信息论与最优编码

李照止 陈须城 编著

信息论与最优编码

章照止
林须端 编著

上海科学技术出版社

(沪)新登字 108 号

责任编辑　赵序明

信息论与最优编码

章照止　林须端　编著

上海科学技术出版社出版

（上海瑞金二路 450 号）

新华书店上海发行所发行　江苏如东印刷厂印刷

浙江省上于县科技外文印刷厂排版

开本 850×1156　1/32　印张 11.75　插页 4　字数 294,000

1993年 3 月第 1 版　1993年 3 月第 1 次印刷

印数 1—1,700

ISBN7-5323-2809-0/TP·26

定价：9.60元

(沪)新登字108号

内 容 提 要

本书系统介绍香农(Shannon)信息论及其有关分支的基本理论与方法.内容包括熵与交互信息、信源的无错编码与率失真理论、信道的容量与编码定理、线性分组码与线性卷积码的概率译码理论、多用户信息论以及密码学.除绪论外每章末尾还介绍了与本章内容有关的发展方向和参考文献.

本书以尽可能少的篇幅和尽可能初等的数学工具完整地介绍信息论各分支的严谨理论与方法,包括近年来发展的新成果和作者的部分工作,并阐明了这一理论的工程实用背景.

本书适用于应用数学、信息科学、通信等有关专业的教师和科研人员学习参考,也可作为这些专业的研究生教材,并可供有关工程技术人员学习提高之用.

前 言

从广泛的意义上说，信息论是一门研究信息传输和信息处理系统中一般规律的科学，它是现代信息科学与技术的理论基础之一。本书着重讨论与最优编码问题有关的理论。由于这一理论是在香农(Shannon)的卓越著作[1]的基础上发展起来的，因此也被称为香农信息最优编码论或香农信息论。此外还简要介绍了近年来迅速发展起来的与这一理论密切相关的密码学理论。

信息论是在总结通信系统模型和解决信息传输问题的基础上发展起来的。一方面，它产生了许多很有意思的新颖的数学问题，形成了完整的、严谨的、有自己特色的数学理论和方法；另一方面，它又有深刻的工程实用背景，与许多科学和技术问题，如数字通信、雷达、导航、模式识别、自动控制、计算机科学、系统科学、统计科学、投资决策、保密学以及信息的存储与处理技术等有密切的联系和广泛的应用。

本书旨在用较小的篇幅对香农信息论及其有关分支的基本理论和方法作一比较系统和深入的介绍，尽可能地反映近年来在这方面的新进展，同时尽可能地阐明这一理论的工程应用背景。

全书共分十一章。第一章简述信息论(包括密码学)所研究的基本问题，是本书内容的一个框架。第二章讨论信息的度量问题，引入熵与交互信息等基本概念及其性质。第三章和第九章分别介绍信源的无错编码和率失真理论。第四章和第五章分别介绍信道的容量和编码定理。第六章和第七、八章分别介绍线性分组码和线性卷积码理论，着重介绍卷积码的维特比译码和序贯译码。第十章介绍多用户的信源和信道模型以及它们的编码定理。第十一章介绍密码学理论。其中六、七、八章主要叙述编译码技术，跳过它

们，不会影响对第九、十、十一这三章的阅读．§4.3、§4.4、§5.4、§9.5分别讨论信道容量、错误界指数和率失真函数的计算问题，可以看作附录。

本书试图通过对信息论中最基本的和最简单的信源和信道模型的深入研究，使读者掌握信息论观点和方法的实质，从而容易运用所学的方法，进一步学习或研究更复杂的模型．基于同样的考虑，本书在介绍线性分组码和卷积码理论时仅研究了二元码．而将结果推广到一般q元码的情形是轻而易举的．为了使读者对信息论的发展有更全面的了解，便于读者进一步学习或选择研究方向，除绪论外在各章的末尾，写了“本章有关文献注释”，书末附有参考文献．阅读本书所必需的有关初等概率论和代数学方面的预备知识，已有许多好书可以参考，例如概率论方面可参看谢衷洁的书[2a]或王梓坤的书[2b]中有关章节；代数学方面可参看万哲先的书[3]中前两章，本书不再另作介绍．只是要说明一点：本书所用的术语“随机变量”，其含意是广义的，它可以是在任意抽象集(空间)中取值的随机变量，不同于通常初等概率论书中所定义的仅在实数集(空间)中取值的随机变量。

近年来，国内陆续出版了几本信息论方面的书[4~6]。本书与它们比较有下列特点：1）取材着重于信息论中的新理论、新方法和与工程应用关系较密切的理论；2）包含了作者长期研究信息论的若干心得，许多定理的证明都作了整理，使它们在数学上更严格和简明；3）由浅入深，注重提高读者的数学修养和对信息论方法的理解和掌握；4）篇幅小，但仍保持较好的系统性和较广的知识面，便于读者在较短时间内学完，且学完后能顺利地阅读信息论方面的有关文献。

本书前十章由章照止执笔，第十一章由林须端执笔，章照止作了一些修改．许文源审阅了本书前十章，沈世镒仔细审阅了全书，他们提出了许多宝贵的意见．李中华、郑玉颖和吴淑华对本书的清稿工作给予了大力的帮助．上海科学技术出版社赵序明同志参与了最后定稿的工作，提供了不少有益的意见．作者谨在此一并

表示衷心的感谢。

本书的编著工作得到国家自然科学基金和国家教委高等学校科学技术基金的部分资助。

作　　者

1991年4月于中国科学院系统科学研究所

北京邮电学院

基本符号说明

$\boldsymbol{p}=\{p_1,p_2,\cdots p_J\}=\{p_j,1\leqslant j\leqslant J\}$　离散(概率)分布，在强调它同时是概率向量（J 维实欧氏空间中的点）时用 $\boldsymbol{p}=(p_1,p_2,\cdots,p_J)$。

$\boldsymbol{Q}=\{Q_{kj};1\leqslant j\leqslant J,1\leqslant k\leqslant K\}$　离散条件(概率)分布。

$\boldsymbol{p}=\{p(x);x\in\mathscr{X}\}$　$\mathscr{X}$ 中取值的随机变量 X 的分布或 $\mathscr{X}$ 上分布，简记为 $\{p(x)\}$，在容易发生混淆时，记作 $\{p_X(x);x\in\mathscr{X}\}$。

$\boldsymbol{Q}=\{Q(y|x);x\in\mathscr{X},y\in\mathscr{Y}\}$　（$\mathscr{Y}$ 中取值的）随机变量 Y 关于（$\mathscr{X}$ 中取值的）随机变量 X 的条件分布或 $\mathscr{Y}$ 上关于 $x\in\mathscr{X}$ 的条件分布，简记为 $\{Q(y|x)\}$，在容易发生混淆时，记作 $\{P_{Y|X}(y|x);x\in\mathscr{X},y\in\mathscr{Y}\}$。

$\boldsymbol{p}\times\boldsymbol{q}$　分布 $\boldsymbol{p}$ 和 $\boldsymbol{q}$ 的乘积分布（若 $\boldsymbol{p}=\{p_j;\ 1\leqslant j\leqslant J\}$，$\boldsymbol{q}=\{q_k;\ 1\leqslant k\leqslant K\}$，则 $\boldsymbol{p}\times\boldsymbol{q}=\{p_jq_k;1\leqslant j\leqslant J,1\leqslant k\leqslant K\}$）。

$\boldsymbol{p}^N$　$\boldsymbol{p}\times\boldsymbol{p}\times\cdots\times\boldsymbol{p}$，分布 $\boldsymbol{p}$ 的 N 次乘积分布。

$\boldsymbol{Q}^N$　条件分布 $\boldsymbol{Q}$ 的 N 次乘积条件分布。

$P\{\cdots\}$　括号内…所示事件发生的概率。

$p(\mathscr{A})$　$\sum\limits_{x\in\mathscr{A}}p(x)$ 或 $P\{x\in\mathscr{A}\}$。

$\overline{p}(x)$　密度函数或密度函数在 x 的值。

EX　随机变量 X 的均值。

$E[Y|X=x]$　在 $X=x$ 的条件下随机变量 Y 的条件均值。

x^N 或 $\boldsymbol{x}=(x_1,x_2,\cdots,x_N)$　N 长符号序列或数字序列或数组，有时简记为 $x_1x_2\cdots x_N$，如 01011。

X^N 或 $\boldsymbol{X}=(X_1,X_2,\cdots,X_N)$　N 个随机变量的序列（N 维随机变量）。

$x^\infty=(x_1,x_2,\cdots)$　无限长符号序列或数字序列。

$X^{\infty}=(X_1,X_2,\cdots)$　无限个随机变量的序列（随机序列）。

$\mathscr{X}\times\mathscr{Y}$　$\{(x,y);x\in\mathscr{X},y\in\mathscr{Y}\}$，集$\mathscr{X}$和$\mathscr{Y}$的笛卡尔乘积。

$\mathscr{X}^N$　$\mathscr{X}\times\mathscr{X}\times\cdots\times\mathscr{X}$，集$\mathscr{X}$的$N$次笛卡尔乘积（$\mathscr{X}$中元素的一切$N$长序列组成的集）。

$\mathscr{A}^*$　$\mathscr{A}$中元素的一切有限长序列组成的集。

$|\mathscr{A}|$　集$\mathscr{A}$中的元素个数。

$GF(2)$　二元域。

$GF(2)^N$　$GF(2)$上的N维向量空间。

$\oplus$　模2加（$GF(2)$或$GF(2)^N$中的加法运算）或模K加。

$((v(1),\cdots,v(t-1)),v(t))$　等价于$(v(1),\cdots,v(t-1),v(t))$。

$[u]$　不超过实数u的最大整数。

$\vec{ab}$　点a到b的联线或分枝。

$H(\boldsymbol{p})$　分布$\boldsymbol{p}$的熵。

$H(p)$　密度函数$p(x)$的微分熵。

$h(\varepsilon)$　分布$\{\varepsilon,1-\varepsilon\}$的熵。

$H(X)$　随机变量X的熵或连续随机变量X的微分熵。

$I(X;Y)$　随机变量X与Y的交互信息。

$I(X;Y,Z)$　等价于$I(X;(Y,Z))$。

$d(\boldsymbol{x},\boldsymbol{y})$　序列$\boldsymbol{x}$与$\boldsymbol{y}$之间的汉明距离。

$d(u,z)$　失真函数。

$d_N(\boldsymbol{u},\boldsymbol{z}),N\geqslant 1$　消息的失真度或消息$\boldsymbol{u}$与$\boldsymbol{z}$之间的失真。

$d(f,\varphi)$　消息$\boldsymbol{U}$关于编译码(f,φ)的译码失真。

$e(f,\varphi)$　消息$\boldsymbol{U}$关于（信源）编译码(f,φ)的译码错误概率或消息关于（信道）编译码(f,φ)的平均译码错误概率。

(f^k,φ^k)　重复使用编译码$(f,\varphi)$$k$次所得的编译码。

f_0　卷积编码器或布尔函数。

$\mathscr{C}^{\perp}$　码$\mathscr{C}$的对偶码。

f_L、G_L、$\mathscr{C}_L$　卷积编码f的L结尾卷积编码、f_L的矩阵表示、卷积码$\mathscr{C}$的L结尾卷积码（f_L编出的码）。

$f_{T(L)}$、$G_{T(L)}$、$\mathscr{C}_{T(L)}$　　卷积编码 f 的 L 截尾卷积编码、$f_{T(L)}$ 的矩阵表示、卷积码 $\mathscr{C}$ 的 L 截尾卷积码（$f_{T(L)}$ 编出的码）。

目 录

第一章

绪　论

信息论是一门研究信息传输和信息处理系统中一般规律的科学，在正式阐述信息论的严谨理论之前，先对信息论所研究的基本问题作一概述，使读者对本书所讨论的问题及其意义先有一个粗略的了解，也许是有益的。

信息是系统传输和处理的对象，它载荷于语言、文字、数据、信号等消息之中．信息的度量是信息论研究的基本问题之一．从当前的研究情况看，要对通常含意下的信息（知识、情报）给出一个统一的度量是很困难的．至今最成功和最普及的信息度量是香农(Shannon)在他的光辉著作《通信的数学理论》[1]中提出的建立在概率模型基础上的信息度量．根据人们长期的实践经验，一个事件给予人们信息的多少是与该事件发生的概率大小有关的．例如"今天某地发生七级以上地震"这个事件使人感到意外，它给人们的信息就很多．相反地，"今天北京不下雪"这个事件是意料之中的，它给人们的信息就较少．因此用 $I(A)=-\log P(A)$ ($P(A)$ 表示事件 A 发生的概率)来度量事件 A 提供的信息是与人们的直观相符合的．称 $I(A)$ 为事件 A 的自信息．进一步，若一个随机试验有 M 个可能结果(事件)，或一个随机消息可能取 M 个消息值(事件)，设它们出现的概率分别为 $p_1,p_2,\cdots,p_M$，则用各事件的自信息的平均值 $H=-\sum_{i=1}^{M} p_i \log p_i$，来度量一个随机试验或一个随机消息所提供的(平均)信息也是合理的．信息度量的单位依赖于对数取什么为底：当对数取 2 为底时，单位为比特；当对数取 e 为底时，单位为奈特．H 即称为熵．香农在上述著作中，应用熵作为信息的基本度量，成功地解决了信息传输的基本问题．从

此信息论发展成为一门独立的学科。

信息论所研究的信息传输系统的基本模型如图 1.1 所示。

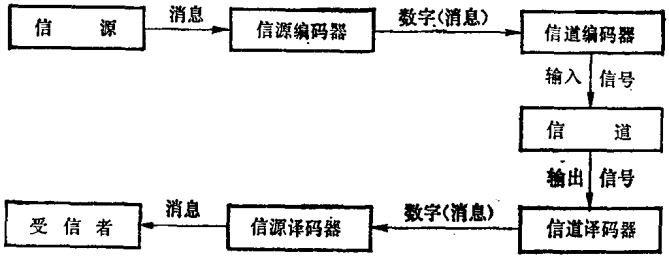

图 1.1　信息传输系统的基本模型

信源是产生消息(或消息序列)的源，消息通常是符号序列或时间函数。例如在电报系统中，消息是由文字、符号、数字组成的报文(符号序列)，称为离散消息；在数字电话系统中，消息是语声波形(时间函数)，称为连续消息。消息取值服从一定的统计规律，故信源的数学模型是一个在信源符号集中取值的随机变量序列或随机过程。

信源编码器将信源产生的消息变换为一个数字序列（通常为二进制数字序列)。

在离散情形，设信源可能产生 M^N 个消息，每个消息由 N 个信源符号组成，各个消息出现的概率分别为 $p_1^{(N)}$，$p_2^{(N)}$，…，$p_{M^N}^{(N)}$，这时信源编码器可以做成从消息到数字序列的一一变换。设第 i 个消息对应的数字序列长为 l_i（包含 l_i 个数字)。定义 $R=\frac{1}{N}\sum_{i=1}^{M^N} p_i^{(N)} l_i$ 为信源编码速率，它表征每个信源符号平均要用多少个数字来表示。若信源译码器做成信源编码器的逆变换，则在无噪信道(即信源译码器的输入为信源编码器的输出)的情形，消息可以正确无误地传送。这时信源编码理论要回答两个问题：(1) 对给定的信源，可能达到的最小编码速率是多少？(2) 如何构造实现这一速率的最优编码。这两个问题在信息论发展的最初年代里就已获得解决。对相当广泛的信源类，若允许 N 无限增大，则

编码速率的最大下限为 $\inf R=\overline{H}=\lim\limits_{N\to\infty}-\frac{1}{N}\sum\limits_{i=1}^{M^N}p_i^{(N)}\log p_i^{(N)}$。$\overline{H}$ 称为信源的熵率，是信源的一个重要参数。当 N 固定时，最优编码就是霍夫曼(Huffman)编码。

在连续情形，信源可能产生的消息是一个无穷集。故信源编码器不可能是消息到数字序列的一一变换，从而信源译码器也不可能是信源编码器的逆变换，通常的方法是先对连续消息进行采样和量化，变为离散消息，再将离散消息变换为数字序列。信源译码器则先将输入的数字序列逆变换为离散消息，再用适当的内插法复制出连续消息。可以想象用这样的方法传送消息，即使在无噪信道的情形，收到的消息也不会与发送消息完全相同，它们之间必然存在误差，称为消息的失真。可以选择一个适当的非负函数 $d(u,z)$ 来度量消息 u、z 之间失真的大小。现在信源编码理论要回答的问题是：

(1) 对给定的信源，在保证消息的平均失真不超过给定的允许限 D 的条件下，可能达到的最小编码速率是多少？

(2) 如何构造实现这一速率的最优编码？

为了求解上述问题，导致熵到率失真函数的推广，发展了信源编码的率失真理论。这一理论主要是从 60 年代到 70 年代初发展起来的。

在实际的通信系统中，信道是指传输信号的媒质或通道，如架空明线、电缆、射频波束、人造卫星等。在信息论的模型里，有时为了研究方便，可以将发送端和接收端的一部分如调制器和解调器归入信道，而且将系统各部分的噪声和干扰都归入信道中考虑。根据噪声和干扰的统计特性，信道有多种模型。最简单的是离散无记忆(恒参)信道。数学上可以用一个信道入口符号集 $\mathscr{X}$、出口符号集 $\mathscr{Y}$ 以及一组条件概率（或称转移概率）$\boldsymbol{Q}=\{Q(y|x);\ x\in\mathscr{X},\ y\in\mathscr{Y}\}$ 来描述。它具有下述特性：若信道输入信号为 $\boldsymbol{x}=(x_1,x_2,\cdots,x_N)$，则相应的输出信号(受扰信号)为 $\boldsymbol{y}=(y_1,y_2,\cdots,y_N)$ 出现的概率为 $Q^{(N)}(\boldsymbol{y}|\boldsymbol{x})=\prod\limits_{i=1}^{N}Q(y_i|x_i)$ $(N=1,2,$

…)。

信道编码器将信源编码器输出的数字序列，每 l 个一组地变换为 N 长的信号（包含 N 个信道入口符号），亦称码字。这样的编码称为分组编码。若数字和信道符号都是二进制的（可用 0,1 表示），则定义 $R=l/N$ 为信道编码速率，它表征每个信道符号载荷多少个数字。$N-l$ 称为码多余度。信道编码（或称纠错编码）的基本思想就是通过编码引进多余度以提高信息传送的可靠性。更确切地说，信道译码器有可能利用这种多余度，将受扰的错误信号仍译为正确的发送数字序列。下面举一个最简单的例子来说明上述思想。设信道编码器将每个输入数字重复 3 次，如将 01 01 1 变换为 00 01 11 00 01 11 11 1。信道译码器用门限译码，即先将输入译码器的信道符号每 3 个一组地相加，再将所得结果逐个与门限 2 比较，小于门限的译为 0，否则译为 1。这样若受扰信号为 01 01 10 10 00 11 01 1，虽然错了 5 个符号，但译码器译出的仍为 01 01 1 与发送数字序列完全相同。信息论的最重要结论就是：对于一个有噪信道，只要在信道编码中引入足够但为有限的多余度，换句话说，只要编码速率足够小，就能通过信道渐近无误地传送消息。“渐近无误”的含意是：只要编码的数字序列充分长，就可使其译码错误概率任意小。信道编码理论要回答的问题是：

（1）对给定的信道，保证信道渐近无误地传送信息所能达到的最大编码速率是多少？（2）对给定的编码速率 R，其最优编码的译码错误概率随编码长度 N 的变化规律怎样？

（3）如何构造实现最大速率传输的最优编码？

对问题(1)和(2)已经获得相当满意的解决。例如对于离散无记忆信道，编码速率的最小上限为

$$\sup R = C = \max_{\boldsymbol{p}} I(\boldsymbol{p},\boldsymbol{Q}),$$

其中极大是对信道入口符号集 $\mathscr{X}$ 上的一切概率分布 $\boldsymbol{p}$ 而取。$\boldsymbol{Q}$ 为信道的转移概率，

$$I(\boldsymbol{p},\boldsymbol{Q}) = \sum_{x}\sum_{y} P(x)Q(y|x)\log\frac{Q(y|x)}{\sum_{x} P(x)Q(y|x)},$$