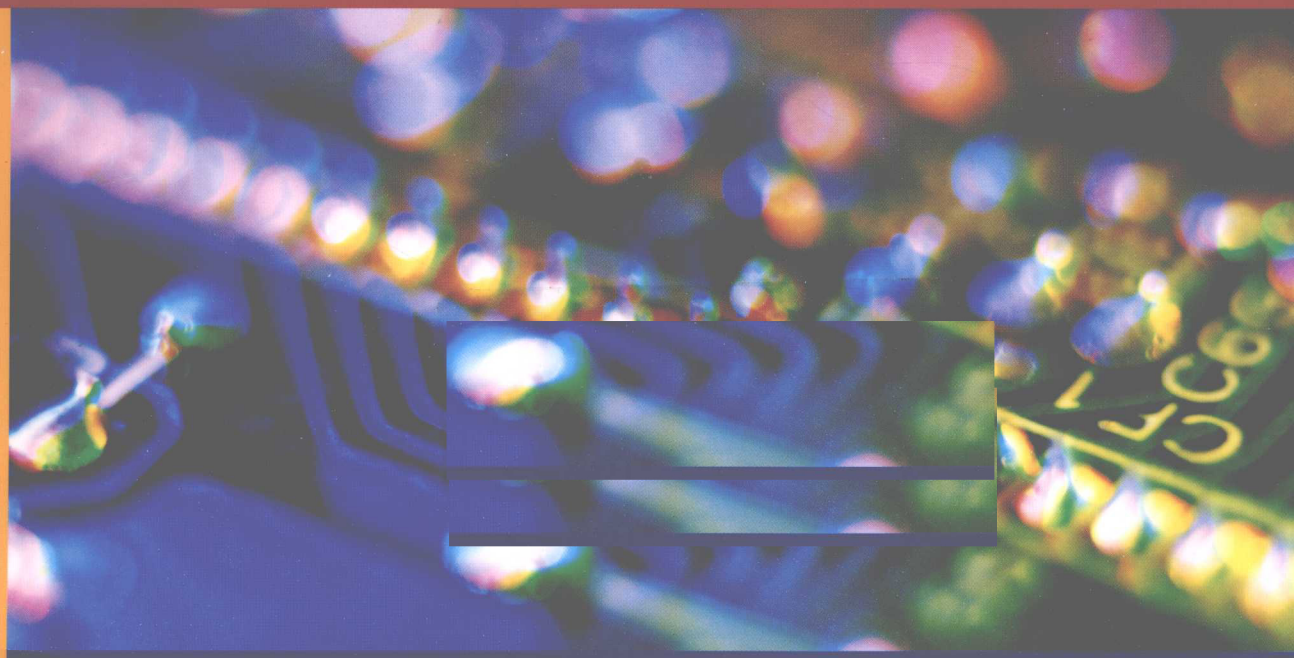
高等院校信息与通信工程系列教材

信息论与编码理论



姜楠 王健 编著

清华大学出版社

高等院校信息与通信工程系列教材

信息论与编码理论

姜楠 王健 编著

清华大学出版社
北京

内 容 简 介

本书系统地讨论了香农信息理论中的基本概念和相关问题，介绍了信源、信道、信源编码、信道编码的一般原理和基本方法。全书分为8章，包括绪论、信息的统计度量、离散信源、离散信道、连续信源和连续信道、无失真信源编码、限失真信源编码、信道编码。

本书内容深入浅出，适合作为信息工程、通信工程、信息安全、计算机应用等相关专业本科生的教材，也可作为研究生的教材或教学参考书，以及从事信息理论、信息技术、通信系统、信息安全研究的科研和工程技术人员的参考用书。

本书配有电子教案、出题系统和实验系统，便于教学和自学。

图书在版编目(CIP)数据

信息论与编码理论/姜楠，王健编著. —北京：清华大学出版社，2010.9
(高等院校信息与通信工程系列教材)
ISBN 978-7-302-23371-8

Ⅰ. ①信… Ⅱ. ①姜… ②王… Ⅲ. ①信息论 ②信源编码—编码理论 ③信道编码—编码理论
Ⅳ. ①TN911.2

中国版本图书馆CIP数据核字(2010)第147051号

责任编辑：文　怡
责任校对：焦丽丽
责任印制：孟凡玉
出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦A座
http://www.tup.com.cn　　**邮　　编**：100084
社　总　机：010-62770175　　**邮　购**：010-62786544
投稿与读者服务：010-62795954，jsjjc@tup.tsinghua.edu.cn
质　量　反　馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：清华大学印刷厂
经　　销：全国新华书店
开　　本：185×260　**印　张**：11.5　**字　数**：268千字
版　　次：2010年9月第1版　　**印　　次**：2010年9月第1次印刷
印　　数：1～3000
定　　价：20.00元

产品编号：036121-01

出版说明

信息与通信工程学是信息科学与技术的重要组成部分。改革开放以来，我国在发展通信系统与信息系统方面取得了长足的进步，形成了巨大的产业与市场，如我国的电话网络规模已占世界首位，同时该领域的一些分支学科出现了为国际认可的技术创新，得到了迅猛的发展。为满足国家对高层次人才的迫切需求，当前国内大量高等学校设有信息与通信工程学科的院系或专业，培养大量的本科生与研究生。为适应学科知识不断更新的发展态势，他们迫切需要内容新颖又符合教改要求的教材和教学参考书。此外，大量的科研人员与工程技术人员也迫切需要学习、了解、掌握信息与通信工程学科领域的基础理论与较为系统的前沿专业知识。为了满足这些读者对高质量图书的渴求，清华大学出版社组织国内信息与通信工程国家级重点学科的教学与科研骨干以及本领域的一些知名学者、学术带头人编写了这套高等院校信息与通信工程系列教材。

该套教材以本科电子信息工程、通信工程专业的专业必修课程教材为主，同时包含一些反映学科发展前沿的本科选修课程教材和研究生教学用书。为了保证教材的出版质量，清华大学出版社不仅约请国内一流专家参与了丛书的选题规划，而且每本书在出版前都组织全国重点高校的骨干教师对作者的编写大纲和书稿进行了认真审核。

祝愿《高等院校信息与通信工程系列教材》为我国培养与造就信息与通信工程领域的高素质科技人才，推动信息科学的发展与进步做出贡献。

北京邮电大学
陈俊亮

前　言

信息论和编码理论是20世纪40年代末期由美国数学家香农等人创立的，经过几十年的发展，现已成为信息科学的基础理论。

信息论和编码理论是从工程实践中抽象概括出来的理论知识，既具有很强的理论性，又有广泛的工程实践背景。初学者往往由于缺乏这种实践背景，很难理解其中的理论知识。本书力图通过读者身边看得见、摸得着的例子来解释这些理论问题。讲解深入浅出，重点在于对理论知识含义的说明，而非枯燥的证明。

本书共分8章。第1章是绪论，介绍信息、通信系统模型、离散与连续等内容。第2章介绍信息的统计度量，也是信息论的基本概念，包括自信息量、互信息量、平均自信息(熵)、平均互信息等，这一章是后续章节的基础。第3、4章分别讨论离散信源和离散信道。第5章概要介绍连续信源和连续信道。第6章和第7章分别讨论无失真信源编码和限失真信源编码。第8章讨论了信道编码。

本书作者一直从事“信息论与编码理论”的教学工作，为了满足信息安全、计算机、通信工程等相关专业人才培养的教学需要，在已有教学经验的基础上编写了本书。并开发了一套“信息论与编码理论实验系统”，该系统能够直观演示信源、信道、信源编译码、信道编译码对数据的处理过程，便于学生建立感性认识，加深对理论知识的理解。本书中多次用到实验系统的输出结果来说明问题。

本书由姜楠、王健共同完成。苏桂莲、李川编写了书中用到的Matlab程序，姜志云、田秀珍、黄海波、杨晓燕绘制了书中用到的部分插图，王树更、马玉英、蔡智文设计了书中部分例子，张明子、秦国玲、葛永德、杨红林帮助整理了习题和习题答案，姚雅欣、王勇、陈丹威、刘伟参与了试题库管理系统的编写工作。清华大学出版社的陈志辉、文怡编辑为本书的出版做了大量的工作。本书出版得到了北京工业大学重点课程(群)优秀教学团队建设项目的资助。

由于作者水平和时间所限，书中难免有不妥之处，诚恳期望读者赐教和指正。

作　者

2010年5月

目　录

第1章 绪 论

信息科学以扩展人类的信息处理能力为主要研究目标，是现代科学技术进步的主要标志之一。作为信息科学的重要理论基础——信息论，是在长期的信息与通信工程实践中，与概率论、随机过程和数理统计等近代数学学科相结合而建立并发展起来的，它主要研究各类电子信息系统和通信系统中信息的描述、传输和处理的一般规律与基本关系，研究信息系统的有效性和可靠性。

1.1 信 息

1.1.1 信息的概念

信息论是应用近代数理统计方法研究信息的传输、存储与处理的科学。因此信息论的研究对象是“信息”，那什么是信息呢？先来看一个例子。

【例 1-1】 张三给李四发送一条短信，报告了一条新闻“某沿海地区发生了海啸”，李四看过之后非常吃惊。

说明：这个例子中涉及信号、消息、信息三个概念，如图 1-1 所示。

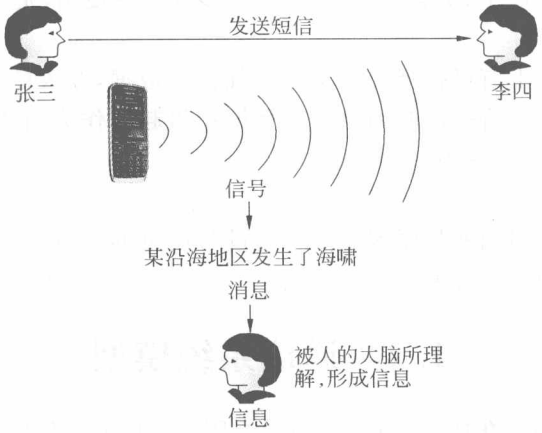


图 1-1 信号、消息、信息

(1) 短信首先被手机编码成无线电波发送出去，该无线电波是短信的载体，是实际存在的无线电“信号”。

(2) 李四的手机接收到信号之后在屏幕上显示“某沿海地区发生了海啸”，这是一条“消息”。

(3) 李四看到消息之后，会在大脑中形成自己的理解，有自己的感受，这是“信息”。

由此可以看出，信号是消息的表现形式，是物理的，例如电信号、光信号等。消息是信息

的载荷者，是信号的具体内容，不是物理的，但是又比较具体，例如语言、文字、符号、图片等。信息包含在消息中，是通信系统中被传送的对象，消息被人的大脑所理解就形成了信息。

但是信息看不见摸不着，我们通过什么来研究它呢？要回答这个问题，大家先考虑一下例 1-1 中，李四看到短信之后为什么会吃惊呢？如果他收到的短信是“张三今天吃饭了”，他还会吃惊吗？

之所以看到不同的消息会有不同的感受，是因为消息所描述的事件发生的概率不同。“某沿海地区发生了海啸”发生的概率很小，属于小概率事件；“张三今天吃饭了”发生的概率很大，属于大概率事件。小概率事件一旦发生会引起人们的关注，而针对大概率事件的发生，人们往往会视而不见。因此信息论中所指的信息是“概率信息”，即用概率来定义信息。事件发生的概率越大，它发生后提供的信息量越小；事件发生的概率越小，一旦该事件发生，它提供的信息量就越大。

概率信息是由美国数学家香农(C. E. Shannon)提出的，故又称香农信息。

1.1.2 信息的性质

1. 信息是无形的

信息看不见、摸不着，不具有实体性。

2. 信息是可以共享的

信息易于复制，能以极快的速度传播，是一种可以共享的重要的社会资源。信息的交流不但不会使信息的持有者失去原有的信息，而且可以获得新的信息。

3. 信息是无限的

信息是无限的，有两个含义：

一是说信息永远在产生、更新和演变中，可以多人共享使用，是一个取之不尽、用之不竭的知识源泉。

二是说信息在时空上有可扩展性。例如天气预报数据，今天的天气预报只对今天起作用，明天就失去价值，但是将一段时间之内的数据积累起来作为历史资料，又可成为关于气候演变的重要信息，给人类造福。

4. 信息是可度量的

信息论中一个重要问题就是要解决信息数量与质量的度量。在香农的信息定义中，信息量与事件发生的不确定性有关。

1.2 通信系统模型

信息论要研究信息，那在什么环境下研究信息呢？信息论是在通信系统环境下研究信息，信息论又称为通信的数学理论。因此通信系统模型(如图 1-2 所示)是信息论研究的基础。

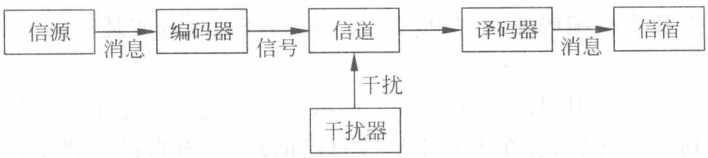


图 1-2 通信系统模型

1.2.1　信源和信宿

信源是产生消息和消息序列的源头，它可以是人、生物、机器或其他事物。信源发出的消息有语音、图像、文字等。信源发出的是消息而不是信息，这是因为信息看不见摸不着，只能通过消息来研究它。

信宿是消息的接收方，即接收消息的人或物。

1.2.2　编码器和译码器

所谓编码是指将消息从一种表示形式变换为另外一种表示形式。根据编码的目的不同，编码器可以分为信源编码、保密编码、信道编码、调制编码四种。

信源编码的目的是压缩消息的数据量，使得消息能够被更经济地传送出去，即提高信息传送的有效性。

保密编码的目的是保证消息的安全性。

信道编码的目的是消除信道上噪声的影响，保证发送的消息不发生错误，即提高信息传送的可靠性。

调制编码的目的是将消息变为能够传送的信号，例如光信号、电信号等。

编码理论主要研究信源编码和信道编码。

译码器的作用与编码器的作用正好相反，可以将接收到的信号恢复为原始的消息。

1.2.3　信道和噪声

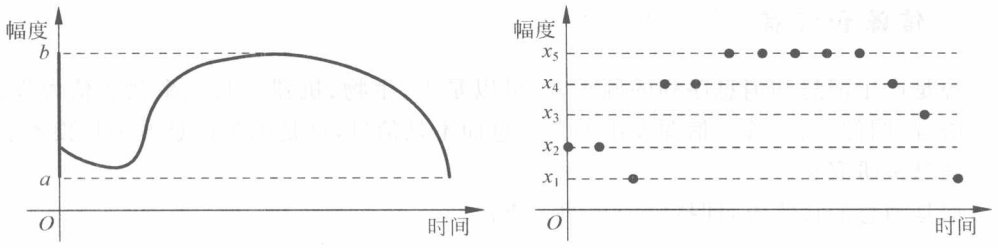
信道是把信号从发射端传送到接收端的通道。

信道上总是存在干扰，通常这种干扰来自通信系统的外部，是通信系统所不能控制的，因此是不可避免的，例如卫星通信的信道经常受到太空中各种辐射的干扰。由于干扰的存在，使得发送的信号经常会发生错误，这也是信道编码产生的原因。

1.3　离散与连续

在 1.2.1 节中已经提到，因为信息看不见摸不着，只能通过消息来研究它。消息有两种形式：离散的和连续的。这两种消息有共同点，都是时间（或者空间）的函数，例如声音是时间的函数，图像是空间的函数（从信息论的角度，不区分时间和空间这两个概念）。离散消息和连续消息在数学模型、数学工具、信息量计算方法、编码方法、工程应用等方面都存在很大的不同。因此有必要对两者做简单介绍。

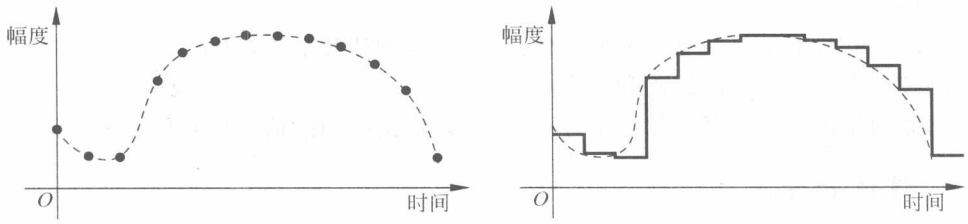
离散消息和连续消息最主要的差别体现在值域。离散消息的值域取自于集合$\{x_1, x_2, \cdots, x_n\}$，该集合是可数的。例如，二进制消息取自于集合{0，1}，英文取自于集合{a，b，…，y，z}，中文取自于汉字集合{我，信，随，大，…}。连续消息的值域取自于区间$[a, b]$，区间是不可数的。图 1-3 所示是离散消息和连续消息的例子，图(a)中所示的曲线是连续消息，因为消息的取值范围为区间$[a, b]$，图(b)中所示的黑点是离散消息，因为消息的取值范围为集合$\{x_1, x_2, x_3, x_4, x_5\}$。



(a) 连续消息　　(b) 离散消息

图 1-3　离散消息和连续消息

还有一种介于两者之间的消息，在时间上是离散的，在幅度上是连续的，这种称为离散时间消息。如图 1-4(a)所示的黑点就是离散时间消息，时间轴上是离散的，而幅度轴上的取值范围仍然是区间$[a, b]$，即可以在区间$[a, b]$内任意取值。图 1-4(a)和图 1-4(b)所示是离散时间消息的两种表示形式。离散时间信号仍然是连续信号，信号的幅度只能在特定的时刻变化。



(a)　　(b)

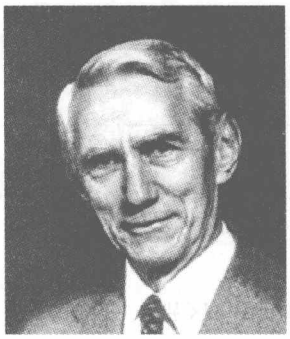
图 1-4　离散时间消息

本书以离散消息为主。

1.4　信息论和编码理论的形成和发展

1948 年 10 月，香农在《贝尔系统技术学报》上发表了一篇题为《通信的数学理论》(*The Mathematical Theory of Communication*)的论文，如图 1-5 所示。在这篇论文中，香农阐述了信息论的关键概念和方法，奠定了信息论的理论基础。这篇论文的发表标志着信息论的形成，因此信息论又称为香农信息论。

其实，早在 1948 年之前，很多信息论和编码理论中的概念和研究方法就已经出现了，像带宽、信息率、随机过程和数理统计的研究方法等。这些概念和方法的出现，与 19 世纪末到 20 世纪 40 年代通信技术的大发展和两次世界大战密不可分，如图 1-6 所示。电报、电话、电视、传真、调频、扩频等通信成果和技术接连出现，一方面迫切需要产生一套理论来指导技术的进一步发展，另一方面为理论的产生积累了经验、奠定了基础。两次世界大战促进了现代密码学的形成，保密编码得以快速发展。因此 1948 年信息论的产生是通信技术和密码学发展的必然结果。



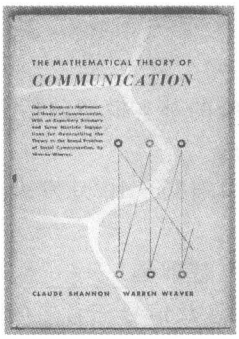


图 1-5　香农和《通信的数学理论》

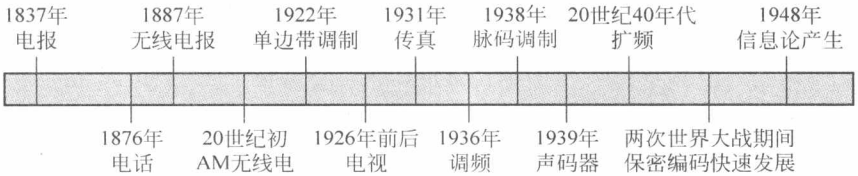


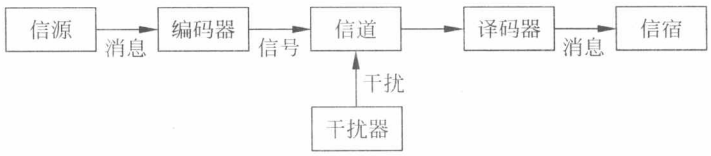
图 1-6　19 世纪末到 20 世纪 40 年代通信技术的大发展

此后，在信息论的理论框架下，信源编码、信道编码、保密编码基本平行发展，理论日趋完善，编码方法不断出现，效果越来越好。目前，信息论和编码理论的发展已经比较成熟，基本能够处理通信系统中遇到的各种情况。现在的发展主要集中在进一步研究信源和信道的特点，以及改善各种编码方法的效果等方面。

小　　结

本章介绍了信息、通信系统模型、离散消息与连续消息、信息论和编码理论的形成和发展，如表 1-1 所示。

表 1-1　本章小结

信息	概率信息
	性质：无形的、可共享的、无限的、可度量的
	
离散消息的值域取自于集合，连续消息的值域取自于区间	
1948 年 10 月，香农在《贝尔系统技术学报》上发表了一篇题为《通信的数学理论》（*The Mathematical Theory of Communication*）的论文	

习 题

1.1 一个通信系统的基本模型包括_____、_____、_____、_____、_____、_____六个组成部分。

1.2 _____于1948年10月发表了论文《通信的数学理论》，奠定了概率信息论的基础。

1.3 信息论的研究基础是_____。

1.4 说明信息、消息及信号三者之间的联系与区别。

1.5 分别举出日常生活中大信息量和小信息量事件的例子。

第 2 章　信息的统计度量

信息有大有小，可以定量度量。本章就要解决信息的定量度量问题，这是后续研究的基础。讨论时以离散消息为主，简要介绍连续消息的度量。

2.1　自信息和条件自信息

2.1.1　自信息的定义与含义

从 1.1 节得知，信息论中所指的信息是“概率信息”，事件发生的概率越大，它发生后提供的信息量越小；事件发生的概率越小，一旦该事件发生，它提供的信息量就越大。那么一个事件所包含的信息量与概率之间的函数关系到底是怎样的呢？

【定义 2-1】　一个事件的自信息量定义为该事件发生概率的对数的负值。

假设事件 $x_i \in \{x_1, x_2, \cdots, x_n\}$，发生的概率为 $p(x_i)$，则其自信息定义式为

$$I(x_i) = -\log p(x_i) \tag{2-1}$$

自信息量有单位，它的单位与所取对数的底有关。通常取对数的底为 2，此时信息量的单位为比特(bit)，在本书中为了书写方便，将底数 2 省略，即 $\log_2(\cdot)$ 写为 $\log(\cdot)$ 的形式。例如 $p(x_i) = 1/2$，则 $I(x_i) = -\log(1/2) = 1$ 比特。自信息量的函数图形如图 2-1 所示，从图中可以看出，由于 $0 \leqslant p(x_i) \leqslant 1$，因此 $I(x_i) \geqslant 0$，且自信息量和概率成反比。规定 $0\log 0 = 0$。

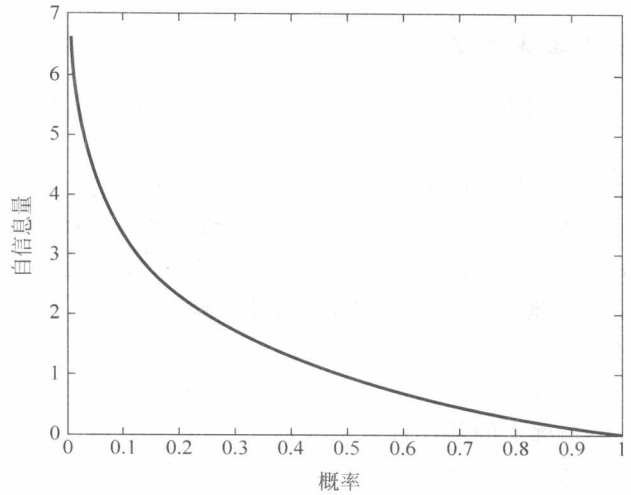


图 2-1　自信息量函数图形