21世纪全国本科院校电气信息类创新型应用人才培养规划教材

离散信息论基础

范九伦
谢 勰 编著
张雪锋

赠送 电子课件 www.pup6.com

赠送 习题参考答案 www.pup6.com

北京大学出版社
PEKING UNIVERSITY PRESS

21 世纪全国本科院校电气信息类创新型应用人才培养规划教材

离散信息论基础

范九伦　谢　勰　张雪锋　编著

北京大学出版社
PEKING UNIVERSITY PRESS

内 容 简 介

本书从离散概率入手，对离散信息论的基本知识进行了介绍，主要内容包括：绪论，离散信息的度量，数据压缩，离散信源，数据纠错，离散信道，数据保密，算法信息论与通用信源编码，微分熵与最大熵原理。为拓宽读者视野，培养学习兴趣，提高人文素养，本书融入了一些历史知识，还补充了信息论实验内容。

本书可供信息安全、信息与计算科学、计算机科学与技术等本科专业的高年级学生使用，也可供从事相关专业的教学、科研和工程技术人员参考。

图书在版编目(CIP)数据

离散信息论基础/范九伦，谢勰，张雪锋编著. —北京：北京大学出版社，2010.8
(21世纪全国本科院校电气信息类创新型应用人才培养规划教材)
ISBN 978-7-301-17382-4

Ⅰ. ①离… Ⅱ. ①范…②谢…③张… Ⅲ. ①离散—信息论—高等学校—教材 Ⅳ. ①O158②TN911.2

中国版本图书馆CIP数据核字(2010)第118436号

书　　名：离散信息论基础
著作责任者：范九伦　谢　勰　张雪锋　编著
策 划 编 辑：李　虎
责 任 编 辑：程志强
标 准 书 号：ISBN 978-7-301-17382-4/TN·0060
出 版 者：北京大学出版社
地　　址：北京市海淀区成府路205号　100871
网　　址：http://www.pup.cn　http://www.pup6.com
电　　话：邮购部62752015　发行部62750672　编辑部62750667　出版部62754962
电 子 邮 箱：pup_6@163.com
印 刷 者：三河市北燕印装有限公司
发 行 者：北京大学出版社
经 销 者：新华书店
787毫米×1092毫米　16开本　12.75印张　294千字
2010年8月第1版　2010年8月第1次印刷
定　　价：25.00元

前　言

人们通常将 Shannon 在 1948 年 10 月发表于 *Bell System Technical Journal* 上的论文 *A Mathematical Theory of Communication* 作为信息论研究的开端。信息论发展至今，可分为狭义信息论和广义信息论，本书则主要介绍狭义信息论的一些基本知识。狭义信息论运用概率论与数理统计方法研究信息的表示、度量、存储、传递等问题，是高等院校很多本科专业的一门专业基础课，我国众多高校也在相关专业开设了信息论课程。

在为信息安全、信息与计算科学等本科专业讲授信息论课程时，我们深刻体会到，要使大学生较好地理解和领会信息论的基本概念，诸如熵、互信息、熵率、信道容量，有很多困难。在多年的教学中，我们一直被两个问题所困惑：一是鉴于信息论不仅具有理论性，也具有实践性，如何保持信息论基本概念、方法在理论叙述上的严谨性，使得学生对信息论有一个清晰的认识，同时又能使学生通过解决实际问题，达到运用信息论的目的；二是鉴于信息论不仅在本科生阶段开设，也在研究生阶段开设，如何将本科讲授内容和研究生讲授内容进行合理切割，尽量避免教学内容重复，使得知识深度与思想广度在不同阶段有所区别。为了较好地解决上述问题，我们萌发了写作本书的念头。在本书的写作中，我们力求达到以下几点。

(1) 图文并茂、循序渐进。本书按照教学目标、教学要求、教学内容的格式进行编写，以叙事、问题的方式展开，改变工科教材艰深古板的固有面貌，具有较强亲和力，使学生初次翻阅就对其产生浓厚兴趣，不会因其理论的抽象而产生敬畏之感。既加强了学生的融会贯通能力，又提高了学生的人文素养。

(2) 凸显信息论的“离散”内容。信息论的研究和应用丰富多样，为了扩大教材的受益面，避免涉足过多的专业领域知识，本书重点围绕离散随机变量(过程)介绍信息论的基本知识，主线明晰，增强了教材的可读性。考虑到信息论的介绍离不开概率论和数理统计知识，本书弱化了数学证明，强化了来龙去脉的讲授，使之显得通俗易懂，同时又给学有余力者留下充足的探求空间。

(3) 强化学生的实际操作训练。对于内容实用性和技巧性较强的章节，如编码理论部分，本书精心设计了相关实验，以实际操作训练加深对理论知识的理解，激发学生对工程实践的兴趣，全方位锻炼学生对知识的掌握程度。

在教材写作中，我们努力将最新的知识、内容和理念传授给学生。本书以离散随机变量(过程)为出发点进行展开，力求以亲切易读的面貌，帮助初学者熟悉必要的理论知识，掌握其思想方法，了解其应用前景，为后续课程和进一步深入学习打下坚实基础。本书共分为 9 章：第 1 章和第 9 章由范九伦和谢勰共同编著；第 7 章由张雪锋编著；其余部分由谢勰编著。全书由范九伦进行统稿和润色。

本书的编著参考了国内外的一些权威教材、相关专著和经典论文，书中的图片大多

来自维基百科，在此向原作者表示感谢。

由于作者学识有限，书中的疏漏和不足之处在所难免，恳请大家不吝赐教。

范九伦

2010 年 3 月

于西安邮电学院

目　录

第 1 章　绪论 …… 1

1.1　基本概念 …… 2
1.1.1　信息的含义 …… 2
1.1.2　信息的表达 …… 3
1.1.3　信息的处理 …… 6
1.2　信息论概览 …… 7
1.2.1　Shannon 与信息论 …… 7
1.2.2　通信系统的数学模型 …… 8
本章小结 …… 9
习题 …… 9

第 2 章　离散信息的度量 …… 10

2.1　基本概念 …… 11
2.1.1　离散熵的定义 …… 11
2.1.2　联合熵与条件熵 …… 16
2.1.3　相对熵与互信息 …… 21
2.2　离散熵的性质 …… 25
2.2.1　离散熵的基本性质 …… 25
2.2.2　链式法则 …… 29
2.2.3　有关离散熵的不等式 …… 31
2.3*　离散熵的形式唯一性 …… 34
本章小结 …… 37
习题 …… 37

第 3 章　数据压缩 …… 39

3.1　基本概念 …… 40
3.1.1　语言与编码 …… 40
3.1.2　唯一可译码 …… 45
3.1.3　即时码与前缀码 …… 48
3.2　数据压缩的性质 …… 50
3.2.1　前缀码的码长约束 …… 50
3.2.2　唯一可译码的码长约束 … 53
3.2.3　最佳码 …… 55
3.3　典型编码 …… 57
3.3.1　Huffman 编码 …… 57
3.3.2　Fano 编码 …… 60
3.3.3　Shannon-Fano-Elias 编码 …… 61
本章小结 …… 65
习题 …… 65

第 4 章　离散信源 …… 67

4.1　基本概念 …… 68
4.1.1　离散信源模型 …… 68
4.1.2　Markov 信源 …… 71
4.1.3　Markov 链 …… 74
4.2　信源编码 …… 78
4.2.1　随机变量扩展 …… 78
4.2.2　变长信源编码定理 …… 80
4.2.3　熵率 …… 83
4.3　渐近均分性 …… 87
4.3.1　典型集 …… 87
4.3.2　信源编码定理 …… 89
本章小结 …… 93
习题 …… 93

第 5 章　数据纠错 …… 95

5.1　基本概念 …… 96
5.1.1　离散信道模型 …… 96
5.1.2　典型信道 …… 100
5.1.3　信道扩展 …… 102
5.2　信道纠错 …… 105
5.2.1　译码准则 …… 105
5.2.2　错误概率估计 …… 108
5.2.3　分组码 …… 110
5.3　线性分组码 …… 113
5.3.1　码字距离 …… 113

5.3.2 纠错能力 …… 115
5.3.3 Hamming 码 …… 117
本章小结 …… 120
习题 …… 120

第 6 章 离散信道 …… 122

6.1 基本概念 …… 123
6.1.1 互信息 …… 123
6.1.2 特殊信道的容量 …… 127
6.1.3 一般信道的容量 …… 131
6.2 数据处理 …… 134
6.2.1 码率 …… 134
6.2.2 数据处理不等式 …… 136
6.2.3 信源信道定理 …… 139
6.3 信道编码 …… 142
6.3.1 联合典型集 …… 142
6.3.2 信道编码定理 …… 143
6.3.3 信道编码逆定理 …… 145
本章小结 …… 147
习题 …… 147

第 7 章 数据保密 …… 149

7.1 信息的保密传输 …… 150
7.1.1 密码学简介 …… 150
7.1.2 保密系统模型 …… 152
7.1.3 几种典型的密码体制 … 153
7.2 密码体制的信息论分析 …… 156
7.2.1 完全保密性 …… 156
7.2.2 唯一解距离 …… 158
本章小结 …… 161
习题 …… 161

第 8 章 算法信息论与通用信源编码 … 162

8.1 基本概念 …… 163
8.1.1 统计编码 …… 163
8.1.2 自适应编码 …… 165
8.2 描述复杂性 …… 169
8.2.1 Kolmogorov 复杂度 …… 169
8.2.2 通用概率 …… 171
8.3 通用信源编码 …… 174
8.3.1 算术编码 …… 174
8.3.2 字典方法 …… 179
本章小结 …… 183
习题 …… 183

第 9 章 微分熵与最大熵原理 …… 185

9.1 基本概念 …… 186
9.1.1 微分熵 …… 186
9.1.2 信息不等式 …… 188
9.2 信息量最大化 …… 189
9.2.1 最大熵问题 …… 189
9.2.2 最大熵分布 …… 191
本章小结 …… 192
习题 …… 192

信息论实验 …… 193

参考文献 …… 196

第1章 绪 论

教学目标

理解信息的含义，了解信息的常见表达方式并对信息的处理实例有直观的认识；了解 Shannon 对于信息论的历史贡献，掌握通信系统的数学模型。

教学要求

知识要点	能力要求	相关知识
信息	(1) 理解信息的含义 (2) 了解信息的表达方式 (3) 了解信息的处理	(1) 信息概念溯源 (2) 自然语言与形式语言 (3) 信号与信息处理
信息论	(1) 了解信息论的初期历史 (2) 掌握通信系统的数学模型	(1) Shannon 生平 (2) 通信与通信系统

引言

从古至今，**信息**(Information)在人类社会中扮演着相当重要的角色，在众多领域中发挥着巨大的作用。对于信息的研究，既可从哲学层面进行思索，也可从科学层面进行展开。对于普通大众而言，对信息的理解更是多种多样：有人认为信息就是消息，也有人认为信息即内容，更有人认为信息等同于知识。一般来说，信息的含义是相当丰富的，若要对信息进行全面、深入、系统的研究，从目前积累的研究成果来看还相距甚远。

如果不宽泛地去研究信息为何物，而将信息概念的内涵缩小(即限定信息于某个狭义的领域)，则有可能对信息开展较为深入的研究和定量的分析。1948 年，Claude Elwood Shannon(图 1.1)完成了一篇划时代的论文：通信的数学理论(*A Mathematical Theory of Communication*)。在通信领域这个“狭义”的限制之下，Shannon 给出了信息的一系列定义、模型和框架，从而奠定了**信息论**(Information

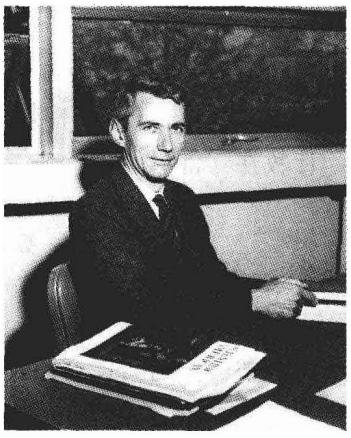

图 1.1 Claude Elwood Shannon

Theory)这门学科的基础。随后，信息论不断发展壮大，至今已走过了60余年的历程。信息论不仅在理论上日渐成熟，而且对实际问题能给予强有力的指导，业已成为一门重要的学科。

《通信的数学理论》所界定的理论一般被称为Shannon理论或Shannon信息论，尽管信息论的发展已超越了Shannon原始论文中所涉及的范围，但该文所提出的思想仍然是经典信息论的主要部分。1998年IEEE信息论分会为纪念信息论诞生50周年（图1.2），专门出版了一系列综述作为纪念，其中Sergio Verdú的综述 *Fifty Years of Shannon Theory* 专门回顾了Shannon理论的发展历史。鉴于Shannon理论在信息论中的基础性地位，本书主要讨论Shannon理论。

图1.2 信息论50周年纪念

Richard Blahut在2000年Shannon雕像落成时所说的一番话为Shannon和信息论在科学史上的地位给出了最好的注解：

"In my opinion, two or three hundred years from now, when people look back to our time, they won't remember who was president of the United States. They won't remember who were the movie stars or the rock stars. But the name Claude Shannon will still be recognized at that time. It will still be taught in schools."

1.1 基本概念

1.1.1 信息的含义

"信息"这个词可谓无所不在，但它究竟是什么，却很难说清楚。在汉语中，"信息"这个词出现得很早，在许多诗词中都可见到。

崔备的《清溪路中寄诸公》提到了"信息"：

偏郡隔云岑，回溪路更深。
少留攀桂树，长渴望梅林。
野笋资公膳，山花慰客心。
别来无信息，可谓井瓶沉。

杜牧在《寄远》也提到了"信息"：

两叶愁眉愁不开，独含惆怅上层台。
碧云空断雁行处，红叶已凋人未来。
塞外音书无信息，道傍车马起尘埃。
功名待寄凌烟阁，力尽辽城不肯回。

而李中所作的《暮春怀故人》长久以来更是被奉为"信息"一词的源头[①]：

① 至于"信息"的源流，学界仍有不同的观点，且各有论证。

池馆寂寥三月尽，落花重叠盖莓苔。
惜春眷恋不忍扫，感物心情无计开。
梦断美人沈信息，目穿长路倚楼台。
琅玕绣段安可得，流水浮云共不回。

上述诗句中“信息”大多依从《现代汉语词典》中的解释，即“音信、消息”[①]，这主要是指获知信息的人所了解的内容。

信息论中所讨论的“信息”则是英语中的 Information，Merriam Webster Dictionary Online 对 Information 给出了诸多解释[②]，这表明 Information 的意义是相当丰富的。不过 Information 的本意仍与汉语中的“信息”相一致，这也是 Information 被译为“信息”的原因。随着时代的发展，汉语中信息一词的含义也日益丰富，基本上与 Information 对等。

需要指出的是，信息与其载体有一定的区分，一般来说信息可以认为是其载体所承载内容的一种抽象形式。而信息还可以进一步升华为知识，著名诗人 T. S. Eliot 在 *The Rock* 这首诗中探讨了信息与知识、智慧等概念之间的关系：

Where is the Life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?

事实上，信息究竟是什么，至今仍未得到一个满意的定义，但可从一些特性上来了解信息，即信息的若干特征，具体如下。

(1) 只要是实体，就必然存在信息。由于实体的状态可以度量，因此这些状态量就是某种形式的信息。

(2) 一方面，信息是客观存在的，例如关于物体颜色的信息；另一方面，不同的认知实体对信息可能存在不同的感知，例如“见仁见智”。

(3) 信息必须依附于一定的载体，不存在虚无缥缈、无所依靠的信息。

(4) 信息可以传输、保存、复制。信息可以在不同实体之间传递，而且实体可以保存所接收的信息，而最令人惊叹的就是信息可以“无限”地被复制[③]。

尽管从信息的特征中难以给出信息的定义，但对于信息的了解可通过考查信息的特征而不断完善，或许 Norbert Wiener 的著名论断是信息最好的定义：

信息就是信息，不是物质也不是能量。

Information is information not matter or energy.

——*Cybernetics*(1948, p. 155)

1.1.2 信息的表达

人们从各种事物中都能获取信息，这也意味着信息的表述方式也千变万化。从接收信息的来源而言，人所能看到、听到、触到的都在发出信息。自然界的这些信息大多可

① 《现代汉语词典》中也给出了信息论中的信息的“定义”，但过于浅显直白。

② http://www.merriam-webster.com/dictionary/information

③ 在物理世界的范围内，这种“无限”是成立的，但它与数学上的无限还是有所差别的。

用连续的函数来表达，即定义在时间和空间上的函数，人类社会自身所拥有的信息也是如此。

一般而言，能够表达信息的方式很多：可以采用文字方式，也可以采用话语形式，还可以采用形象的图形、图像。不同的形式可以表示相同的信息，而这些形式都是服务于信息表达的需要，具体场合需采用合适的表达形式。不过，即便信息内容相同，表达形式也相同，仍可能对不同的人产生不同的效果，正如苏轼的名句：

横看成岭侧成峰，
远近高低各不同。
不识庐山真面目，
只缘身在此山中。

事实上，对于这个问题的理解还是要回到信息的定义中，人类自身所生活的世界，必然有一些难以解释和自相矛盾的地方。因此，本书回避对信息理解的偏差问题。

随着科学技术的发展，信息表达从传统的模拟信号逐渐演变成现在的数字信号。以书籍为例，最早人类阅读的是竹简上刻出的文字，尔后又进化到印刷书，现在直接就是数字形式的电子书。在网络时代，电子书更演变为超链接形式，其撰写、阅读都大不相同。不过，无论信息的表达形式如何变换，其目的是为了人与人之间的交流，因此必须遵循一定的标准。以英文的通信为例，最初以 Morse 电码（图 1.3）来表达，而现在则常用 ASCII 码。事实上，对于任何类型的信息，在进行交流时都必须遵循事先约定的规范标准，也只有规定了信息表达的标准，才能极大地促进信息的交流，使得信息的效用最大化。

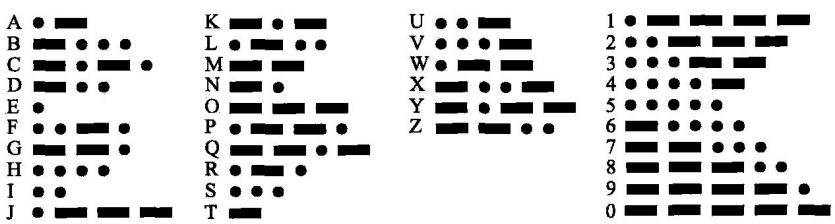

图 1.3　Morse 电码

随着电子技术的发展，人类社会已进入到数字时代。从信息表达的标准演变看，鉴于离散的信息表达方式具有许多优点，如稳定性高、处理方便等，信息表达趋于离散化。目前，离散的信息逐渐成为研究和应用的重点。

人类社会中信息的离散表达实例相当丰富，例如语言是由基本元素形成的字符串，又比如图像数字化即成为像素点的二维矩阵。更令人惊叹的是，自然界的许多信息也需要以离散形式来抽象，而其中最典型的例子就是 DNA 序列。最关键的一类 DNA 片段是传递生物遗传信息的基因，其实质是指导遗传时的生物发育，它以 A、T、C、G 这 4 种碱基作为信息表达的基本要素。这意味着，自然界选择以离散的形式传递信息的目的是要保证信息传递的稳定性，并尽可能减少信息传递的差错。除此之外，自然界还有许多其他的离散信息表达方式。从这些实例可以看出，研究离散的信息不但很有必要，而且也非常有意义。

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

图 1.4 DNA 双螺旋模型的原始论文

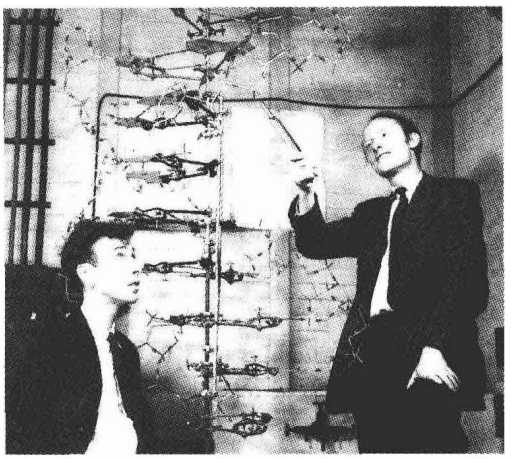

图 1.5 Watson 和 Crick

尽管离散的信息表达方式不如连续的信息表达方式丰富，但仍然有足够多的选择，因为可以利用符号序列来提高所表达信息的种类。例如，当用语言表示信息时，可用非常长的叙述并配以复杂的逻辑体系来表示人类所要传达的大部分信息。

利用一些基本部件的组合表达信息的威力相当强大，例如生产著名的 LEGO 积木玩具(图 1.6 和图 1. 7)的 LEGO 公司在 1974 年宣称：可用 6 块 8 个凸起的长方体 LEGO 积木砌出至少 102981500 款组合①。当然，在现实中，如果表示信息的序列过于冗长，其效率可能不高；如果表示信息的序列过短，其鲁棒性可能不好。因此在实际生活中，需要以合理、高效的序列来有效地表示信息。

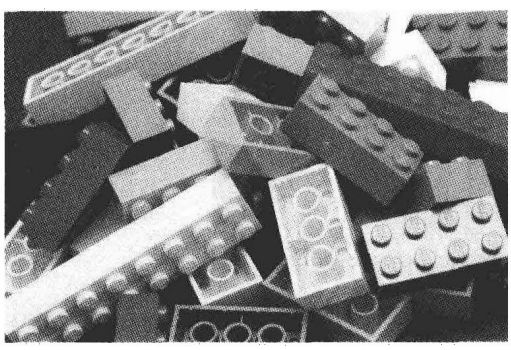

图 1.6 LEGO 积木

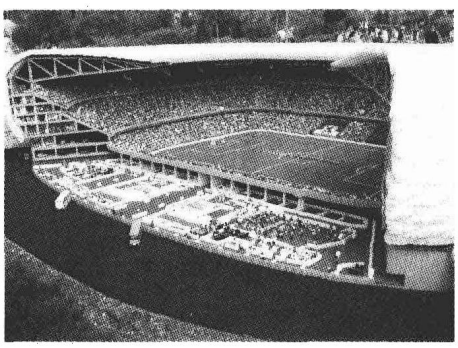

图 1.7 LEGO 创意——德国安联体育场

信息的离散化表示能力取决于信息表示的基本元素。表达信息的基本元素个数不同，其表达能力也有所差异。以语言为例，由于汉字结构复杂，使得汉字的基本元素相当丰富。而英语的基本元素相当简单，所有的英语单词只由 26 个英语字母决定。基本元素的个数并无优劣之分，只因信息的载体不同，而信息表示的关键在于能否充分利用信息的载体来有效地表达信息。

① 当然，实际上能砌出更多种组合。至于究竟能有多少种，这就是 A LEGO Counting Problem.

1.1.3 信息的处理

研究信息，自然要涉及**信息处理**(Information Processing)，即对信息的存储、传输、复制、加工、修改等操作。可以数字图像为例来简单认识信息处理过程。从图像输入到计算机中的那一刻起，就需要对图像进行存储，至于采用何种方式则要看具体需求：如果使用者对画质要求比较高，就必须采用能获得高分辨率的格式；如果使用者要求存储量小，就必须采用压缩性能比较好的格式。而从源头上看，经过数码相机等设备将真实的物理世界图景转换成图像，已经对信息作了一定的处理，即从模拟信号转换成数字信号。从信息加工的角度看，某些时候需要对图像进行恢复，例如修复某些损坏的局部；有些时候需要对图像进行修改，例如去掉某些需要保密的信息。此外，将图像从计算机发到其他计算机或设备中，便涉及信息的传输，不但要保证传输效率高，还要保证尽量少出错。在实际中，还需要提取一部分有用的信息，例如**图像分割**(Image Segmentation)，它是为后续工作有效进行而将图像划分为若干个有意义区域的一种技术，图 1.8 是对经典的 Lena 图像进行二值化分割的结果。

一般而言，信息处理既包括各种具体信息的处理，如文本、图像和视频等的处理，还包括对抽象意义下的信息处理，如**编码理论**(Coding Theory)、**模式识别**(Pattern Recognition)等。抽象的信息处理包括信息的压缩、分类、识别、选择等过程，而在这些处理过程中，信息不会增加，只可能丢失，因此要在尽可能保存原始信息的前提下进行处理。更重要的是，对于同一目的下的信息处理，必须有一些基本要求来衡量处理过程的优劣，具体如下。

图 1.8　图像处理示例——对 Lena 图像进行分割

(1) 效果评价。其中最常见的是正确性的要求，即与原信息之间地差别程度，最好是完全一样。例如从模拟信息转换成数字信息，不可避免地要丢失一些信息。事实上，差错是普遍存在的，人们应做的是尽可能追求较低的差错率，这是对正确性的重要衡量。此外，还有一些与主观因素相融合的效果评价指标，例如在播放视频时为不伤眼需要色彩“柔和”，显然这种指标随个体而不同。不过大多数效果评价还是有一定的客观衡量标准的，这也为研究的开展提供了可能。

(2) 性能评价。其中最常见的是处理时间和存储空间的要求，算法的语言描述需要衡量**时间复杂度**(Time Complexity)和**空间复杂度**(Space Complexity)，而一般均采用这些复杂度的**渐进记号**(Asymptotic Notation)，即以**渐进复杂度**(Asymptotic Complexity)衡量信息处理的性能。时间复杂度对应着信息处理的速度，如果时间复杂度较低，那么信

息处理能适应实时性要求高的场合，不过除少数效果和速度俱佳的处理方法，大多数情况下这种信息处理的效果稍差。传统信息处理以串行处理为主，而目前的趋势则是利用并行处理方式，这样能大大加快信息处理的速度。空间复杂度则对应处理的存储需求量，如果空间复杂度较低，那么信息处理则能适应存储量不大的设备，如手持设备。

(3) 稳定评价。其中最常见的是信息处理能否适应不同环境的指标，如**鲁棒性**(Robustness)。良好的信息处理方法不会随环境的变化而导致性能和效率的巨大变化，当然这种特性是以增加信息处理过程的复杂程度为代价的。

信息处理技术仍在不断发展，但其基本原理和技术必须遵循一定的原理，即信息论中的基本理论。

1.2 信息论概览

1.2.1 Shannon 与信息论

人类社会进入 20 世纪以后，通信方式有了新的突破，主要是大量采用了无线电技术，例如电话、电报、电视、雷达等众多新设备。新技术不断发展的同时，对于理论基础的呼唤则是非常自然的事。

在通信技术的最初发展过程中，需要解决的问题与信息处理一样，也是效果、性能和稳定这 3 个方面。从效果上讲，如何更好地提高通信质量是主要问题；从性能上讲，如何快速而且大量地传输信息是主要问题；从稳定上讲，如何提高通信的抗干扰性是主要问题。在 20 世纪 20 年代，Harry Nyquist 和 Ralph Hartley 对这些问题作出了一些基础性的探讨。20 多年后，Norbert Wiener 进一步在其经典名著《控制论》(*Cybernetics: or the Control and Communication in the Animal and the Machine*)中给出信息的度量。

1948 年，Shannon 以其开创性论文《通信的数学理论》完成了信息论的奠基性工作，他完整、系统地叙述了经典信息论的基本框架。在 Shannon 信息论中，最引人注目的则是**熵**(Entropy)这个概念，Shannon 后来回忆道：

My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.

可以看出，Shannon 回避了“信息”这个名词，而从“熵”来巧妙地解决问题。事实上，在当时通信研究遇到一些无法解决的基础问题的大背景下，Shannon 对熵的定义不但有效地回答了人们的一些疑问，还具有相当大的新意，因此信息论这个学科迅速地发展起来。

不过究其本质，信息论这门学科的出发点还是从 Information 的动词形式 Inform 开始的，Inform 则意味着人与人之间的信息传递。人类最常见的信息表示方式是语言形式，在信息论的研究中，也是类比语言的表达、处理来分析问题的，本书在后续章节中将详

细给出讲解。

人与人之间的信息传递可推广到一般的通信过程，下一节给出通信的数学模型。

1.2.2 通信系统的数学模型

Shannon 对通信过程给出了一个简要的模型，如图 1.9 所示。在 Shannon 所给的通信系统中，**信源**(Source)不断发出**消息**(Message)，这是信息传递的开始。由于消息的形式多种多样，需要将其转化成电信号的形式，而且还要进行一定的处理以便高效的传输。于是，消息经过**编码**(Coding)之后变为**信号**(Signal)，并经过**信道**(Channel)传输。在传输过程中，由于信道的物理局限性，一般存在一些干扰信号传递正确性的**噪声**(Noise)，需要特殊处理。一旦在信道中接收到信号后，先要消除噪声的影响，再将信号转成消息，这些步骤就是**译码**(Decoding)过程，其结果最终交给**信宿**(Destination)。

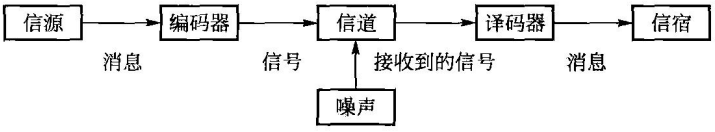

图 1.9 通信系统模型

将上述通信系统细化，并考虑到保密性的需求，则可得到图 1.10 中的细化模型。信源首先利用信源编码将消息转化为数字信号，其主要目的是为了高效地表示消息，随后将编码加密以防止窃取、篡改，最后再考虑到信道的噪声情况予以信道编码，以保证错误出现时尽可能多地恢复信息。经过信道传输后，编码可能会发生一些变化，首先要利用信道编码译码恢复编码，再进行解密以得到原有的信源编码，最后利用信源编码译码转换回原有的消息[①]以交给信宿。

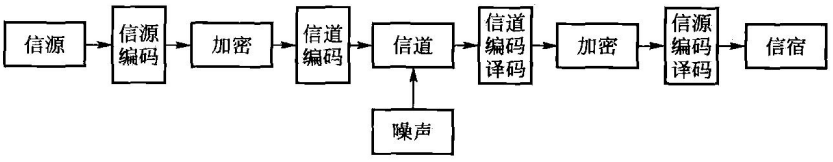

图 1.10 通信系统的细化模型

Shannon 正是如庖丁解牛般地将通信过程分解，再对各个环节给出详细地论证，最后完整地建立了通信的信息理论，从而为信息论学科奠定了坚实的基础。

在 Shannon 所论述的信息论范围内，一般是对通信系统的各个组成部分进行研究，即信源编码问题、信道编码问题、信息的保密问题等。而贯穿于通信系统模型则是熵这个概念，它不但阐述了理论上的合理性，而且也给出了实际通信所必须遵循的规则，本书围绕通信系统的细化模型进行叙述，并给出理论上的分析。

这里应强调的是，Shannon 信息论只是信息论的主要部分，多年来信息论和相关学科的交叉融合使得信息论的研究范围不断扩大。例如传统通信系统模型是两用户的简单情

① 在理想情况下，信宿所收到的消息内容(信息)与信源相同，但信宿的消息形式未必与信源的消息形式相同。

况，而现代的通信则是在多用户的情况下进行，因此传统的信息论必须加以改进以适应这种新情况。又比如传统上是以概率论的角度来考查熵，而现在许多领域需要从模糊集角度来描述熵。还比如在计算机科学中也有对信息的描述和度量，而这种方法与熵的思路既有区别又有联系。事实上，有关信息的研究发展非常迅速，必须从**信息科学**(Information Science)的高度来考查、分析和理解信息，而这仍是一片正在开垦的无尽领域。

以已故物理学家 John Archibald Wheeler 所说的一段精妙言辞作为本章的结束：

I think of my lifetime in physics as divided into three periods

In the first period ... I was in the grip of the idea that

Everything is Particles.

... I call my second period

Everything is Fields.

... Now I am in the grip of a new vision, that

Everything is Information.

本章小结

本章首先给出了信息的若干种含义，从信息的多样化来论证研究信息的难度。随后给出了一些常见的信息表达形式，并强调了离散元素组合的威力。人类社会中常常需要对信息进行有效的处理，为此本章还讨论了若干类常见的信息处理模式。

从信息的上述讨论可以看出，要给出全面、系统的信息理论是相当困难的。Shannon 仅关注通信问题，在此意义上利用熵的概念构建了 Shannon 信息论的基本雏形。为简化问题，必须介绍通信系统的数学模型，不过对此模型的略加细化更有利于问题的分析解决。

本章仅介绍了一些粗浅的概念，所提到的信息论的内容也是形象的、不严格的，后续章节将对 Shannon 信息论进行全面的介绍，并对其中重要的概念和定义给出严格定义和证明。当然，理解信息论的概念和思想更为重要，这才是学习本课程所要达到的最终目标。

习 题

(一) 填空题

1. 在 Shannon 信息论中，最重要的概念是_____。

2. 通信系统一般由_____5 个部分组成。

(二) 综述题

1. 阅读 Shannon 的 *A Mathematical Theory of Communication*，整理出 Shannon 信息论的大纲。

2. 阅读 Sergio Verdú 的综述 *Fifty Years of Shannon Theory*，写出阅读报告。

第2章

离散信息的度量

教学目标

从理论和实践的角度掌握离散熵、联合熵与条件熵、相对熵与互信息等基本概念，并能应用它们解决相关问题；掌握离散熵的性质，尤其是链式法则；理解 Jensen 不等式的意义并能证明关于离散熵的不等式；了解离散熵的形式唯一性。

教学要求

知识要点	能力要求	相关知识
离散熵	(1) 准确理解离散熵的概念 (2) 掌握离散熵的性质	(1) 熵与描述复杂性 (2) 熵与划分
联合熵与条件熵	(1) 准确理解联合熵与条件熵的概念 (2) 掌握链式法则	(1) 条件熵的物理意义 (2) 对系统的分步考察
相对熵与互信息	(1) 准确理解相对熵与互信息的概念 (2) 理解互信息和条件熵的关系	(1) 相对熵的物理意义 (2) 信息不等式的应用

引言

著名的奥地利物理学家 Ludwig Eduard Boltzmann 为世人留下了一座不朽的丰碑，如图 2.1 所示。其上并无多余的溢美之词，仅有他为人类留下的伟大公式：

$$S=k\log W$$

它所描述的概念称为**熵**(Entropy)，这个公式如此简单却又优美，仿佛神来之笔。Boltzmann 在描述气体动力论时，用上述公式刻画了系统的熵 S 和 W，其中 W 是指与一个系统的宏观状态对应的可能出现的微观状态数，而 k 则是 Boltzmann 常数。