信息论基础

XINXILUN JICHU

杨孝先　杨　坚　编著

中国科学技术大学出版社

信息论基础

杨孝先　杨　坚　编著

中国科学技术大学出版社

前　言

“信息论基础”是中国科学技术大学数学系、少年班与零零班应用数学专业开设的一门选修课. 参阅 T. Cover 等著的经典文献[1]，总结编者的教学成果，我们编写了《信息论基础》讲义. 该讲义自 2002 年的 1999 级开始已经有 9 届学生使用，他们提供了很多宝贵的建议，并评论称此讲义对定理给予的证明，包括种种算法，有很多技巧，其中一些是非常巧妙而有趣的，因此很享受这个学习的过程. 本次成书即以此讲义为蓝本编写而成，并进行了较大幅度的修订和充实.

本书除了介绍 Shannon 信息论的基本概念、基本分析方法与主要结论之外，还注意到了与后续的另一门选修课——“数学控制论”的联系.

本书配备了较多的例题与习题. 书末还附有习题的提示与答案，供读者参考.

本书以少而精、把主要问题讲深讲透而不是面面俱到为原则，行文力求通俗易懂，便于读者自学.

编著者

2010 年 8 月于合肥

目　录

第1章 概 论

从DVD到个人电脑，从卫星通信到文件传输网络再到条形码，在当今人们的生活中，信息几乎在每个领域都扮演着重要的角色，几乎所有需要迅速准确地传输数据的领域，都以被称为“伟大思想”的香农(C. E. Shannon)信息论作理论基础.它是一个业已成熟的科学体系，是应用概率论与数理统计方法来研究信息传输、交换、存储与处理的一门学科，也是起源于通信实践而发展起来的一门新兴应用科学.本书着重介绍Shannon信息论的基本概念、基本分析方法、主要结论以及编译码理论等.这种信息理论是基于概率统计意义上的信息理论，它对信息技术的发展产生了持久而又深刻的影响.

1.1 信息论的基本内容——信息理论所关注的信息特征

信息已成为现代社会一种重要的资源，信息的要领已被广泛地采用，信息科学技术在近几十年得到迅速发展.信息、材料、能源是现代科学技术的三大支柱.信息、物质、能量是构成一切系统的三大要素.这些都说明了人类对信息重要性的认识.信息科学是研究信息的产生、获取、度量、变换、检测、传输、识别、处理、安全等及其应用的一门学科.

信息的产生与获取是相当广泛的.其一是包括来自物理、化学、宇宙空间、地学、生物学等的自然信息，获取的主要工具是传感器与传感设备.有物理型的，例如热、光、磁、电、声、力等；化学型的，例如气体、化合等；生物型的，如神经、感觉、视觉、听觉、触觉等.其二是来自社会的信息，包含政治、军事、经济、管理、金融、商情、文化及各种情报等.获取途径主要是靠社会调查，并利用统计方法进行整理.其三

是知识信息，包括古今中外记录下来的种种知识，以及专家的经验等．获取途径是靠学习与交流或上网．总之，人们都知道：电话、网络、电视、雷达、信函、书刊、报纸等传输的声音、图像、磁信号、文字、数字、符号等能引起人们产生兴趣或使得人们获得认知的东西就是信息．

但是，获取的信息究竟是否可靠呢？以网络为例，网络已让知识免费流通，信息基本上可自由传播，意见免除审查，并逐渐成为人类的集体记录库，正如以往人类的历史记载一样，其中的信息难免充满了人为的捉弄与虚伪的记载，影响到知识的纯度．网络上充满了种种来历不明、未能查实的信息．特别是当今处于一个“暴露狂”时代，“暴露自己，也暴露别人”是暴露狂者的时代精神．“我秀故我在”仿佛已取代了“我思故我在”，成为人类存在的本质．主动暴露的信息就如同隐藏的信息那样，都需要获取者认真地积极追究与冷静思考，千万不要盲目信任．当然，获取信息从来不是知识的终极目标，获取信息的目的只是为了形成自己的价值判断，从而做出有益的决定．人的生存离不开信息，人的五官在不停地接收信息，人的神经系统在不断地传递信息，人的大脑在不停地处理、存储与利用信息．

信息具有抽象性，即指它是构成一切系统的三大要素之一．物质是具体的、最基本的，能量可看作是物质的运能形式，而信息既不是物质又不是能量，但又离不开物质与能量，它是人类认识世界与改造世界的新层次．信息可由一个人掌握，也可以大家共享．信息是有寿命的，且是普遍存在的．故信息的本质与它的科学定义是当前科学界甚至哲学界都热衷研究的课题．

信息具有不确定性，恰恰是这种不确定性引起了人们对信息的注意．信息的核心问题是其度量问题，要对它给出一个统一的度量是困难的．至今最为成功的，也是最为普及的信息度量，是由信息论的创始人 Shannon 在他的著名论文“通信的数学理论”中给出的，是建立在概率统计模型上的信息度量．他把信息定义为“用来消除不确定性的东西”．例如人类的通信过程就是一种消除不确定，即判断对与错的过程，随着不确定性的消除，就获得了信息．原来的不确定性消除得越多，获得的信息就越多．若原来的不确定性全部消除，就获得了全部信息；若只消除了部分不确定性，就只能获得部分信息；若原来的不确定性没有消除，就不能获得任何信息．故 Shannon 又把信息定义为：信息是事物运动状态或存在方式的不确定性的描述．

在通信中，信息表达有三个层次：信号、消息与哲学．其中信号最具体，它是一个物理量，可测量、可显示、可描述，同时它又是载荷信息的实体，称为信息的物理表达层．消息又称符号，如语言、文字、声音、图像、图片等，信息就蕴含在消息之中．同一个消息如一封家书，用文字形式写在信纸上，对收信人而言获得的信息可抵万

金，而对别人来说也许只是废纸一张. 对这些载有信息的符号，为了从数学上进一步描述或表达不同形式的信息，将消息分为离散(数字)消息与连续消息两大类型，以便用随机变量、随机序列与随机过程来进行分析，称为信息的数学表达层. 它也是信息论中主要的描述形式. 三个层次中最抽象的是哲学表达层的信息. 信息是具体的信号与消息的内涵，是信号与消息载荷的内容，也是消息描述的对象，而信号是信息在物理表达层上的外延. 同样一个信息可用不同形式的物理量载荷，也可使用不同的数学描述方式(数字或模拟). 同样，同一个类型的信号或消息也可以代表不同内容的信息. 这里介绍的信息论常称为概率信息论或狭义信息论，是指数学上的信息论与通信中的信息论. 它是由 C. E. Shannon 提出来的，是关于信息处理与可靠通信中的数据压缩、广播、电视、卫星通信、信息与计算机科学各专业、计算机存储及因特网通信理论的数学基础. 因此，通信系统的基本模型也即信息论的基本模型如图 1.1 所示.

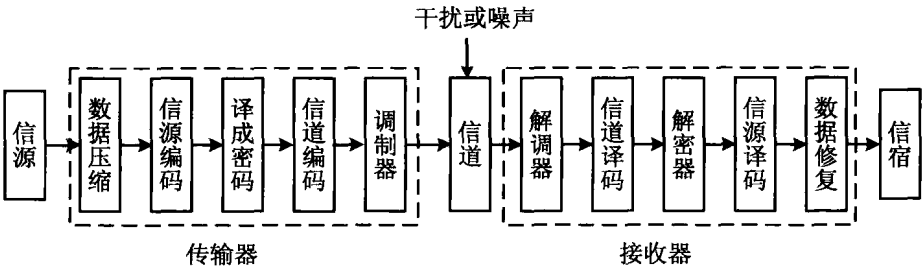

图 1.1　通信系统基本模型

信源是通信的起点，是产生消息或消息序列的源泉. 消息一般是某种序列或时间参数的函数，消息的取值服从一定的统计规律，故信源的数学模型可以是离散的随机序列或连续的随机过程. 编码是用符号来表征消息，即进行信源编码. 此外，还要将符号转换成信道所要求的信号，即进行信道编码，而译码则是编码的反变换. 信道就是传输信号的媒介或通道，即将信源发出的消息经过压缩再编成信号后，需要通过某种渠道传送给接收者，这种渠道就是信道，如电缆、光缆、电离层、人造卫星等. 在信息论的模型中，把系统中各个部分的干扰与噪声都归入到信道中. 在信道的输入、输出模型中，按干扰与噪声的统计特性，用输入、输出的条件转换概率或称前向转移概率来描述信道的特征. 信道编码是把信源编码输出的数字序列变换成适合于信道传输的信道输入符号组成的序列，主要作用是对其输出序列提供保护，以抵抗信道干扰与噪声. 信宿就是消息的接收者. 总之，编码是把消息变换成便于传输的信号的方法、措施，而信道则是传递、存储信号的具体物理设施. 从信源得

到的消息，经过编码进入信道.信道是通信系统的关键部分，它将信源的输出输入通信系统，然后再将其输出并经过加工复原成消息后送给用户.前者称为编码或调制，后者称为译码或解调.C. E. Shannon 正是在研究通信系统的基础上，于 1948 年创立了信息论这一伟大思想.

因计算机与信息的联系非常密切，故常把与计算机有关的知识、技术与学科都加上“信息”二字.如把以计算机科学为学习与研究主要对象的学院叫做信息学院，而计算机科学又叫信息科学.信息技术、信息科学、信息时代中的“信息”，以及人们从手机、电视、网络、信函、书刊等中获取的声音、图像、文字等能被感知的所有形态，与 Shannon 信息论中所关注的信息是有区别的.Shannon 信息论中的信息是有确定的含义的，是可度量的，是真正意义上的信息.它考虑信源发出某种字母集合中的字母组成的序列发生的概率现象，仅着眼于该序列出现的概率，信息量是由选择信息的概率所决定的.前面指的信息大多都兼有计算机与真正意义上的信息的双重意义.因此，信息时代是指人类已处于一个需要大量使用计算机，并处理各种信息的时代.若仅把信息时代理解为研究 Shannon 信息论的时代是不对的，当然认为就是研究计算机的时代也不完全正确.Shannon 信息论有非常明确、具体的研究内容与研究工具，不像信息技术与信息科学那样还存在许许多多人们还不清楚的内容.

1.2 信息论简史

目前都公认信息论的建立，开始于 C. E. Shannon 研究通信系统时所发表的著名论文.然而，从历史上来看，早在 1925 年，信息作为科学名词就出现在数学之中了.数学家 R. A. Fisher 从古典统计理论的角度定义了信息量，通常被称为 Fisher 信息量，至今仍在估计问题中有着重要的价值.1948 年 Shannon 的论文仅讨论了无记忆信源与无记忆信道.此后，数学家们纷纷把 Shannon 的基本概念与编码定理推广到更一般的信源、更一般的编码结构与性能度量，并给出了严格的证明.例如 A. N. Kolmogorov 于 1956 年提出了信息量的一般定义，其后又指出熵相等是动力系统同构的必要条件，并开辟了遍历理论的一个新方向.在此基础上，G. J. Chaitin 于 1987 年建立了算法信息论.另外，数学家 S. K. Kullback 在 1959 年给出了鉴别信息的概念及其与 Fisher 信息量、Shannon 熵的关系.由于 Shannon

熵在连续随机变换时失去了意义，故这时鉴别信息特别重要.上述这些工作不仅对数学本身，而且对信息技术也产生了重大的影响.例如，动力系统同构问题的研究，使人们对信源编码问题有了更深刻的认识，并得到了一些新的结果与编码方法，在生物信息论中已取得较大的成果；而鉴别信息的概念为估计问题、识别问题提供了理想的数学工具，并在信号处理学中获得了重要的应用.数学家们还提出了诸如 ε -熵、α -阶熵、β -次熵及 γ -型熵等概念.这些熵在模糊识别及模糊信息论中都有一定的应用，但其重要性远不及 Shannon 熵.

1.3 关于控制论、信息论与系统论

20 世纪 40 年代初，数学家 N. Wiener 在研究计算机时发现机器的控制系统与人脑的功能有相似之处，即动物与机器中的控制与通信过程存在着许多共性，尤其是在信息的传输、变换处理过程中有许多共同规律.虽然在物质构造与能量转换方面，动物与机器有显著不同，但在信息传输、变换与处理方面有着惊人的一致之处.因此感到有必要、也有可能建立一门新的综合性的边缘学科，研究动物与机器、社会经济等的控制过程的共同规律与方法. 1948 年，N. Wiener 的名著《控制论——关于动物与机器中控制与通信的科学》一书面世了，为控制论的产生奠定了基础.N. Wiener 的研究有如下的特点：其一，摆脱了 Newton、Laplace 的机械决定论，建立在统计理论的基础上；其二，抛开对象的物质与能量的形态观点，着重于以信息观点来讨论系统的功能；其三，抛开一时一地固定的观点，而着重于所有可能的行动方式与状态，重视变化的趋势；其四，把系统观点、信息观点与反馈观点结合起来，从而形成一门新的科学技术.从以上特点可以看出，控制论与信息论的关系非常密切，控制论是一门典型的横向学科，着重研究控制过程的数学关系，而不涉及控制过程的内在的物理、化学、生物或其他方向的现象.它是自动控制、通信工程、计算技术、神经生理学、神经病理学、数学等有关学科相互结合而产生的，它突破了工程技术与生物科学之间的传统界限，跨越了两大领域之间的“鸿沟”.

系统论是生物学家 L. V. Bertalanffy 首先提出的.他发现一切生物体都是在一定的时间与空间中，呈现出复杂的有层次的结构，是由各要素组成的有机整体，整体的功能大于组成它的各部分的功能的总和.例如人的身体作为一个整体，由脑、心、肺、肝、肾等部分组成，而各个部分功能的总和，明显小于整体的功能，又如

社会、一个企业、一个工厂也是如此.另外,系统工程与系统论虽然有联系,但并不完全一样.系统工程是用系统论的观点来解决系统的分析和设计问题.例如,1911年工程师 Tagler 研究了合理安排工序问题,通过分析工人的动作,提高工作效率,探索管理规律,并提出了被称为 Tagler 系统的企业管理方法与体制;又如第二次世界大战期间,运用运筹学制定出护航编队与作战计划等.1948 年,由 Bowles 领导的科学家小组提出了许多用于分析大系统的数学方法,就形成了系统工程的数学方法.20 世纪 70 年代以后,系统工程得到很大的发展.美国以运筹学为基础,发展了系统工程;日本以质量管理为出发点发展了系统工程;俄罗斯则是在控制论的观点上建立起了系统工程.我国在钱学森先生的推动下,对系统工程给予了很大重视,系统工程得到较快发展.

人们都重视信息论、控制论与系统论的综合应用,它们都于 1948 年左右形成,其基本思想、基本方法有许多类似之处.其一,都是综合的整体观点,即从个别分析到综合研究;其二,都是从机械的静止的观点到动态的观点与方法;其三,都是从物质、能量的交换发展到物质、能量与信息相互的交换;其四,信息反映了系统的重要性、反映了系统的组织化与复杂化,系统越复杂,信息就越重要.人脑是由 $10^{10}\sim10^{11}$ 个神经细胞组成的,而消耗的功率不到 1 瓦特.总之,信息与控制是不可分割的,信息论是控制论的基础.

1.4 信息论的应用

对科学家而言,信息论是用来理解和发现自然规律的一门基础学科,而对工程技术人员来说,信息论被看作是各种应用的理解依据,研究信息论可深入到信息传输系统的设计中去,对信息和传输给出更清晰的概念,对技术极限与应用有着更深刻的理解,并为设计出有效的通信系统提供理论依据.信息论还帮助人类发展传送数据包或码流的新技术与方法,且可预知尚有多大的改进余地.

数字多用户通信系统是极为复杂的通信系统,目前设计者们还只能按经验来设计,信息论从理论上可提供一些新理论与新方法,帮助设计者解决困难的问题.

大家知道,在许多技术领域中,都是在有观测的数据之后,要求从中提取出有用的信息.这一提取信息的过程实质上就是信息论的应用过程.如在进行地球物理勘探中,通过人工地震得到大量的测量数据,或利用人造卫星发回的种种图片等,

从中提取关于地层构造、地质性质、地下矿藏的信息；又如通过市场调查得到大量的商品供求数据，从中提取有用的市场信息，以供制造厂家参考.在自然科学、工程技术与社会科学中还会遇到许多这样的例子，都要用到信息论的基本概念与方法.但是在应用中需要注意的是，违背基本概念与方法的套用，往往会出现很大的错误.信息的基本概念在于它的不确定性，任何已确定的事物都不会再含有信息.

信息论中的编码理论也是应用广泛的一个重要领域.编码的目的是为了优化通信系统中控制系统的有效性、可靠性、安全性与经济性，故常常要使用编码技术.例如电报中常用的莫尔斯码就是按照信息论的基本编码原则设计出来的.又如从超市、百货商店或药店买来的一些商品上面，有一张由粗细条纹组成的标签，从这张标签上可得知该商品的生产厂家、生产日期、价格等信息，这些标签是利用条形码设计出来的，而条形码可看作是一种编码.如今每出版一本图书都会给定一个国际标准书号(ISBN)，大大地方便了图书的销售、编目与收藏工作，国际标准书号也是一种编码.实际上，人们在日常生活与生产实践中，正在越来越广泛地使用编码技术.密码就是以提高通信的安全为目的的编码，一般通过加密与解密来实现.从信息论的观点来看，加密是“增熵”的过程，解密则是“减熵”的过程.

20世纪70年代以来，由于电子计算机的广泛应用，以及通信系统能力的大大提高，极大地推动了信息论的不断发展，其应用领域日益扩大，信息的概念与方法已渗透到了各个科学技术领域.现在的科学与技术中没有用到信息论的地方已经不多了，可以说信息科学技术早已突破了Shannon信息论，即狭义信息论的应用范围，而扩展到科学研究、工程技术、物质生产与社会生活的各个方面.这就是一般信息论与广义信息论.一般信息论研究信息在传输过程中受到干扰时，在接收端如何把信息从干扰中提取出来，这就使我们建立了最优滤波理论(Wiener与Kalman滤波器)等.广义信息论是一门综合性的新兴学科，至今无严格的定义.总之，凡是能用通信系统模型描述的过程或系统，如神经传导系统、市场销售系统等，都能用信息基本理论来研究.除包括一般信息论之外，广义信息论还包括医学、生物学、心理学、遗传学、神经生理学、语言学、量子力学及社会学与经济管理中的有关信息的问题.另外，所有研究信息识别、控制、提取、变换、传输、处理、存储、显示、价值、安全、作用及信息量大小的一般规律和实现的工科学科都属于广义信息论的范围.从而可以看出，人们研究信息论是为了高效、可靠、安全及随心所欲地变换和利用各种各样的信息.

1.5 有关的常用不等式

基本不等式 对任意的实数 $x>0$,则有 $1-\frac{1}{x}\leqslant \ln x\leqslant x-1$,而等号成立的充要条件是 $x=1$.

证 当 $0<t\leqslant 1$ 时,有 $\frac{1}{t}\geqslant 1$,故得

$$1-x=\int_x^1 \mathrm{d}t\leqslant \int_x^1 \frac{1}{t}\mathrm{d}t=-\ln x.$$

于是,有

$$\ln x\leqslant x-1,\quad 0<x\leqslant 1.$$

其次,当 $t\geqslant 1$ 时,有 $\frac{1}{t}\leqslant 1$,故得到

$$\int_1^x \frac{1}{t}\mathrm{d}t\leqslant \int_1^x \mathrm{d}t=x-1,$$

或有

$$\ln x\leqslant x-1,\quad \text{对 } x\geqslant 1.$$

当且仅当 $t=1$ 时,才有 $\frac{1}{t}=1$. 因此,当 $x\neq 1$ 时,有 $\ln x<x-1$. 当 $x=1$ 时,则不等式左右两端都为零,故 $\ln x\leqslant x-1$,对任意的 $x>0$ 成立.

最后,若在 $\ln x\leqslant x-1$ 中,令新的 x 是旧的 x 的倒数,即令 $x=\frac{1}{x}$,就有

$$-\ln x\leqslant \frac{1}{x}-1\quad 或\quad 1-\frac{1}{x}\leqslant \ln x,\quad \forall x>0$$

成立.

从而得到 $1-\frac{1}{x}\leqslant \ln x\leqslant x-1$,对任意的 $x>0$ 成立,且等号成立的充要条件是 $x=1$.

注意:若规定 $x=0=0^+$,且 $\ln 0=-\infty$,上述基本不等式 $\forall x\geqslant 0$ 成立,今后使用它时,总设 $x\geqslant 0$,不再一一赘述.

对数和不等式 对任意的 $a_i \geqslant 0$ 与 $b_i \geqslant 0 (i = 1,2,\cdots,M)$，则有 $\sum_{i=1}^{M} a_i \cdot \ln \frac{\sum_{i=1}^{M} a_i}{\sum_{i=1}^{M} b_i} \leqslant \sum_{i=1}^{M} a_i \ln \frac{a_i}{b_i}$，且等号成立的充要条件是：$\frac{a_i}{b_i} = \frac{\sum_{i=1}^{M} a_i}{\sum_{i=1}^{M} b_i}$ = 常数 $(i = 1,2,\cdots,M)$，并规定：$0 \ln \frac{0}{\beta} = 0 (\beta \geqslant 0)$，$\gamma \ln \frac{\gamma}{0} = +\infty (\gamma > 0)$，$0 \ln \frac{0}{0} = 0$.

证 按规定，可令 $\alpha = \frac{\sum_{i=1}^{M} a_i}{\sum_{i=1}^{M} b_i} \geqslant 0$，并将 $x = \frac{\alpha b_i}{a_i} \geqslant 0$ 代入基本不等式，得到

$$\ln \frac{\alpha b_i}{a_i} \leqslant \frac{\alpha b_i}{a_i} - 1,$$

将其两边乘以 a_i，并关于 i 求和，有

$$\sum_{i=1}^{M} a_i \ln \frac{\alpha b_i}{a_i} \leqslant \alpha \sum_{i=1}^{M} b_i - \sum_{i=1}^{M} a_i = 0.$$

故有

$$\sum_{i=1}^{M} a_i \ln \frac{b_i}{a_i} + \sum_{i=1}^{M} a_i \ln \alpha \leqslant 0,$$

或有

$$\sum_{i=1}^{M} a_i \ln \frac{\sum_{i=1}^{M} a_i}{\sum_{i=1}^{M} b_i} \leqslant \sum_{i=1}^{M} a_i \ln \frac{a_i}{b_i},$$

且等号成立的充要条件是 $x = 1$. 即

$$\frac{a_i}{b_i} = \frac{\sum_{i=1}^{M} a_i}{\sum_{i=1}^{M} b_i} = \text{常数} \quad (i = 1,2,\cdots,M).$$

马尔可夫不等式 对任意的 $\alpha > 0$，设 Z 是任意的具有均值为 μ 的非负随机变量，则有

$$P\{Z \geqslant \alpha\mu\} \leqslant \frac{1}{\alpha}.$$

证 记 $s=\{z\in Z\mid z\geqslant\alpha\mu\}$，故有

$$\mu=\sum_{z\in Z}zP_Z(z)=\sum_{z\in s}zP_Z(z)+\sum_{z\in\bar{s}}zP_Z(z)\geqslant\alpha\mu\sum_{z\in s}P_Z(z)=\alpha\mu P_Z(s).$$

于是得

$$P\{Z\geqslant\mu\alpha\}=P_Z(s)\leqslant\frac{1}{\alpha}.$$

切比雪夫不等式 对任意的 $\alpha>0$，设 Z 是具有均值为 μ 与方差为 σ^2 的任意随机变量，则有

$$P\{|Z-\mu|\leqslant\alpha\sigma\}>1-\frac{1}{\alpha^2}.$$

证 若用 $(Z-\mu)^2$ 与 α^2 分别代替马尔可夫不等式中的 Z 与 α，而 $E[(Z-\mu)^2]=\sigma^2$，就得到

$$P\{(Z-\mu)^2\geqslant\alpha^2\sigma^2\}=P\{|Z-\mu|\geqslant\alpha\sigma\}\leqslant\frac{1}{\alpha^2},$$

或有

$$P\{|Z-\mu|<\alpha\sigma\}>1-\frac{1}{\alpha^2}.$$

习 题 1

1.1 说明写信符合基本通信模型.

1.2 信息论关注的信息特征是什么？它与读者本身所具有的对“信息”这一术语的概念有何差别？

1.3 设 $f(x)(x\in K)$ 是可微的下凸($\cup$)函数，$E[Z]<+\infty$，试证：$f[E(Z)]\leqslant E[f(Z)]$(称为 Jensen 不等式).

1.4 应用习题 1.3 的结论证明对数和不等式.这里假设习题 1.3 中的 Jensen 不等式对离散型的随机变量也成立.事实上，读者可用数学归纳法给予证明.