信息技术基础

XINXIJISHUJICHU

周航慈 孙丽华
王仁波 元如林 编著

北京航空航天大学出版社
http://www.buaapress.com.cn

信息技术基础

周航慈　孙丽华　王仁波　元如林　编著

北京航空航天大学出版社
http://www.buaapress.com.cn

内 容 简 介

本书较系统地介绍信息技术的理论基础和信息技术各个环节的关键技术，包括“信息论基础”、“信息压缩编码技术”、“信息容错技术”和“信息安全技术”四部分内容。

第一章为“信息论基础”，介绍了信息论的基本理论知识。第二章到第七章为“信源编码（压缩编码）技术”，论述如何高效地将各种信息（文字、声音和图像）数字化，它是“多媒体技术”的基础。第八章到第十二章为“信息容错技术”，介绍如何在恶劣环境中存储和传输信息，并保持信息的准确性。第十三章到第十五章为“信息安全技术”，对如何保护信息不被非法窃取、信息系统不被破坏、确保信息社会正常运转的关键技术作了进一步阐述。

本教材理论与实际相结合，每章均有习题，并附有综合练习。它是高等院校“信息工程”类（计算机、电子信息工程、自动化和仪器仪表等）专业的必修课。不同专业可根据需要选择教学内容，将课时控制在 40 到 60 学时。也可作为相关专业研究生或从业人员的参考书。

图书在版编目(CIP)数据

信息技术基础/周航慈等编著. —北京：北京航空航天大学出版社，2002.8

ISBN 7-81077-153-1

Ⅰ.信… Ⅱ.周… Ⅲ.信息理论—高等学校教材 Ⅳ.TN911.2

中国版本图书馆 CIP 数据核字(2002)第 048471 号

信 息 技 术 基 础

周航慈 孙丽华 王仁波 元如林 编著

责任编辑 金友泉

*

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(100083) 发行部电话(010)82317024 传真(010)82328026

http://www.buaapress.com.cn

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

河北省涿州市新华印刷厂印制 各地书店经销

*

开本：787 mm×1 092 mm 1/16 印张：17.75 字数：454 千字

2002 年 8 月第 1 版 2002 年 8 月第 1 次印刷 印数：5 000 册

ISBN 7-81077-153-1/TN·004 定价：24.00 元

前　言

当今社会是信息化社会，电子计算机和通信网络已经广泛应用于社会的各个领域，利用这些先进技术建立起来的信息系统改变了人们的生活和工作方式，信息产业也已经成为国民经济的支柱产业。在高等院校中，信息工程类专业是最为热门的专业之一，信息技术已经改变了很多传统电子类专业的知识结构，使得这些专业已经脱胎换骨，投入到信息技术的领域中去了。但是，很多从信息工程类专业（计算机专业、电子信息工程专业、自动化专业和仪器仪表专业等）毕业的大学生缺乏信息技术的基本知识，不知“信息量”是怎么计算的，不知信息处理有哪些环节，各个环节有哪些关键技术？造成这种现象的原因是我们的教学计划跟不上信息时代的发展需要，没有把“信息技术基础”列入教学大纲。这种不正常现象和信息技术的迅猛发展是非常不相称的。我们尝试将信息技术最基本的知识编辑成书，供信息工程类专业作为教材使用，也可以为众多没有系统学习过“信息技术基础”的工程技术人员作为学习参考书。

第一章为“信息论基础”。它介绍信息论的基本理论知识，使读者对“信息论”有一个初步的了解，是全书的理论基础。

第二章到第七章为“信源编码（压缩编码）技术”。它论述如何高效地将各种信息（文字、声音和图像）数字化，这是“多媒体技术”的基础。其中，第二章介绍无失真压缩编码的基本理论；第三章介绍常用的几种无失真压缩编码方法；第四章介绍率失真函数及率失真编码定理；第五章介绍连续信源的数字化与预测编码方法；第六章介绍变换编码的基本原理；第七章介绍图像压缩的若干国际标准。

第八章到第十二章为“信息容错技术”。它介绍如何在恶劣环境中存储和传输信息，并保持信息的准确性。其中，第八章介绍信道容量及信道编码定理；第九章介绍能够发现差错的检错码；第十章介绍纠错编码的代数基础；第十一章和第十二章分别介绍能够纠正错误的线性分组码和循环码。

第十三章到第十五章为“信息安全技术”。对如何保护信息不被非法窃取，信息系统不被破坏，确保信息社会正常运转的关键技术给以阐述。其中，第十三章介绍密码学基础与数据加密标准（DES）；第十四章介绍公开密钥加密体制；第十五章介绍互联网络的安全与对策。

为了配合教学，每一章都有适量的练习题。

在附录一中收录了信息论中的部分定理证明，供感兴趣的读者作进一步地学习和参考。附录二为课程综合练习，通过综合练习，可以加深读者对信息技术主要环节的理解。

本书全部内容计划为60学时，不同专业可根据需要选择教学内容和讲授深度，将实际教学控制在40～60学时。

本书第一章、第二章、第四章、第八章和附录一由孙丽华编写；第三章、第五章、第六章、第七章、第九章的小部分、第十一章的小部分和附录二由周航慈编写；第九章的大部分、第十章、第十一章的大部分和第十二章由元如林编写；第十三章、第十四章和第十五章由王仁波编写。周航慈负责全书的策划、内容安排、文稿修改和审定。

本书在编写过程中，得到北京航空航天大学出版社和何立民教授的大力支持，得到南昌大学和东华理工学院有关部门的关心和资助，在此表示衷心感谢！

由于本书涉及知识领域广泛，而且变化日新月异，限于时间和水平的限制，难免有差错和不足之处，敬请读者指正！

编　者

2001年12月31日

目　录

第一章　信息论的基本概念

1.1　绪　论

信息论是研究信息的传输、存储和处理的学科，亦称“信息论”为“通信的数学理论”。它主要研究在通信系统设计中如何实现信息传输的有效性和可靠性。

所谓通信，即能彼时、彼地精确地或近似地复现原信号。通俗地说，通信就是消息传递的过程。在通信前，对收信者来说，消息存在不确定性；收到消息后，不确定性被部分或全部排除（视是否存在干扰）。所以，通信过程是一种从不确定到确定的过程，不确定性排除了，收信者就获得了信息。

各种通信系统（电报、电话、雷达和计算机网络等），尽管它们的形式与用途不同，但从信息论的角度看，都可概括为如图 1－1 所示的模型。

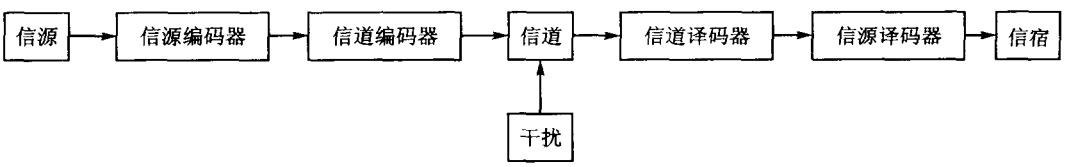

图 1－1　通信系统模型

信源是产生消息的源，消息可以是语言、文字、图像等。它们可以是连续的，也可以是离散的，但都是随机的，即事先不可能知道它们确切的内容，否则通信就无意义了。

信源编码器是把消息变换成适合信道传送的信号的设备，即把文字、语言和图像等这些能够为人们的感觉器官所感知的物理现象变换成各种电信号的设备。它主要以提高信息传输的效率为目的。

信道是信号传递的通道或媒质。它的种类很多，可以是架空明线，可以是电缆、光缆，也可以通过电离层反射实现信号的传输，还有我们日常生活中所用的磁带、磁盘、光盘和书籍等都可看成信道。

由于信号在信道中传播时，不可避免地会加入干扰（来自系统内部的噪声即乘性干扰；来自外界的干扰（它与信道的物理特征无关）即加性干扰），这就给信号的复原带来很大的困难。信道编码器也称纠错编码器，它是为了将信号能够准确地从噪声中分离出来，事先对信号进行编码，以提高信息传输的准确度。编码时，给信码加上一定的冗余度，也就是说牺牲了有效性却换来了可靠性。

例如，有两份中文电报待发：“中华人民共和国”和“母亲病愈，身体健康”。从提高效率的角度出发，可将它们压缩成“中国”、“母病愈”，这样原意未变，但电文变得简洁明了，说明原电文具有冗余度。冗余度大的消息具有较强的抗干扰能力。

如果我们收到的消息为“中×人民×和国”，“母亲病×，身体健康”时，很容易根据上下文

的意思，把它纠正为“中华人民共和国”和“母亲病愈，身体健康”。

如果发的是经压缩后的电文，收到“×国”和“母病×”，就很难确定发的是“中国”还是“美国”、是“母病愈”还是“母病危”，这样就会造成很大的错误。

所以，关于信源编码和信道编码可以理解为：信源编码就是通过减少或消除信源的剩余度来提高传输效率；而信道编码则是通过增加信源的剩余度来提高抗干扰能力。一般说，有效性和可靠性是一对矛盾，很难使二者都达到最佳，总是在一定的条件下达到平衡。

消息通过信道传输到目的地以后，得到的是信号和各种噪声的混和物。信道译码器就是从这些混和物中尽可能无误地将信号提取出来，这就是噪声中的信号检测问题。信源译码器将信号还原成消息，最后送到信宿，即人、计算机和机器等。

综上所述，通信就是把消息从此岸传到彼岸。为尽可能提高通信的有效性和可靠性，必须研究信源、信道和信宿的内在特性。而消息及噪声都是随机的，对于它们的内在特性只能用统计特性来描述。

1948 年，美国工程师 C. E. Shannon(香农)首次在它的著作中，用概率统计的观点来研究通信理论问题，奠定了 Shannon 信息论的基础。但香农信息论只为设计有效而可靠的通信系统提供了理论依据，它不是构造性的，即只指明了方向和证明了实现的可能性，但没有提出实现的具体方法。后人沿着他指出的方向，寻求有效可靠的通信系统，这就产生了纠错编码理论、调制理论、信号检测理论等，这些都是信息论的分支。

1.2 信 源

信源是产生消息的源，实际中的信源是非常复杂的，但它们的共同特征是都能产生随机的输出信号。信源发布消息的过程可以看成是一个随机过程，对于这种随机过程，可用概率统计模型来描述它。

记一个信源为$\{x,q(x)\}$。其中，x 为随机过程，$q(x)$是它对应的概率统计分布，根据 x 的不同情况，信源可以分成许多种类型。

如果 x 取值的字符集 X 为离散集合，则称信源为离散信源，否则就是连续信源。它们分别对应空间离散的和空间连续的随机过程。

如果随机过程 x 在时间上是离散的，则称为时间离散信源；若在时间上是连续的，就称为波形信源。

从信源的统计特性来看，又可以分成几种情况。若 x 的取值在各时刻相互独立，称为无记忆信源；若 x 在各时刻的随机变量相互关联，就称为有记忆信源。

根据前面介绍的通信系统模式，该课程主要研究的是对各种信源及信道的编译码问题。下面介绍几种常见的信源。

1.2.1 离散信源

1. 一维离散信源

信源的输出是数目有限的单个符号(如文字及字母等)，其数学模型是离散型的概率空间，即$\begin{bmatrix} x \\ q(x) \end{bmatrix}=\begin{bmatrix} a_0 & a_1 & \cdots & a_{k-1} \\ q(a_0) & q(a_1) & \cdots & q(a_{k-1}) \end{bmatrix}$。当满足 $\sum_{i=0}^{k-1} q(a_i)=1$ 时，则信源的每次输出且必

须输出 $X=\{a_0,a_1,\cdots,a_{k-1}\}$ 中的一个。

[例 1]　南昌地区春季天气预报如下列表：

天　气	晴	多　云	阴	雨	雪
事　件	a_0	a_1	a_2	a_3	a_4
概　率	0.08	0.12	0.27	0.52	0.01

天气预报可视为一信源，其概率统计模型为：

$$\begin{bmatrix} x \\ q(x) \end{bmatrix}=\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 \\ 0.08 & 0.12 & 0.27 & 0.52 & 0.01 \end{bmatrix}$$

[例 2]　二进制对称信源，用 0 和 1 分别代表两个事件，0 出现的概率为 p，1 出现的概率为 $1-p$，该信源的概率统计模型应为：

$$\begin{bmatrix} x \\ q(x) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$$

2. 一维离散信源的扩散信源

信源的输出是一符号序列（L 个符号）$X_1,X_2,\cdots,X_L$，而每一个 X_l 都取自同一符号集，即 $X=\{a_0,a_1,\cdots,a_{k-1}\}$，则 $X_l\in X=\{a_0,a_1,\cdots,a_{k-1}\}$，其中 $l=1,2,\cdots,L$。

记符号序列 $\boldsymbol{X}=X_1,X_2,\cdots,X_L\in X^L$ 为 L 维矢量，其数学模型是 L 维概率空间（含 k^L 个元素），即：

$$\begin{bmatrix} \boldsymbol{X} \\ p(\boldsymbol{X}) \end{bmatrix}=\begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_{k^L} \\ p(\boldsymbol{X}_1) & p(\boldsymbol{X}_2) & \cdots & p(\boldsymbol{X}_{X^L}) \end{bmatrix}$$

上式称为单符号离散信源的 L 维扩展信源。若序列 $\boldsymbol{X}=X_1,X_2,\cdots,X_L$ 的各符号两两统计独立，则有 $p(\boldsymbol{X})=\prod_{i=1}^{L}P(X_i)$，对应的信源称为 L 维离散无记忆扩展信源。

[例 3]　如果 $X=\{0,1\}$，$k=2$，$L=3$，则三维扩展信源共含有 $2^3=8$ 个元素：000，001，010，011，100，101，110 和 111。

3. 平稳 Markov（马尔可夫）信源

一般情况下，信源前后发出的符号总是相互依赖的，这种信源称为有记忆源。实际上，信源发出的信号往往只与前面几个符号关系较强，与更前面的符号依赖关系就弱。

当信源每次发出的符号只与前 m 个符号有关，就称为 m 阶 Markov 信源。为了描述这种关系，引入状态的概念，m 阶 Markov 过程的每个状态由 m 个符号组成。若符号种类有 k 种 $(a_0,a_1,\cdots,a_{k-1})$，则有 k^m 种状态，信源每输出一个符号就转入另一种状态。

用 $s_1,s_2,\cdots$ 表示随机状态（每个状态由 m 个符号决定），则：

（1）若状态转移概率 $p(s_j|s_i)$（表示由 s_i 转向 s_j 的概率，$i,j=1,2,\cdots$）与时间起点无关，仅与状态 s_i 和 s_j 有关，则称该信源为平稳信源。

（2）若转向每一状态的概率只与前一时刻的状态有关，而与更早时刻的状态无关，即 $p(s_i|s_{i-1},s_{i-2},\cdots)=p(s_i|s_{i-1})$，则称该信源为 Markov 信源。

符合上述两个条件的称为平稳 Markov 信源。下面是一个二阶平稳 Markov 信源的例子。

[例 4]　一个二阶 Markov 信源，取自符号集 (a_0,a_1)，则有 $k^m=2^2$ 个状态，每个状态由 $m=2$ 个符号决定，设状态编码为：s_0：00　s_1：01　s_2：10　s_3：11。