21世纪

高等院校规划教材

英文版

梁建武　郭　迎　罗喜英　刘军军　编　著

Information Theory & Coding

信息论与编码

注重学科体系的完整性，兼顾考研学生需要

强调理论与实践相结合，注重培养专业技能

中国水利水电出版社
www.waterpub.com.cn

21世纪高等院校规划教材

Information Theory & Coding
信息论与编码（英文版）

梁建武　郭　迎　罗喜英　刘军军　编著

中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

本书重点介绍经典信息论的基本理论，并力图将信息论的基本理论和工程应用的编码理论联系起来，介绍一些关于这些理论的实际应用。全书分为 7 章，内容包括信息度量的基本理论、无失真信源编码、限失真信源编码、信道编码及其应用等。

本书注重基本概念，并且用通俗易懂的语言对它们加以诠释。在当前信息、通信系统飞速发展的大背景下，本书力图用较多的例子和图表来阐述概念和理论，同时尽量避免纠缠于烦琐难懂的公式证明之中。为了加深读者对所讲述知识的理解，每章最后都配有适量的练习题供读者选用。

本书可作为高等院校电子信息类学生双语教学的教材或参考书，也可作为通信、电信、电子等领域从业人员的参考资料。

本书配有电子教案，读者可以从中国水利水电出版社网站免费下载，网址为：http://www.waterpub.com.cn/softdown/。

图书在版编目（CIP）数据

信息论与编码=Information Theory & Coding：英文 /
梁建武等编著. —北京：中国水利水电出版社，2008
21 世纪高等院校规划教材
ISBN 978-7-5084-5569-3

Ⅰ. 信… Ⅱ. 梁… Ⅲ. ①信息论—高等学校—教材—英文②信源编码—编码理论—高等学校—教材—英文③信道编码—编码理论—高等学校—教材—英文 Ⅳ. TN911.2

中国版本图书馆 CIP 数据核字（2008）第 064948 号

书　　名	Information Theory & Coding 信息论与编码（英文版）
作　　者	梁建武　郭迎　罗喜英　刘军军　编著
出版发行	中国水利水电出版社（北京市三里河路 6 号　100044） 网址：www.waterpub.com.cn E-mail：mchannel@263.net（万水） sales@waterpub.com.cn 电话：（010）63202266（总机）、68367658（营销中心）、82562819（万水）
经　　售	全国各地新华书店和相关出版物销售网点
排　　版	北京万水电子信息有限公司
印　　刷	北京市天竺颖华印刷厂
规　　格	184mm×260mm　16 开本　13.25 印张　323 千字
版　　次	2008 年 7 月第 1 版　2008 年 7 月第 1 次印刷
印　　数	0001—3000 册
定　　价	24.00 元

序

随着计算机科学与技术的飞速发展，计算机的应用已经渗透到国民经济与人们生活的各个角落，正在日益改变着传统的人类工作方式和生活方式。在我国高等教育逐步实现大众化后，越来越多的高等院校会面向国民经济发展的第一线，为行业、企业培养各级各类高级应用型专门人才。为了大力推广计算机应用技术，更好地适应当前我国高等教育的跨跃式发展，满足我国高等院校从精英教育向大众化教育的转变，符合社会对高等院校应用型人才培养的各类要求，我们成立了“21 世纪高等院校规划教材编委会”，在明确了高等院校应用型人才培养模式、培养目标、教学内容和课程体系的框架下，组织编写了本套“21世纪高等院校规划教材”。

众所周知，教材建设作为保证和提高教学质量的重要支柱及基础，作为体现教学内容和教学方法的知识载体，在当前培养应用型人才中的作用是显而易见的。探索和建设适应新世纪我国高等院校应用型人才培养体系需要的配套教材已经成为当前我国高等院校教学改革和教材建设工作面临的紧迫任务。因此，编委会经过大量的前期调研和策划，在广泛了解各高等院校的教学现状、市场需求，探讨课程设置、研究课程体系的基础上，组织一批具备较高的学术水平、丰富的教学经验、较强的工程实践能力的学术带头人、科研人员和主要从事该课程教学的骨干教师编写出一批有特色、适用性强的计算机类公共基础课、技术基础课、专业及应用技术课的教材以及相应的教学辅导书，以满足目前高等院校应用型人才培养的需要。本套教材消化和吸收了多年来已有的应用型人才培养的探索与实践成果，紧密结合经济全球化时代高等院校应用型人才培养工作的实际需要，努力实践，大胆创新。教材编写采用整体规划、分步实施、滚动立项的方式，分期分批地启动编写计划，编写大纲的确定以及教材风格的定位均经过编委会多次认真讨论，以确保该套教材的高质量和实用性。

教材编委会分析研究了应用型人才与研究型人才在培养目标、课程体系和内容编排上的区别，分别提出了 3 个层面上的要求：在专业基础类课程层面上，既要保持学科体系的完整性，使学生打下较为扎实的专业基础，为后续课程的学习做好铺垫，更要突出应用特色，理论联系实际，并与工程实践相结合，适当压缩过多过深的公式推导与原理性分析，兼顾考研学生的需要，以原理和公式结论的应用为突破口，注重它们的应用环境和方法；在程序设计类课程层面上，把握程序设计方法和思路，注重程序设计实践训练，引入典型的程序设计案例，将程序设计类课程的学习融入案例的研究和解决过程中，以学生实际编程解决问题的能力为突破口，注重程序设计算法的实现；在专业技术应用层面上，积极引入工程案例，以培养学生解决工程实际问题的能力为突破口，加大实践教学内容的比重，增加新技术、新知识、新工艺的内容。

本套规划教材的编写原则是：

在编写中重视基础，循序渐进，内容精炼，重点突出，融入学科方法论内容和科学理念，反映计算机技术发展要求，倡导理论联系实际和科学的思想方法，体现一级学科知识组织的层次结构。主要表现在：以计算机学科的科学体系为依托，明确目标定位，分类组织实施，兼容互补；理论与实践并重，强调理论与实践相结合，突出学科发展特点，体现

学科发展的内在规律；教材内容循序渐进，保证学术深度，减少知识重复，前后相互呼应，内容编排合理，整体结构完整；采取自顶向下设计方法，内涵发展优先，突出学科方法论，强调知识体系可扩展的原则。

本套规划教材的主要特点是：

（1）面向应用型高等院校，在保证学科体系完整的基础上不过度强调理论的深度和难度，注重应用型人才的专业技能和工程实用技术的培养。在课程体系方面打破传统的研究型人才培养体系，根据社会经济发展对行业、企业的工程技术需要，建立新的课程体系，并在教材中反映出来。

（2）教材的理论知识包括了高等院校学生必须具备的科学、工程、技术等方面的要求，知识点不要求大而全，但一定要讲透，使学生真正掌握。同时注重理论知识与实践相结合，使学生通过实践深化对理论的理解，学会并掌握理论方法的实际运用。

（3）在教材中加大能力训练部分的比重，使学生比较熟练地应用计算机知识和技术解决实际问题，既注重培养学生分析问题的能力，也注重培养学生思考问题、解决问题的能力。

（4）教材采用“任务驱动”的编写方式，以实际问题引出相关原理和概念，在讲述实例的过程中将本章的知识点融入，通过分析归纳，介绍解决工程实际问题的思想和方法，然后进行概括总结，使教材内容层次清晰，脉络分明，可读性、可操作性强。同时，引入案例教学和启发式教学方法，便于激发学习兴趣。

（5）教材在内容编排上，力求由浅入深，循序渐进，举一反三，突出重点，通俗易懂。采用模块化结构，兼顾不同层次的需求，在具体授课时可根据各校的教学计划在内容上适当加以取舍。此外还注重了配套教材的编写，如课程学习辅导、实验指导、综合实训、课程设计指导等，注重多媒体的教学方式以及配套课件的制作。

（6）大部分教材配有电子教案，以使教材向多元化、多媒体化发展，满足广大教师进行多媒体教学的需要。电子教案用 PowerPoint 制作，教师可根据授课情况任意修改。相关教案的具体情况请到中国水利水电出版社网站 www.waterpub.com.cn 下载。此外还提供相关教材中所有程序的源代码，方便教师直接切换到系统环境中教学，提高教学效果。

总之，本套规划教材凝聚了众多长期在教学、科研一线工作的教师及科研人员的教学科研经验和智慧，内容新颖，结构完整，概念清晰，深入浅出，通俗易懂，可读性、可操作性和实用性强。本套规划教材适用于应用型高等院校各专业，也可作为本科院校举办的应用技术专业的课程教材，此外还可作为职业技术学院和民办高校、成人教育的教材以及从事工程应用的技术人员的自学参考资料。

我们感谢该套规划教材的各位作者为教材的出版所做出的贡献，也感谢中国水利水电出版社为选题、立项、编审所做出的努力。我们相信，随着我国高等教育的不断发展和高校教学改革的不断深入，具有示范性并适应应用型人才培养的精品课程教材必将进一步促进我国高等院校教学质量的提高。

我们期待广大读者对本套规划教材提出宝贵意见，以便进一步修订，使该套规划教材不断完善。

21 世纪高等院校规划教材编委会

2004 年 8 月

Brief Introduction

This book is mainly about foundation of information theory and the coding theory. Also, some practical applications of these theories are mentioned in some occasions. On the basis of introducing the measure of information, it mainly focuses on the introduction of the theory of lossless source coding, limited loss source coding, channel coding and their applications. It pays attention to basic concepts and uses popular text to explain them. In the background of the current communication system, it includes many examples and figures to elaborate these concepts and theorems. At the same time, it does its best to avoid getting entangled in proving the equations. To deepen the understanding to the knowledge it touches upon, there are some exercises attached at the end of every chapter.

This book can be used as the teaching material for college students, and also, it can be the reference for career man in the area of communication, telecommunication and electronic.

Preface

Information theory is a mathematical theory of communication theory. It is science to study information measure and coding using mathematical statistics method. It can be divided into generalized information theory and narrow information theory which is also called classical theory of information or probabilistic information theory.

This book is on the foundation of probabilistic information measures and their application to coding theorems for information sources and noisy channels. It pays attention to elaborate the basic theories, concepts and methods. Under this guideline, there are many examples throughout this book to make it clearer about these concepts and theorems. Too mathematical a book has the potential danger of scaring away beginner. So, it makes every effort to use simple mathematics tools in the argument of information theory, which may lead beginners to the way of mastering it.

It can be used as teaching material in Electronics Department for college students of upper grade or for graduate students. The whole book is composed by seven chapters. They are both lively introductions and fairly detailed references to the fascinating world of information theory.

Chapter 1 is introduction, a general illustration from the angle of history about the basic problems this book dealing with. For example, the concept of information and the measure method of it, the formation and development of information theory are all included.

Chapter 2 deals with some basic concepts on information measurement. Through the study of this chapter you may gain the ability of knowing how much information you can get from a trial of tossing a coin and read the result of it. Also, this chapter introduces some theorems on continuous source.

Chapter 3 is on discrete source information. The source model and the calculation of source entropy are introduced in this part. Through the study of this part, the reader can understand why there is space for compression.

Chapter 4 is about channels. For channel, it uses channel capacity to measure the transfer ability. The channel model and the capacity calculation of some typical channels discrete or continuous ones are the focus. After the study of this part, readers can see the real situation of information transmission.

Chapter 5 deals with lossless source coding theorem. In chapter 3, it knows that source needs to be compressed to suit for the transmission. In this part, it shows how to compress the information of source keeps.

Chapter 6 deals with limited distortion source coding. Former chapters are all in condition of without distortion, but in real world, distortion in some degree can't be realized by people's sense, so the information in transmission or records can be less. This chapter is about the theorems and practice of this idea.

Chapter 7 introduces the coding theory. Some famous and useful coding method is included in it, such as the CRC code and Hamming code.

The arrangement of content for this book is from simple to complex. To improve the ability of analyzing and solving problems, there are some questions and exercises at the end of each chapter.

This book is the harvest of hard work and wisdom of a group of people. The members participating the writing of the book are LiangJianwu(Chapter1,2,3,4), GuoYing(Chapter7), LuoXiying(from Hunan University of Science and Technology, Chapter5),LiuJunjun(Chapter6), LongXiaomei(verify), TianYe(proofreading), ZhouYuanyuan(typesetting). Otherwise, HeZhibin, LiHuawei, FuShifeng, TanHailong, WenZheng, all did great favors for the writing of this book. We have tried our best to free the book from all errors, but unfortunately there does not exist a foolproof error control technique for that. So, you are welcome to point out our mistakes and give us your opinions.

Authors of this book in Central South University

March, 2008

目　录

Chapter 1 Introduction

Contents

- Information
- Origin and development of Information Theory
- The application and achievement of Information Theory methods
- Formation and Development of Information Theory

In this chapter, it shows the basic problems the whole book will talk about. What is information? What is information theory? The achievements, application, development and formation of information theory are also the topics of this chapter.

Before it starts, there is something must be known

First, the main content will show those: how to measure the information, this question can be answered after the way of Shannon measures information is understood. As a student of communication engineering, the concepts such as information source and channel, have been contracted with before. The distortionless way of coding for source and channel are included in the Shannon's first theorem, Shannon's second theorem and third theorem which will be shown later. In addition, many examples for the applications of Information Theory will be involved in the course of this study.

Second, the importance of studying the Information Theory also needs to be emphasized. The Information Theory is the elementary theory of Information Science and Technology. Without the foundation of Information Theory, one cannot be engaged in the communication domain research and innovation, nor can he touch the edge of this field. In brief, the Information Theory is the essential elementary knowledge to master for one who has high level information technology.

The course of Fundamental Information Theory is the foundation curriculum for communication and the information field. Only when it is mastered, can it learn the succeeding curriculum and occupy in the scientific research and the innovation in the information field. So, cherish this study opportunity, study hard to raise the scientific research abilities gradually and set up the consummated personality foundation, thus to lay solid foundation for further study and scientific research.

Third, goal of this learning must be clear. Fully understands the basic concepts and theorems of Information Theory, skilled grasp the content of the materials present for you, skilled grasp in the mathematical description of the fundamental research object in information Engineering. Achieve

certain ability to solve problem and analysis through doing some exercises. With the base of all above, the ultimate objective thing is get the ability to think and describe with certain scientific method with Information Theory ideas and viewpoints.

1.1 What is Information

Although there is no accurate definition of Information, there are two obvious characteristics for information: widespread and abstract.

Widespread: The objective world is filled with information. Humanity cannot leave without information. Knowledge in books is the useful information accumulated.

Abstract: There are three essential factors composed the objective world and promoted the development of society: matter, energy and information. In these three essential factors, matter is the foundation and entity; energy is the motion of matter form. Through the famous formula: $E = mc^2$, the matter may be transformed into energy, but the energy is the main power to change the objective world. Information is the third essential composition.

To make it more clearly, we'd better to discriminate three concepts:

Information: It attaches to the matter and the energy, but also is different from matter and energy. Without information, we cannot use the matter and the energy well. Human beings use information and knowledge to change the real world, create new matter, and enhance the efficiency of energy usage and to discover new energy forms.

The information is objective. It is the main power which man uses to contact, and change the objective world. Also it is in a higher level that the men understood the objective world.

Maybe it can be explained as the asymency degree in space and time field for matter and engrey.

Signal: It is the physical expression level of the information and the most concrete level. Also, it is a physical quantity, an entity to bear information, may be surveyed, be described and demonstrated.

Message (Also named symbol): It is the mathematical expression level of the information. Although it is not a physical quantity, it can be described in the form of quantity. It is the further mathematics abstract from concrete physical signal. The concrete physical signal is divided into two types:

The discrete (digital form) message is a group of unknown quantities, may be described with the random sequence: $U = (U_1 \cdots U_i \cdots U_L)$.

The continuous (analog form) news, is a group of unknown quantity, it may be described with the stochastic process: $U(t, \omega)$.

Through the comparison, it's clear to see the relationships among signal, message and information:

Information is in higher level of philosophy abstract, i.e. it is the higher expression level of signal and message. In these three levels, signal is the most concrete and information is the most

abstract. They are in the philosophy connotation and extension relations. Information may be seemed to be the connotation of concrete physical signal and the mathematical description of the message, and the meaning of concrete supporting body for content of signal and message description. Signal is the extension of abstract information in the physical level; message is the extension of abstract information in mathematical level expression.

The same information may be carried by different signal forms (e.g. word, language, image and so on); the same information also may be described with different mathematical expression forms (e.g. discrete or continuous). Similarly, the identical signal form, for instance "0" and "1" may be expressed in different forms, without or with, breaks or passes, low or high (level) and so on.

Main characteristics of information

The information comes from the matter, but it is not the matter itself. It is produced from the material movement, also may be separated from the source element and exist independently. The information comes from the inner world, but it does not limit to the energy domain. The information and the energy are closely linked, but they have essential difference with the energy. The information has the nature of knowledge, but it has more widespread connotation than knowledge, and can be gained and used.

Important nature of information

So much has been talked about information, if anyone wants you to describe it with some words, the following maybe all right:

Universal existence, sequence, relativity, measurable, extendable, can be saved, transported and portable, compressed, vicarious, diffused, shared.

The information will play a vital role in our society which is a highly information-oriented one. It will be a more precious wealth than material and energy.

Shannon Information (Probability information)

Shannon is the creator of Probability Information Theory, which is the main topic of this book. The information with the help from Probability Theory can be used to measure of the state of motion or the indefinite existence way of the matter. Consider the basic model of communication system as following:

It seemed that the message is transmitted in the communication system, and actually the information contained in these messages is transmitted at the same time.

To understand Shannon's theory, learning some concepts may be inevitable.

Sample space: since things may have various appearances, the message set of all possibility is called sample space.

Regarding the discrete message set, the probability measure is to assign a probability to each possibility. A sample space and its probability measure are called a probability space. Expression is: $[X,P]$, X is the possibility, and P is its probability. It is expressed as below formula:

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & \cdots, & a_q \\ p(a_1), & p(a_2), & \cdots, & p(a_q) \end{bmatrix} \tag{1.1}$$

In this expression, a_i is the selected symbol, $p(a_i)$ is the probability of the message accordingly, called pre-test probability. With the expression we can calculate some index about information.

Self-Information of the message (symbol):

$$I(a_i) = \log \frac{1}{P(a_i)} = -\log P(a_i) \tag{1.2}$$

After-test probability: also called conditional probability $p(a_i \mid b_j)$ —the probability that the receiver received message (symbol) b_j while the transmitter has transmitted a_i.

Mutual information: the information of pre-test minus the information lost through the transmission (information content which the receiver obtains):

$$I(a_i;b_j) = \log \frac{1}{P(a_i)} - \log \frac{1}{p(a_i \mid b_j)} \tag{1.3}$$

1.2 What's Information Theory?

It is the theory that C. E. Shannon summarized from the martial intelligence transmission question in late 1940's. The theory regards the objective probability information as the research objects. Its main research question is:

- Source description, information ration measurement, analysis and computation.
- Channel description, the channel transmission ration measure, analysis and computation.
- Statistical match between source, channels and communication systems, as well as optimization of communication systems — i.e. the three Shannon coding theorems.

The Information Theory has been developed for 50 years, until now, it is still the basic theory to instruct the development of communication technology, and it is also the fountain-head of innovating new communication systems.

1.2.1 Origin and Development of Information Theory

In the historical perpetual flow of human being, the information transmission and the dissemination methods have experienced five reformations:

- The first reformation: the production of language.
- The second reformation: the production of writing.
- The third reformation: the invention of printing method.
- The fourth reformation: the invention of telegram and telephone.
- The fifth reformation: computer technology and the communication have been unified, which promoted the development of networks.

Contribution of Shannon

In 1948, Shannon published *A Mathematical Theory of Communication* in *The Bell System Technical Journal*, which symbolized the birth of Information Theory. In 1949, he published

Communication Theory of Secrecy Systems. In this theory, he first put forward the viewpoint of Information Theory to make comprehensive elaboration on information security. In 1959, he published an article about discrete source coding theorem with the fidelity criterion. In this paper, he proposed the data rate loss theory, and laid the theory foundation for the source compression coding research. In 1961, he published another paper about bi-channel communication, which propelled the development of multi-user information theory's research.

Research categories of Information Theory

As is shown in Fig. 1.1, the foundational information theory is only a small part of the whole theory system. Because of this reason, people who engage in information technology may have a foot in the door from this part.

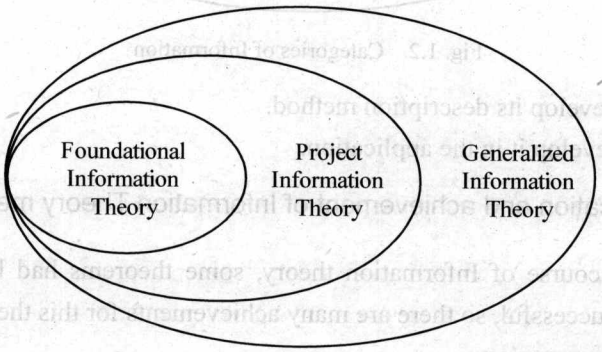

Fig. 1.1 Categories of Information Theory

Foundational Information Theory

It mainly includes researching mathematical description and quantitative analysis of the communication systems. And it also includes researching the optimizations of the system, in other words, it is research on theoretically potential power of the communication system and the mathematical limit. Since it takes the existence as the main research body, we also call it mathematical Information Theory.

Project Information Theory

It takes the project as the background, mainly studies on the best working rule of each part in communication systems. With the best principle of design, it takes the structure as the main body, primarily engaged in the technical questions.It mainly includes:

- Source coding, decoding theory and methods to structure it.
- Channel coding, decoding theory and methods to structure it.
- Best modulation and demodulation theory and their realizations.
- Best examination, estimate value and best receive theory and their realizations.
- Best information processing theory, method and algorithm.

Generalized Information Theory

The core question is how to develop it. It has triple meanings:

First is the widening of information meaning. There are several kinds of information with

different meanings below in Fig. 1.2, from which it shows that the Shannon information talking about in this book is just a basic form of it.

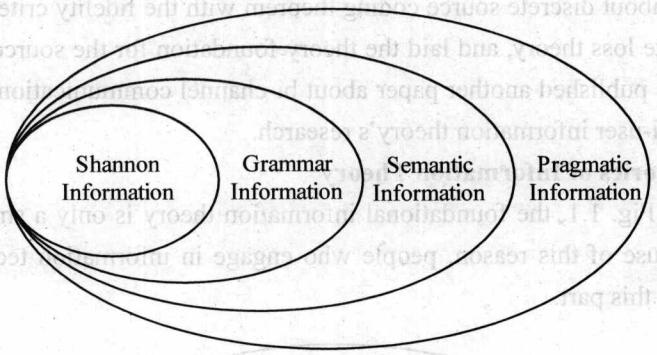

Fig. 1.2 Categories of Information

Second is how to develop its description method.

Finally is how to develop it in the application.

1.2.2 The application and achievement of Information Theory methods

In the developing course of Information theory, some theorems had been proved failed but some had been proved successful, so there are many achievements for this theory.

(1) Voice signal compression

Voice signal is the earliest form of information to be transmitted. The speed of the transmission of compressed voice signal becomes important. In the project of real world, PCM sampling is still in use, its speed of sampling is 64kbit/s, with which the voice signal can be transmitted from the sender and recovered well in the receive end. As far as Information Theory concerned, the speed of coding can be far lower than the speed which we now using: 64kbit/s to 100bit/s. It is the target to lower the speed of sampling in projects.

(2) Image information compression

It is more significant than other source for image to do a **thoroughgoing** compression, which is because the huge redundant information in it. Here are several standards and their compression ratios:

H.261	48:1 (TV conference)
JPEG	24:1 (Static image)
MPEG-1	50:1 (Dynamic image)

(3) File compression

To solve the problem of space shortage for storage, it is usually compress files with many algorithms. Until now there are more than 20 kinds of compress algorithms. Source programs, destination codes, papers, novels, scientific data and books can all be compressed by 30 percent. They are all lossless compressions which make them can be unzipped to read them.

(4) Enhancement of data transmission speed in analog telephone channel