



教育部高等学校电子信息类专业教学指导委员会规划教材  
高等学校电子信息类专业系列教材

省级精品课程“信息论与编码技术”配套教材

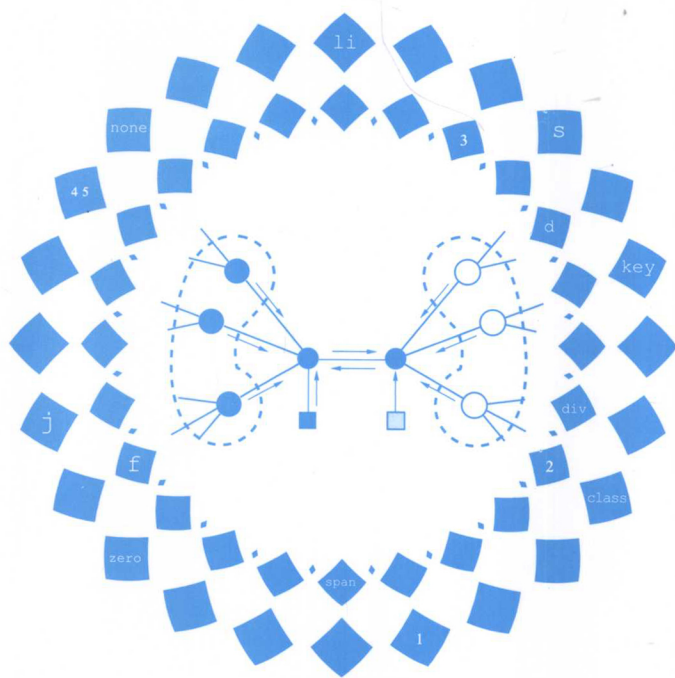
信息与通信工程

**I**nformation Theory and Coding

# 信息论与编码

冯桂 周林 编著

Feng Gui Zhou Lin



清华大学出版社





教育部高等学校电子信息类专业教学指导委员会规划教材  
高等学校电子信息类专业系列教材

---

Information Theory and Coding

# 信息论与编码

---

冯桂 周林 编著

Feng Gui Zhou Lin

清华大学出版社  
北京

## 内 容 简 介

本书系统地介绍了两部分内容：香农信息论的三个基本概念(信源熵、信道容量和信息率失真)，以及对应这三个概念的香农三大编码定理；信源编码和信道编码的基本原理与经典方法，同时对接近香农极限的 Turbo 码和 LDPC 码也作了介绍。为了便于教学和加深读者对概念的理解及自检，本书每章后面都附有思考题与习题。

本书尽量用通俗、生动的语言描述信息理论与信源、信道编码的基本概念，并通过例题和图表来说明基本概念及原理，因而比较适合工科相关专业的教学和读者自学(已掌握工科高等数学和工程数学的读者都能读懂本书)。

本书可作为高等院校通信工程、电子信息工程、信息工程及相关专业的本科生、研究生的教材或教学参考书，也可供从事相关专业的科研和工程技术人员参考。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

### 图书在版编目(CIP)数据

信息论与编码/冯桂,周林编著.--北京:清华大学出版社,2016

高等学校电子信息类专业系列教材

ISBN 978-7-302-42427-7

I. ①信… II. ①冯… ②周… III. ①信息论—高等学校—教材 ②信源编码—高等学校—教材  
IV. ①TN911.2

中国版本图书馆 CIP 数据核字(2015)第 306756 号

责任编辑：盛东亮 赵晓宁

封面设计：李召霞

责任校对：李建庄

责任印制：宋 林

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>, <http://www.wqbook.com>

地 址：北京清华大学学研大厦 A 座 邮 编：100084

社 总 机：010-62770175 邮 购：010-62786544

投稿与读者服务：010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈：010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载：<http://www.tup.com.cn>, 010-62795954

印 装 者：北京鑫海金澳胶印有限公司

经 销：全国新华书店

开 本：185mm×260mm 印 张：16.5

字 数：410 千字

版 次：2016 年 8 月第 1 版

印 次：2016 年 8 月第 1 次印刷

印 数：1~2000

定 价：39.00 元

---

产品编号：061110-01

# 高等学校电子信息类专业系列教材

## 一 顾问委员会

谈振辉	北京交通大学 (教指委高级顾问)	郁道银	天津大学 (教指委高级顾问)
廖延彪	清华大学 (特约高级顾问)	胡广书	清华大学 (特约高级顾问)
华成英	清华大学 (国家级教学名师)	于洪珍	中国矿业大学 (国家级教学名师)
彭启琮	电子科技大学 (国家级教学名师)	孙肖子	西安电子科技大学 (国家级教学名师)
邹逢兴	国防科学技术大学 (国家级教学名师)	严国萍	华中科技大学 (国家级教学名师)

## 一 编审委员会

主任	吕志伟	哈尔滨工业大学		
副主任	刘旭	浙江大学	王志军	北京大学
	隆克平	北京科技大学	葛宝臻	天津大学
	秦石乔	国防科学技术大学	何伟明	哈尔滨工业大学
	刘向东	浙江大学		
委员	王志华	清华大学	宋梅	北京邮电大学
	韩焱	中北大学	张雪英	太原理工大学
	殷福亮	大连理工大学	赵晓晖	吉林大学
	张朝柱	哈尔滨工程大学	刘兴钊	上海交通大学
	洪伟	东南大学	陈鹤鸣	南京邮电大学
	杨明武	合肥工业大学	袁东风	山东大学
	王忠勇	郑州大学	程文青	华中科技大学
	曾云	湖南大学	李思敏	桂林电子科技大学
	陈前斌	重庆邮电大学	张怀武	电子科技大学
	谢泉	贵州大学	卞树檀	第二炮兵工程大学
	吴瑛	解放军信息工程大学	刘纯亮	西安交通大学
	金伟其	北京理工大学	毕卫红	燕山大学
	胡秀珍	内蒙古工业大学	付跃刚	长春理工大学
	贾宏志	上海理工大学	顾济华	苏州大学
	李振华	南京理工大学	韩正甫	中国科学技术大学
	李晖	福建师范大学	何兴道	南昌航空大学
	何平安	武汉大学	张新亮	华中科技大学
	郭永彩	重庆大学	曹益平	四川大学
	刘缠牢	西安工业大学	李儒新	中科院上海光学精密机械研究所
	赵尚弘	空军工程大学	董友梅	京东方科技集团
	蒋晓瑜	装甲兵工程学院	蔡毅	中国兵器科学研究院
	仲顺安	北京理工大学	冯其波	北京交通大学
	黄翊东	清华大学	张有光	北京航空航天大学
	李勇朝	西安电子科技大学	江毅	北京理工大学
	章毓晋	清华大学	张伟刚	南开大学
	刘铁根	天津大学	宋峰	南开大学
	王艳芬	中国矿业大学	靳伟	香港理工大学
	苑立波	哈尔滨工程大学		
丛书责任编辑	盛东亮	清华大学出版社		

# 序

## FOREWORD

我国电子信息产业销售收入总规模在 2013 年已经突破 12 万亿元,行业收入占工业总体比重已经超过 9%。电子信息产业在工业经济中的支撑作用凸显,更加促进了信息化和工业化的高层次深度融合。随着移动互联网、云计算、物联网、大数据和石墨烯等新兴产业的爆发式增长,电子信息产业的发展呈现了新的特点,电子信息产业的人才培养面临着新的挑战。

(1) 随着控制、通信、人机交互和网络互联等新兴电子信息技术不断发展,传统工业设备融合了大量最新的电子信息技术,它们一起构成了庞大而复杂的系统,派生出大量新兴的电子信息技术应用需求。这些“系统级”的应用需求,迫切要求具有系统级设计能力的电子信息技术人才。

(2) 电子信息系统的功能越来越复杂,系统的集成度越来越高。因此,要求未来的设计者应该具备更扎实的理论基础知识和更宽广的专业视野。未来电子信息系统的设计越来越要求软件和硬件的协同规划、协同设计和协同调试。

(3) 新兴电子信息技术的发展依赖于半导体产业的不断推动,半导体厂商为设计者提供了越来越丰富的生态资源,系统集成厂商的全方位配合又加速了这种生态资源的进一步完善。半导体厂商和系统集成厂商所建立的这种生态系统,为未来的设计者提供了更加便捷却又必须依赖的设计资源。

教育部 2012 年颁布了新版《高等学校本科专业目录》,将电子信息类专业进行了整合,为各高校建立系统化的人才培养体系,培养具有扎实理论基础和宽广专业技能的、兼顾“基础”和“系统”的高层次电子信息人才给出了指引。

传统的电子信息学科专业课程体系呈现“自底向上”的特点,这种课程体系偏重对底层元器件的分析与设计,较少涉及系统级的集成与设计。近年来,国内很多高校对电子信息类专业课程体系进行了大力度的改革,这些改革顺应时代潮流,从系统集成的角度,更加科学合理地构建了课程体系。

为了进一步提高普通高校电子信息类专业教育与教学质量,贯彻落实《国家中长期教育改革和发展规划纲要(2010—2020 年)》和《教育部关于全面提高高等教育质量若干意见》(教高【2012】4 号)的精神,教育部高等学校电子信息类专业教学指导委员会开展了“高等学校电子信息类专业课程体系”的立项研究工作,并于 2014 年 5 月启动了《高等学校电子信息类专业系列教材》(教育部高等学校电子信息类专业教学指导委员会规划教材)的建设工作。其目的是为推进高等教育内涵式发展,提高教学水平,满足高等学校对电子信息类专业人才培养、教学改革与课程改革的需要。

本系列教材定位于高等学校电子信息类专业的专业课程,适用于电子信息类的电子信

息工程、电子科学与技术、通信工程、微电子科学与工程、光电信息科学与工程、信息工程及其相近专业。经过编审委员会与众多高校多次沟通,初步拟定分批次(2014—2017年)建设约100门课程教材。本系列教材将力求在保证基础的前提下,突出技术的先进性和科学的前沿性,体现创新教学和工程实践教学;将重视系统集成思想在教学中的体现,鼓励推陈出新,采用“自顶向下”的方法编写教材;将注重反映优秀的教学改革成果,推广优秀的教学经验与理念。

为了保证本系列教材的科学性、系统性及编写质量,本系列教材设立顾问委员会及编审委员会。顾问委员会由教指委高级顾问、特约高级顾问和国家级教学名师担任,编审委员会由教育部高等学校电子信息类专业教学指导委员会委员和一线教学名师组成。同时,清华大学出版社为本系列教材配置优秀的编辑团队,力求高水准出版。本系列教材的建设,不仅有众多高校教师参与,也有大量知名的电子信息类企业支持。在此,谨向参与本系列教材策划、组织、编写与出版的广大教师、企业代表及出版人员致以诚挚的感谢,并殷切希望本系列教材在我国高等学校电子信息类专业人才培养与课程体系建设中发挥切实的作用。

吕志伟 教授

# 前言

## PREFACE

1948年,香农(C. E. Shannon)发表的开创性的文章《通信的数学理论》为信息论和编码技术奠定了坚实的理论基础。信息论是信息科学中最成熟、最完整、最系统的一部分,它以活跃、新颖的思路和高效解决问题的方法显示出独特的魅力,在此基础上发展起来的数据通信和计算机技术,反过来又为信息编码技术的发展和應用创造了有利的环境。随着社会信息化的不断深入,信息论和编码技术已经渗透许多应用领域,展示出勃勃生机和巨大的发展前景。

信息论不仅在方法论的层面上解决通信的有效性和可靠性问题,而且在认识论的层面上帮助人们认识事物的本质。学完信息论之后,再重新审视周围的事物时,会产生许多新的看法和认识。用信息论的方法可以宏观地认识某些政治问题,也可以定量地解决某些经济问题,还可以分析、解释学习中存在的问题。总之,信息论是高层次信息技术人才所需掌握的、必不可少的基础知识,因而目前各高等院校的电子信息类专业的本科生、研究生都把信息论和编码技术作为一门重要的专业基础理论课。

由于信息论牵涉众多学科,需要广泛的数学基础,许多读者虽然认识到信息论和编码技术的重要性,但在繁杂的公式面前往往望而却步。针对这种情况,作者根据多年的教学经验,在编写过程中强调基本原理的理解,选材时充分考虑其实用性,把信息论涉及的数学知识限制在工科高等数学和工程数学的范畴内,尽量以通俗形象的语言描述定义、性质和结论的物理概念,叙述中重概念描述、少理论推导,在每章结尾还附有相应的思考题与习题以加深认识。因此,本书适于作为通信、信息工程专业本科生、研究生的教材,也可作为其他专业学生及有关科技人员的参考书。

本书主要内容包括经典信息论的基本内容和主要结论、信息压缩编码的基本原理、提高通信可靠性的纠错编码理论和方法。全书共分8章,遵照由浅入深、循序渐进的教学规律,系统地组织教学内容。第1章绪论,介绍信息论与编码的基本概念、数字通信系统模型,以及信息论、信道编码和信源编码理论的主要发展历程和意义;第2章信源及其熵,介绍信源的数学模型和分类、离散信源的信息熵及其性质、连续信源的信息熵、信源的冗余度等;第3章信道及其容量,介绍信道的数学模型与分类、信道疑义度与平均互信息、离散信道的信道容量、连续信道的信道容量、信源与信道的匹配以及信道编码定理等;第4章信道编码,介绍信道编码的基本概念和经典信道编码理论,主要包括线性分组码、循环码和卷积码的概念及编码、译码算法等;第5章信源编码,介绍信源编码器及相关概念、无失真信源编码定理、变长编码和实用信源编码方法等;第6章信息速率失真函数,介绍失真测度、信息率失真函数、等概率与对称失真信源的信息率失真计算和保真度准则下的信源编码定理等;第7章现代信道编码技术,介绍近年来信道编码领域最重要的两种新技术——Turbo码和LDPC

码,分别介绍二者的概念、基本原理和编、译码算法等;第8章 MATLAB 在信息论与编码分析中的应用,介绍 MATLAB 的使用基础、应用 MATLAB 分析离散信源和离散信道、信源和信道编码技术的 MATLAB 分析和仿真。本书第1、2、3、5章由冯桂编写,第4、6、7章由周林编写,第8章由冯桂、周林共同编写,全书由冯桂统稿。

本书获华侨大学教材建设基金资助。在编写过程获得了华侨大学通信工程系老师和研究生的支持与协助,在此表示感谢。

限于编者的水平,又加上时间比较仓促,书中难免有欠妥之处,殷切希望读者指正,将不胜感激。本书责编的 E-mail: shengdl@tup.tsinghua.edu.cn。

编 者

2016年4月



# 目录

## CONTENTS

---

<b>第 1 章 绪论</b> .....	1
1.1 基本概念 .....	1
1.1.1 信息的一般概念 .....	1
1.1.2 香农信息定义 .....	4
1.1.3 信息论与编码技术发展简史 .....	6
1.2 数字通信系统模型 .....	8
1.3 信息论与编码理论研究的主要内容和意义 .....	9
1.3.1 信息论研究的主要内容 .....	9
1.3.2 香农信息论对信道编码的指导意义 .....	9
1.3.3 香农信息论对信源编码的指导意义 .....	9
思考题与习题 .....	10
<b>第 2 章 信源及其熵</b> .....	11
2.1 信源的数学模型和分类 .....	11
2.1.1 信源的数学模型 .....	11
2.1.2 信源的分类 .....	12
2.2 离散信源的信息熵及其性质 .....	16
2.2.1 自信息 .....	16
2.2.2 信源的信息熵 .....	17
2.2.3 熵的基本性质 .....	19
2.3 离散无记忆信源的扩展信源 .....	22
2.4 离散平稳信源 .....	24
2.4.1 平稳信源的概念 .....	24
2.4.2 二维平稳信源 .....	25
2.4.3 一般离散平稳信源 .....	28
2.5 连续信源的信息熵 .....	29
2.5.1 单符号连续信源的熵 .....	29
2.5.2 波形信源的熵 .....	32
2.5.3 最大熵定理 .....	32
2.6 信源的冗余度 .....	33
2.6.1 信源效率 .....	34
2.6.2 信源冗余度 .....	34
思考题与习题 .....	35

<b>第 3 章 信道及其容量</b> .....	38
3.1 信道的数学模型与分类 .....	38
3.1.1 信道的分类 .....	39
3.1.2 信道的数学模型 .....	40
3.1.3 单符号离散信道 .....	41
3.2 信道疑义度与平均互信息 .....	43
3.2.1 信道疑义度 .....	43
3.2.2 平均互信息 .....	44
3.2.3 平均互信息的性质 .....	47
3.3 离散无记忆的扩展信道 .....	49
3.4 离散信道的信道容量 .....	51
3.4.1 信道容量的定义 .....	51
3.4.2 简单离散信道的信道容量 .....	52
3.4.3 对称离散信道的信道容量 .....	54
3.4.4 离散无记忆 $N$ 次扩展信道的信道容量 .....	56
3.5 连续信道的信道容量 .....	57
3.5.1 连续单符号加性高斯噪声信道的信道容量 .....	57
3.5.2 多维无记忆加性连续信道的信道容量 .....	58
3.5.3 限频限时限功率的加性高斯白噪声信道的信道容量 .....	62
3.6 信源与信道的匹配 .....	64
3.7 信道编码定理(香农第二定理) .....	65
思考题与习题 .....	66
<b>第 4 章 信道编码</b> .....	69
4.1 信道编码的概念 .....	69
4.1.1 信道编码的分类 .....	69
4.1.2 与纠错编码有关的基本概念 .....	71
4.1.3 检错与纠错原理 .....	76
4.1.4 检错与纠错方式和能力 .....	77
4.2 线性分组码 .....	79
4.2.1 线性分组码的基本概念 .....	79
4.2.2 生成矩阵和一致校验矩阵 .....	81
4.2.3 线性分组码的译码 .....	87
4.2.4 线性分组码的纠错能力 .....	90
4.2.5 汉明码 .....	92
4.3 循环码 .....	94
4.3.1 循环码的多项式描述 .....	95
4.3.2 循环码的生成矩阵 .....	96
4.3.3 系统循环码 .....	99
4.3.4 多项式运算电路 .....	99
4.3.5 循环码的编码电路 .....	101
4.3.6 循环码的译码电路 .....	103
4.4 常用的循环码 .....	106
4.4.1 循环冗余校验码 .....	106

4.4.2	BCH 码*	108
4.4.3	RS 码*	111
4.5	卷积码	113
4.5.1	卷积码的编码	113
4.5.2	卷积码的译码	121
	思考题与习题	126
<b>第 5 章</b>	<b>信源编码</b>	129
5.1	信源编码器和无失真信源编码定理	129
5.1.1	码的分类	131
5.1.2	码树	133
5.1.3	Kraft 不等式	135
5.1.4	无失真信源编码定理(香农第一定理)	136
5.2	变长编码	137
5.2.1	香农码	138
5.2.2	费诺码	139
5.2.3	霍夫曼码	141
5.3	实用信源编码方法	146
5.3.1	游程编码	146
5.3.2	算术编码	149
5.3.3	预测编码	155
5.3.4	变换编码	157
	思考题与习题	164
<b>第 6 章</b>	<b>信息率失真函数</b>	169
6.1	失真测度	170
6.1.1	系统模型	170
6.1.2	失真度和平均失真度	170
6.2	信息率失真函数及其性质	172
6.2.1	信息率失真函数的定义	172
6.2.2	信息率失真函数的性质	172
6.3	等概率、对称失真信源的信息速率失真函数	175
6.4	保真度准则下的信源编码定理	177
6.5	限失真信源编码(香农第三定理)	178
	思考题与习题	178
<b>第 7 章</b>	<b>现代信道编码技术</b>	180
7.1	Turbo 码	180
7.1.1	Turbo 码的提出	180
7.1.2	Turbo 码编码器	181
7.1.3	Turbo 码译码器	181
7.2	LDPC 码	183
7.2.1	LDPC 码的提出	183
7.2.2	LDPC 码基本概念	183
7.2.3	规则 LDPC 码	185

7.2.4 非规则 LDPC 码 .....	186
7.2.5 准循环 LDPC 码 .....	187
7.2.6 重复累积 LDPC 码 .....	189
7.2.7 LDPC 码译码算法 .....	192
思考题与习题 .....	195
<b>第 8 章 MATLAB 在信息论与编码分析中的应用 .....</b>	<b>196</b>
8.1 MATLAB 基础 .....	196
8.1.1 MATLAB 语言特点 .....	197
8.1.2 MATLAB 运行环境简介 .....	198
8.1.3 MATLAB 基础 .....	210
8.2 MATLAB 在信息理论分析中的应用 .....	225
8.2.1 离散信源的 MATLAB 分析 .....	225
8.2.2 离散信道的 MATLAB 分析 .....	225
8.2.3 应用 MATLAB 进行信息理论分析的实例 .....	226
8.3 MATLAB 在编码技术分析中的应用 .....	227
8.3.1 信源编码技术的 MATLAB 分析 .....	227
8.3.2 信道编码技术的 MATLAB 仿真 .....	228
8.3.3 应用 MATLAB 进行编码技术分析的实例 .....	245
思考题与习题 .....	249
<b>参考文献 .....</b>	<b>250</b>

现代科学技术的飞速发展使人们对周围世界的认识和理解不断加深,特别是 20 世纪 60 年代以来,计算机技术的迅猛发展、计算机及相关设备的迅速更新换代和个人微型计算机的普及,极大地提高了人们处理信息、存储信息、控制和管理信息的能力,人类社会进入了信息时代。作为现代科学技术基础理论之一的信息论在各个领域的应用和推广使人们对许多经典的概念有了全新的解释,使过去曾经不确切地描述有了精确的定量分析方法。

信息论是在长期通信工程的实践中,由通信技术、概率论、随机过程和数理统计等相结合逐步发展起来的一门学科。通常公认信息论的奠基人是美国科学家香农(C. E. Shannon),他于 1948 年发表的著名论文《通信的数学理论》,为信息论的诞生和发展奠定了理论基础。

信息理论在学术界引起了巨大的反响,在香农信息论的指导下,为提高通信系统信息传输的有效性和可靠性,人们在信源编码和信道编码两个领域进行了卓有成效的研究,取得了丰硕的成果。随着信息理论的迅猛发展和信息概念的不断深化,信息论所涉及的内容早已超越了通信工程的范畴,进入了信息科学这一更广、更新的领域,并渗透许多学科,得到多个领域的科学工作者的重视。

#### 本章重点内容:

- 信息的定义及其基本特性,信息论与编码的发展历史;
- 数字通信系统模型;
- 信息论与编码技术所研究的主要内容;
- 香农信息论对信源编码和信道编码研究的指导意义。

## 1.1 基本概念

### 1.1.1 信息的一般概念

当今社会,在各种生产、科学研究和社会活动中,无处不涉及信息的交换和利用。可以说,人们正处在“信息社会”中。通过电话、电报、传真和电子邮件,可以自由地交流信息;通过报纸、书刊、电子出版物和互联网等媒介,可以有选择地获取大量信息;通过电台、电视台等视听媒体,可以“身临其境”地感受最新信息。但以上所述还远不能概括信息的全部含义:四季交替透露的是自然界的消息,而牛顿定律揭示的是物体运动内在规律的消息。信息含义之广几乎可以涵盖整个宇宙,且内容庞杂、层次混叠、不易理清。因此,迅速获取信息,正

确处理信息,充分利用信息,既能促进科学技术和国民经济的飞跃发展,又能在各种形式的竞争中占得先机。

如今有关信息的新名词、新术语层出不穷,信息产业在社会经济中所占份额也越来越大,信息基础设施建设速度之快成了当今社会的重要特征之一,物质、能源、信息构成了现代社会生存发展的三大基本支柱。

信息的价值在于它为人们能动地改造外部世界提供了可能,信息所揭示的事物运动规律为人们应用这些规律提供了可能,而信息所描述的事物状态也为人们推动事物向有利的方向发展提供了可能。掌握的资源 and 能量越多,面对同样的信息时人们能用以改造世界的可能性也越大。今天人们所掌握的物质力量比过去增大了不知多少倍,因此,信息对于当今社会的发展和人们生活的重要性较之几百年前、几十年前甚至十几年前都有很大的提高。这是信息社会的一个重要特征。

信息的重要性不言而喻,那么,如此神通广大、无处不在而又无所不能的信息究竟是什么呢?

信息是信息论中最基本、最重要的概念,既抽象又复杂。关于信息的科学定义,到目前为止,国内外已提出近百种,它们从不同的侧面和不同的层次来揭示信息的本质。从本质的意义上说,信息是人类社会活动所产生的各种状态和消息的总称,信息是人们对客观事物运动规律及其存在状态的认识。

在信息论和通信理论中经常会遇到信息、消息和信号这三个既有联系又有区别的名词。在学习信息论与编码技术之前,先介绍这几个基本概念。

对信息、消息和信号的定义比较如下。

**信息:**信息是任何随机事件发生所包含的内容。人们在对外部世界的观察中获得信息,信息是抽象的意识或知识,它是看不见、摸不着的。而且信息仅仅与随机事件的发生相关,非随机事件的发生不包含任何信息。从这一点上可以得知,信息量的大小与随机事件发生的概率有直接的关系,概率越小的随机事件一旦发生,它所包含的信息量就越大,而出现概率大的随机事件一旦发生,它所包含的信息量就越小。

**消息:**消息是信息的载体。在世界各地的人要想知道其他地方发生事情的内容,只能从各种各样的消息中得到,这些消息可以是广播中的语言、报纸上的文字、电视中的图像或互联网上的文字与图像等。可见,消息是具体的,它载荷信息,但它不是物理性的。信息只与随机事件的发生有关。每时每刻在世界上的每个地方,都会有各种事件发生,这些事件的发生绝大多数是随机的,即这些随机事件的消息中含有信息;如果事件的发生不是随机而是确定的,那么该消息中就不含信息,该消息的传输也就失去了意义。

**信号:**信号是消息的物理体现。为了在信道上传输消息,就必须把消息加载(调制)到具有某种物理特征的信号上去。信号是信息的载体,是具有物理性的,如电信号、声信号、光信号等。以人类的语言为例,当人们说话时,发出声信号,这种声信号经过麦克风的转换变成了电信号。这里的声信号和电信号都是所指的信号。

按照信息论的观点,信息不等于消息。在日常生活中,人们往往对消息和信息不加区别,认为得到了消息,就是获得了信息。例如,当人们收到一封电报,接到一个电话,收听了广播或看了电视等以后,就认为获得了“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,信息与消息有着密切的联系。但是,信息与消息并不等同。人

们收到消息后,如果消息告诉了原来不知道的新内容,会感到获得了信息,而如果消息是基本已经知道的内容,得到的信息就不多。所以信息应该是可以测度的。

在电报、电话、广播、电视(也包括雷达、导航、遥测)等通信系统中传输的是各种各样的消息。这些被传送的消息有着各种不同的形式,如文字、数据、语言、图像等。所有这些不同形式的消息都是能被人们的感觉器官所感知的,人们通过通信,接收到消息后,得到的是关于描述某事物状态的具体内容。例如,电视中转播球赛,人们从电视图像中看到了球赛的进展情况,而电视的活动图像则是对球赛运动状态的描述。当然,消息也可用来表述人们头脑里的思维活动。例如,朋友给您打电话说:“我想去北京”,您从这条消息得知了您的朋友的想法,该语言消息反映了人的主观世界——大脑物质的思维运动所表现出来的思维状态。

因此,用文字、符号、数据、语言、音符、图形、图像等能够被人们的感觉器官所感知的形式,把客观物质运动和主观思维活动的状态表达出来就成为消息。可见,消息中包含信息,是信息的载体,得到消息,进而获得信息。

同一则信息可用不同的消息形式来载荷,如前所述的球赛进展情况可用电视图像、广播语言、报纸文字等不同消息来表述。而一则消息也可载荷不同的信息,它可能包含非常丰富的信息,也可能只包含很少的信息。因此,信息与消息是既有区别又有联系的。

在各种实际通信系统中,为了克服时间或空间的限制而进行通信,必须对消息进行加工处理。把消息变换成适合于信道传输的物理量(如声、光、电等),这种物理量即为信号。信号携带着消息,是消息的运载工具。如前例中,携带球赛进展情况的电视图像转换成电信号,电信号经过调制变成高频调制电信号,才能在信道中传输;在通信系统的接收端,通过解调还原出原始电信号,在电视屏幕中呈现给观众,从而使观众获得信息。

同样,同一消息可用不同的信号来表示,同一信号也可表示不同的消息。例如,红、绿灯信号:若在十字路口,红、绿灯信号表示能否通行的信息;而在电子仪器面板上,红、绿灯信号却表示仪器是否正常工作或者表示高低电压等信息。所以,信息、消息和信号是既有区别又有联系三个不同的概念。

从以上的讨论中可以看到,信息、消息和信号之间有着密切的关系。信息是一切通信系统所要传递的内容;而消息作为信息的载体可能是一种“高级”载体;信号作为消息的物理体现,是信息的一种“低级”载体。作为系统设计人员,所接触的只是信号,而这种信号最终要变成消息的形式才能被大众接受。

信息的基本概念在于它的不确定性,任何已确定的事物都不含有信息。信息具有以下特征。

- 信息是可以识别的。信息离不开物理载体,人们可以通过对这些物理载体的识别来获得信息。有些可以用人的感官直接识别信息,例如承载于语言、文字中的信息可以直接用耳、目接收进而识别;而有些则需借助于各种传感器间接识别信息,例如在遥感测量中要利用对电磁波敏感的传感器来间接进行。
- 信息是可以存储的。信息可以用多种方式存储起来,在需要的时候把存储的信息调取出来。相同的信息可以用文字的形式记录在书刊笔记中,也可以用录音、录像的方式存储在磁性介质中,或者利用计算机存储设备存储起来。
- 信息是可以传递的。信息可以通过多种途径进行传递,人与人之间的信息传递,既

可以通过语言、文字,也可以通过体态、动作或表情;社会规模的信息传递,常通过报纸、杂志、电话、广播、电视和网络等。从原则上来说,各种物质的运动形式都可以用于信息的传递。

- 信息是可以量度的。信息量有大小的差别,出现概率越大的随机事件一旦发生,它所包含的信息量就越小;反之,出现概率越小的随机事件一旦发生,它所包含的信息量就越大。
- 信息是可以加工的。人们在收到各种原始信息之后,经过各种方式的加工可以产生新的信息,如研究人员通过收集资料或实验获得的原始信息,经过加工处理可能提出新的见解;计算机对输入的信息通过加工处理,可为人们提供更有意义的结果。
- 信息是可以共享的。信息可以像实物一样作为商品出售,但信息知识特性使其交易又不同于一般的实物交易,信息交易后,信息出售者与信息购买者共同享有信息。
- 信息的载体是可以转换的。同样内容的信息,可以有不同的形态,可以被包含在不同的物体变化之中,可以从一种形态转换到另一种形态。如用感官识别出来的声音、味道、颜色等信息可以转换成语言、文字等形式。在这种转换中,信息的物理载体发生了变化,但信息的内容可以保持完好无损。信息的这个特性,为人们借助于仪器间接地识别信息提供了基础,也为信息的传递、存储和处理带来了方便。

### 1.1.2 香农信息定义

信息仅仅与随机事件的发生相关,用数学的语言来说,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——概率论和随机过程来测度不确定性的。若从直观概念来说,不确定性的。大小可以直观地看作事先猜测某随机事件是否发生的难易程度。

某一事物状态的不确定性的。大小,与该事物可能出现的不同状态数目和各状态出现的概率大小有关。既然不确定性的。大小能够测度,那么信息也是可以测度的。

信息如何测度?当人们收到一封电报,或者听了广播、看了电视,到底能得到多少信息量?由于信息量与不确定性消除的程度有关,可用消除不确定性的多少来测度信息量。

**【例 1.1.1】** 假设有 A、B 两个布袋,在袋内各装有大小均匀、手感完全一样的 100 个球。A 袋内有红、白球各 50 个,B 袋内有红、白、蓝、黑 4 种球各 25 个。如果随意从 A 袋或 B 袋中取出一个球,猜测取出的球是什么颜色。该事件当然具有不确定性,按给定条件,显然,“从 A 袋中取出是红球”要比“从 B 袋中取出是红球”要容易得多。因为,在 A 袋中只是在“红”、“白”两种颜色中选择一种,而且“红”与“白”机会均等,即概率为  $1/2$ 。而在 B 袋中,红球只占  $1/4$ ,取出红球的可能性就小。因此,事件“从 A 袋中取出的是红球”比事件“从 B 袋中取出的是红球”发生的不确定性要小。从这个例子可知,不确定性的。大小与可能发生的不同状态数目和各状态发生的概率有关。

**【例 1.1.2】** 足球的魅力在于其比赛结果的不确定性。如果实力接近的两个队进行比赛,在比赛之前,很难预测谁能获得胜利,所以这个事件的不确定性很大,当得知比赛结果时,就会获得较大的信息量。如果实力相差悬殊的两个队进行比赛,一般结果是强队取得胜利,所以当得知比赛结果是强队获胜时,人们并不觉得奇怪,因为结果与猜测是一致的,所以消除的不确定性较小,获得的信息量也较小;当得知比赛结果是弱队取胜时,人们会感到非



常惊讶,认为出现了“黑马”,这时将获得很大的信息量。

由上述两个例子可知:某一事物状态的不确定性的尺寸,与该事物可能出现的不同的状态数目和各状态出现的概率的大小有关。某一事物状态出现的概率越小,其不确定性越大,一旦出现,带来的信息量就越大;反之,某一事物状态出现的概率接近于1,即预料中肯定会发生的事件,那它的不确定性就接近于0,如果出现,带来的信息量就很小。

香农信息反映的就是事物的不确定性。在香农著名的论文《通信的数学理论》中,香农根据概率测度和数理统计学系统地研究了通信中的基本问题,并给出了信息的定量表示,得出了带有普遍意义的重要结论,由此奠定了现代信息论的基础。

把某事物各种可能出现的不同状态,即所有可能选择的消息的集合,称为样本空间,用 $X$ 表示。在样本空间中,每个可能选择的消息是这个样本空间的一个元素。对于离散消息的集合,概率测度就是对每一个可能选择的消息指定一个概率(非负,且总和为1)。一个样本空间和它的概率测度一起构成一个概率空间。

一般概率空间用 $[X, P]$ 来表示。在离散情况下, $X$ 可写成 $\{a_1, a_2, \dots, a_q\}$ 。在样本空间中,选择任一元素 $a_i$ 的概率用 $p_X(a_i)$ 表示,这里下标 $X$ 表示所考虑的概率空间是 $X$ 。如果不会引起混淆,下标可以略去,写成 $p(a_i)$ 。所以在离散情况下,概率空间表示为

$$\left[ \begin{array}{c} X \\ p(x) \end{array} \right] = \left[ \begin{array}{cccc} a_1 & a_2 & \dots & a_q \\ p(a_1) & p(a_2) & \dots & p(a_q) \end{array} \right]$$

其中, $p(a_i)$ 为选择符号 $a_i$ 作为消息的概率,称为先验概率。在接收端,对是否选择这个消息(或符号) $a_i$ 的不确定性是与 $a_i$ 的先验概率成反比的,即对 $a_i$ 的不确定性可表示为先验概率 $p(a_i)$ 的某一函数。

在概率空间中, $a_i$ 本身携带的信息量(香农信息)定义为

$$I(a_i) = \log \frac{1}{p(a_i)}$$

$I(a_i)$ 称为消息(符号) $a_i$ 的自信息。

由上式可知: $a_i$ 出现的先验概率 $p(a_i)$ 越大,则其自信息 $I(a_i)$ 越小;反之, $a_i$ 出现的概率越小,则自信息越大。因此,自信息可描述消息 $a_i$ 出现的先验不确定性,其关系如图1.1.1所示。

由于在信道中存在干扰,假设接收端收到的消息(符号)为 $b_j$ , $b_j$ 可能与 $a_i$ 相同,也可能与 $a_i$ 不同,则条件概率 $p(a_i/b_j)$ 反映接收端收到消息 $b_j$ 而发送端发出的是 $a_i$ 的概率,此概率称为后验概率。这样,接收端收到 $b_j$ 后,发送端发送的符号是否为 $a_i$ 尚存在的不确定性应是后验概率的函数,即

$$\log \frac{1}{p(a_i/b_j)}$$

于是,收信者在收到消息 $b_j$ 后,已经消除的不确定性为先验的不确定性减去尚存在的不确定性。

这就是收信者获得的信息量,即

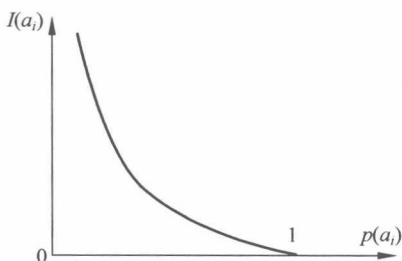


图 1.1.1 自信息与其对应的先验概率的关系