高等院校“十一五”规划教材

信息技术基础

主　编　齐浩亮　郑晓霞

中国水利水电出版社
www.waterpub.com.cn

高等院校“十一五”规划教材

信息技术基础

主　编　齐浩亮　郑晓霞

中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

本书全面、系统地介绍了信息与熵、信息采集、存储技术、信息压缩与校验、信息检索与利用、信息发布、信息安全、熵的理论应用等内容，采用循序渐进的方式描述所有概念，并以结构清晰的图示和例子以及严谨的证明满足不同层次的学生的需求。

本书注重理论与实际相结合，强调实用性和实践性，以适应现代信息技术理论体系不断更新和发展的要求，使学生能够紧跟信息技术的快速变化趋势。全书内容丰富、结构完整，讲述过程循序渐进、深入浅出，理论论述精辟，重点突出。

本书可作为普通高等学校计算机、信息管理与信息系统等相关专业的教材，也可以作为计算机应用人员的培训教材或参考书。

本书配有电子教案，读者可以从万水书苑或中国水利水电出版社网站下载，网址为：http://www.wsbookshow.com，http://www.waterpub.com.cn/softdown/。

图书在版编目（CIP）数据

信息技术基础 / 齐浩亮，郑晓霞主编. －北京：中国水利水电出版社，2009

高等院校“十一五”规划教材

ISBN 978-7-5084-6363-6

Ⅰ. 信… Ⅱ. ①齐…②郑… Ⅲ. 电子计算机－高等学校－教材 Ⅳ. TP3

中国版本图书馆 CIP 数据核字（2009）第 040527 号

策划编辑：石永峰　责任编辑：李　炎　加工编辑：徐　雯　封面设计：李　佳

书　　名	高等院校“十一五”规划教材 **信息技术基础**
作　　者	主　编　齐浩亮　郑晓霞
出版　发行	中国水利水电出版社 （北京市海淀区玉渊潭南路 1 号 D 座　100038） 网址：www.waterpub.com.cn E-mail：mchannel@263.net（万水） sales@waterpub.com.cn 电话：（010）68367658（营销中心）、82562819（万水）
经　　售	全国各地新华书店和相关出版物销售网点
排　　版	北京万水电子信息有限公司
印　　刷	北京蓝空印刷厂
规　　格	184mm×260mm　16 开本　15 印张　365 千字
版　　次	2009 年 4 月第 1 版　2009 年 4 月第 1 次印刷
印　　数	0001—3000 册
定　　价	25.00 元

前　言

随着计算机软、硬件技术和网络通信技术的进步与发展，计算机技术和信息技术在各行各业得到了广泛应用，并且已渗透到人们的日常生活中。由于其应用的广泛性，我们强调信息技术课程要与其他学科的教学活动整合在一起，要“淡化”计算机（信息技术）课程的“学科性”，强调其“综合性”和“应用性”。

为了适应我国高等教育发展的新形势，深化教学改革，提高教学质量，符合应用型本科教学的总体思路，进行信息技术教育的主要目的是信息技术的应用，而不是为了学习而学习，是以信息技术为工具，处理教师和学生在各学科教与学过程中的应用，并让学生掌握基本的信息处理能力，使学生做到学以致用。

信息技术是高等院校计算机专业和信息管理专业学生所学的专业基础课程。随着信息技术应用的普及，计算机教育也不断地深化，教学要求也从计算机一般操作过渡到了信息技术基础知识的应用。

本书的作者长期工作在教学一线，具有丰富的教学经验，并不断进行教学改革，从教学方法、教学手段、教学内容到实践教学，不断创新、发展。课题组成员首先在培养学生兴趣上下功夫，把握学生的兴奋点，有效调动学生的求知欲望。其次，把学习的主动权交给学生，让学生在亲身实践中品尝艰辛和乐趣。

本书的作者具有良好的科研经历，能够在教材中体现信息技术的最新进展，使学生能够紧跟信息技术的快速变化。本书将从基础理论、基础应用着手，每个环节节节相扣，便于学生的理解与学习。

全书分为 8 章，内容包括信息与熵、信息采集、存储技术、信息压缩与校验、信息检索与利用、信息发布、信息安全、熵的理论应用。本书采用循序渐进的方式描述所有概念，并以结构清晰的图示和例子以及严谨的证明满足不同层次的学生的需求，并在课程中以实例生动地解释重要概念，增强了课程的生动性。

本书由齐浩亮、郑晓霞任主编。其中第 2 章和第 8 章由齐浩亮编写，第 3 章和第 7 章由郑晓霞编写，第 1 章和第 5 章由孔蕾蕾编写，第 6 章由孙栩编写，第 4 章由邹钰编写。

在本书的编写过程中，参考了很多同行的著作，在此深表感谢！

由于作者水平有限，书中错误和不当之处在所难免，希望广大读者批评指正。

编　者

2009 年 3 月

目　录

第 1 章　信息与熵

1.1　信息

1.1.1　信息的一般概念

1. 什么是信息

信息是最古老的概念之一，与人们的生活长期相关，是人们相互交流的工具。历史上的结绳记事就是在文字创造之前，古代人们传递信息最早的形式之一。语言的产生是人类文明发展的主要标志，语言是一种信息，而且是复杂的信息。古代的遗迹也为我们研究当时社会的文化和经济提供了可靠的信息。可以说，人类是通过了解自然信息来了解自然，通过了解社会信息来了解人类社会的过去，人们的生活离不开信息。

人类的社会生活是不能离开信息的，人类的社会实践活动不仅需要对周围世界的情况有所了解，做出正确的反应，而且还要与周围的人群沟通，才能协调地行动，这就是说，人类不仅时刻需要从自然界获得信息，而且人与人之间也需要进行沟通，交流信息。人类需要随时获取、传递、加工、利用信息，否则就不能生存。人们获得信息的方式有两种：一种是直接的，利用自己的感觉器官，通过耳闻、目睹、鼻嗅、口尝、体触等直接了解外界情况；一种是间接的，即通过语言、文字、信号等工具传递消息，从而获得信息。通信是人与人之间交流信息的手段，语言是人类通信的最简单、最基础的要素。人类早期只是用语言和手势直接进行通信，交流信息。

虽然人们长期与信息打交道，但人们只关心具体的、单独的信息，特别是只关心信息的内容，以此进行相关的工作。人们每天都在制造信息，但从来不从整体上关心信息，也未能从更基础的方面认识信息。

社会的发展使人们与信息的关系越来越密切，对信息的认识也越来越深入、越来越全面。电磁波的发现使复杂信息的传递可以通过更简单的工具来实现，这就为研究信息最简单的形式之一——信号的传递性质提供了可能。

信息在各个领域被广泛使用，不同学科领域的信息研究者也都在试图从自己学科的角度来研究信息，阐述信息的概念。50 多年来，虽然信息科学——将信息作为主要研究对象的独立学科得到了前所未有的发展，但是对信息的定义仍是众说纷纭，未能达成共识。一个能在各个学科通用的信息定义仍在探讨之中。

1928 年，哈特莱（L.R.V.Hartley）在《贝尔系统电话》杂志上发表了一篇题为“信息传输（Transmission of Information）”的论文，区分了消息和信息。他认为“信息是指有新内容、新知识的消息”，将信息理解为选择通信符号的方式，并用选择的自由度来计算这种信息的大小。

1975 年，意大利学者郎高（G.Longo）出版了专著《信息论：新的趋势与未决问题》，并在序言中指出“信息是反映事物的形成、关系和差别的东西，它包含在事物的差异之中，而

不是在事物本身”。

1996 年，中国学者钟义信在《信息科学原理》中详尽阐述了信息的概念。他指出，在信息概念的诸多层次中，最重要的有两个层次：一个是没有任何约束条件的本体论层次；另一个是受主体约束的认识论层次。从本体论的层次上考察，信息可被定义为“事物运动的状态以及它的状态改变的方式”。在此，“事物”泛指一切可能的研究对象，包括外部世界的物质客体和主观世界的精神现象；“运动”泛指一切意义上的变化，包括机械运动、物理运动、化学运动、生物运动、思维运动和社会运动等；“运动方式”是指事物运动在时间上所呈现的过程和规律；“运动状态”则是事物运动在空间上所展示的性状与态势。由于宇宙间的一切事物都在运动，都有一定的运动状态和状态改变的方式，因而一切事物都在产生信息。

从认识论的角度考察，信息是主体所感知或者主体所描述的事物的运动状态及其状态变化的方式。认识论层次的信息概念涵盖以下内容：

（1）语法信息。由于主体具有观察力，能够感知事物运动状态及其变化方式的外在形式，由此获得的信息可称为语法信息。

（2）语义信息。出于主体具有理解力，能够领悟事物运动状态及其变化方式的逻辑含义，由此获得的信息可称为语义信息。

（3）语用信息。由于主体具有明确的目的性，能够判断事物运动状态及其变化方式的信息可称为语用信息。

语法信息、语义信息和语用信息三者综合在一起构成认识论层次上的全部信息，即全信息。

2. 信息的类型

信息分类有许多不同的准则和方法，根据不同的分类方法，信息可分为不同的类型。常见的分类方法有：

（1）以信息的性质为依据，信息可以分为语法信息、语义信息、语用信息。

（2）以认识主体为依据，信息可以分为客观信息（关于认识对象的信息）、主观信息（经过认识主体思维加工的信息）。

（3）以观察的过程为依据，信息可以分为实在信息、先验信息、实得信息。

（4）以信息的载体性质为依据，信息可以分为电子信息、光学信息、生物信息等。

（5）以信息的逻辑意义为依据，信息可以分为真实信息、虚假信息、不定信息。

（6）以信息的应用部门为依据，信息可以分为工业信息、农业信息、军事信息、政治信息、科技信息、文化信息、经济信息等。

（7）以信息源的性质为依据，信息可以分为语音信息、图像信息、文字信息、数据信息、计算信息等。

（8）以信息的运动状态为依据，信息可以分为连续信息、离散信息、半连续信息等。

（9）以信息的生成领域为依据，信息可以分为宇宙信息、自然信息、社会信息、思维信息等。

（10）以信息的价值为依据，信息可以分为有害信息、无害信息。

还有许多不同的分类原则和方法，这里就不再一一列举。

广义信息论把信息定义为物质在相互作用中表征外部情况的一种普遍属性，它是一种物质系统的特性以一定形式在另一种物质系统中的再现。信息概念具有普遍意义，它已经广泛地渗透到各个领域；信息科学是具有方法论性质的一门科学；信息方法具有普适性。

所谓信息方法就是运用信息观点，把事物看做是一个信息流动的系统，通过对信息流程的分析和处理，达到认识事物复杂运动规律的一种科学方法。它的特点是撇开对象的具体运动形态，把它作为一个信息流动过程加以分析。信息方法着眼于信息，揭露了事物之间普遍存在的信息联系，对过去难于理解的现象从信息观点作出了科学的说明。信息论为控制论、自动化技术和现代通信技术奠定了理论基础，为研究大脑结构、遗传密码、生命系统和神经病理象开辟了新的途径，为管理的科学化和决策的科学化提供了思想武器。信息方法是认识论在当代以电子计算机和现代通信技术为中心的新技术革命中的浪潮，是认识论的研究和发展，将进一步提高人类认识与改造自然界的能力。

3. 信息的特征

（1）普遍性。信息是事物具有的一种属性，广泛存在于自然界、人类社会和人们的思维领域中，与客观事物一样，无所不在，无时不在。只要有物质及其运动的存在，就会有信息的产生。物质及运动的普遍性，决定了信息存在的普遍性。

（2）寄载性。信息的产生、存储以及传递必须依附一定的物质载体。当信息依附于一定的载体后，才能被保存、接收和运用。信息不能独立存在和交流，只能借助磁介质、电磁波、纸张等不同的载体，采用、文字、符号、图像、代码等不同的表现方式来呈现。相同的信息可以负载在不同的物质载体上，同一物质载体也可以承载不同类型的信息。

（3）可传递性。信息可以在时间和空间上从一点转移到另一点。在时间上的转移称为存储；在空间中的转移称为通信。信息也可以从一种形态转换成另一种形态，在各种物质和能量形式之间转换。信息总是处于一定的传递过程中，并且在传递中发挥其价值。可传递性使信息能够积累和传播。

（4）可识别性。信息是事物本质特征和运动规律的反映，是具体的、真实的、可知的，人类可以通过自己的感觉器官来感知信息、接收信息、利用信息，信息可以被收集、加工、整理、归纳、综合、记忆。信息的可识别性是人类能够认识客观世界的基础。

（5）可共享性。信息不同于物质和能量资源，信息可以被复制、传播或分配给众多的用户。同一内容的信息可以在同一时刻、同一地域被两个以上的使用者分享，并且在共享的过程中，信息不仅不会损失或减少，而且还会出现信息内容的扩增。

（6）时效性。信息的时效性是指信息经过接收、加工、整理、传递、利用的时间间隔及其效率。信息的内容随着时间的变化而不断地变化更新。失去了时效的信息，不能反映变化中事物的新的运动状态和方式的信息，虽然其信息载体依然存在，但其内容已经老化，效用就会降低，甚至失去价值。

信息除了上述的基本特征以外，还有客观性、可度量性、层次性、累积性等特征，这些特征共同构成了信息的复杂性。

1.1.2　香农信息定义

香农（C.E.Shannon）是现代通信理论——信息论的创始人，也是一位影响人类社会进程的科学家。

香农在美国密执安大学和麻省理工学院学习时，修过布尔代数课，并在布尔的指导下使用了微分分析仪，这使他对继电器电路的分析产生兴趣。他认为这些电路的设计可用符号逻辑来实现，并意识到分析继电器的有效数学工具正是布尔代数。

1938年，香农发表了著名的论文《继电器和开关电路的符号分析》，首次用布尔代数进行开关电路分析，并证明布尔代数的逻辑运算可以通过继电器电路来实现，明确地给出了实现加、减、乘、除等运算的电子电路的设计方法。这篇论文成为了开关电路理论的开端。

香农在贝尔实验室的工作中进一步证明，可以采用能实现布尔代数运算的继电器或电子元件来制造计算机，香农的理论还为计算机具有逻辑功能奠定了基础，从而使电子计算机既能用于数值计算，又具有各种非数值应用功能，使得以后的计算机几乎在任何领域中都得到了广泛的应用。

1948年香农发表的《通信的数学理论》与《在噪声中的通信》奠定了狭义信息论的基础。这一理论认为通信就是信息传输，是将消息由发信者送给收信者的过程，因而给出了一般通信系统的模型。他还利用统计数字的方法，正确处理信息的形式和内容的辩证关系，解决了信息量问题，给出了信息量的数学公式。

香农的狭义信息论第一个给予信息以科学定义：信息是人们对事物了解的不确定性的消除或减少。这是从通信角度上下的定义，即信源发出了某种情况的不了解的状态，即消除了不定性。并且香农采用概率统计数学方法来度量为定性被消除的量的大小：$H(x)$为信息熵，是信源整体的平均不确定度。

在香农寻找信息量的名称时，数学家冯·诺依曼建议称其为熵，理由是不确定性函数在统计力学中已经应用在“熵”的测量中了。在热力学中，熵是物质系统状态的一个函数，它表示微观粒子之间无规则的排列程度，即表示系统的紊乱度。维纳说：“信息量的概念非常自然地从属于统计学的一个古典概念——熵。正如一个系统中的信息量是它的组织化程度的度量，一个系统的熵就是它的无组织程度的度量；这一个正好是那一个的负数。”这说明信息与熵是一个相反的量，信息是负熵，所以在信息熵的公式中有负号，它表示系统获得信息后无序状态的减少或消除，即消除不确定性的大小。

香农信息论是信息科学发展史上的里程碑。他以电报信号为基础，总结归纳了信息的作用，消除人们认识的不确定性，利用概率论数学工具，用定量的形式对最简单情况下的不确定性进行了描述，并在此基础上给出了信息在消除不确定性上所遵从的规律。

1.2 熵

1.2.1 自信息

事件发生的不确定性与事件发生的概率大小有关，概率越小，不确定性越大，事件发生后所含有的信息量就越大。小概率事件的不确定性较大，一旦出现必然使人感到意外，因此产生的信息量就大，特别是几乎不可能出现的事件一旦出现，必然产生极大的信息量；大概率事件是预料之中的事件，不确定性小，即使发生也没什么信息量，特别是概率为 1 的确定事件发生以后，不会给人以任何信息量。因此随机事件的自信息量 $I(x_i)$是该事件发生概率 $p(x_i)$的函数，并且 $I(x_i)$应该满足以下公理化条件：

（1）$I(x_i)$是 $p(x_i)$的严格递减函数。当 $p(x_1)<p(x_2)$时，$I(x_1)>I(x_2)$。概率越小，事件发生的不确定性越大，事件发生以后所包含的自信息量越大。

（2）极限情况下，当 $p(x_i)=0$ 时，$I(x_i)\to\infty$；当 $p(x_i)=1$ 时，$I(x_i)=0$。

（3）从直观概念上讲，由两个相对独立的不同的消息所提供的信息量应等于它们分别提供的信息量之和，即自信息量满足可加性。

可以证明，满足以上公理化条件的函数形式是对数形式。

定义 1.1　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。设事件 x_i 的概率为 $p(x_i)$，则它的自信息量定义为

$$I(x_i)=-\log 2p(x_i)=\log\frac{1}{p(x_i)} \tag{1.1}$$

自信息量的单位与所用对数的底有关。

（1）通常取对数的底为 2，信息量的单位为比特（binaryunit，bit）。当 $p(x_i)$=1/2 时，$I(x_i)$=1bit，即概率等于 1/2 的事件具有 1bit 的自信息量。例如，一枚均匀硬币的任何一种抛掷结果均含有 1bit 的信息量。比特是信息论中最常用的信息量单位，当取对数的底为 2 时，2 常省略。

注意：计算机术语中 bit 是位的单位（binarydigit，bit），与信息量单位不同但有联系，1 位的二进制数字最大能提供 1bit 的信息量。

（2）若取自然对数（以 e 为底），自信息量的单位为奈特（naturalunit，nat）。理论推导中或用于连续信源时用以 e 为底的对数比较方便。

$$1\text{nat}=\log_2 e\ \text{bit}=1.443\text{bit}$$

（3）工程上以 10 为底较方便。若以 10 为对数底，则自信息量的单位为哈特（hartley，hat），以纪念哈特莱首先提出用对数来度量信息。

$$1\text{Hat}=\log_2 10\text{bit}=3.322\text{bit}$$

（4）如果取以 r 为底的对数（r>1），则 $I(x_i)=-\log r\ p(x_i)$。

$$1r\ \text{进制单位}=\log_2 r\ \text{bit}$$

例 1.1

（1）英文字母中“a”出现的概率为 0.064，“c”出现的概率为 0.022，分别计算它们的自信息量。

（2）假定前后字母出现是互相独立的，计算“ac”的自信息量。

（3）假定前后字母出现不是互相独立的，当“a”出现以后，“c”出现的概率为 0.04，计算“a”出现以后，“c”出现的自信息量。

解：

（1）I(a)=$-\log_2 0.064=3.96$bit

I(c)= $-\log_2 0.022=5.51$bit

（2）由于前后字母出现是互相独立的，“ac”出现的概率为 0.064×0.022，所以

$$I(\text{ac})=-\log_2(0.064\times 0.022)=-(\log_2 0.064+\log_2 0.022)=I(\text{a})+I(\text{c})=9.47\text{bit}$$

即两个相对独立的事件的自信息量满足可加性，也就是由两个相对独立的事件的积事件所提供的信息量应等于它们分别提供的信息量之和。

（3）“a”出现的条件下，“c”出现的概率变大，它的不确定性变小。

$$I(\text{c}|\text{a})=-\log_2 0.04=4.64\text{bit}$$

例 1.2

同时掷 2 颗骰子，事件 A、B、C 分别表示：(A)仅有一个骰子是 3；(B)至少有一个骰子是 4；(C)骰子上点数的总和为偶数。试计算事件 A、B、C 发生后所提供的信息量。

解：

掷 2 颗骰子，共有 6×6=36 种结果。记投掷结果为(x,y)，其中 x 和 y 是 2 颗骰子分别掷出的点数，则事件 A 对应于 10 种结果：(1,3)，(2,3)，(4,3)，(5,3)，(6,3)，(3,1)，(3,2)，(3,4)，(3,5)，(3,6)；事件 B、C 分别对应于 11 种结果和 18 种结果。因此它们的概率分别为

p(A)=10/36

p(B)=11/36

p(C)=18/36

A、B、C 所提供的信息量分别为

I(A)=$-\log_2 p$(A)=1.85bit

I(B)=$-\log_2 p$(B)=1.71bit

I(C)=$-\log_2 p$(C)=1bit

1.2.2 互信息

定义 1.2 一个事件 y_j 所给出关于另一个事件 x_i 的信息定义为互信息，用 $I(x_i;y_j)$ 表示。

$$I(x_i;y_j)=I(x_i)-I(x_i\mid y_j)=\log 2\frac{p(x_i\mid y_i)}{p(x_i)} \tag{1.2}$$

互信息 $I(x_i;y_j)$ 是已知事件 y_j 后所消除的关于事件 x_i 的不确定性，它等于事件 x_i 本身的不确定性 $I(x_i)$ 减去已知事件 y_j 后对 x_i 仍然存在的不确定性 $I(x_i;y_j)$。互信息的引入使信息的传递得到了定量的表示。

例 1.3

某地二月份天气出现的概率分别为晴 1/2，阴 1/4，雨 1/8，雪 1/8。某一天有人告诉你：“今天不是晴天”，把这句话作为收到的消息 y_1，求收到 y_1 后，y_1 与各种天气的互信息量。

解：

把各种天气记作 x_1(晴)，x_2(阴)，x_3(雨)，x_4(雪)。收到消息 y_1 后，各种天气发生的概率变成了后验概率：

$$p(x_1\mid y_1)=\frac{p(x_1y_1)}{p(y_1)}=0$$

$$p(x_2\mid y_1)=\frac{p(x_2y_1)}{p(y_1)}=\frac{1/4}{1/4+1/8+1/8}=\frac{1}{2}$$

$$p(x_3\mid y_1)=\frac{p(x_3y_1)}{p(y_1)}=\frac{1/8}{1/4+1/8+1/8}=\frac{1}{4}$$

同理，$p(x_4\mid y_1)=\frac{1}{4}$

根据互信息量的定义，可计算出 y_1 与各种天气之间的互信息：

$$I(x_1;y_1)=\log_2\frac{p(x_1\mid y_1)}{p(x_1)}=\infty$$

$$I(x_2;y_1)=\log_2\frac{p(x_2\mid y_1)}{p(x_2)}=1\text{bit}$$

$$I(x_3;y_1)=\log_2\frac{p(x_3\mid y_1)}{p(x_3)}=1\text{bit}$$

$$I(x_4;y_1)=\log_2\frac{p(x_4\mid y_1)}{p(x_4)}=1\text{bit}$$

例 1.4

已知信源发出 a_1 和 a_2 两种消息，且 $p(a_1)=p(a_2)=1/2$。此消息在二进制对称信道上传输，信道传输特性为 $p(b_1|a_1)=p(b_2|a_2)=1-\varepsilon$，$p(b_1|a_2)=p(b_2|a_1)=\varepsilon$。求互信息量 $I(a_1;b_1)$和 $I(a_1;b_2)$。

解：

由 $p(a_1)=p(a_2)=1/2$，$p(b_1|a_1)=p(b_2|a_2)=1-\varepsilon$，$p(b_1|a_2)=p(b_2|a_1)=\varepsilon$ 可知

$$p(a_1b_1)=p(b_1|a_1)\cdot p(a_1)=\frac{1}{2}(1-\varepsilon)$$

$$p(a_1b_2)=p(b_2|a_1)\cdot p(a_1)=\frac{1}{2}\varepsilon$$

$$p(a_2b_1)=p(b_1|a_2)\cdot p(a_2)=\frac{1}{2}\varepsilon$$

$$p(a_2b_2)=p(b_2|a_2)\cdot p(a_2)=\frac{1}{2}(1-\varepsilon)$$

$$p(b_1)=\sum_{i=1}^{2}p(a_ib_1)=1/2$$

$$p(b_2)=\sum_{i=1}^{2}p(a_ib_2)=1/2$$

所以

$$I(a_1;b_1)=\log_2\frac{p(a_1\mid b_1)}{p(a_1)}=\log_2\frac{p(a_1b_1)}{p(a_1)p(b_1)}=\log_2\frac{1/2(1-\varepsilon)}{1/2\times1/2}=1+\log_2(1-\varepsilon)\text{bit}$$

$$I(a_1;b_2)=\log_2\frac{p(a_1\mid b_2)}{p(a_1)}=\log_2\frac{p(a_1b_2)}{p(a_1)p(b_2)}=\log_2\frac{1/2\varepsilon}{1/2\times1/2}=1+\log_2\varepsilon\text{bit}$$

1.2.3　平均自信息

自信息量是信源发出某一具体消息所含有的信息量，发出的消息不同，其自信息量就不同，所以自信息量本身为随机变量，不能用来表征整个信源的不确定度。我们用平均自信息量来表征整个信源的不确定度。平均自信息量又称为信息熵、信源熵，简称熵。

因为信源具有不确定性，所以信源常用随机变量来表示，用随机变量的概率分布来描述信源的不确定性。通常把一个随机变量的所有可能的取值和这些取值对应的概率 $[X,P(X)]$ 称为它的概率空间。

假设随机变量 X 有 q 个可能的取值 x_i，$i=1,2,\ldots,q$，各种取值出现的概率为 $p(x_i)$，$i=1,2,\ldots,q$，它的概率空间表示为

$$\begin{bmatrix}X\\P(X)\end{bmatrix}=\begin{bmatrix}X=x_1\ldots X=x_i\ldots X=x_q\\p(x_1)\ \ldots\ p(x_i)\ \ldots p(x_q)\end{bmatrix}$$

这里要注意，$p(x_i)$满足概率空间的基本特性：非负性$0 \leqslant p(x_i) \leqslant 1$和完备性$\sum_{i=1}^{q} p(x_i)$。

定义 1.3 随机变量 X 的每一个可能取值的自信息$I(x_i)$的统计平均值定义为随机变量 X 的平均自信息量。

$$H(X) = E[I(x_i)] = -\sum_{i=1}^{q} p(x_i)\log_2 p(x_i) \tag{1.3}$$

这里 q 为 X 的所有可能取值的个数。

熵的单位也与所取的对数底有关，根据所取的对数底不同，可以是比特/符号、奈特/符号、哈特/符号或者是 r 进制单位/符号，通常用比特/符号为单位。

熵这个名词是香农从物理学中热熵的概念借用过来的，热熵是表示分子混乱程度的一个物理量，因此，香农用熵来描述信源的平均不确定性。但是在热力学中，任何孤立系统的演化都会使热熵只能增加不能减少，而在信息论中，信息熵正相反，只会减少，不会增加，所以有人称信息熵为负热熵。

信息熵 $H(X)$是对信源的平均不确定性的描述。要对信源输出的消息进行无失真的编码，平均每个信源符号至少需要用 $H(X)$个码符号。

一般情况下，信息熵并不等于收信者平均获得的信息量。只有在无噪情况下，收信者才能正确无误地接收到信源所发出的消息，获得的平均信息量就等于$H(X)$，而一般情况下，因为干扰和噪声的存在，收信者不能全部消除信源的平均不确定性，获得的信息量将小于信息熵。

1.2.4 熵函数的性质

信息熵 $H(X)$是随机变量 X 的概率分布的函数，所以又称为熵函数。如果把概率分布 $p(x_i)$，i=1,2,...,q，记为 $p_1,p_2,...,p_q$，则熵函数又可以写为概率矢量 $p_1,p_2,...,p_q$ 的函数形式，记为 $H(p)$。

$$H(X) = -\sum_{i=1}^{q} p_i \log_2 p_i = H(p_1, p_2, \&, p_q) = H(p) \tag{1.4}$$

因为概率空间的完备性，即 $\sum_{i=1}^{q} p_i = 1$，所以，$H(p)$是$(q-1)$元函数。当 q=2 时，因为 $p_1 + p_2 = 1$，若令其中一个概率为 p，则另一个概率为$(1-p)$，熵函数可以写成$H(p)$。

熵函数$H(p)$具有以下性质：

1. 对称性

$$H(p_1, p_2, \ldots, p_q) = H(p_2, p_1, \ldots, p_q) = \ldots = H(p_q, p_1, \ldots, p_{q-1}) \tag{1.5}$$

也就是说概率矢量 $p = (p_1, p_2, \ldots, p_q)$ 各分量的次序可以任意变更，熵值不变。对称性说明熵函数仅与信源的总体统计特性有关。

2. 确定性

$$H(1,0) = H(1,0,0) = H(1,0,0,0) = \ldots = H(1,0,\ldots,0) = 0 \tag{1.6}$$

在概率矢量 $p = (p_1, p_2, \ldots, p_q)$ 中，只要有一个分量为 1，其他分量必为 0，它们对熵的贡献均为 0，因此熵等于 0，也就是说确定信源的平均不确定度为 0。

3. 非负性

$$H(p)=H(p_1,p_2,\ldots,p_q)\geqslant 0 \tag{1.7}$$

对于确定信源，等号成立。

信源熵是自信息的数学期望，自信息是非负值，所以信源熵必定是非负的。离散信源熵才有这种非负性，以后讲到的连续信源的相对熵则可能出现负值。

4. 扩展性

$$\lim_{\varepsilon\to 0}H_{q+1}(p_1,p_2,\ldots p_q-\varepsilon,\varepsilon)=H_q(p_1,p_2,\ldots p_q) \tag{1.8}$$

这是因为 $\lim\limits_{\varepsilon\to 0}\varepsilon\log_2\varepsilon=0$。

这个性质的含义是：增加一个基本不会出现的小概率事件，信源的熵保持不变。虽然小概率事件出现给予收信者的信息量很大，但在熵的计算中，它占的比重很小，可以忽略不计，这也是熵的总体平均性的体现。

5. 连续性

$$\lim_{\varepsilon\to 0}H(p_1,p_2,\ldots p_{q-1}-\varepsilon,p_q+\varepsilon)=H(p_1,p_2,\ldots p_q) \tag{1.9}$$

即信源概率空间中概率分量的微小波动，不会引起熵的变化。

6. 递增性

$$H(p_1,p_2,\ldots,p_{n-1},q_1,q_2,\ldots,q_m)=H(p_1,p_2,\ldots,p_n)+p_nH\left[\frac{q_1}{p_n},\frac{q_2}{p_n},\ldots,\frac{q_m}{p_n}\right] \tag{1.10}$$

这个性质表明，假如有一信源的 n 个元素的概率分布为 $(p_1,p_2,\ldots,p_n)$，其中某个元素 x_n 又被划分成 m 个元素，这 m 个元素的概率之和等于元素 x_n 的概率，这样得到的新信源的熵增加了一项，增加的一项是由于划分产生的不确定性。

7. 极值性

$$H(p_1,p_2,\ldots,p_n)\leqslant H\left[\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}\right]=\log_2 n \tag{1.11}$$

式中 n 是随机变量 X 的可能取值的个数。

极值性表明离散信源中各消息等概率出现时熵最大，这就是最大离散熵定理。连续信源的最大熵则还与约束条件有关。

直观来看，随机变量的不确定程度并不都是一样的。例如，抛掷一枚均匀硬币结果所得到的信息量会比抛掷一枚偏畸硬币所得到的信息量大；抛掷一颗均匀骰子的试验比抛掷一枚均匀硬币的试验所得到的信息量大。怎么度量这种不确定性呢？香农指出，存在这样的不确定性的度量，它是随机变量的概率分布的函数，而且必须满足 3 个公理性条件：

（1）连续性条件：$f(p_1,p_2,\ldots,p_n)$ 应是 p_i，$i=1,2,\ldots,n$ 的连续函数。

（2）等概时为单调函数：$f(1/n,1/n,\ldots,1/n)$ 应是 n 的增函数。

（3）递增性条件：当随机变量的取值不是通过一次试验而是若干次试验才最后得到时，随机变量在各次试验中的不确定性应该可加，且其和始终与通过一次试验取得的不确定程度相同，即

$$F(p_1,p_2,\ldots,p_n)=f[(p_1+p_2+\ldots+p_k),p_{k+1},\ldots,p_n]+(p_1+p_2+\ldots+p_k)f(p_1'+p_2'+\ldots+p_k') \tag{1.12}$$

其中，$p'_k = p_k/(p_1, p_2, \ldots, p_n)$。

香农根据这 3 个公理性条件于 1948 年先提出了熵的概念，他当时并没有像我们现在这样把熵看成自信息的均值。后来，Feinstein（范恩斯坦）等人从数学上严格地证明了当满足上述条件时，信息熵的表达形式是唯一的。

1.2.5 联合熵与条件熵

一个随机变量的不确定性可以用熵来表示，这一概念可以方便地推广到多个随机变量。

定义 1.4 二维随机变量 XY 的概率空间表示为

$$\begin{bmatrix} XY \\ P(XY) \end{bmatrix} = \begin{bmatrix} x_1y_1 & \cdots & x_iy_i & \cdots & x_ny_m \\ p(x_1y_1) & \cdots & p(x_iy_i) & \cdots & p(x_ny_m) \end{bmatrix}$$

其中，$p(x_iy_i)$ 满足概率空间的非负性和完备性：$0 \leqslant p(x_iy_i) \leqslant 1$，$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_i) = 1$。

二维随机变量 XY 的联合熵定义为联合自信息的数学期望，它是二维随机变量 XY 的不确定性的度量。

$$H(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_i)\, I(x_iy_i) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_iy_i)\log_2 p(x_iy_i) \tag{1.13}$$

考虑在给定 $X = x_i$ 的条件下，随机变量 Y 的不确定性为

$$H(Y \mid x_i) = -\sum_{j} p(y_j \mid x_i)\log_2 p(y_j \mid x_i) \tag{1.14}$$

由于对于不同的 x_i，$H(Y \mid x_i)$ 是变化的，对 $H(Y \mid x_i)$ 的所有可能值进行统计平均，就得出给定 X 时，Y 的条件熵 $H(Y|X)$。

定义 1.5

$$\begin{aligned} H(Y \mid X) &= \sum_{i} p(x_i)H(Y \mid x_i) \\ &= -\sum_{i}\sum_{j} p(x_i)p(y_j \mid x_i)\log_2 p(y_j \mid x_i) \\ &= -\sum_{i}\sum_{j} p(x_iy_j)\log_2 p(y_j \mid x_i) \end{aligned} \tag{1.15}$$

其中，$H(Y|X)$表示已知 X 时，Y 的平均不确定性。

同理，

$$H(X \mid Y) = -\sum_{i}\sum_{j} p(x_iy_j)\log_2 p(x_j \mid y_i)$$

下面讨论各类熵的关系。

（1）联合熵与信息熵、条件熵的关系。

$$H(XY) = H(X) + H(Y \mid X)$$

即两个随机变量 X 和 Y 的联合熵等于 X 的熵加上在 X 已知条件下 Y 的条件熵，这个关系可以方便地推广到 N 个随机变量的情况，即

$$H(X_1X_2\ldots X_N) = H(X_1) + H(X_2 \mid X_1) + \& + H(X_N \mid X_1X_2\ldots X_{N-1}) \tag{1.16}$$