表示这个离散信源的概率空间是一个完备集, 信源输出的消息只可能是符号集  $\{a_1, a_2, a_3, \dots, a_6\}$  中的任何一个, 而且每次必定选取其中一个。

信源输出的消息数是有限的或可数的, 而且每次只输出符号集中的一个消息, 这样的信源称为单符号的离散信源, 可用一维离散型随机变量来描述这些消息。其数学模型就是离散型的概率空间:

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & \dots, & a_q \\ p(a_1), & p(a_2), & \dots, & p(a_q) \end{bmatrix}$$

并满足

$$\sum_{i=1}^q p(a_i) = 1$$

此式表示信源可能取的消息(符号)只有  $q$  个:  $\{a_1, a_2, \dots, a_q\}$ , 而且每次必定取其中一个。

可见, 当信源给定, 其相应的概率空间就已给定; 反之, 如果概率空间给定, 这就表示相应的信源给定。所以, 概率空间能表征离散信源的统计特性。

在实际情况下, 存在着许多这种消息数有限或可数的信源, 如计算机的代码、电报符号、阿拉伯数字码等。

### 1.1.2 单符号的连续信源

信源的输出是单个符号(代码)的消息, 但其可能出现的消息数是不可数的无限值, 即输出消息的取值是连续的, 这样的信源称为单符号的连续信源, 可用一维连续型随机变量来描述这些消息。其数学模型为连续型的概率空间:

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} (a, b) \\ p(x) \end{bmatrix} \text{ 或 } \begin{bmatrix} R \\ p(x) \end{bmatrix}$$

并满足

$$\int_a^b p(x) dx = 1 \text{ 或 } \int_R p(x) dx = 1$$

其中  $R$  表示实数集  $(-\infty, \infty)$ , 而  $p(x)$  是随机变量  $X$  的概率密度函数。

在实际情况下, 存在着许多这种消息数是不可数的无限值的信源, 如语音信号、遥控系统中测得的电压、温度、压力等连续数据。