国外信息科学经典教材

Prentice Hall

工程应用编码与信息理论

Applied Coding and Information Theory for Engineers

(美) Richard B. Wells 著

尹长川 罗 涛 藤 勇 等译

乐光新 审

机械工业出版社
CHINA MACHINE PRESS

国外信息科学经典教材

工程应用编码与信息理论

（美） Richard B. Wells 著

尹长川 罗 涛 滕 勇 等译

乐光新 审

机 械 工 业 出 版 社

本书从工程应用角度，用大量实例，系统地讲述了编码与信息理论。主要内容包括：离散信源和熵、信道和信道容量、游程长度受限码、线性分组纠错码、循环码、卷积码、网格编码调制、信息论和密码学、仙农编码定理等。

作者在讲述深奥理论的过程中，尽量避开以往理论化、形式化、论证式的组织形式，以不拘泥形式的、富于趣味性的、易于接受的陈述方法，紧密结合工程实践，将编码与信息理论的知识呈现在读者面前。

本书适合作为高等院校通信、信息、电子等相关专业的教学用书，也可供相关工程技术人员参考。

ISBN:0-13-961327-7

北京市版权局著作权合同登记号：图字：01-2002-6558

图书在版编目(CIP)数据

工程应用编码与信息理论/(美)韦尔斯(Wells, R. B.)著；尹长川等译. —北京：机械工业出版社，2003.3
书名原文：Applied Coding and Information Theory for Engineers
(国外信息科学经典教材)
ISBN 7-111-11818-9

Ⅰ.工… Ⅱ.①韦…②尹… Ⅲ.①通信工程—编码理论—高等学校—教材②通信工程—信息论—高等学校—教材 Ⅳ.TN911.2

中国版本图书馆 CIP 数据核字(2003)第 018078 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)
北京中加印刷有限公司印刷·新华书店北京发行所发行
责任编辑：刘 青 责任印制：付方敏
2003 年 4 月第 1 版·第 1 次印刷
787mm×1092mm 1/16·16.5 印张·402 千字
0 001—5 000 册
定价：32.00 元

出版说明

在人类迈入信息时代的今天，信息技术的应用无处不在，我国对信息技术的重视和鼓励也达到了空前的程度。信息技术的发展速度很快，可谓日新月异，尤其在一些发达的国家更是如此。我国信息技术起步较晚，但发展速度惊人，这正是改革开放的具体体现。在我国大力发展信息产业的今天，为了能融入国际的潮流和掌握最新的技术，从国外引进先进的知识和技术就显得格外重要。为此，我们决定引进一系列国外信息技术领域有代表性的优秀教材，将它们献给我们的学子、教师和 IT 业的有志之士，藉此为我国的信息产业贡献一份微薄之力。

随着我国加入 WTO，国际间的竞争将越来越激烈，而国际间的竞争实际上就是人才的竞争、教育的竞争。为了加快培养具有国际竞争力的高水平技术人才，加快我国教育改革的步伐，使我国的高等教育尽快与国际接轨，这就需要引进先进的教学思想和教学方法，而引进国外优秀的教材无疑是一种很好的途径。同时，引进国外的优秀教材也有利于提高我国自编教材的水平，让我们的教育工作者从中得到启发。

我们这套丛书遵循“新、优、特”的原则，做到知识新、质量优和内容有特点。这套丛书涵盖了计算机、通信、电子技术等领域，每一本书都是精心挑选，在某个领域或学科内具有很强的代表性和很高的价值，很多在国外也被作为大学的教科书，由国际知名的出版公司出版。在引进过程中，我们邀请有关专家对书稿的整体水平进行了评定；在翻译过程中，我们聘请国内相关领域内的有很高学术水平的专家和学者，以保证书籍的水平和质量，做到对读者负责，为读者着想。

相信这套丛书的出版对正在苦读和即将面临挑战的学子们会有很大的帮助和提高，也能让我们的教学工作者从中得到启发，同时对从事 IT 行业的工程技术和研究人员而言也是很难得的工具书。

译者的话

本书是一本有关编码与信息理论的入门性教科书，与传统的相同体裁的教科书相比有很大不同。本书考虑的读者对象是面向应用的工程师和不具有很多近代代数知识的大学本科高年级学生或者低年级的研究生。本书在写作手法上抛弃了传统的“定理—证明”的讲述方式，而是代之以更多的实际例子，向面向工程的学生们讲述编码与信息理论的基础知识。本书内容较为丰富，涵盖了信源编码、信道编码（分组码、卷积码和 TCM 码）、数据变换码、密码学和仙农信息论等方面的内容，适合作为相关专业高年级本科生和低年级研究生的学习用书，也可作为工程技术人员的参考工具。

本书由北京邮电大学电信工程学院尹长川副教授主持翻译。其中尹长川负责前言及第1、3章的翻译，罗涛负责第5、9章和习题解答，秦升平负责第2章，吴军力负责第4章，滕勇负责第6、7章，王健康负责第8章的翻译。尹长川、罗涛、吴军力和侯晓林负责全书译稿的校对，并由尹长川负责全书的统稿。乐光新教授审阅了全部译稿并提出了若干改进意见。

由于时间仓促，加之译者水平有限，错误和不当之处在所难免，敬请各位读者批评指正。

译　者

前　言

欢迎学习信息与编码理论。正如大家所说，我们正生活在信息时代的黎明，这是一个被比作下一次工业革命的时代。虽然这种看法经常被提及，以致于显得有些老生常谈，但是这个新时代对于商业、工业和社会的潜在重要性是不管怎么说都不会夸大的。当工业革命到来的时候，随之带来的新需求要求人们在技术科学、艺术和技能方面都要熟悉。类似的道理，随着信息革命带来的是需要大量的人员通晓和掌握应用于多种用途的信息技能。

本书是为初学者写的。其取材来源于我为电气与计算机工程、计算机科学以及数学专业的本科生所开设的一门课程的课堂讲义。这本书的讲述水平是面向三年级或四年级的在校本科学生和很少或从未接触过本专题的有实践经验的工程师。这本书的目的是帮助大家在信息工程的实践方面开始起步。

近些年来，有关本专题的入门性的教科书实际上已完全绝迹。市场上确实有一些很好的研究生水平的教科书，但是对于那些需要掌握面向市场能力、任务繁重的新学生，或者正在寻找入门性的资料以便可以从紧急的新项目上起步的专业人员来说，这些教科书经常是有点太理论化而较少论及实践了。考虑到这些读者，我故意放弃了在有关本专题的多数课本上所看到的那种传统的“定理—证明”的方式。虽然本教科书的材料里也有定理出现，但是我已经试着以这样的方式组织本书的结构，即将所需的数学推导紧接在“如何”应用这些方法和定理的前面。这样，书中没有浓缩所有数学定理的大章节，各数学专题只有在需要的时候才以“救急”的方式进行介绍。

这本教材里的材料足够用做大学三年级或者四年级一学期的课程。本教材假定读者以前具有初等线性代数和基本概率论方面的知识背景。数字逻辑设计或者初步的通信系统方面的先修课程，对于本课程的学习是有帮助的，但不是必需的。

第 1 章首先对数字通信系统进行了讲述，并引入了信息的概念。我发现当学生们学到“信息”和“数据”是不同的概念时经常会感到惊奇。本章介绍了离散信源以及熵和联合熵的基本概念，由此引出了用于数据压缩的信源编码的介绍，我们将理论应用于霍夫曼(Huffman)编码、莱姆培尔-兹夫(Lempel-Ziv)编码和算术编码。对信源建模和自适应编码的内容只进行了简单介绍，但是为希望深入学习这些重要专题的读者提供了参考文献。

在第 2 章我们继续进行信息理论的学习，引入了离散无记忆信道的概念。在描述和定义了这些信道之后，我们介绍了互信息和信道容量，描述了用于计算离散无记忆信道的信道容量的阿里莫托-布拉哈特(Arimoto-Blahut)算法。我们还比较详细地介绍了非常重要的二进制对称信道，由此引出了分组编码的思想和著名的仙农(Shannon)第二定理。本章还介绍了马尔可夫(Markov)过程和有记忆信道，并由此向大家引出了许多重要概念。接下来介绍了受限信道，以及序列的自相关函数和功率谱的重要概念。在本章的最后，我们将理论应用于数据变换码(data translation codes)，并介绍了游程受限(d, k)码。

整个第 3 章都是有关应用方面的内容，如果教学时间有限，教师可以跳过这一章而不影响内容的连续性。本章是有关一类特殊的数据变换码，该类编码具有多种名称，分别被称做线路

码、调制码或游程长度受限码。在本章里综述了前置分组编码技术，包括状态独立的固定码率/固定分组码、用于固定长度分组码依赖于状态的编码、可变长度/固定码率的分组码、前视码(look-ahead codes)。在本章的最后，简单介绍了无直流电平码。

第4章介绍了线性分组纠错码的一般理论。本章首先讨论了编码问题和有噪信道上的错误概率计算，接下来讨论了使用二进制重复码进行错误纠正以及一些重要的界和线性分组码必须遵循的限制。接着给出了与二元域和二元矢量空间有关的一些简单背景，为更理论化的导出代数编码作准备。随后介绍了汉明(Hamming)距离、汉明重量和汉明立方体的基本思想以及一些重要的数学定义和概念。介绍了采用标准阵的译码方法，定义了系统分组码。由此引出了对汉明码的深入讨论，这是我们最先遇到的"重要的"实用码。在讨论基本汉明码的同时，我们还讨论了一些有用的"变体"，包括对偶码和扩展汉明码。讨论了用于纠错和检错的编码。在本章的最后，讨论了线性分组码的差错率以及用于纠错码和自动请求重传系统的码的性能。

第5章继续讨论线性分组码，介绍了循环分组码。在讨论了基本定义和性质之后，介绍了循环码的多项式表示，并讨论了用于由二元域构造的多项式的求模算术。接着将注意力转移到了用于循环码的产生和译码的有效方法上，给出了许多用于实现编码器和译码器的实用电路。在本章的最后，我们给出了几种有用且重要的标准码，包括汉明码(本章再次提到)、一些简单的BCH码以及一些好的纠突发错误码。本章还讨论了使用循环冗余校验(CRC)码的错误检测以及一些有用的"变体"，包括交织和截短码。

在第6章中，我们从分组码转到介绍卷积码。在讨论了基本编码器之后，我们考察了卷积码的一些结构特征以及这些码的状态图和结构图表示，讨论了码的传递函数表示及其用途。我们还深入讨论了维特比(Viterbi)算法。这里的讲法与多数的课本和文章里的讲法不同，在描述为什么该算法能工作之前，我们首先描述了该算法是什么以及是如何工作的。据我的学生们反映，这种讲法上的次序颠倒比传统的教学法更容易被接受。我们讨论了硬判决和软判决维特比译码，并比较和讨论了这两种方法的性能差异。接着以列表的形式给出了一些已知的好卷积码。我们从维特比算法回到讨论一些实际的实现问题，包括译码的回溯(traceback)方法和使用凿孔(punctured)卷积码以获得更高的码率。

第7章是对网格编码调制的简单介绍。我们介绍了二维 *I-Q* 信道以及用于这些信道的发送机和接收机，讨论了用于相位调制和正交幅度调制系统的编码信道的错误概率特性。然后介绍了系统递归卷积编码器及其网格图表示，讲述了昂格尔博克(Ungerboeck)的典型编码器，并使用奇偶校验多项式给出了TCM码的八进制表示。接下来讨论了集合分割以及如何使用集合分割来构造使用昂格尔博克编码器的TCM码。在本章的最后扼要地给出了用于相位调制和正交幅度调制的一些好码。

第8章简要介绍了信息论在密码学方面的应用。本章首先介绍了一些基于密码的简单保密系统，并简要描述了一些可以用来攻击保密系统的方法。接下来介绍了仙农(Shannon)的完善保密性的定义，并导出了达到完善保密的条件。随后讨论了自然语言的熵率以及在密码分析学中如何利用自然语言的冗余，由此引出了虚假密钥和惟一解距离的重要概念。接下来讨论了计算安全性问题，描述了仙农的扩散和混淆技术，由此引出了乘积加密系统的重要技术。最后，本章简要描述了编码、公开密钥保密系统以及某些其他问题。公开密钥保密系统虽然很重要，但是因为公开密钥保密系统的理论更牵涉到数论而不是信息论，所以本章并未对其进行

深入描述，但是向读者提供了几本好的参考文献，以供随后进一步了解公开密钥加密的理论和实践。

第9章是本书的最后一章。在这一章中，给出了二进制对称信道这一特殊情况下仙农第二定理的证明，介绍了随机编码理论的思想，并讨论了该定理向我们说明和没有说明的内容。接下来我们将注意力转到对仙农无噪声编码定理和用于进行信源压缩的前置码存在性的推导。在本章的最后用几句话对信息论进行了总结，并向读者指出了进一步学习的方向。

虽然在讲述编码与信息理论时一定程度上的循规蹈矩是必要的，但是在本书的讲述中，我尝试着有时使讲解尽可能的不那么正规，读者会不时地读到作者的一些轻松愉快的评注。作者相信，“如果课本写作时有其轻松愉快的时刻，难道阅读的时候就不能也有这样的时刻吗?”在有关此类专题的传统研究生水平的课本中，编码与信息理论的一个更重要的方面是缺少趣味性。在本书对数学理论的陈述中，我已经试着不舍弃这种趣味性。

很显然，作者在本书的编写过程中所起的作用是重要的，但是课本的价值不在于作者，而是在于其所要服务的读者。此外，任何一本教材（特别是本书）的存在都应归功于为其作出过贡献的许多人。想到此处，我首先要感谢我的学生们，他们对手稿、课后练习和题解手册向我提供了大量的反馈意见。我还要感谢艾劳恩·布雷南（Aaron Brennan）先生，感谢他帮助设计了一些课后练习。我要特别感谢滑铁卢（Waterloo）大学的乔治·弗里曼（George Freeman）博士，他富有洞察力的评价和建议在很大程度上提高了本书的质量。最后，我要感谢 Prentice-Hall 出版社快乐班子（merry band）的爱丽丝（Alice）、汤姆（Tom）和其他的人，虽然他们很少被作者提及，但是正是他们的努力使得本书比手稿有了很大提高。

Richard B. Wells

目　录

出版说明
译者的话
前言
第 1 章　离散信源和熵 …… 1
1.1　数字通信和存储系统概述 …… 1
1.2　离散信源和熵 …… 2
1.2.1　信源符号集和熵 …… 2
1.2.2　联合熵和条件熵 …… 4
1.2.3　符号块的熵和链准则 …… 6
1.3　信源编码 …… 8
1.3.1　映射函数和效率 …… 8
1.3.2　互信息 …… 9
1.3.3　短暂的离题——关于加密 …… 11
1.3.4　本节小结 …… 13
1.4　霍夫曼(Huffman)编码 …… 13
1.4.1　前置码和即时译码 …… 13
1.4.2　霍夫曼码的构造 …… 14
1.4.3　硬件实现方法 …… 16
1.4.4　霍夫曼编码效率的稳健性 …… 17
1.5　词典码和莱姆培尔-兹夫(Lempel-Ziv)编码 …… 18
1.5.1　动态词典编码的基本原理 …… 18
1.5.2　链接表 LZ 算法 …… 19
1.5.3　译码过程 …… 21
1.5.4　LZ 压缩的大数据块要求 …… 22
1.6　算术编码 …… 23
1.6.1　码字长度和渐近均分性质 …… 23
1.6.2　算术编码方法 …… 25
1.6.3　算术码的译码 …… 27
1.6.4　算术编码的其他问题 …… 28
1.7　信源模型和自适应信源编码 …… 28
1.8　小结 …… 29
1.9　习题 …… 30
参考文献 …… 32
第 2 章　信道和信道容量 …… 34

2.1 离散无记忆信道模型 …… 34
2.1.1 转移概率矩阵 …… 34
2.1.2 输出熵和互信息 …… 35
2.2 信道容量和二进制对称信道 …… 37
2.2.1 互信息的最大化和信道容量 …… 37
2.2.2 对称信道 …… 39
2.3 分组编码和仙农(Shannon)第二定理 …… 41
2.3.1 疑义度(Equivocation) …… 41
2.3.2 熵率(Entropy Rate)和信道编码定理 …… 42
2.4 马尔可夫(Markov)过程和有记忆信源 …… 43
2.4.1 马尔可夫过程 …… 43
2.4.2 稳态概率和熵率 …… 46
2.5 马尔可夫链和数据处理 …… 47
2.6 受限信道 …… 49
2.6.1 调制理论和信道约束 …… 49
2.6.2 线性时不变信道 …… 50
2.7 序列的自相关和功率谱 …… 52
2.7.1 时间序列的统计特性 …… 52
2.7.2 功率谱 …… 54
2.8 数据变换码 …… 56
2.8.1 对数据序列的限制 …… 56
2.8.2 码的状态空间和网格图描述 …… 58
2.8.3 数据变换码的容量 …… 60
2.9 (d,k)序列 …… 61
2.9.1 游程长度受限码和最大熵序列 …… 61
2.9.2 最大熵序列的功率谱 …… 63
2.10 小结 …… 67
2.11 习题 …… 68
参考文献 …… 72
第3章 游程长度受限码 …… 73
3.1 数据变换码的一般考虑 …… 73
3.2 前缀码和分组码 …… 74
3.2.1 固定长度分组码 …… 74
3.2.2 可变长度分组码 …… 75
3.2.3 前缀码和克拉夫特(Kraft)不等式 …… 78
3.3 状态依赖固定长度分组码 …… 79
3.4 可变长度固定码率码 …… 82
3.5 前视(look-ahead)码 …… 85
3.5.1 码字的级联 …… 85

3.5.2 k的限制 …… 87
3.5.3 非规范和规范的设计方法 …… 87
3.6 无直流码 …… 90
3.6.1 连续数字和(Running Digital Sum)与数字和偏差 …… 90
3.6.2 状态分裂和谱零点匹配码 …… 91
3.7 小结 …… 96
3.8 习题 …… 97
参考文献 …… 97
第4章 线性分组纠错码 …… 99
4.1 一般考虑 …… 99
4.1.1 用于纠错的信道编码 …… 99
4.1.2 二进制对称信道中的差错率和错误分布 …… 100
4.1.3 错误检测和纠错 …… 102
4.1.4 最大似然译码原理 …… 104
4.1.5 汉明距离和码的能力 …… 105
4.2 二元域和二元矢量空间 …… 107
4.2.1 二元域 …… 107
4.2.2 矢量空间中线性码的表示 …… 110
4.3 线性分组码 …… 111
4.3.1 矢量空间的基本性质 …… 111
4.3.2 汉明重量、汉明距离和汉明立方体 …… 112
4.3.3 汉明球和冗余度要求的界 …… 113
4.4 线性分组码的译码 …… 114
4.4.1 完备译码器和限定距离译码器 …… 114
4.4.2 伴随式译码器和一致校验定理 …… 116
4.5 汉明码 …… 117
4.5.1 汉明码的设计 …… 117
4.5.2 汉明码的对偶码 …… 120
4.5.3 扩展汉明码 …… 120
4.6 线性分组纠错码的差错率性能界 …… 122
4.6.1 分组差错率 …… 122
4.6.2 比特差错率 …… 124
4.7 采用请求重传的限定距离译码器的性能 …… 127
4.7.1 近似差错性能 …… 127
4.7.2 ARQ系统的有效码率 …… 128
4.7.3 ARQ协议 …… 129
4.8 小结 …… 130
4.9 习题 …… 131
参考文献 …… 132

第5章 循环码 …… 133
5.1 循环码的定义和性质 …… 133
5.2 循环码的多项式表示 …… 134
5.3 多项式模运算 …… 136
5.3.1 多项式环 …… 136
5.3.2 一些重要的代数恒等式 …… 137
5.4 循环码的生成和译码 …… 140
5.4.1 生成式、奇偶校验和伴随多项式 …… 140
5.4.2 系统循环码 …… 140
5.4.3 系统循环码编码器的硬件实现 …… 142
5.4.4 循环码译码器的硬件实现 …… 144
5.4.5 梅吉特译码器 …… 145
5.5 错误捕获(Error-Trapping)译码器 …… 148
5.5.1 纠错过程中伴随式的更新 …… 148
5.5.2 突发错误图样和错误捕获 …… 149
5.6 一些标准循环分组码 …… 153
5.6.1 汉明码 …… 153
5.6.2 BCH码 …… 154
5.6.3 纠突发差错码 …… 155
5.6.4 循环冗余校验码 …… 156
5.7 循环码的简单改进 …… 157
5.7.1 码的扩展 …… 158
5.7.2 码的截短 …… 158
5.7.3 截短码的非循环性 …… 161
5.7.4 交织 …… 161
5.8 小结 …… 164
5.9 习题 …… 164
参考文献 …… 166
第6章 卷积码 …… 167
6.1 卷积码的定义 …… 167
6.2 卷积码的结构特性 …… 170
6.2.1 状态图和网格图表示 …… 170
6.2.2 卷积码的传递函数 …… 172
6.3 维特比(Viterbi)算法 …… 174
6.4 维特比算法的工作原理Ⅰ——硬判决译码 …… 178
6.4.1 采用硬判决的最大似然译码 …… 178
6.4.2 错误事件概率 …… 180
6.4.3 比特差错率的界 …… 181
6.5 一些已知的好卷积码 …… 183

6.6 维特比算法的工作原理Ⅱ——软判决译码 …… 185
6.6.1 欧几里德(Euclidean)距离与最大似然 …… 185
6.6.2 结(tie)的消除与信息量损失 …… 187
6.6.3 似然度量的计算 …… 188
6.7 维特比译码的回溯(Traceback)方法 …… 189
6.8 凿孔(Punctured)卷积码 …… 193
6.8.1 凿孔 …… 193
6.8.2 好的凿孔卷积码 …… 194
6.9 小结 …… 196
6.10 习题 …… 197
参考文献 …… 198
第7章 网格编码调制 …… 200
7.1 多幅度/多相位离散无记忆信道 …… 200
7.1.1 *I-Q* 调制 …… 200
7.1.2 *n* 进制 PSK 信号星座 …… 201
7.1.3 PSK 的差错率 …… 202
7.1.4 正交幅度调制 …… 203
7.2 系统递归卷积编码器 …… 205
7.3 信号映射与集合分割 …… 206
7.4 已知 PSK 和 QAM 的好网格码 …… 209
7.5 小结 …… 212
7.6 习题 …… 212
参考文献 …… 213
第8章 信息论与密码学 …… 215
8.1 密码系统 …… 215
8.1.1 密码系统的基本组成 …… 215
8.1.2 一些简单的密码体制 …… 216
8.2 对密码系统的攻击 …… 220
8.3 完善保密性 …… 220
8.4 语言熵和成功密文攻击 …… 222
8.4.1 密钥疑义度定理 …… 222
8.4.2 虚假密钥和密钥疑义度 …… 223
8.4.3 语言冗余和惟一解(Unicity)距离 …… 224
8.5 计算安全性 …… 225
8.6 扩散与混淆 …… 226
8.7 乘积加密系统 …… 228
8.7.1 可交换、不可交换和幂等乘积加密 …… 228
8.7.2 混合变换与好的乘积加密 …… 229
8.8 编码 …… 231

8.9 公共密钥系统 …… 231
8.10 其他问题 …… 232
8.11 小结 …… 232
8.12 习题 …… 233
参考文献 …… 234
第 9 章 仙农编码定理 …… 236
9.1 随机编码 …… 236
9.2 平均随机码 …… 237
9.3 对仙农第二定理的讨论 …… 239
9.4 仙农-费诺编码 …… 240
9.5 仙农无噪声编码定理 …… 241
9.6 最后的话 …… 242
参考文献 …… 243
附录 部分习题答案 …… 244

第 1 章　离散信源和熵

1.1　数字通信和存储系统概述

通信或信息存储系统在日常生活中是经常见到的。从广义上说，通信系统就是能够将信息从一个地方发送到另一个地方的系统。这样的例子有很多，例如：电话网络、无线电、电视、蜂窝电话、本地计算机网络等。存储系统是用于存储并且随后能够重现信息的系统。在某种意义上，这样的系统也可以看作是通信系统，其将信息从现时（现在）传送到过后（未来）。存储系统的例子包括磁盘、光盘驱动器、磁带录音机和放像机等。

这两类系统都可以抽象地用图 1-1-1 所示的框图来表示。在所有情况下，都有一个信息产生源，来自信源的信息被一个对信息进行编码和调制的系统进行处理。编码器 / 调制器将信息加工为某种形式的信号，这样的信号使我们更便于以物理形式传送（或存储）信息。在通信系统中，这种功能常被称做发送机，而在存储系统中，被称做记录器或写入器。

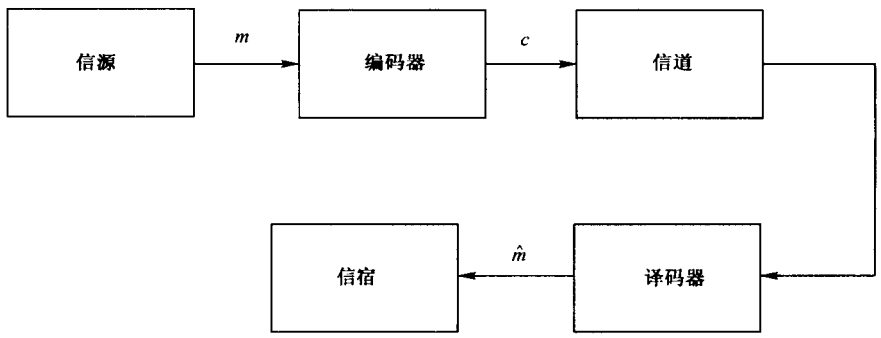

图 1-1-1　基本信息处理系统

编码系统的输出接下来通过某种物理通信信道进行传送（在通信系统中），或者存储在某种物理存储媒质中（在存储系统中）。前者的例子如：使用电磁波的无线传输和使用铜电话电缆或光缆的有线传输，后者的例子如：磁盘（像在软盘驱动器中使用的磁盘）、磁带和光盘（像 CD-ROM 或者 CD 播放机中所使用的光盘）。不论媒质的外观形式如何，我们都将其称之为“信道”。

通过信道传送（或在信道中存储）的信息必须在目的地进行恢复和处理以复原为信息的原始表示，这是译码器 / 解调器要完成的工作。在通信系统中，这种装置经常被称为接收机。在存储系统中，这种模块经常被称为回放系统或读出装置。译码器所进行的信号处理可以看作是编码器完成功能的逆过程。译码器的输出接着提供给最终用户或目的地，我们将信息的目的地称为信宿。

一般情况下，物理信道将输出一个与原始输入信号 c 不同的接收信号 r，这是由于信道引

入了信号失真和噪声的缘故。因此，译码器只能产生原始消息 m 的估计值 $\hat{m}$。所有经过精心设计的系统的目标都是使 m 的重现尽量可靠，同时，要使每单位时间内发送（对于通信系统）或者每单位存储单元内存储（对于存储系统）的信息尽可能的多。

在典型情况下，信源消息 m 是由信息源发出的一个时间符号序列组成。如果该序列在时间上连续，则该信源被称为连续时间信源。否则，该信源被称为离散时间信源。连续时间信源的一个例子是话音的波形。离散时间信源的例子包括计算机产生的数据序列或者印在本页面上的文本（这也是数据存储系统的一个例子）。

由信源产生的符号其特征可以表现为连续幅度或者离散幅度。话音是一个连续幅度信源的例子，因为话音波形的模型中包含了实数值信号幅度。本页上的文字又是离散幅度信源的例子，因为其字符均取自有限的符号表。

在这本入门性的教材中，我们主要涉及离散时间／离散幅度的信源，因为这些信源数学上处理最简单，而且实际上所有新型的通信或存储系统当前都属于此种类型。本书中所讲述的理论也可以推广到连续信源情况，但有关连续信源的理论更适合于在较高级的课程中来介绍。

1.2 离散信源和熵

1.2.1 信源符号集和熵

信息论在很大程度上是建立在概率论的概念和数学之上的，这是因为“信息”这一术语本身就具有传送消息中的不可预测性的含义。消息中的信息含量直接与消息所传达的“令人惊奇”的量有关。例如，假定某人对你说，“美国的首都是华盛顿”，如果是在你小的时候，那时你还不知道美国的首都在哪里，当你第一次听到这句话时，它是有信息量的。可是现在当你看到前面这句话时，这条消息没有任何信息量（至少对于美国公民来说是如此）。从信息论的观点来说，以上的语句对你来说所含信息量为零。

信息和知识是有区别的，虽然由两者中的一个可以产生另一个，但是它们两者并不是相同的东西。在韦氏（Webster）词典中知识的定义为：

（1）通过经验或联想获得的对某件事情所知道的事实或状态。

（2）意识到某件事情的事实或状态。

（3）所有已知的总和。

凡是能够增长我们知识的那些知识可以说是含有信息的，因此信息可以用是否能够增加我们知识这一特性来区分。凡是不能增加我们知识的那些知识不提供任何信息。因此，含有信息的知识本身具有令人惊奇或者不确定性的成分。从信息的这一特点也可以得出结论，资料（data）和信息也不是相同的东西。当你阅读我为你所写的这本书中的内容时，我向你提供了资料，但是除非你能够读懂书中的内容，否则我没有向你提供任何信息。

一个信源是由它能产生的输出符号集和控制这些符号输出的概率规则来定义的，一个有限离散信源是具有有限个不同符号的信源，该符号集经常被称为信源符号集。对于具有 M 个可能符号的信源符号集，我们可以将该符号集表示为一个集合

$$A = \{a_0, a_1, \cdots, a_{M-1}\} \tag{1-2-1}$$

集合中元素的数量被称为该集合的基数，记为

$$M = |A|$$

在一个时间序列中的信源输出符号可以表示为

$$\bar{a} = (s_0 s_1 \cdots s_t \cdots) \tag{1-2-2}$$

其中，$s_t \in A$ 是信源在时间 t 所发出的符号。在本书中如果没有另外说明，我们均假定时间 t 取自整数时间值。在任意给定的时刻，信源发出符号 a_m 的概率记为 $p_m = \Pr(a_m)$。如果概率集合

$$P_A = \{p_0, p_1, \cdots, p_{M-1}\} \tag{1-2-3}$$

不是时间的函数，我们说该信源具有平稳性。因为信源发出的符号一定取自符号集 A，因此我们有

$$\sum_{m=0}^{M-1} p_m = 1 \tag{1-2-4}$$

在数学上最容易处理的信源是同步信源，该信源是在固定的时间间隔内发出新的符号。异步信源是指其发出的符号之间的时间间隔不固定的信源，这样的信源可以通过定义其输出的符号之一为空(null)字符的方式来近似地处理。如果该信源在时间 t 不发出任何字符，我们称该信源在时间 t 产生空字符。因此，空字符实际是一种“虚”符号，它是指在时间 t 没有“实际”符号产生。采用这种方式，我们可以将很多离散时间异步信源近似为同步信源。

信源发出的符号在物理通信系统中必须以某种方式来表示。在数字通信系统中，信源发出的符号一般是用二进制方式来表示的。例如，$M = 4$ 的信源其输出符号可以用一对二进制数字来表示。在此情况下，a_0 可以表示为 00，a_1 表示为 01…… 通常将以这种方式表示的符号称为信源数据。

在信息论中，数据和信息有很大的区别。两者并不是同一个概念，一般来说，数据不等同于信息。为了表明这两个概念之间的区别，我们考虑符号集中只有一个符号的信息源。该信源只能发出惟一的符号，而不能发送其他的符号。该符号的表示是“数据”，但是显然该数据没有任何信息。无论该数据代表什么，我们都对其没有任何“疑问”。因为信息的含义是符号中的不确定性(在下文中将要讲到)，因此该信源的信息含量为零。

信源的信息含量是信源的一种重要属性，并且该信息含量可以进行测量。在仙农(Shannon)的原始论文中(仙农创立了信息论这门科学)，给出了每个信源符号所传送的平均信息量的精确数学定义。该量度被称为信源的熵，定义为

$$H(A) = \sum_{m=0}^{M-1} p_m \log_2(1/p_m) \tag{1-2-5}$$

式(1-2-5)告诉我们由单个符号的概率所决定的信源的信息含量。应注意对于 $M = 1$ 的无信息含量的信源，其符号的概率为 1，而 $H = 0$。

在以上给出的熵的定义中，我们使用了概率倒数的以 2 为底的对数。也可以使用任何其他的底数来定义熵。对于用式(1-2-5)定义的熵，熵的度量单位为“比特”(bit)。如果我们使用自然对数(以 e 为底的对数)代替式(1-2-5)中以 2 为底的对数，则熵的度量单位被称为“奈特”(nat)。最常用的度量熵的单位是比特。

【例 1-2-1】 已知某四进制信源，其符号的概率分布为

$$P_A = \{0.5, 0.3, 0.15, 0.03\}$$