

高等学校通用教材

信息理论基础

(修订版)

周荫清 主编

北京航空航天大学出版社

内 容 简 介

本书重点介绍经典信息论的基本理论,力图将信息论和它的工程应用联系起来。全书共分 10 章。内容包括信息的统计度量;离散信源和连续信源;信道与信道容量;信源与信宿之间的平均失真度以及信息率失真函数;信源编码与信道编码;信息论在信号处理中的应用。

本书深入浅出,概念清晰,系统性强;可作为理工科高等院校电子类相关专业的教材,亦可供从事通信、雷达、导航、生物工程、系统工程、管理工程等有关专业的科技人员参考。

图书在版编目(CIP)数据

信息理论基础/周荫清主编. —2 版(修订版)

北京:北京航空航天大学出版社,2002.2

ISBN 7-81077-120-5

I. 信… II. 周… III. 信息论著 IV. G201

中国版本图书馆 CIP 数据核字(2001)第 074856 号

信息理论基础(修订版)

周荫清 主编

责任编辑 曾昭奇

责任校对 陈 坤

*

北京航空航天大学出版社出版发行

北京市海淀区学院路 37 号(邮编 100083) 发行部电话:82317024 传真:82328026

<http://www.buaapress.com.cn>

E-mail: pressell@publica.bj.cninfo.net

北京朝阳科普印刷厂印装 各地书店经销

*

开本:787×1 092 1/16 印张:16 字数:410 千字

2002 年 2 月第 2 版 2002 年 2 月第 1 次印刷 印数:5 000 册

ISBN 7-81077-120-5/TN·001 定价:24.00 元

前 言

信息论是通信的数学理论,是应用近代数理统计方法研究信息的量度、编码和通信的科学。

信息论是 20 世纪中叶从通信中发展起来的理论,是数学中的概率论与通信技术相结合的边缘学科。自香农在 1948 年发表奠定信息论基础的“通信的数学理论”一文以来,信息论学科有了很大的发展并已延伸到许多领域中。人们已经认识到,在现代科学技术高度发展的过程中,学习和掌握信息论日益成为人们的一种需要。

从信息论涉及的内容和研究对象而言,人们常分为狭义信息论和广义信息论两种。狭义信息论是在香农信息论基础上发展起来的,又称为经典信息论。它仍然是当今研究信息理论的基石,也是本书重点讲述的内容。广义信息论是在更为一般的基础上建立的,是以广义信息作为主要研究对象,亦称为信息科学。这一学科虽尚未十分成熟,但前景异常广阔。

近十多年来,作者为北京航空航天大学电子工程系本科高年级学生和研究生开设了《信息理论基础》课程;同时编写了相应的教材《信息论基础》,于 1993 年北京航空航天大学出版社出版。本教材是在该教材基础上修改而成,是开设《信息理论基础》课程以来教学经验的总结。作为教材,有意注重基本理论、基本概念和基本方法的阐述,以及能力的培养。同时在数学工具的运用上力求准确、简明、适中,尽量使读者应用较浅的数学工具能够准确、系统地认识和掌握信息理论的基本概念和分析方法。

本书可作为电子类有关专业高年级大学生和研究生教材。全书共分 10 章。首先介绍信息的统计度量、离散信源、离散信道和信道容量;然后介绍无失真信源编码、有噪信道编码,以及限失真信源编码。这些内容是香农信息论的核心部分。在上述基础上引出并论述连续信源和波形信道的信息测度以及有关理论。最后介绍纠错码和信息论方法在信号处理中的应用。

本书力图在内容编排上由浅入深、深入浅出,以最易接受的方式介绍信息理论的基本内容及其应用。为了提高分析问题和解决问题的能力,各章后面配有一些难易程度不等的习题,可根据实际需要选用。

本书第 1 至 4 章由周荫清编写,第 5、6 章由刘玉战编写,第 7、10 章由李景文编写,第 8、9 章由李春升编写,最后由周荫清统编全书。

参加本书修订工作的有博士生燕英、徐华平、段世忠、孙娜。他们为本书出版做了大量工作,付出了许多辛勤劳动。

在编写和修改过程中曾得到中国科技大学研究生院傅祖芸教授的热情帮助,并提出了许多宝贵意见,编者在此表示致谢。

限于水平,书中难免存在一些缺点和错误,殷切希望广大读者批评指正。

周荫清

2001年10月1日于北京

目 录

第1章 绪 论

1.1 信 息	(1)
1.2 通信系统模型	(3)
1.3 信息论的形成和发展	(4)

第2章 信息的统计度量

2.1 自信息量和条件自信息量	(7)
2.1.1 自信息量	(7)
2.1.2 条件自信息量	(8)
2.2 互信息量和条件互信息量	(9)
2.2.1 互信息量	(9)
2.2.2 互信息量的性质	(10)
2.2.3 条件互信息量	(11)
2.3 离散集的平均自信息量	(12)
2.3.1 平均自信息量(熵)	(12)
2.3.2 熵函数的数学特性	(13)
2.3.3 条件熵	(21)
2.3.4 联合熵	(22)
2.3.5 各种熵的性质	(22)
2.3.6 加权熵	(27)
2.4 离散集的平均互信息量	(28)
2.4.1 平均条件互信息量	(28)
2.4.2 平均互信息量	(29)
2.4.3 平均互信息量的性质	(30)
2.5 连续随机变量的互信息和相对熵	(32)
2.5.1 连续随机变量的互信息	(32)
2.5.2 连续随机变量的熵	(34)
习 题	(36)

第3章 离散信源

3.1 信源的数学模型及其分类	(39)
3.1.1 信源的数学模型	(39)

3.1.2	信源的分类	(40)
3.2	离散无记忆信源	(42)
3.3	离散无记忆信源的扩展信源	(43)
3.3.1	最简单的离散信源	(44)
3.3.2	N 次扩展信源	(44)
3.3.3	N 次扩展信源的熵	(45)
3.4	离散平稳信源	(47)
3.4.1	平稳信源	(47)
3.4.2	平稳信源的熵	(48)
3.4.3	极限熵	(49)
3.5	马尔可夫信源	(51)
3.5.1	有限状态马尔可夫链	(51)
3.5.2	马尔可夫信源	(55)
3.6	信源的相关性和剩余度	(58)
	习 题	(59)

第 4 章 离散信道及其容量

4.1	信道的数学模型及其分类	(65)
4.2	离散无记忆信道	(67)
4.2.1	离散信道的数学模型	(67)
4.2.2	单符号离散信道	(68)
4.2.3	信道疑义度	(70)
4.2.4	平均互信息	(72)
4.2.5	各种熵、信道疑义度及平均互信息量之间的相互关系	(77)
4.3	离散无记忆扩展信道	(78)
4.3.1	N 次扩展信道	(78)
4.3.2	定 理	(81)
4.4	信道的组合	(85)
4.5	信道容量	(91)
4.5.1	信道容量的定义	(91)
4.5.2	离散无噪信道	(92)
4.5.3	离散对称信道	(94)
4.5.4	一般离散信道	(98)
4.5.5	离散无记忆 N 次扩展信道	(103)
4.5.6	独立并联信道	(104)
4.5.7	信源和信道匹配	(105)
	习 题	(106)

第 5 章 无失真信源编码

5.1	编码器	(111)
-----	-----	-------

5.2	分组码	(113)
5.3	定长码	(116)
5.4	变长码	(123)
5.4.1	码的分类和主要编码方法	(123)
5.4.2	克拉夫特不等式和麦克米伦不等式	(124)
5.4.3	唯一可译码判别准则	(128)
5.4.4	变长编码定理	(128)
5.4.5	变长码的编码方法	(134)
习 题		(142)

第 6 章 有噪信道编码

6.1	噪声信道的编码问题	(145)
6.1.1	错误概率和译码规则	(145)
6.1.2	译码规则	(146)
6.2	错误概率与编码方法	(150)
6.2.1	简单重复编码	(151)
6.2.2	消息符号个数	(152)
6.2.3	(5.2)线性码	(154)
6.2.4	汉明距离	(155)
6.3	有噪信道编码定理	(157)
6.4	错误概率的上界	(161)
习 题		(162)

第 7 章 限失真信源编码

7.1	失真测度	(164)
7.1.1	失真函数	(164)
7.1.2	平均失真	(166)
7.2	信息率失真函数	(167)
7.2.1	D^* 允许信道(试验信道)	(167)
7.2.2	信息率失真函数的定义	(167)
7.2.3	信息率失真函数 $R(D)$ 的性质	(168)
7.3	限失真信源编码定理和逆定理	(171)
7.3.1	限失真信源编码定理	(171)
7.3.2	限失真信源编码逆定理	(175)
7.4	信息率失真函数的计算	(177)
7.4.1	$R(D)$ 参量表示法求解	(177)
7.4.2	应用参量表示式计算 $R(D)$ 的例题	(180)
7.4.3	$R(D)$ 的迭代计算方法	(185)
习 题		(188)

第8章 连续信源和波形信道

8.1 连续信源的特征	(189)
8.1.1 连续信源	(189)
8.1.2 连续信源的熵	(189)
8.1.3 连续信源的最大熵	(191)
8.1.4 联合熵、条件熵和平均交互信息量	(193)
8.1.5 连续信源的熵速率和熵功率	(194)
8.2 连续信道的信道容量	(197)
8.2.1 时间离散信道的容量	(197)
8.2.2 时间连续信道的容量	(200)
8.3 连续信道的信道编码定理	(202)
8.4 连续信源的信息率失真函数	(202)
习 题	(206)

第9章 纠错编码

9.1 纠错码的基本概念	(209)
9.2 纠错码分类	(211)
9.3 线性分组码	(212)
9.3.1 校验矩阵与生成矩阵	(212)
9.3.2 线性分组码的纠、检错能力	(216)
9.3.3 校验矩阵与最小距离的关系	(219)
9.3.4 线性分组码的伴随式	(219)
9.3.5 线性分组码的译码	(220)
9.3.6 汉明码	(221)
9.4 几种重要的纠错码	(222)
9.4.1 循环码	(222)
9.4.2 卷积码	(224)
习 题	(228)

第10章 信息论方法在信号处理中的应用

10.1 最大熵谱估计	(231)
10.1.1 最大熵谱估计的机理及伯格递推算法	(231)
10.1.2 最大熵谱估计	(233)
10.2 最小误差熵估计与卡尔曼滤波	(241)
10.2.1 最小均方误差准则与最小误差熵准则	(241)
10.2.2 最小误差熵准则推导卡尔曼滤波方程	(243)

参考文献

第1章 绪论

1.1 信息

信息论或称为通信的数学理论,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。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。

信息这一概念是在人类社会互通情报的实践过程中产生的。信息概念是十分广泛的。由于信息科学比起其他学科,如物理学、数学、化学、生物学等还显得很年轻,人类对信息的认识还远远不够,迄今为止,信息并没有形成一个很完整的系统的概念。不同的研究学派对信息的本质及其定义还没有形成统一的意见和认识。

信息论的发展对人类社会和科学技术的进步有着相当深刻的影响。信息作为一种资源,如何开发、利用、共享是人们普遍关注的问题。

信息的通俗概念 人们认为信息是一种消息。这是一种最普通的概念,是目前社会上最流行的概念。例如,当人们收到一封电报,接到一个电话,收听了广播或看到电视以后,就说得到了信息。这个概念好像使人一听就明白,其实极不准确。确切地说,这种概念把消息当成了信息。的确,人们从接收到的电报、电话、广播和电视的消息中能获得各种信息。但是,信息和消息并不是一回事,不能等同。举例来说,有人告诉你一条消息,这条消息告诉了许多原来不知道的新内容,这条消息就很有意义,信息量就大;反之,如果这条消息告诉的是原来就已经知道的内容,那么这条消息意义就不大,信息量就小。

信息的广义概念 人们认为信息是对物质存在和运动形式的一般描述。1975年 Lango 提出“一旦您理解了是信息触发了行为和能力,信息是含于客体间的差别中而不是客体本身中,您就意识到在通信中所被利用的(亦即携带信息的)实际客体是不重要的,仅仅差别关系重要。”在这里,客体是指消息,差别是指信息,差别不是客体,信息不同于消息。例如,妻子给丈夫邮寄一包衣物。那么,衣物是客体,但客体是不重要的,重要的是含于客体中的、寄去的思念和情感,这是信息。

物质、能量和信息是构成客观世界的三大要素。信息是物质和能量在空间和时间中分布的不均匀程度。信息不是物质。信息是事物的表征,它表征事物的状态和运动形式。信息存在于任何事物之中,有物质的地方就有信息,信息充满着整个物质世界。

信息是一个十分抽象的概念。信息本身是看不见、摸不着的,它必须依附于一定的物质形式,如文字、声波、电磁波等。这种运载信息的物质,称为信息的载体。一切物质都有可能成为信息的载体。

概率信息 概率信息是由美国数学家香农(C. E. Shannon)提出来的,故称香农信息或狭义信息。他是从不确定性(随机性)和概率测度的角度给信息下定义的。香农从信息源具有随机性不定度出发,为信源推出一个与统计力学的熵相似的函数,称为信息熵;而这个熵就是信源的信息选择不定度的测度,从而我们可以认为信息表征信源的不定度,但它不等同于不定

度,而是为了消除一定的不定度必须获得与此不定度相等的信息量。

我们可以从下面这个例子来理解概率信息的直观意义。设甲袋中有 100 个球,其中 50 个是红球,另外 50 个是白球;乙袋中也有 100 个球,其中有 25 个红球,25 个白球,25 个蓝球,25 个黑球。今从甲、乙袋中各取出一个球。当被告知,从甲袋中取出的球是红球,从乙袋中取出的球也是红球时,那么这两个消息包含的信息量是不相同的。由于从甲袋中取出一个红球的概率大,不确定性小,因此信息量小;而从乙袋中取出一个红球的概率小,不确定性大,故信息量大。

至此,我们已经了解到信息不是消息,而消息也不同于信号。下面我们概括一下信息、消息和信号三者的含义及其差异。

信 息 信息是一个十分抽象而又复杂的概念。它包含在消息之中,是通信系统中传送的对象。信息作为客观世界存在的第三要素,与物质、能量相比,具有一些特殊的性质。

(1) **信息是无形的** 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) **信息是可共享的** 信息是可以共享的。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。信息可以由甲传递给乙,又可以由乙传递给丙,依次类推等等。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。但是,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。信息虽具有共享性,原占有者不会因信息传递而丢失这个信息,但占有者和获得者可以利用同一个信息进行竞争和对抗。因此,在信息的占有和传播方面,存在着斗争。为了限制信息的共享,加设密码、数据库保安措施等就是这种斗争中产物。

(3) **信息是无限的** 信息像物质、能量一样,对于人类也是一种资源。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。信息如浩瀚天空,永远在产生、更新、演变,是一种取之不尽、用之不竭的源泉。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,给人类创造幸福。

(4) **信息是可度量的** 信息论中最重要的问题,就是要解决信息数量与质量的度量。信息度量应满足信息的 3 个基本方向:结构的、统计的和语义的。

结构理论是研究大量信息的离散构造的。它通过简单计算信息元方法,或者用大量信息简易编码所提供的组合方法对信息进行测量。

统计理论是利用熵的概念,作为统计发生概率的不确定性度量,从而得出这些或那些消息的信息量。

消 息 消息是比较具体的概念,但是它不是物理的。消息是信息的载荷者。消息具有不同的形式,例如语言、文字、符号、数据、图片等,所有这些形式都是能够被人们感觉到的。构成消息的条件有二个:一是构成的消息能够被通信双方所理解;二是可以在通信中进行传递和交换。在日常生活中,从电报、电话、电视等通信系统中得到的是一些描述各种主客观事物运动状态或存在形式的具体消息。需要指出的是,同一个消息可以含有不同的信息量,而同一信息可以用不同形式的消息来载荷。

信号 信号是消息的表现形式,消息则是信号的具体内容。信号是消息的载体,是表示消息的物理量。一般指随时间而变化的电压或电流称为电信号。电信号与非电信号可以比较方便地互相转换。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。

1.2 通信系统模型

信息论研究的主要问题是通信系统设计中如何实现信息传输、存储和处理的有效性和可靠性。我们将通信系统定义为信息的传输系统,例如电报、电话、图像、计算机和导航等系统。实际的通信系统虽然形式和用途各不相同,但从信息传输的角度来看,在本质上有许多共同之处,它们均可概括为如图 1.1 所示的基本模型。

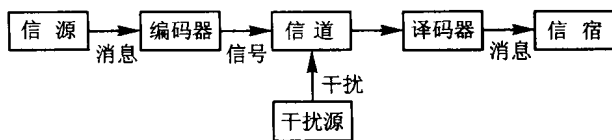


图 1.1 通信系统基本模型

信源 产生消息的源。消息可以是文字、语言、图像等。它可以是离散序列,也可以是连续形式,但都是随机发生的,亦即在未收到这些消息之前不可能确切地知道它们的内容。这些消息可以用随机变量或随机过程来描述。信源研究的主要内容是消息的统计特性和信源产生信息的速率。

编码器 将信源发出的消息变换成适于信道传送的信号的设备。它包含下述三个部分。

(1) 信源编码器。在一定的准则下,信源编码器对信源输出的消息进行适当的变换和处理,其目的在于提高信息传输的效率。

(2) 纠错编码器。纠错编码器是对信源编码器的输出进行变换,用以提高对于信道干扰的抗击能力,亦即提高信息传输的可靠性。

(3) 调制器。调制器是将纠错编码器的输出变成适合于信道传输要求的信号形式。纠错编码器和调制器的组合又称为信道编码器,如图 1.2 所示。在实际系统中不一定每个编码器都含有这三个部分,有的只有其中的两个部分或一个部分组成。

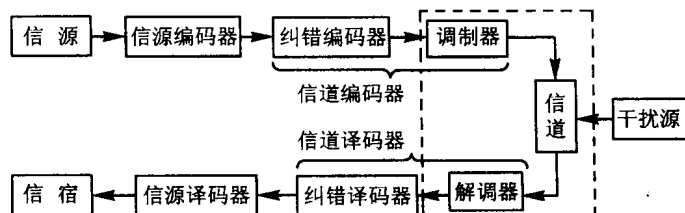


图 1.2 通信系统模型

信道 把载荷消息的信号从发射端传到接收端的媒质或通道,是包括收发设备在内的物理设施。在狭义的通信系统中,实际信道有架空明线、电缆、波导、光纤、无线电波传播空间等。当然,对广义的通信系统来说,信道还可以是其他传输媒介。

干扰源 在信道中引入噪声和干扰,这是一种等效的表达方式。为了分析方便,把在系统中其他各部分产生的噪声和干扰都等效成信道干扰,并集中作用于信道。由于噪声和干扰往往具有随机性,因此它是划分信道的重要因素,并且是决定信道传输能力的决定因素。研究信道的中心课题是它的统计特性和传输能力。

实际干扰可以分成以下两大类。

(1) 加性干扰。由外界引入的随机干扰,如天电干扰以及设备内部噪声,它们与信道的输入信号统计无关。信道的输出是输入信号和干扰的和。

(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。

译码器 编码的逆变换。它要从受干扰的信号中最大限度地提取出有关信源输出消息的信息,并尽可能地复现信源的输出。译码器也可分为信源译码器和信道译码器。译码器的输出送给信宿。

信 宿 信息传送过程中的接收者,即接收消息的人或物。信宿和信源可处于不同的地点或存在于不同时刻。

通信系统的模型不是不变的,这要视实际情况而定。图 1.1 和图 1.2 给出的模型只适用于收发两端单向通信的情况,它只有一个信源和一个信宿,信息传输也是单向的。在网通信情况下,可能有很多分开的信源、信道和信宿进行信息交换。例如,广播通信是一个输入、多个输出的单向传输通信,而卫星通信网则是多个输入、多个输出和多向传输的通信。要研究这些通信系统中的信息传输和处理问题,只要对两端单向通信系统模型作适当修正,引出多用户通信系统模型,并将单路通信的信息理论发展成为多用户信息理论。这是近 20 多年来信息理论研究中的一个十分活跃的课题。

在通信系统中,信源发出的消息可以是连续消息,也可以是离散消息。由连续消息变换成连续信号,在时间上是连续的,又称为模拟信号。采用模拟信号作为传输信号的通信系统称为模拟通信系统,如广播、电视、载波长话等。由离散消息变换成离散信号,在时间上是离散的,又称为数字信号。采用数字信号作为传输信号的通信系统称为数字通信系统,如电报、数据传输、数字电话等。数字通信有着许多突出的优点,如抗干扰能力强、可用纠错技术提高系统可靠性等。数字通信是当前通信系统的重要发展方向。

1.3 信息论的形成和发展

香农信息论的基本任务是为设计有效而可靠的通信系统提供理论依据。

信息论是信息科学的主要理论基础之一,它是在长期通信工程实践和理论基础上发展起来的。信息论自诞生到现在不过 40 多年,这在人类的历史长河中是十分短暂的,但它的诞生和发展对科学技术的影响是相当深刻的。现在它已成为一门独立的理论学科。回顾其发展历史,我们将会认识到,在现代科学技术的高度发展过程中,学习和掌握信息理论日益成为一种必要。

通信系统对人类社会的发展有着十分重要的作用。日常生活、工农业生产、科学实践等一切都离不开信息传递和流动。电的通信系统(电信系统)已有 150 年的历史,它在信息论的发展过程中起到了积极推动作用。

1924年奈奎斯特(H. Nyquist)解释了信号带宽和信息率之间的关系。他指出,如果以一个确定的速率来传输电报信号,就需要一定的带宽。他将信息率和带宽联系起来了。

1928年哈特莱(R. V. Hartley)引入了非统计(等概率事件)信息量概念。他提出信息量等于可能消息数的对数。他的工作对后来香农的思想是有影响的。

1936年阿姆斯特朗(E. H. Armstrong)提出在传输过程中增大带宽可以增强抑制干扰的能力。根据这一思想,他提出了宽频移的频率调制方法。

1936年达得利(H. Dudley)发明了声码器。他提出的概念是通信所需要的带宽至少同所要传送的消息的带宽应该一样。

20世纪40年代初期,由于军事上的需要,维纳(N. Wiener)在研究防空火炮的控制问题时,提出了《平稳时间序列的外推、内插与平滑及其工程应用》的论文。他把随机过程和数理统计的观点引入通信和控制系统中来,揭示了信息传输和处理过程的统计本质。他还利用早在20世纪30年代初他本人提出的“广义谐波分析理论”对信息系统中的随机过程进行谱分析。这就使通信系统的理论研究引起了质的飞跃,取得了突破性进展。

1948年香农在贝尔系统技术杂志上发表了两篇有关“通信的数学理论”的文章。在这两篇论文中,他用概率测度和数理统计的方法,系统地讨论了通信的基本问题,得出了几个重要的而带有普遍意义的结论,并由此奠定了现代信息论的基础。

从20世纪50年代开始,信息论在学术界引起了巨大的反响。1951年美国IRE成立了信息论组,并于1955年正式出版了信息论汇刊。在此期间,一些科学家(包括香农本人)做了大量工作,发表了许多重要文章。他们将香农已得到的数学结论作了进一步的严格论证和推广。其中,1954年范恩斯坦(A. Feinstein)的论著是有很大贡献的。1959年香农发表了“保真度准则下的离散信源编码定理”,系统地提出了信息率失真理论。这一理论是频带压缩、数据压缩的数学基础,为各种信源压缩编码的研究奠定了理论基础,一直到今天仍然是信息论领域的重要研究课题。

在整个20世纪50年代,维纳理论也有很大进展。维纳的工作是从研究处在统计平衡的时间序列开始的。维纳证明:在一定条件下,处在统计平衡的时间序列的时间平均等于集平均。基于此点,维纳提出了他的著名的滤波和预测理论。维纳理论在滤波理论中的开拓作用是毋庸置疑的,他在滤波方法上的创见仍然直接影响着后来的科学工作者。

20世纪50年代中期,空间技术飞速发展,要求对卫星轨道进行精确测量。为此,人们将滤波问题以微分方程表示,提出了一系列适应空间技术应用的精确算法。20世纪60年代初卡尔曼(R. E. Kalman)和布西(R. S. Bucy)提出了递推滤波算法,成功地将状态变量引入到滤波理论中来,用消息与干扰的状态空间模型代替了通常用来表示他们的协方差函数。将状态空间描述与离散时间更新联系起来,更适于计算机直接进行运算,而不是去寻求滤波器冲击响应的明确表示式。这种算法得出的是表征状态估值及其均方误差的微分方程,给出的是递推算法。这就是著名的卡尔曼滤波理论,或称为卡尔曼-布西滤波。

20世纪70年代以后,卡拉思(T. Kailath)等人发展了信息过程理论。这一理论不仅可以用来解决白高斯型检测问题和线性最小平方估计问题,而且也可以用来解决非高斯型检测和非线性最小平方估计问题。

1961年香农发表的论文《双路通信信道》开拓了多用户理论的研究。随着卫星通信、计算机通信网络的迅速发展,多用户理论的研究取得了许多突破性进展。从20世纪70年代以后,

人们从经典的香农单向通信的信息论推广到多用户信息理论。多用户信息理论成为当前信息论的中心研究课题之一。

从 20 世纪 40 年代开始,信息理论与技术在人类历史长河中已经取得了长足的进展,它已形成一门综合性的新兴学科,在人们面前展示出光辉灿烂的前景。现在,信息理论与技术不仅直接应用于通信、计算机和自动控制等领域,而且还广泛渗透到生物学、医学、语言学、社会学和经济学等领域。特别是通信技术与微电子、光电子、计算机技术等相结合,使现代通信技术的发展充满生机与活力。能够不受时间、空间、地点的限制,可以随时随地进行各种各样的信息交换。人们追求的目标是实现宽带综合业务数字信息网(B-ISDN),使人类进入高度发展的信息科学时代。

第2章 信息的统计度量

2.1 自信息量和条件自信息量

2.1.1 自信息量

通过某个过程或手段,获得了关于随机信息源一定的了解,减少了不确定性,即从这个信息源获得了信息。从信息源获取信息的过程就是其不确定性缩减的过程。可见信息源包含的信息与其不确定性是紧密相关的。在统计分析中,使用概率作为衡量不确定性的一种指标。可以推论,随机事件包含信息的度量应是其概率的函数。

定义 2.1.1 任意随机事件的自信息量定义为该事件发生概率的对数的负值。

设该事件 x_i 的概率为 $p(x_i)$,那么,它的自信息定义式为

$$I(x_i) \stackrel{\text{def}}{=} -\log p(x_i) \quad (2-1)$$

自信息量是取其概率的对数的负值, $I(x_i)$ 为非负。自信息量的单位与所用对数的底有关。通常取对数的底为 2,信息量的单位为比特(bit)。若 $p(x_i)=1/2$,则 $I(x_i)=1$ 比特,即该事件 x_i 具有 1 比特的自信息量。比特是信息理论中最常用的信息量单位。若取自然对数(对数底为 e),此时,自信息量的单位为奈特(nat)。

$$1 \text{ 奈特} = \log_2 e \text{ 比特} \approx 1.443 \text{ 比特}$$

若以 10 作为对数的底,此时自信息量的单位为哈脱来(haitely)。

$$1 \text{ 哈脱来} = \log_2 10 \text{ 比特} \approx 3.322 \text{ 比特}$$

由于各种随机事件发生的概率不同,它们所包含的不确定性也有大小的差别。一个随机事件的出现概率接近 1,说明该事件发生的可能性很大,它所包含的不确定性就很小。反之,对于小概率事件,它所包含的不确定性就很大。

可以看出,小概率事件所包含的不确定性大,其自信息量大;出现概率大的随机事件所包含的不确定性小,其自信息量小。在极限情况下,概率为 1 的确定性事件,其自信息量为零。

随机事件的不确定性在数量上等于它的自信息量,也可以用式(2-1)计算。

定义 2.1.2 二维联合集 XY 上的元素 (x_i, y_i) 的联合自信息量定义为

$$I(x_i, y_i) \stackrel{\text{def}}{=} -\log p(x_i, y_i) \quad (2-2)$$

式中, x_i, y_i 为积事件, $p(x_i, y_i)$ 为元素 x_i, y_i 的二维联合概率。

例 2.1.1 设在甲袋中放入 n 个不同阻值的电阻,如果随机地取出一个,并对取出的电阻值进行事先猜测,其猜测的困难程度相当于概率空间的不确定性,概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$$

其中, x_i 代表阻值为 i 的电阻, $i=1, 2, \cdots, n$; $p(x_i)$ 表示取出电阻值为 i 的电阻的概率。为简

便起见,假定取出电阻是等概的,即

$$p(x_i) = \frac{1}{n} \quad i = 1, 2, \dots, n$$

那么,被告知“取出的阻值为 i 的电阻”所获得的信息量为

$$I(x_i) = -\log_2 p(x_i) = \log_2 \frac{1}{p(x_i)} = \log_2 n$$

由于取出电阻值为 i 的电阻是等概分布的,因此随意取出任一阻值的电阻所获得的信息量都是相等的。

如果在甲袋中放入 $\frac{1}{2}n(n+1)$ 个不同阻值的电阻,其中阻值为 $1\ \Omega$ 的 1 个, $2\ \Omega$ 的 2 个, $\dots, n\ \Omega$ 的 n 个。若从中随意取出一个,并对取出的电阻值进行事先猜测,其猜测的困难程度相当于概率空间的不确定性,概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ 1 & 2 & \dots & n \\ \frac{1}{2}n(n+1) & \frac{1}{2}n(n+1) & \dots & \frac{1}{2}n(n+1) \end{bmatrix}$$

其中, x_i 代表阻值为 i 的电阻, $i=1, 2, \dots, n$; 概率 $p(x_i) = i / \frac{1}{2}n(n+1)$, $i=1, 2, \dots, n$ 。

那么被告知“取出的阻值为 i 的电阻”所获得的信息量为

$$I(x_i) = -\log p(x_i) = \log \frac{n(n+1)}{2i}$$

当 $i=1\ \Omega$ 时, $I(x_1) = \log n(n+1)/2$; 当 $i=n\ \Omega$ 时, $I(x_n) = \log(n+1)/2$ 。这样,被告知“取出的阻值为 $1\ \Omega$ 的电阻”比“取出的阻值为 $n\ \Omega$ 的电阻”所获得的信息量要大。

2.1.2 条件自信息量

定义 2.1.3 联合集 XY 中,对事件 x_i 和 y_j ,事件 x_i 在事件 y_j 给定的条件下的条件自信息量定义为

$$I(x_i | y_j) \stackrel{\text{def}}{=} -\log p(x_i | y_j) \quad (2-3)$$

由于每一个随机事件的条件概率都处在 $0 \sim 1$ 范围内,所以条件自信息量均为非负值。

例 2.1.2 设在一正方形棋盘上共有 64 个方格,如果甲将一粒棋子随意地放在棋盘中的某方格且让乙猜测棋子所在位置:

(1) 将方格按顺序编号,令乙猜测棋子所在方格的顺序号;

(2) 将方格按行和列编号,甲将棋子所在方格的行(或列)编号告诉乙之后,再令乙猜测棋子所在列(或行)的位置。

由于甲是将一粒棋子随意地放在棋盘中某一方格内,因此棋子在棋盘中所处位置为二维等概率分布。二维概率分布函数为 $p(x_i y_j) = 1/64$,故

(1) 在二维联合集 XY 上的元素 $x_i y_j$ 的自信息量为

$$I(x_i y_j) = -\log_2 p(x_i y_j) = -\log_2 \frac{1}{64} = \log_2 2^6 = 6 \text{ 比特}$$

(2) 在二维联合集 XY 上,元素 x_i 相对 y_j 的条件自信息量为

$$I(x_i | y_j) = -\log_2 p(x_i | y_j) = -\log_2 \frac{p(x_i y_j)}{p(y_j)} = -\log_2 \frac{1/64}{1/8} = 3 \text{ 比特}$$

2.2 互信息量和条件互信息量

2.2.1 互信息量

设有两个离散的符号消息集合 XY , X 表示信源发出的符号消息集合, Y 表示信宿接收到的符号消息集合。由于接收者信宿事先不知道信源发出的是哪一个符号消息, 因此每个符号消息相当于一个随机事件。信源发出的符号消息通过信道传递给信宿, 如图 2.1 所示。有时也把信源发出的信息说成是信道的输入消息, 而把信宿收到的消息说成是信道的输出消息。

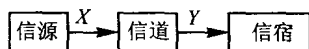


图 2.1 简化的通信系统模型

通常预先知道信源集合 X 包含的各个符号消息 $X_1, X_2, \dots$ 以及它们的概率分布 $p(x_i)$, $i=1, 2, \dots$, 亦即预先知道信源集合 X 的概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots \\ p(x_1) & p(x_2) & \cdots \end{bmatrix}$$

其中 $x_i, i=1, 2, \dots$ 为集合 X 中各个消息 $X_i, i=1, 2, \dots$ 的取值; 概率 $p(x_i), i=1, 2, \dots$ 称为先验概率。

信宿收到的符号消息集合 Y 的概率空间为

$$\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots \\ p(y_1) & p(y_2) & \cdots \end{bmatrix}$$

其中 $y_j, j=1, 2, \dots$ 是集合 Y 中各个消息符号 $Y_j, j=1, 2, \dots$ 的取值; 概率 $p(y_j), j=1, 2, \dots$ 为消息符号 $Y_j, j=1, 2, \dots$ 出现的概率。当信宿收到集合 Y 中的一个消息符号 Y_j 后, 接收者重新估计关于信源各个消息 X_i 发生的概率就变成条件概率 $p(x_i | y_j)$, 这种条件概率又称为后验概率。

定义 2.2.1 对两个离散随机事件集 X 和 Y , 事件 y_j 的出现给出关于事件 x_i 的信息量定义为互信息量。其定义式为

$$I(x_i; y_j) \stackrel{\text{def}}{=} \log \frac{p(x_i | y_j)}{p(x_i)} \quad (2-4)$$

互信息量的单位与自信息量的单位一样取决于对数的底。当对数底为 2 时, 互信息量的单位为比特。由式(2-4)又可得到

$$I(x_i; y_j) \stackrel{\text{def}}{=} \log \frac{1}{p(x_i)} - \log \frac{1}{p(x_i | y_j)}$$

上式意味着互信息量等于自信息量减去条件自信息量。或者说互信息量是一种消除的不确定性的度量, 亦即互信息量等于先验的不确定性 $\log(1/p(x_i))$ 减去尚存在的不确定性 $\log(1/p(x_i | y_j))$ 。