“十三五”普通高等教育规划教材

信息理论基础

Xinxi Lilun Jichu

臧鸿雁　李国东　范修斌　编著

北京邮电大学出版社
www.buptpress.com

五”普通高等教育规划教材

信息理论基础

臧鸿雁　李国东　范修斌　编著

北京邮电大学出版社
·北京·

内 容 提 要

本书重点介绍经典信息论基本理论，全书共 7 章，内容包括：信息的统计度量；信源的数学模型及其信息量的度量；无失真信源编码理论和方法；信道的数学模型和信道容量；信道编码理论及信道编码方法；信息论在密码学中的应用。

本书通俗易懂，概念清楚，表述简洁。可作为高等院校信息与计算科学、信息工程、通信工程等相关专业的本科生、研究生少学时的信息论教材或参考书，也可供相关领域科研工作者参考。

图书在版编目(CIP)数据

信息理论基础 / 臧鸿雁，李国东，范修斌编著．--北京：北京邮电大学出版社，2017.12
ISBN 978-7-5635-5326-6

Ⅰ. ①信… Ⅱ. ①臧… ②李… ③范… Ⅲ. ①信息论—教材 Ⅳ. ①G201

中国版本图书馆 CIP 数据核字(2017)第 306570 号

书　　名　信息理论基础
编 著 者　臧鸿雁　李国东　范修斌
责任编辑　张保林
出版发行　北京邮电大学出版社
社　　址　北京市海淀区西土城路 10 号(100876)
电话传真　010-82333010　62282185(发行部)　010-82333009　62283578(传真)
网　　址　www.buptpress3.com
电子信箱　ctrd@buptpress.com
经　　销　各地新华书店
印　　刷　北京九州迅驰传媒文化有限公司
开　　本　787 mm×960 mm　1/16
印　　张　11.75
字　　数　229 千字
版　　次　2017 年 12 月第 1 版　2017 年 12 月第 1 次印刷

ISBN 978-7-5635-5326-6　　定价：39.00 元

如有质量问题请与发行部联系

序

信息论是20世纪40年代发展起来的学科，是研究通信的数学理论，是应用数理统计的方法研究信息的度量、编码和通信的科学，主要研究提高通信系统有效性和可靠性的理论和方法。近年来，随着信息理论和编码技术的长足发展，它已经发展成为一门与多领域交叉的综合性学科，广泛应用于通信、计算机、电子信息、信息安全等领域，广泛渗透于生物、语言学、经济、社会科学等众多领域。因此，越来越多的高校针对不同专业的本科生纷纷开设信息理论基础、信息工程基础、信息论与编码等相关课程。

作者面向信息与计算科学专业学生，从事信息理论基础课程教学15年，在长期教学实践中，深知初学者接受这门学科时的障碍和难点。本书为数学学科的读者编写，适用于不超过48学时课堂教学的本科课程或研究生课程。本书的特点是注重基本概念的阐述，数学建模的思想方法，数学证明的逻辑性。在教材内容上，并不求全，比如：本书内容仅针对离散信息，未涉及连续信息；本书放弃了限失真新源编码部分，而是力求将信息理论中最基本的概念、理论、思想、方法讲解透彻，并在阐述基本概念时加入了一些通俗易懂的例子。学生如果能透彻理解本书内容，限失真新源编码部分和连续信息的相关概念与理论便很容易自学获得。

本书的先修课程是概率论与数理统计，本书对随机过程相关概念按照需要做了适当补充。本书内容共包括7章。第1章是绪论。第2章是信息的度量。第3章是信源的数学建模和在此基础上信源的度量问题。第4章是信源编码，介绍无失真信源编码理论和方法。第5章是信道的数学模型及其信道容量。第6章是信道编码理论和方法。第7章是信息论在密码学中的应用。其中，第1～6章由臧鸿雁、李国东编写，第7章第1、2节由范修斌编写，第7章第3节由臧鸿雁编写。

本教材的出版得到了北京科技大学教材建设经费的资助。感谢北京科技

大学研究生柴宏玉同学，信息与计算科学系12级同学何嘉鑫、袁恬、段鑫磊、汪伦、赵晓婷和李改静，14级韦心元、吴得泱等同学对本书的贡献。

由于作者水平有限，书中难免有不妥之处，请读者赐教和指正。

编者

2017年8月

目　录

第1章 绪 论

信息论亦称为通信的数学理论，产生于20世纪40年代末50年代初，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。信息论的创始人是美国贝尔电话研究所的数学家香农(C. E. Shannon，1916—2001)。信息论的产生和发展是信息时代的迫切需要。

1.1 信息传输中的若干问题

高质量、高效率、大信息量的通信是信息时代的迫切需要。比如：远在国外的小明在傍晚的时候跟女朋友视频聊天。这个简单的通信过程涵盖了很多关于通信的基本问题。

Shannon说："通信的基本问题就是在一点重新准确地或近似地再现另一点所选择的消息。"这也是通信的最基本要求。

通信的发展目标是实现无论任何人在任何时候、在任何地方与另一个人进行任何类型的信息交换。那么问题来了，即人们对通信的要求越来越高，传输的信息量越来越大，要求通信系统能支持的用户量也越来越大，而各种资源有限，比如频谱资源有限，信道的信息传输能力有限，处理能力有限，等等。通信的发展就是不断地解决日益增长的高需求和有限资源的矛盾。

在上述问题的解决过程中有几个最基本的问题。

(1) 信息需要度量，如何度量？

(2) 信源有冗余，如何利用？

(3) 信道有干扰，如何度量信道传输信息的能力？即信道容量的问题：如何在信道有干扰的情况下检错和纠错？

这些问题正是信息论这门学科的主要研究内容。

1.2 信息论研究内容

信息论是通信的数学理论，主要研究在通信系统的设计中，如何通过编码实现信息传输的有效性和可靠性。通常情况下的信息论包括狭义信息论、一般信息论和广义信息论。

狭义信息论，也称经典信息论，主要研究信息的测度、信道容量及信源和信道编码理论等问题。又称香农基本理论。

一般信息论，也称工程信息论，主要研究信息传输和处理问题。除香农基本理论以外，还包括编码理论、噪声理论、信号滤波和预测、统计检测和估计、调制理论、信息处理理论以及保密理论等。

广义信息论不仅包括上述两方面内容，而且包括所有与信息有关的自然和社会领域，如模式识别、计算机翻译、心理学、遗传学、神经生理学、语言学、语义学，甚至包括社会学中有关信息的问题。

本书只讨论经典信息论，主要包括如下内容。

第二章是信息的度量；第三章是信源的数学描述，信源的数学建模问题；第四章是信源编码，介绍无失真信源编码理论和方法；第五章是信道的数学描述及其信道容量；第六章是信道编码理论和方法；第七章是信息论在密码学中的应用。

本书的主要知识结构图如图 1.1 所示。

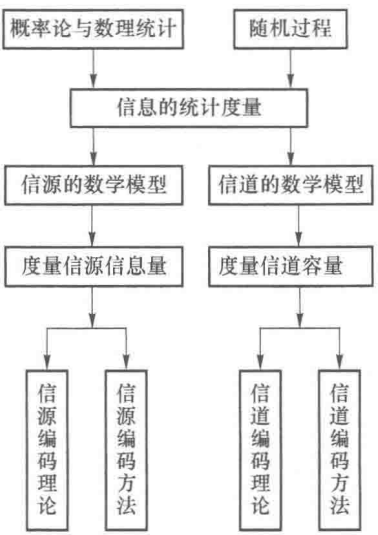

图 1.1 本书的主要知识结构图

1.3　信息论基本概念

1.3.1　信息的定义

信息是信息论中最基本、最重要的概念，是一个既复杂又抽象的概念。不同领域的学者从不同角度给出了信息的不同的定义，下面仅列出如下几种定义。

(1) 信息可以界定为由信息源(自然界、人类社会等)发出的被使用者接受和理解的各种信号。

(2) 作为一个社会概念，信息可以理解为人类共享的一切知识，或社会发展趋势以及从客观现象中提炼出来的各种消息之和。被认为等同于消息、情报、知识。

(3) 信息是对物质存在和运动形式的一般描述。物质、能量、信息是构成客观世界的三大要素。信息不是物质，但存在于任何事物中，有物质的地方就有信息，信息必须依赖一定的物质形式。一切事物，包括自然界和人类社会，都在发出信息。我们每个人每时每刻都在接收信息。在人类社会中，信息往往以文字、图像、图形、语言、声音等形式出现。

(4) 概率信息定义(Shannon 信息)：信息是事物运动状态或存在方式的不确定性的描述。事物是确定的，则无信息量。

【例 1.3.1】 袋中有 100 个球，若从中随机抽一个，抽到的是红球。针对以下三种情况讨论该消息含有的信息量的大小。

(1) 袋中全为红球。

获得消息：抽到的是红球。则此消息不含任何信息量。

(2) 袋中有 99 个红球、1 个白球。

获得消息：抽到的是红球。则此消息发生概率大，不确定小，信息量小。

(3) 袋中有 1 个红球、99 个白球。

获得消息：抽到的是红球。则此消息发生概率小，不确定大，信息量大。

1.3.2　信息的特殊性质

信息有如下性质：

1. 无形的

信息不同于物质和能量，它是看不见、摸不着的。信息不具有实体性。

2. 可共享的

(1) 信息的共享推动人类社会的发展。信息的交流，不会使交流者失去原有

的信息，而且还可以获得新的信息。信息的共享是无限的。信息的共享性对人类社会的发展起到了积极推动作用。信息扩散越快、越广，就会越加速人类社会的文明进程。

(2) 人类社会的竞争现象阻碍了信息性质的发挥，现实的人类社会在各方面都存在着激烈的竞争。例如军事中的电子综合战、商业活动中的市场竞争等，这些现象阻碍了信息性质的发挥。

(3) 需保密的要避免共享，涉及信息加密技术，为了限制信息的共享，可以加设密码等保护措施。

3. 无限的

信息作为事物运动状态和存在状态的一般描述，和事物及它们的运动一样是永恒的、无限的。信息的无限性还表现在时空上的可扩展性。例如，今天气象台报告的气象数据所包含的信息，明天就失去价值，明天又会产生新的信息。如果将所有这些信息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. 可度量的

信息量是能够度量其大小的，这是信息论的基础。

1.3.3 自然信息、社会信息和知识信息

通常情况下，信息包括自然信息、社会信息和知识信息。

(1) 自然信息：是指一切自然物发出的信息，它包括来自自然界的信息，如我们观察到的宇宙间星球的运动变化，地球上的各种自然现象，我们欣赏的风景等，都是自然信息。

(2) 社会信息：是指人类社会在生产和交往活动中所交流或交换的信息。即除人的生物信息和生理信息以外的，与人类的社会活动有关的一切信息。

(3) 知识信息：包括书本，前人的经验等。

1.3.4 信息、消息、信号的比较

信息、消息、信号是几个比较相近、容易混淆的概念，下面对这几个概念加以比较。

消息是信息的载体，它具有不同的形式，例如语言、文字、符号、数据、图片等。同一个消息可以含有不同的信息量，而同一信息可以由不同的消息来载荷，比如报纸、电视、网络等。

信号是消息的表现形式，消息是信号的具体内容。信号是表示消息的物理量。

一般把随时间而变化的电压或电流称为电信号。在实际应用中常常将各种物理量，如声波动、光强度、机械运动的位移或速度等，转变为电信号，以利于传输。

在通信系统中，传输的本质内容是信息（消息），系统中实际传输的是信号，信息包含在信号中，信号是载体，信号到了接收端，通过处理变成文字、语音或图像，人们再从中得到有用的信息。

1.4　通信系统模型

通信系统是通信中所需要的一切技术设备和传输媒质构成的总体。研究通信系统的目的是找到信息传输过程中的一般规律，以提高信息传输的可靠性、有效性、保密性和认证性，以达到信息传输系统最优化。

（1）可靠性高，就是信源发出的消息经过信道传输后，尽可能准确地、不失真地再现在接收端。

（2）有效性高，就是经济效果好，用尽可能短的时间和尽可能少的设备来传输一定数量的信息。提高可靠性和有效性常常会发生矛盾，需要统筹兼顾。

（3）所谓保密性，就是隐蔽和保护通信系统中传输的信息，使它只能被授权接收者获取，而不能被未授权者接收和理解。

（4）所谓认证性，就是指接收者能正确判断所接收的消息的正确性和完整性，而不是伪造的和被篡改的。

以上四点构成现代通信系统对信息传输的全面要求，靠编码来实现。编码可分为信源编码、信道编码、密码编码三类。

通信系统基本模型见图1.2。

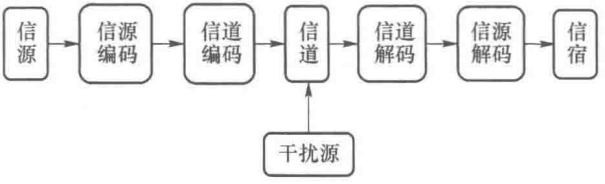

图1.2　通信系统基本模型

上述基本模型中包含以下几个要素。

1. 信源

即产生消息的源，可能连续（如语音），也可能离散（如文字，图像等）。其核心问题是它包含的信息到底有多少，如何将信息定量表示出来。

信息是抽象的，信源则是具体的。例如，人们交谈，人的发声系统就是语声信源；人们看书、读报，书和报纸本身就是文字信源；常见的信源还有图像信源、数字

信源等等。信号的产生物被称为信源，相对应的概念应该是信号的接收物被称为信宿。

2. 编码

按规则将信息的意义用符码编排起来的过程就是编码过程，这种编码通常被认为是编码的第一部分。编码的第二部分则是针对传播的信道，把编制好的符码又变换成适于信道中传输的信号序列，以便于在信道中传递，如声音信号、电信号、光信号等等。如信息源产生的原始讯息是一篇文章，用电报传递的时候，就要经过编码，转换成电报密码的信号，然后才能经过信道传播。

3. 编码器

将信号(如比特流)或数据进行编制、转换为可用以通信、传输和存储的信号形式的设备。

(1) 信源编码器：将信源发出的消息变换成二进制或者多进制数字序列，同时要求每条消息的数字序列的平均长度越小越好。即除去原来消息的冗余度，减少信道的占用时间，提高传输的有效性。

(2) 信道编码器：信道编码器是针对信道对传输信号的损伤而设置的一个功能部件。

信道是通信系统中传输信号的重要组成部件。信道的损伤会造成数字通信系统接收机输出误码率增加，使信息的恢复受到恶化。信道编码器的作用就是通过对信息序列进行编码的方式来提高接收机识别差错的能力，从而降低误码率以改善恢复信息的质量。

信道编码器包括纠错编码器和调制器。纠错编码器在信源码中加入纠错码，使消息在传递过程中具有一定的抗干扰能力，提高传输的可靠性。调制器是把基带信号调制到高频信号，以利于信号传输。

4. 信道

即信息传递的通道，是将信号进行传输、存储和处理的媒介。信道的关键问题是它的容量大小，要求以最大的速率传送最大的信息量。信道有光纤、同轴电缆、双绞线、微波和红外等。

5. 干扰源

在无线电通信系统中，干扰源是产生干扰的发射、辐射或感应。也就是产生妨碍无线电接收信号的那些杂乱的电波。干扰源可分为加性干扰和乘性干扰。

(1) 加性干扰：由外界引入的随机干扰，如设备内部噪声，它们与信道的输入信号统计无关。信道的输出是输入信号与干扰的和。

(2) 乘性干扰：信号在传播过程中由于物理条件的变化引起信号参量的随机变化而构成的干扰。此时信道的输出信号是输入信号与某些随机参量相乘的结果。

6. 译码

即编码的逆变换，是最大限度提取信号并加以恢复。译码器是进行译码的设备，可分为信源译码器和信道译码器。

7. 信宿

信宿是相对于信源而言的。信宿是信息动态运行一个周期的最终环节。其功能是接收情报信息，并选择对自身有用的信息加以利用，直接或间接地为某一目的服务。信宿可以把信息资源转化为人类的巨大物质财富，在信息的再生产过程中，还可以起到巨大的反馈作用。

1.5　信息论创立的意义

在20世纪中叶，人类终于对三个非常重要的概念——质量、能量、信息量——都有了定量的计量办法。为阐明质量概念做出伟大贡献的是发现物体力学定律的牛顿；为阐明能量概念做出伟大贡献的是热力学第一定律的发现者们：迈耳、焦耳、赫尔姆霍兹、开尔文。而为阐明信息概念做出伟大贡献的就是香农。

20世纪中期，随着原子弹的出现，物理学成为最荣耀的科学学科。在随后的50年里，晶体管、人造卫星、集成电路、电脑的飞跃发展无不与物理学知识的应用有关。但是我们也惊奇地发现这些新技术都是为提高信息的处理能力服务。光荣的物理学家们忙了半个世纪，终于发现自己仅是给信息科学当仆人。信息量能进入物理学吗？但“信息不是物质”！在物理学的版图中人们不知道把信息论放到哪里合适。人类知识体现的这种新的混乱局面需要我们不断地澄清。

信息论是信息科学领域第一次脱离物理层面，仅仅只在数学层面验证了数据传输的极限和压缩率的极限。后续很多分支领域的研究都在这个基础展开，比如无线网络信道分析、图像编码压缩、多媒体数据流传输等等，它们的研究成果还将服务于更多的学科领域。

华为创始人任正非在2016年全国科技创新大会上表示，“华为现在的水平尚停留在工程数学、物理算法等工程科学的创新层面，尚未真正进入基础理论研究。随着逐步逼近香农定理、摩尔定律的极限，而对大流量、低时延的理论还未创造出来，华为已感到前途茫茫，找不到方向。”这句话正说明信息论广泛服务于通信领域，同时，通信领域的实际问题会对信息理论提出更多的要求和问题，这必将进一步促进信息理论自身的发展。

第2章　信息的统计度量

从心理学认知角度来讲，信息是有大小之分的。一个很少发生的事情突然发生时，给人的冲击是很大的。反之，经常发生的事情给人的印象则是很淡的。信息量有大小之分，这是信息理论与编码的基础。

信源通常有如下几种形式：文本、静止图像、运动图像、声音。不同形式的信源含有的信息量是不同的。比如，相对于文本来讲，图像含有的信息量更大，对于一篇新闻报道，有时一幅图片所能传达的信息胜过千言万语。而相同形式的信源含有的信息量也可能不同，比如，同尺寸的图片，信息量也是大有不同的。与静止图像相比，运动图像的信息量更大，因为运动的图像是由多幅静止图像构成的。

生活中我们有这样的经验，对于爆炸新闻，人们总是津津乐道，而在爆炸新闻背后总能挖掘出更多的新闻。爆炸新闻属于小概率事件，小概率事件的发生往往是大量原因的积累才导致的。那些大量的原因以及更多的新闻都是小概率事件本身所含有的信息量。比如地震是小概率事件，对我们来讲是突发事件，只是我们看不到地壳内部的运动而已，而实际上地震的发生一定是地壳长期运动的结果。总之，我们可以直观地感受到，小概率事件含有的信息量更大。

我们可以直观地感受到不同的事件、不同的信源含有的信息量不同，需要用数学方法对信息量进行量化。1948年，香农(Claude Elwood Shannon，1916.4.30—2001.2.26)以概率统计为基础给出了信息量大小的量化体系。

2.1　事件的自信息量

设 x_i 是一个事件，x_i 含有的信息量记为 $I(x_i)$，首先任何事件所含有的信息量是非负的。记 $P(x_i)$ 为事件 x_i 发生的概率，若 $P(x_i)$ 较大，则事件 x_i 的不确定性小，含有的信息量也小；若 $P(x_i)$ 较小，则事件 x_i 的不确定性大，含有的信息量也大。已经发生的事件含有的信息量应该为0；另外，若两个事件 x_i 和 y_j 相互独立，则它们的积事件所包含的信息量应该等于各自所含有的信息量之和。即事件的自信息量应该满足如下5个条件：