

信息编码技术 及其应用大全

黄新亚 米央 编著

周爱民 审校



电子工业出版社

信息编码技术及其应用大全

黄新亚 米 央 编著

周爱民 审校

电子工业出版社

(京)新登字 055 号 内容简介

本书在信息编码理论的基础上,介绍了信息编码技术,其中包括等长码和变长码、汉字编码、信息分类编码、检错和纠错码(冗余编码)和信息压缩编码(去冗余编码)。最后两章介绍了磁卡编码和图形码(条码和矩阵码)的编码技术以及它们的应用。

本书适合于从事数据通信、计算机应用和信息工程等方面的工程技术人员使用或自学参考,也可作为大专院校这类专业学生的参考书。

信息编码技术及其应用大全

著者 黄新亚 米央

责任编辑 林木森

*

电子工业出版社出版

北京市海淀区万寿路 173 信箱 (100036)

电子工业出版社发行 各地新华书店经销

河北三河华燕印刷厂印刷

*

开本: 787×1092 毫米 1/16 印张: 6 125 字数: 112 千字

1994 年 8 月第一版 1994 年 8 月第一次印刷

印数: 10000 册 定价: 10 元

ISBN7-5053-2376-8/TP·678

前 言

四十年代末,美国 C. E. Shannon 的开创性文章——通信的数学理论,为信息编码技术奠定了理论基础。数据通信和计算机技术的发展离不开信息编码,而它们的相互结合又反过来为信息编码技术的发展和应 用创造了有利的环境。随着社会信息化的不断深入,信息编码技术已经渗透到许多应用领域,并且正在扩大。

本书选材时充分考虑其实用性。各章内容安排既相对独立,又互有联系,以便各取所需。叙述重概念,少理论推导。注意引用国际上近年来有关这方面的新成果。

本书共设八章。第一章简单介绍了信息编码的理论基础。第二章以计算机代码体系为背景,描述了等长码的编码原理;在变长码中,重点介绍了 Huffman 最佳编码原理。第三章在比较全面地介绍了汉字编码的基础上,引述了中、日、韩汉字统一编码的最新成果。第四章为信息分类编码。第五章和第六章,安排了检、纠错码(冗余编码)和信息压缩编码(去冗余编码)的内容。最后两章,描述了磁卡编码和图形码(条码和矩阵码)的技术以及它们的应用。

本书编写过程中,得到室里同志的广泛支持。张怡清高工提供了许多有引用价值的资料和数据。王琨、成十周同志承担了图表的设计和制作。在此,向他们表示深切的感谢。

中国电子学会《电子世界》编辑部周爱民同志对全书进行了仔细审阅,并提出了许多宝贵意见,在此特予致谢。

书中一些内容取材于公开发表的文献资料中,未能一一征询,深表歉意。

由于时间仓促,水平有限,书中难免有欠妥和错误之处,敬请读者指教,不胜感谢。

编著者

目 录

第一章 信息度量	(1)
§ 1. 1 何谓信息	(1)
1. 1. 1 信息在计算机中的含义	(1)
1. 1. 2 信息在通信系统中的含义	(1)
§ 1. 2 信息的特征和作用	(3)
1. 2. 1 信息的特征	(3)
1. 2. 2 信息的作用	(5)
§ 1. 3 信息度量	(5)
1. 3. 1 信息量的定义	(6)
1. 3. 2 信息量的可加性	(7)
§ 1. 4 信息熵	(8)
1. 4. 1 熵的表示形式	(8)
1. 4. 2 熵的基本性质	(9)
第二章 信息编码	(13)
§ 2. 1 有关信息编码的术语	(13)
2. 1. 1 信源	(13)
2. 1. 2 信源编码	(14)
2. 1. 3 编码空间	(14)
§ 2. 2 计算机代码	(14)
2. 2. 1 ASCII 代码	(15)
2. 2. 2 EBCDIC 代码	(19)
2. 2. 3 ISO 646 信息代码字符集	(20)
2. 2. 4 ISO 2022 信息代码字符集	(21)
2. 2. 5 ISO/IEC 10646 通用多八位编码字符集	(22)
§ 2. 3 变长码	(24)
2. 3. 1 信息冗余度和编码效率	(24)

2. 3. 2	Morse 编码	(24)
2. 3. 3	唯一可译码	(26)
2. 3. 4	Shannon 编码	(27)
2. 3. 5	Huffman 编码	(29)
2. 3. 6	Huffman 编码与 Shannon 编码的比较	(33)
2. 3. 7	改进 Huffman 编码	(34)
2. 3. 8	黑白游程编码	(34)
2. 3. 9	双字长编码	(35)
第三章	汉字编码	(37)
§ 3. 1	汉字编码的结构模型	(37)
§ 3. 2	外部码	(38)
3. 2. 1	汉字的特点	(38)
3. 2. 2	汉字编码原则	(39)
3. 2. 3	汉字编码的熵	(40)
3. 2. 4	汉字编码效率	(40)
3. 2. 5	外部码的形式	(42)
3. 2. 6	汉字编码方案举例	(42)
3. 2. 7	汉字输入代码的转换	(45)
3. 2. 8	汉字编码前景	(46)
§ 3. 3	交换码	(47)
3. 3. 1	区位码字符	(48)
3. 3. 2	汉字的排列	(49)
§ 3. 4	内部码	(49)
3. 4. 1	双字节汉字内部码	(50)
3. 4. 2	三字节汉字内部码	(50)
3. 4. 3	加引导符方式	(51)
§ 3. 5	多文种编码体系	(51)
§ 3. 6	统一的中日韩汉字编码	(52)

3. 6. 1	基本多文种平面	(52)
3. 6. 2	统一编码的原则和步骤	(53)
3. 6. 3	UCS 对中文信息处理的影响	(58)
第四章	信息分类编码	(60)
§ 4. 1	信息代码的特点	(60)
§ 4. 2	信息分类原则	(61)
§ 4. 3	信息分类方法	(62)
4. 3. 1	线分类法	(62)
4. 3. 2	面分类法	(64)
§ 4. 4	信息代码设计	(65)
4. 4. 1	代码编制基本原则	(65)
4. 4. 2	代码的种类	(66)
4. 4. 3	代码的设计	(69)
§ 4. 5	代码标准化	(70)
4. 5. 1	标准化的作用	(71)
4. 5. 2	标准化的特点	(71)
§ 4. 6	信息代码的应用	(71)
4. 6. 1	居民身份证的编号	(71)
4. 6. 2	中国公民出境理由分类编码	(74)
第五章	检错码与纠错码	(76)
§ 5. 1	信息检错与纠错的必要性	(76)
§ 5. 2	检错编码	(77)
5. 2. 1	奇偶校验	(77)
5. 2. 2	突发性检错码	(78)
5. 2. 3	信息代码校验码	(79)
5. 2. 4	条码校验	(82)
5. 2. 5	信用卡帐号校验码的设置	(82)
§ 5. 3	纠错码	(84)

5. 3. 1	Hamming (7, 4) 纠错编码	(84)
5. 3. 2	Hamming 距离	(86)
5. 3. 3	利用 Hamming 距离纠错	(87)
5. 3. 4	循环码	(89)
5. 3. 5	循环 Hamming 纠错码	(90)
第六章 信息压缩编码		(93)
§ 6. 1	字符信息压缩	(93)
6. 1. 1	ASCII 码数字字符的二进制表示	(94)
6. 1. 2	信息的代码化表示	(94)
6. 1. 3	空格 (Null) 压缩编码	(95)
6. 1. 4	位映象 (Bit Mapping) 压缩编码	(96)
6. 1. 5	游程长度编码 (Run Lenth Code) 压缩	(96)
6. 1. 6	Huffman 编码用于信息压缩	(97)
6. 1. 7	前端压缩编码技术	(98)
§ 6. 2	图形信息压缩	(98)
6. 2. 1	点阵汉字的信息压缩	(98)
6. 2. 2	矢量汉字的信息压缩	(100)
6. 2. 3	线条图形信息压缩	(102)
第七章 磁卡信息编码		(108)
§ 7. 1	磁卡的几何图形	(108)
§ 7. 2	磁道信息代码	(110)
7. 2. 1	第一磁道信息编码	(110)
7. 2. 2	第二磁道信息编码	(111)
7. 2. 3	第三磁道信息编码	(112)
7. 2. 4	磁道编码信息的有关特性	(112)
§ 7. 3	磁卡应用实例	(113)
7. 3. 1	信用卡	(113)
7. 3. 2	储值卡	(117)

7. 3. 3	金融交易卡	(119)
7. 3. 4	进出口控制	(124)
7. 3. 5	护照	(124)
第八章	图形码	(129)
§ 8. 1	条码的优点	(129)
§ 8. 2	条码术语	(131)
§ 8. 3	条码编码和结构类型	(133)
8. 3. 1	编码方法	(133)
8. 3. 2	编码容量	(133)
8. 3. 3	条码类型	(134)
§ 8. 4	条码举例	(134)
8. 4. 1	交替 25 条码	(134)
8. 4. 2	通用商品条码	(137)
8. 4. 3	39 条码	(145)
8. 4. 4	CODABAR 条码	(149)
8. 4. 5	128 条码	(152)
§ 8. 5	其他类型条码	(158)
8. 5. 1	二维条码	(158)
8. 5. 2	单宽条码	(159)
§ 8. 6	条码信息阅读	(162)
§ 8. 7	条码的应用	(163)
8. 7. 1	商店条码信息管理系统	(163)
8. 7. 2	条码在邮电系统的应用	(165)
§ 8. 8	矩阵码	(166)
8. 8. 1	Vericode 码	(166)
8. 8. 2	Data Matrix 码	(169)
8. 8. 3	Phillips Dot 码	(170)
8. 8. 4	Array Tag 码	(171)

8. 8. 5 矩阵图形码的阅读.....	(172)
附录一 改进 Huffman 编码表	(174)
附录二 信息代码校验码标准系统	(178)
F2. 1 校验码标准系统	(178)
2. 1. 1 校验码标准系统的内容.....	(178)
2. 1. 2 校验码标准系统的表示.....	(179)
2. 1. 3 校验码标准系统代码字符的值.....	(179)
2. 1. 4 校验码标准系统中纯系统的计算方法.....	(181)
F2. 2 校验码标准系统选择原则	(183)

第一章 信息度量

信息是人类社会活动所产生的各种状态和消息的总称。在信息学还没有形成一门科学之前,只限于对信息进行定性描述。自从 Shannon 于 1948 年发表《通信的数学理论》一书后,才建立起对信息定量描述的理论。本章主要介绍信息的定义、信息的度量和信息熵的概念,它们是信息编码的理论基础。

§ 1. 1 何谓信息

所谓信息,目前还没有一个确切的统一定义,而只能从某个研究领域去说明其内含和外延。从信息的分类编码来说,我们可以把具有一定含义的事物或概念称之为信息。下面就信息在计算机和通信系统中的含义进行说明。

1. 1. 1 计算机中的信息含义

在计算机系统中,把信息定义为数据、消息中所包含的内容。在这种意义下,信息是人们进行运算和处理所需的条件、内容和结果,并常常表现为数字、数据、图表和曲线等形式。

在用作计算的场合,需要根据计算的方法和计算的条件(构成原始信息),编制出计算程序,经过运算处理,得到相应的解答,并以数字或图表(输出信息)的形式提供给用户。

作为控制用计算机,首先要将理想的控制目标和被控对象的实际原始状态(输入信息)送给计算机,经过处理得到相应的输出信息,并由它去控制被控对象。然后,将理想控制目标与实际控制误差作为补充信息(效果)反馈回去,修正和调整相应的控制程序,直到控制误差达到允许的范围。

1. 1. 2 通信系统中的信息含义

对于一个通信系统来说,信息要经过发送、传输和接收三个步骤,由此引出下面的定义:

(1) 从产生信息的客观对象来定义

“信息是客观世界各种事物变化和特征的反映”，“信息是客观世界中各种事物运动状态和方式的表征”，“信息是被反映事物的属性”，“信息是事物的内容、形式及其发展变化的反映”。这类定义说明客观世界是“形”，信息就是反映它的“影”。“形”与“影”之间，既互相依存，又互相制约。

一份电报，首先得拟出报文才能发送，报文内容即“形”，表征报文内容的通信码型则为“影”。

(2) 从接收信息的认识主体来定义

Shannon 在 1948 年发表的《通信的数学理论》一书中，指出一个系统所接收的“信息是能够用来消除不确定性的东西”。与此相类似，“信息就是收信者事先所不知道的内容”。这类定义说明认识是由于得到信息而增进了对客观事物的了解。

收报方得到了报文，方知报文的内容。

(3) 从传输中信息所依附的载体来定义

日常，人们通常把信息视作信号、数据、资料、情报、消息、新闻、知识等的总称。客观对象对认识的主体作用，或认识主体对客观对象的反映，通过传输的信息来表现，而信息常常包含在各种信号及其组合之中。信息包含语言、文字、电波等，而数据、资料等实际上就是复合信号。这种定义说明信息作为内容有其外在的表现形式。

可见，信息的任何表示形式，都离不开其所依附的载体。电报的发送、传输和接收，离不开纸张和电波载体。

(4) 从信息的发送、传输、接收的统一中，客观和主体之间的相互作用来定义

早期信息论从通信理论的角度认为信息是信源、信道、信宿之间的联系中生成的东西；控制论奠基人 N. Wiener 在《人有人的用处》一书中认为“信息这个名称的内容，就是我们对外

界进行调节，并使我们的调节为外界所了解时而与外界交换来的东西”。“信息就是变异度”，“信息就是差异”。这类定义说明信息作为作用与反作用的过程存在于事物反映特性的相互联系之中。

载有电报信息电波的波形是随时间变化的，未调制电波是不携带任何信息的。

信息定义的多样性，有三方面的原因：信息本身的复杂性；目前信息科学尚不发达；实际需要的不同。人们出于不同的研究目的，从不同的角度出发，因而对信息作出不同的理解和解释，这是非常自然的。

§ 1. 2 信息的特征与作用

理解了信息的含义，我们就可以对信息的特征进行考察，进而说明信息的作用。

1. 2. 1 信息的特征

(1) 客观性

信息不是无本之木，无源之水，它是客观事物的反映，反映是以客观为前提的。即使主观信息，如决策、指令、计划等，也是以其客观内容为依托的。以客观信息为基础并接受客观实际的检验。

(2) 普遍性

信息是无所不在的，客观物质的普遍性，决定了信息的普遍存在。

可以说，只要对某一事物有所认识，就可以对其进行描述，而描述本身则离不开以信息来表征。

(3) 无限性

物质世界是无穷尽的，因此表征物质的信息也是无穷尽的。当代信息正以爆炸性的速度增长，就充分说明了这一点。

• (4) 动态性

信息不是一成不变的，它是随时间而变化，需要不断更新，才能得到有用的信息。每天每时金融股市行情的变化，就充分说明信息的动态变化趋势。

(5) 依附性

信息依附于载体而存在，它自身不能独立存在和交流，而载体可以因不同需要而变换。

新闻报纸的文字载体是纸张，程序软件的信息载体是磁带、磁盘或内存储器。

(6) 计量性

信息是可以度量的，这便是在本章中将要重点介绍的内容。

(7) 变换性

信息通过处理或加工，可以使其内容和形式发生变化，以适应某个方面的需要。

信息虽然不能离开客观对象而凭空产生，但是它可以通过处理，变换其表现形式。例如一种产品的月增长数字信息，为了直观形象起见，可用直方图表示。

信息的产生离不开物质，信息的产生需要能量。有效地使用信息可使信息转化为物质与能量，还可以减少时间的开销。在当今材料、能源和信息三大科学领域中的相互关系中，充分体现了这三者相互依存的关系。计算机技术的发展，在对它们的相互之间转换过程中，起到了十分巨大的变革作用。

(8) 传递性

信息在时间上的传递就是记忆或存储，空间的传递便是扩散。

人们通过交谈，以达到思维信息交流的目的。有线（电缆或光缆）无线广播和通信，是当今信息传递和扩散的主要形式。

(9) 系统性

信息是一种集合，各种信息在相互联系中形成统一的体系，

信息系统是物质世界系统的再现。

信息系统是信息“有序化”的过程，这正是信息处理的基础和依据。

1. 2. 2 信息的作用

由于信息具有上述诸特征，因此，它在人类认识和改造世界中起着十分重大的作用，这表现在：

(1) 认识的依据

认识世界，首先要获得客观世界的有关信息，对它们进行分析，从中得出正确的结论。

(2) 实践的指南

改造世界，要依靠决策，对外部环境和内部活动进行规划和调节，对未来发展和变化进行预测，这都需要借助于信息。

(3) 实现有序的保证

信息是系统组织化的重要因素，它作为“黏合剂”，“联系纽带”，“神经中枢”，能使社会和经济机体协调地发展。

(4) 开发资源的条件

物质提供材料，能源提供动力，这是人类发展的有形资源，或称第一资源。信息提供智力，为人类社会的发展开发了无形的第二资源。

(5) 开发智慧的源泉

智慧是知识的结晶和运用，它是社会发展的强大动力，而信息的积累和升华，是开发智慧的源泉。智慧是以信息的增值和知识的创新为基础的，这在科技蓬勃发展的今天，显得尤为重要。

§ 1. 3 信息度量

在信息论没有作为一门科学建立起来之前，信息的度量是一个长期没有得到很好解决的问题。自然科学研究的对象，通常都要作定量描述。定性描述只能解释一些现象和表面的东西，定

量描述才有可能解释事物的本质。由于信息必须适应客观世界各类实体的统一描述,因此,信息的度量就不可能以量纲的形式出现,而必须跟世界万物普遍存在的规律性的东西相联系。下面请看两个例子。

例 1: 北京地区天气预报

- (a) 7 月份“ 明天有雨”;
- (b) 12 月份“ 明天有雪”。

例 2: 新闻报导

- (a) 澳星发射成功;
- (b) 一煤气站投入使用。

现在观察一下上述两个例子中,哪个信息量大些?

显然,例 1 中的 (b) 信息量要比 (a) 的大,这是因为 7 月份北京正值雨季,而 12 月份很少下雪。第 2 例中,澳星上天属重大新闻,煤气站投入使用是经常发生的。由此可以得出结论,信息量的大小是跟事物发生的概率紧密相关的。事物发生的概率正是事物存在的普遍性的东西。也就是说,出现概率小的事件包含的信息量大。因此,信息量应该是概率的单调减函数。

1. 3. 1 信息量的定义

由于信息量是概率的单调减函数,所以很自然地把信息量定义为:

$$-\log p \quad (1. 1)$$

式中, p 为某事物发生的概率,对数以 2 为底(以下凡不注明的,均以 2 为底),这是有其实际意义的。例如,电位的高低,脉冲的有和无,信号灯的明和灭,都是两种状态。目前的数字电子计算机(以下简称计算机),都是以二电平逻辑来工作的。以 2 为底的信息量单位称为“位”(bit),“位”是信息度量的最基本单位。

一个独立事件的信息量可以这样来描述。戈壁滩的天气预

报“明天有雨”，假设在一年中只有 $1/64$ 的下雨天，根据 (1. 1) 式，则“明天下雨”的信息量为：

$$-\log 1/128 = 7 \text{ (位)}$$

可见，戈壁滩难得下雨，概率很小，故信息量大。”明天是晴天”的信息量为

$$-\log 127/128 \approx 0 \text{ (位)}$$

经常是晴天，概率大，信息量几乎小到零。经常出现的，司空见怪，听不听预报，八九不离十，差不多都是一个样。只有偶发事件，才包涵丰富的信息。这便是信息量为概率的单调减函数的充分例证。

1. 3. 2 信息量的可加性

信息量运算中的一个重要特性就是它的可加性。现举例说明之。

设有士兵列队方阵为 9 排 9 列，士兵甲站在第 7 排第 5 列，根据 (1. 1) 式，有

$$\text{信息量} = -\log p$$

式中 p 是信息集合中某事件发生的概率。这里，士兵列队方阵中共有 81 个士兵站位，指定其中一位士兵的概率为 $1/81$ 。所以“士兵甲所在方阵位置”的信息量为：

$$-\log (1/81) = 6.34 \text{ (位)}$$

指定第 9 排中的某一排的概率为 $1/9$ ，故有“士兵甲在第 7 排”的信息量为：

$$-\log (1/9) = 3.17 \text{ (位)}$$

指定 9 列中的一个站位的概率为 $1/9$ ，因此“士兵甲在第 5 列”的信息量为：

$$-\log (1/9) = 3.17 \text{ (位)}$$

因此有： $\log (1/81) = \log (1/9) + \log (1/9) = \log (1/81) = 6.34 \text{ (位)}$