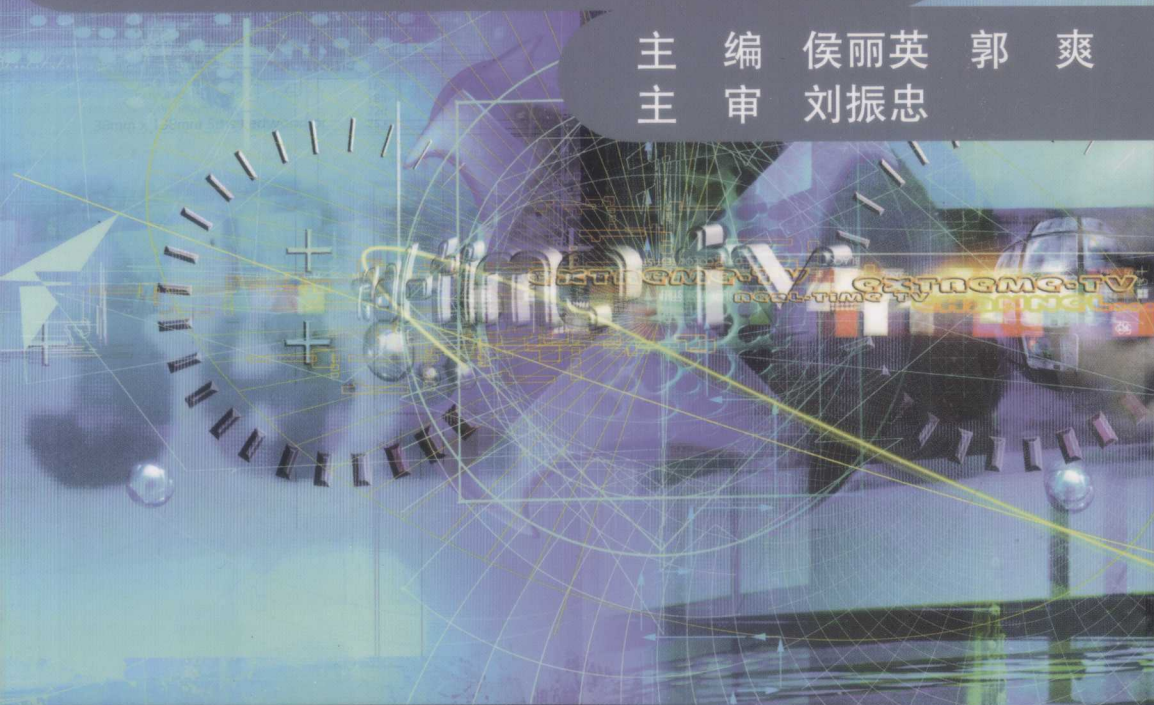


信息论基础

XINXILUN JICHU

主 编 侯丽英 郭 爽
主 审 刘振忠



东北林业大学出版社

信息论基础

主编 侯丽英 郭 爽

主审 刘振忠

东北林业大学出版社

图书在版编目 (CIP) 数据

信息论基础/侯丽英, 郭爽主编. —哈尔滨: 东北林业大学出版社, 2009. 6
ISBN 978-7-81131-474-8

I. 信… II. ①侯…②郭… III. 信息论 IV. TN911.2

中国版本图书馆 CIP 数据核字 (2009) 第 102922 号

责任编辑: 卢 伟

封面设计: 彭 宇



信息论基础

Xinxilun Jichu

主编 侯丽英 郭 爽

主审 刘振忠

东北林业大学出版社出版发行

(哈尔滨市和兴路 26 号)

哈尔滨市工大节能印刷厂印装

开本 787 × 1092 1/16 印张 14.625 字数 335 千字

2009 年 6 月第 1 版 2009 年 6 月第 1 次印刷

印数 1—1 000

ISBN 978-7-81131-474-8

定价: 28.00 元

信息论基础

XINXILUN JICHU

ISBN 978-7-81131-474-8



9 787811 314748 >

定价: 28.00 元

前 言

为了适应高等农林院校信息与计算科学专业和理工科有关专业本科生信息科学基础课程的教学工作,我们组织编写了这部《信息论基础》教材。

《信息论基础》是信息与计算科学专业的一门重要基础课,通过它的学习,我们不但能掌握这一学科的基本概念、基本原理和方法,更重要的是让我们看到数学理论在解决通信技术与通信工程中问题的思想和过程,从而它在提高学生运用数学分析问题、解决问题能力和提高综合素质方面起着十分重要的作用。

本书具有以下几个方面的特点:

(1)在适应性上,本教材针对我国高等教育实行扩招和多层次办学的现状,考虑了普通农林院校学生的实际水平,采取起点适当降低、难度分散处理、注重联系生产和生活实际的原则编写。

(2)在内容上,我们仍然把香农信息论作为本教材的核心,这些内容不仅是近代信息处理的理论基础,具有深刻而又严格的数学描述与证明,而且在通信领域得到重要应用。但在某些定理、公式的推导和证明上,我们尽量选择简便的证明方式,并注重定理含义的解释;另外在保证课程系统性的前提下,本书增加了概率论与随机过程学科中相关理论、结论的预备知识,对部分内容进行了调整,以便适当地降低初学者学习课程的难度,更好地掌握相关知识。

(3)在与实践的结合上,由于信息论是一门理论与实践并重的科学,我们选择了若干个与信息处理过程密切相关的理论问题,给出它们的实验内容;在例题与习题的选择上,也尽量注意到知识点在生产实践和工程等方面的应用,以此来提高学生信息论的兴趣,加深对基础理论内容的理解。

本书由黑龙江八一农垦大学、沈阳农业大学、大庆师范学院等院校有多年教学经验的教师共同编写。主编由黑龙江八一农垦大学的侯丽英和大庆师范学院的郭爽担任。具体编写分工如下:第一章和第二章由邓廷勇(黑龙江八一农垦大学)编写;第三章由王立华(沈阳农业大学)编写;第四章与实验设计由郭庆丰(黑龙江八一农垦大学)编写;第五章和第七章的1~3节由侯丽英编写;第六章和第七章的4~5节由郭爽编写。黑龙江八一农垦大学的代冬岩、野金花、侯杰、魏玉芬和李雨共同完成了全部课后习题和参考答案的编写以及部分书稿的校对工作。

本书由东北农业大学数学系刘振忠教授担任主审。

由于编者能力有限,难免有疏漏之处,敬请读者指正。

编 者

2009年1月

目 录

第一章 信息论概述	(1)
第一节 信息的概念	(1)
一、信息的定义	(1)
二、信息的特征	(2)
第二节 信息论简介	(3)
一、通信系统模型	(4)
二、信息论研究的主要内容	(6)
三、信息论的一些应用成果	(6)
习题一	(9)
第二章 信息论的基本概念	(10)
第一节 预备知识	(10)
一、概率论基础知识回顾	(10)
二、凸函数及其性质	(12)
三、Jensen 不等式	(13)
第二节 离散熵	(13)
一、自信息	(13)
二、信息熵的定义	(15)
三、熵函数的性质	(17)
四、联合熵与条件熵	(21)
第三节 离散互信息	(24)
一、事件的互信息	(24)
二、随机变量的互信息	(26)
三、互信息函数的性质	(27)
四、联合互信息与条件互信息	(31)
第四节 连续随机变量下的信息熵与互信息	(33)
一、连续随机变量下的微分熵	(33)
二、连续随机变量下的互信息	(37)
习题二	(39)

第三章 信源的熵率及冗余度	(44)
第一节 信源的分类及其数学模型	(44)
第二节 离散平稳信源的熵率	(45)
一、离散平稳信源的定义	(45)
二、离散平稳无记忆信源	(47)
三、离散平稳有记忆信源	(47)
第三节 马尔可夫信源的熵率	(50)
一、马尔可夫链的基本概念	(51)
二、马尔可夫信源的熵率	(54)
第四节 信源的冗余度	(58)
一、离散信源的冗余度	(58)
二、连续信源的最大熵与熵功率	(61)
习题三	(64)
第四章 信道及信道容量的计算	(67)
第一节 信道的分类及其数学模型	(67)
一、信道的分类	(67)
二、离散单符号信道的数学模型	(68)
第二节 信道容量及其计算方法	(70)
一、信道容量的概念	(70)
二、几种特殊信道的信道容量	(72)
三、离散对称信道的信道容量	(73)
四、一般离散信道的信道容量	(77)
五、信道容量定理	(81)
六、信道容量的迭代算法	(85)
第三节 离散无记忆扩展信道及其信道容量	(88)
第四节 组合信道及其信道容量	(91)
一、独立并联信道	(91)
二、级联信道	(92)
第五节 连续信道及其信道容量	(92)
一、高斯加性信道的信道容量	(92)
二、多维高斯加性信道的信道容量	(94)
第六节 波形信道及其信道容量	(95)
习题四	(96)

第五章 无失真信源编码	(100)
第一节 信源编码的基本概念	(100)
一、编码器	(100)
二、码的分类	(102)
第二节 渐近等分割性和 ε 典型序列	(104)
一、预备知识	(104)
二、渐近等分割性和 ε 典型序列	(105)
第三节 信源的定长编码定理	(109)
第四节 信源的变长编码定理	(113)
一、Kraft 不等式和 McMillan 不等式	(114)
二、唯一可译码的判别准则	(117)
三、无失真变长编码定理	(118)
四、香农第一编码定理	(120)
第五节 霍夫曼码和其他编码方法	(124)
一、香农编码	(124)
二、香农—费诺—埃利斯编码	(125)
三、霍夫曼编码	(125)
四、费诺编码	(129)
第六节 实用的无失真信源编码方法	(131)
一、游程编码	(131)
二、算术码	(133)
三、LZW 码	(134)
习题五	(137)
第六章 信道编码	(141)
第一节 信道编码的基本概念	(141)
一、错误概率与译码规则	(142)
二、错误概率与编码方法	(146)
第二节 有噪信道编码定理	(153)
一、有噪信道编码定理	(153)
二、有噪信道编码逆定理	(156)
三、错误概率的上界	(157)
第三节 纠错编码	(158)
一、纠错编码分类	(158)

二、纠错编码的基本概念	(160)
第四节 几种重要的纠错码	(161)
一、线性分组码	(161)
二、汉明码	(172)
三、循环码	(173)
四、卷积码	(177)
习题六	(181)
第七章 信息率失真函数与熵压缩编码	(184)
第一节 失真测度	(185)
一、失真函数	(185)
二、平均失真	(187)
第二节 信息率失真函数	(188)
一、 D 失真许可信道	(188)
二、信息率失真函数的定义	(188)
三、信息率失真函数 $R(D)$ 的性质	(189)
第三节 信息率失真函数的计算	(194)
一、应用参量表示式计算 $R(D)$	(194)
二、二元信源和离散等概信源的 $R(D)$ 函数	(200)
第四节 限失真信源编码定理	(202)
一、限失真信源编码定理	(202)
二、限失真信源编码逆定理	(206)
第五节 熵压缩编码具体方法	(207)
一、标量量化	(207)
二、矢量量化	(208)
三、变换编码	(209)
四、预测编码	(209)
习题七	(210)
实验设计	(213)
习题参考答案	(217)
参考文献	(223)

第一章 信息论概述

信息已经成为现代社会的一项重要资源,它在自然科学和社会科学中被广泛采用.信息论是20世纪40年代后期从长期通讯实践中总结出来的一门学科,它是研究信息的有效处理和可靠传输的一般规律的科学.在这一章中,我们将简要地从科学技术的角度讨论什么是信息,信息论的研究对象、方法和主要内容,通过对信息论一些应用成果的回顾,进一步提高对这一学科的总体认识.

第一节 信息的概念

一、信息的定义

信息是一个十分通俗而又广泛的名词,当今信息社会中,常把它作为人类认识世界的向导与智慧的源泉,也是社会与生产力发展的动力与资源.但到底什么是信息?从不同的学科及不同的角度有不同的解释定义,我们把它大体归结为三大类.

第一种类型就是从日常生活认识来看,信息被看做新闻、消息与知识.例如,当人们收到一封电报,或者听了天气预报,就会说得到了信息.在近代,信息一词在英语中被译作information, information在牛津字典里给出的解释是“某人被通知或被告知的内容、情报、消息”.这种定义较符合人们在日常生活中对信息的理解,虽然有着十分广泛的应用,但其含义模糊,不能作为科学名词或技术术语的定义.

第二种类型是从哲学角度上讲,信息、物质和能量被看做是构成物质世界的三大支柱,物质是基础是实体,能量是物质运动的形式,而信息是依附于物质和能量,但又不同于物质和能量而客观存在的,人们利用信息改造物质,创造新物质,提高能量的利用效率,发现新的能量形式,所以信息是人类认识改造客观世界的更高层次.

第三种类型是科学家的论述,他们是在对信息的处理过程中产生对信息的理解和定义,偏重于组成信息的形式、结构,把信息作为事物的联系变化与差异.如艾什比把信息定义为“差异度”;维纳(N. Wiener)把信息定义为一种“组织性”的度量,并把它作为控制系统在进行调节活动时的相互作用、相互交换的内容.

信息作为一个可以用严格的数学公式定义的科学名词首先出现在统计数学中,随后出现在通信技术中.人们收到消息后,如果消息告诉了我们很多原来不知道的新内容,我们会感到获得了很多的信息,而如果消息是我们基本已经知道的内容,就会感觉得到的信息不多,所以信息是可以度量的.那么怎样度量信息呢?人们需要有一个可以用数学模型来表示的信息概念.

1928年,哈特莱(Hartley)首先提出了用对数度量信息的概念,即一个消息所含有的信息用它的所有可能的取值的个数的对数来表示.比如,抛掷一枚硬币可能有两种结果:正面和反面,所以当我们得知抛掷结果后获得的信息量是 $\log_2 2 = 1$ bit. 而一个十进制数字可以表示0~9中的任意一个符号,所以一个十进制数字含有 $\log_2 10 = 3.3219$ bit的信息量.

哈特莱的工作给了香农很大的启示,他进一步注意到信息量不仅与它的可能值的个数有关,还与信息本身的不确定性有关.例如,抛掷一枚偏畸硬币,如果正面向上的可能值是90%,那么当我们得知抛掷结果是反面时得到的信息量会比得知抛掷结果是正面时得到的信息量大.

一个消息之所以会含有信息,正是因为它具有不确定性,一个不具有不确定性的消息是不会含有任何消息的,而通信的目的就是为了消除或部分消除这种不确定性.比如,在得知硬币的抛掷结果前,我们对于结果会出现正面还是反面是不确定的,通过通信,我们得知了硬币的抛掷结果,消除了不确定性,从而获得了信息.因此,在香农(C. E. Shannon)理论中,把信息定义为信息是对事物运动状态或存在方式的不确定性的描述.

从数学的角度来讲,不确定性就是随机性,具有不确定性的事件就是随机事件.因此,可运用研究随机事件的数学工具——概率,来度量不确定性的大小.在香农信息论中,我们把消息用随机事件表示,而发出这些消息的信源则用随机变量来表示.比如,抛掷一枚硬币的试验可以用一个随机变量来表示,而抛掷结果这个具体的消息则用随机事件表示.

这种建立在概率模型上的信息概念不同于日常生活中那种把形式、内容等全部包含在内的广义信息,也排除了信息一词主观上的含义和作用,而只是对信息的统计特性的定量描述,它与内容无关,不随信息具体表达形式的变化(如把文字翻译成二进制码)而变化,它反映了信息表达形式中统计方面的性质,而且与日常生活中信息的概念并不矛盾,因此是一个科学的定义.本书所讨论的信息理论正是基于这种定义形式展开的.

二、信息的特征

了解信息的特征不仅有助于我们对信息这个概念的理解,也有助于我们今后对信息处理的理解.信息有以下主要特征:

(1)信息必须依附于某种载体,这种载体包含确定的物质内容,并通过一定的结构形式表现.如语言文字信息,它所依附的物质载体是纸张、油墨,仅仅纸张、油墨并不能构成信息,油墨在纸张上形成文字后才能构成信息,不同的文字就是油墨在纸张上的不同表现形式.另外,在相同是物质载体条件下,构成信息的结构形式是千变万化的,同样26个英文字母,由它们组合可以产生各种各样的文字、图画、指令等内容.

(2)信息需在交换、交流、分析与比较中发挥作用,因此信息处理中有信息的生成体、载体与受体的区别.在文字的信息交换中有作家、出版物与读者的区别,通话时有打电话、听电话与传电话的区别,这些都是信息的生成体、载体与受体,在信息论中,我们称之为信源、信道与信宿.而且在不同的生成体与受体中信息所发挥的作用与价值是不同的,同一篇文章,对不同的读者可能产生不同的理解与影响.

(3)信息具有可设计、传递、复制、存储、修改与扩展等特性,对这些特性的处理过程

统称为信息处理. 信息科学就为研究信息处理提供理论基础, 其中包括它们的数学模型、基本的度量关系与性质、相关的优化算法等. 信息处理的内容与方式是多样的, 而且同一信息具有不同的表达方式, 因此信息处理的特点就是在不同的领域选择不同的表达方式, 并确定它们的相互转换关系, 这种转换关系在现有的电子、计算机条件下快速实现.

(4) 为使信息处理达到精确化与定量化的要求, 人们经常设定一些指标来考察信息处理的可行性与优劣, 常见的重要指标有:

①正确率. 它表示在经信息处理后与原始信息的差异性, 最严格的要求是无失真的信息处理. 在一般信息处理中, 出现误差总是不可避免, 因此在信息处理中经常使用允许误差的概念.

②处理速度与速率. 这是与正确率相并行的重要指标, 它是在实现正确率要求条件下完成信息处理所需要的时间, 只有使信息处理的速度与信息需求同步才能继续推进信息科学与技术的发展. 信息处理的速度或速率问题同时与软、硬件技术有关.

③处理空间. 这是在实现正确率要求条件下完成信息处理所需要的空间要求, 这在计算机中一般指存储条件, 在其他信息处理中也有相应的存储要求.

④其他指标, 如:

- 容错率. 在实际信息处理中, 干扰与误差不可避免, 因此无论信息的设计、传递、复制与存储中都有容错的要求. 也就是在信息处理过程中, 如果干扰与误差出现, 可以自动消除这些差错.

- 稳定性与自适应性. 由于外部干扰及环境条件的变化要求, 信息处理需要具有稳定性与自适应性的特征, 其中编码的容错性实际上也是一种自适用性, 它是一种对于干扰的自动纠正的特性.

- 技术要求、经济成本等在实现信息处理时也需要考虑.

第二节 信息论简介

信息论的创始人克劳德·艾尔伍德·香农 (Claude Elwood Shannon, 1916—2001 年), 美国数学家. 1948 年香农在《贝尔系统技术杂志》(Bell System Technical Journal) 上连载发表了著名的论文《通讯的数学原理》. 1949 年, 香农又在该杂志上发表了另一影响深远的论文《噪声下的通信》. 在这两篇论文中, 香农阐明了通信的基本问题, 给出了通信系统的模型, 提出了信息量的数学表达式, 并解决了通信容量、信源统计特性、信源编码、信道编码等一系列基本问题. 这两篇论文成为信息论的奠基性著作.

由于现代通讯技术的飞速发展及其他学科的交叉渗透, 信息论的研究已经从香农当年仅限于通信系统的数学理论的狭义范围扩展开来, 而成为现在称之为信息科学的一个庞大体系. 它的研究内容不仅仅包括通信, 而且包括所有与信息有关的自然和社会领域, 如模式识别、机器翻译、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. 但作为大学本科课程的信息论基础, 我们仍然把香农信息论作为本教材的核心, 这些内容不仅是近代信息处理的理论基础, 具有深刻而又严格的数学描述与证明,

而且在通信领域也得到重要应用。

一、通信系统模型

作为通信系统的数学理论,香农在 1948 年的奠基性文章中提出了通信系统的一般模型,由于信息论是源于通信实践发展起来的一门新兴应用科学,故通信系统的基本模型也是信息理论的基本模型,如图 1.1 所示。

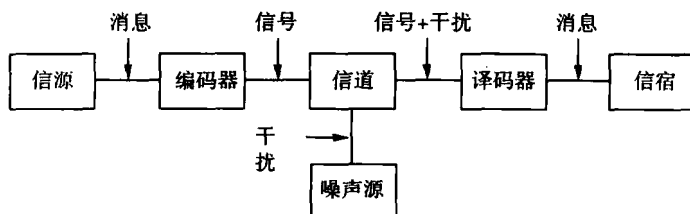


图 1.1 通信系统模型

这个模型不仅包括电话、电报、传真、电视、雷达等狭义的通信系统,还包括生物有机体的遗传系统、神经系统、视觉系统,甚至人类社会的管理系统。信息以消息的形式在这个通信系统中传递,人们通过研究通信系统中消息的传输和处理来得到信息传输和处理的规律,目的是提高通信的可靠性和有效性。

在任何一个信息流通系统中,都有一个发出信息的发送端,有一个接受信息的接收端,以及信息流通的通道。在信息传递的过程中不可避免的会有噪音,所以有一个噪音源。为了把信源发出的消息变成适合在信道中传输的信号,还需要加入编码器,在送到信宿之前要进行反交换,所以要加入译码器。于是这个通信系统主要由五部分组成。

1. 信源

信息的产生和发送者称之为信源,其输出为消息或消息序列。信源可以是人、生物机器,或其他事物。比如,各种气象状态是信源,能够产生独特的气味吸引蜜蜂来采花蜜的花朵是信源,人脑的思维活动也是一种信源。

我们称信源发出的信息为消息,这里并不考虑内含的语言信息,而只考虑它的统计特性,通信的主要目标之一是使接收端能尽可能准确的复制信源发送的信息。消息有着各种不同的形式,例如:文字、符号、语言、图片、图像、气味等。消息以能被通信双方所理解的形式,通过通信进行传递和交换。消息携带着信息,是信息的载体。信源输出的消息是随机的、不确定的,但又有一定的规律性,因此用随机变量或随机矢量等数学模型来表示信源。

2. 信道

信道是指通信系统把载荷消息的信息的信号从发送端送到接收端的媒介或通道,是包括收发设备在内的物理设施。信道除了传播信号以外,还有存储信号的作用。在狭义的通信系统中,实际通信有明线、电缆、光缆、无线电波传播空间、磁盘、光盘等,这些都属于传输电磁波能量的信道。对于广义的通信系统来说,信道还可以是其他的传输媒介。

信道中通常会有噪声干扰,使传输的信息产生失真,如语音通讯系统产生的噪音、电视系统中的“雪花”干扰等。由于噪声的存在,使信道能可靠传输信息的能力受到限制,信道的最大理论信息传输速率称之为信道容量。信道只能用低于信道容量的速率来可靠地

传输信息,如传输速率超过了信道容量,就会出现错误。

为了分析方便起见,把在系统其他部分产生的干扰和噪声都等效地折合成信道干扰,看成是由一个噪声源产生的,它将作用于所传输的信号上。这样,信道输出的已是叠加了干扰的信号。噪声源的统计特征是划分信道的依据,并且是信道传输能力的决定因素。由于干扰或噪声往往具有随机性,所以信道用输入和输出之间的条件概率分布来描述。

3. 编码器

编码就是把消息变成适合在信道传输的物理量,这种物理量称为信号,如电信号、光信号、声信号、生物信号等。信号携带着消息,它是消息的载体。

编码器可分为信源编码器和信道编码器。通常信源发出的信息是有冗余的,比如普通语言与文字是有高冗余度的。为了有效地进行通信,往往对有冗余的信息先进行无冗余或少冗余编码,或称压缩编码,这就是信息源编码器的任务,以提高信息传输的效率,从而提高通信系统的有效性。根据消息的不同特征及信宿对复制消息的不同要求,复制时可以无失真,如对文本消息;也可以允许有一定失真,如对语言、图像、影视信息。由此发展起来了无失真信源编码和允许失真率的限失真信源编码理论,以寻求信息压缩的最优理论极限;同时也寻求压缩率尽可能接近这个理论极限的实用压缩编码技术,这两部分构成了信源编码理论的主体。

由于噪声的存在,使信道能可靠传输信息的能力受到限制。信道编码是为了提高信息传输的可靠性而有目的地对信源编码器输出的代码组添加一些监督码元,使之具有纠、检错能力。这样为了增加传输信息的抗干扰能力,就需增加信息的冗余度,这是信道编码器的任务。

在实际的通信系统中,可靠性和有效性常常是相互矛盾的。那么什么是信息传输的有效性和可靠性呢?比如教师授课的备课阶段,需要把知识进行加工和提炼,以提高信息传输的有效性;而在讲授过程中,为了让学生听得明白,有时又需要适当地重复,这是为了提高信息传输的可靠性。提高有效性必须去掉信源符号的冗余部分,但是这会导致可靠性的下降,而提高可靠性就需要增加监督码元,这又降低了有效性。有时为了兼顾有效性,就不一定要求接收端绝对准确地再现原来的消息,而是可以允许一定的误差或失真,也就是说允许近似地再现原来的消息。

由于只要信源编码的压缩率不超过信道容量,就可以达到既有效又可靠地传输信息的目的,这样就把看似矛盾的信源编码(要减少冗余)和信道编码(要增加冗余)统一在了一起。所以为提高效率,可以把信源与信道编码器合二而一为一个编码器。

4. 译码器

接收端的译码器是把信道输出的已叠加了干扰的编码信号进行编码的逆运算,变成信宿能够理解的消息。译码器也可分成信源译码器和信道译码器。译码器需要尽可能准确地再现信源输出的消息。为了提高效率,也可把信道与信源译码器合二而一为一个译码器。

5. 信宿

信宿是消息传送的对象,即接受消息的人、机器或其他事物。

以上介绍的通信系统是点对点、一对一的通信模型,信息传输也是单向的。在 20 世纪

70 年代以后,随着通信技术的发展,如卫星通信、广播通信、无线通信、计算机网络通信的发展,需要从理论上建立适合网络通信的数学模型,并回答在网络通信中信息压缩和可靠传输的理论极限以及实用编译码技术,由此发展起来的网络信息理论或称多用户信息理论中出现了众多的通信模型,涉及信源编码的有相关信源模型和信源多终端模型,涉及信道的有多址信道、广播信道、双向信道、窄扰信道、中继(转播)信道、防窃听信道等.但由于网络通信模型在数学处理上的困难以及实用编译码技术研究的滞后,上述模型还有相当一部分在理论上没有完全解决或没有得到真正的实际应用.

二、信息论研究的主要内容

信息论研究的是关于点对点的通信系统中最根本、最本质的问题.例如:

(1) 什么是信息? 如何度量信息?

(2) 怎样确定信源的输出中含有多少信息量?

(3) 对于一个信道,它传输信息量的最高极限(信道容量)是多少?

(4) 为了能够无失真地传输信源信息,对信源编码时所需的最少的码符号数是多少? 这是无失真信源编码,即香农第一定理的内容.

(5) 在有噪信道中有没有可能以接近信道容量的信息传输率传输信息而错误概率几乎为零? 这是有噪信道编码,即香农第二定理的内容.

(6) 如果对信源编码时允许一定量的失真,所需的最少的码符号数又是多少? 这是限失真信源编码,即香农第三定理的内容.

毫无疑问,如果我们对这些问题都有了确定的答案,那么在设计通信系统时就有了目标和指导方向,同时也有了评价通信系统优劣的标准.

三、信息论的一些应用成果

信息论从它诞生的那时起就吸引了众多领域学者的注意,他们竞相应用信息论的概念和方法去理解和解决本领域中的问题.50 年过去了,这期间虽有失败的教训但也不乏成功的例子,下面我们就举若干成功的范例.

1. 语音信号压缩

语音信号一直是通信网中传输的主要对象.自从通信网数字化以来,压缩语音信号的编码速率就成为通信中的一个重要问题.根据信息理论的分析,语音信号所需的编码速率可以远远低于仅按 Nyquist 采样定理和量化噪声分析所决定的编码速率.几十年来的工作已在这方面取得巨大的进展:长途电话网标准的语音编码速率已从 1972 年 CCITT G. 711 标准中的 64 kbit/s 降低到降低到 1992 年的 16 kbit/s.在移动通信中,1988 年欧洲 GSM 标准中的语言编码速率为 13.2 kbit/s,而 1989 年美国 CTIA 标准中的语音编码速率仅为 7.95 kbit/s.特别是按音素识别与合成原理构造的声码器的速率可低于 100 bit/s,已接近信息论指出的极限.

2. 图像信号压缩

图像信号的信息量特别巨大,这对图像信号的传输及存储都带来极大的不便.经过多年的研究,到 20 世纪 80 年代时图像压缩逐步进入建立标准的阶段.1989 年 CCITT 提出

电视电话/会议电视的压缩标准 H. 261, 其压缩比达到 25:1 到 48:1. 1991 年 CCITT 与 ISO 联合提出的“多灰度静止图像压缩编码”标准 JPEG, 其压缩比 24:1. 对常规电视信号的压缩在 1991 年提出的国际标准 MPEG-1 中其平均压缩比可达 50:1. 这些巨大的进展为图像信号进入通信领域以及多媒体计算机的发展创造了条件. 此外, 随着全数字高清晰度电视技术的发展, 相应的压缩编码研究也已取得进展, 这就是 1993 年提出的 MPEG-2.

3. 计算机文件的压缩

由于数据库的广泛应用, 存储计算机文件所需的存储量问题日益突出. 在过去的 20 多年中对计算机文件的压缩已发展了 20 余种不同的算法. 1988 年澳大利亚举行的压缩算法对抗赛对各种算法进行了一次大检阅. 竞赛所用的文件内容多种多样, 如各种高级语言的源程序、目标码、论文、小说、科学数据、图书目录等, 压缩工作全在当时比较典型的微机上进行. 测试结果表明, 其中较好的算法都能使各种文件压缩后所需的存储量只为原文件的 25% ~ 50%, 其平均值约为 30%, 压缩所需的时间和存储器开销都不大. 目前, 各种压缩算法已经在计算机中得到广泛的应用.

4. 模拟话路中数据传输速率的提高

20 世纪 50 年代初计算机开始在美国联网, 当时模拟话路是传输计算机数据的几乎唯一可用的信道. 最早的调制器其速率只有 300 bit/s, 而信息论指出, 标称带宽为 4 kHz, 信噪比为 25 dB 的话路信道的极限速率应在 25 kbit/s 左右. 所以在以后的 30 多年中就开始了提高速率的长期的、现在看来是极其成功的工作. 1967 年速率为 4 800 bit/s, 1971 年为 9 600 bit/s, 1980 年开始进入 14.4 kbit/s, 1985 年时利用多维网格编码调制的调制器 Codex2680 使速率达到 19.2 kbit/s, 非常接近于理论极限. 信息论在模拟话路数据传输中获得的巨大成功, 其意义远远超出模拟话路本身. 实际上由于模拟话路中所用的调制器可用 VLSI 电路实现, 价格低廉, 因而实际上成了信息论方法的试验场. 各种在话路调制器中获得成功的调制方法都很快被推广应用到载波的高次群信道及数字微波、数字卫星信道中, 他们都同样获得了成功.

5. 降低信息传输所需的功率

在远距离无线通信, 特别是深空通信中如何降低信息传输所需的功率至关重要, 因为在这种情况下发送设备的功率和天线的尺寸都已成为设备生产和使用中的一个困难问题. 幸运的是正是在这个领域信息论获得了它第一批令人信服的成果. 从 20 世纪 60 年代后期起, NASA 发射的所有深空探测器无一例外地在其通信设备中采取了信道编码措施, 因为根据信息理论的分析, 采用低码率的信道编码可以降低传送单位比特所需的能量 E_b 与噪声功率谱密度 N_0 之比. 现在利用不太复杂的信道编码就可以使同样误码率下所需的 E_b/N_0 比不采用信道编码时低 6 dB 左右. 其中一些好的方案(如用 RS 码作为外码、卷积码作为内码的方案)可以使误码率在 10^{-5} 的情况下所需的 E_b/N_0 降到 0.2 dB, 比不用信道编码时所需的 10.5 dB 降低了近 10 dB.

6. 计算机网中数据传输可靠性的保证

随着计算机技术的发展, 计算机设备的布局变得愈来愈分散, 各种终端及外围设备离主机也越来越远, 这就产生了计算机网. 近年来, 计算机网还与分布式计算机系统相联系而变得更为重要. 在用各种电缆连接而成的计算机网中电噪声和各种外界的电磁干扰是

不能不考虑的,因为它使传输的信息发生差错.一般情况下,局域网中的差错率在 10^{-8} 左右,广域网中的差错率在 $10^{-3} \sim 10^{-5}$. 这样高的差错率在实际应用中是无法接受的.

目前普遍采用的解决办法是带自动重发请求的差错检测码. 差错检测的方法从最简单的奇偶检验到比较复杂的循环冗余检验都被采用,但较大的网一般都用循环冗余检验. 这种方法已被各种网络通信协议采用并成为标准. 例如 ISO 制定的高级数据链路协议 HDLC 就采用 CCITT V. 41 的 CRC 码进行循环冗余检验. HDLC 在全世界已被广泛采用,这一标准有很广的应用领域,许多协议都是从它派生出来的.

7. 计算机的容错问题

计算机的存储器是计算机的重要组成部分. 随着计算机技术的发展无论是内存或外存其存取速度以及存储密度等都在不断提高,如何保证存取的正确性已成为越来越突出的问题. 特别是在外存中,由于存储量大,而且存储体的表面难免有缺损,要保证绝对不发生差错已不可能,现在广泛采取的解决办法是增加适当的检错纠错装置. 例如 IBM4300、Cray - 1 等大型机的内存都有较简单的检错纠错措施. 而在外存中,根据不同的情况和要求从较简单的 Fire 码到 BCH 码以及 RS 码等都被广泛使用,例如在 IBM360 的光盘存储器中就采用了 RS 码.

8. 图像信号的复原与重建

图像的复原与重建是图像信号处理的一个重要内容,在实用中有很大的价值. 20 世纪 80 年代以来,最大熵方法在图像复原与重建中取得了很大的成功. 在退化图像复原中,图像退化的原因是多种多样的,如由于景物的运动、光学系统的不理想、噪声等. 图像重构的形式也很多,如计算机层析图像、合成孔径射电望远镜图像、结晶学研究中用的光学干涉仪或无线电干涉仪的图像、核磁共振波谱仪图像等. 在这些应用中最大熵方法较其他方法优越的主要原因是其合理性,即所得结果是我们可以而且能够期望的最好结果. 同时也有一些派生的好处,如在盲解卷时同时给出卷积函数,在重建图像中可以同时对仪器中的某些参数进行校正等.

虽然最大熵法在这些应用中目前还不能给出性能的解析表达式,但算法已比较成熟,如常用的剑桥算法等.

9. 信息安全

信息在通信过程中的安全问题也是通信理论需研究的重要问题,香农于 1949 年发表了《保密通信的信息理论》,在该论文中他为密码学确立了一系列的基本原则与指标,如加密运算中的完全性、冗余度等指标,这些与信息的度量密切相关. 因此人们常把通信中的编码与密码作为通信工程中的两个不同侧面. 随着网络通信的安全管理要求,各种不同类型的加密体制与协议使近代密码学发展成一个独立的学科分支,但是它仍与信息论的基本理论、编码算法有着密切的关系.

信息论方法的应用领域相当广泛,有时甚至出乎我们的预料. 在语言学、生物学、医学方面早已有大量论文甚至专著讨论信息论方法的应用;在经济学方面,美国经济学家 K. J. Arrow 因“信息经济学”的研究成果而获得诺贝尔经济学奖;美国 Stanford 大学的 T. Cover 首先把信息度量的一些性质引入金融工程的组合投资决策中,如最优策略的递推计算、投资收益的遍历性、边信息在投资策略中的应用等问题. 至于其他领域的种种应