

# 数字媒体传播 技术概论



胡兵 编著



清华大学出版社

# 数字媒体传播 技术概论



胡兵 编著

清华大学出版社  
北京

## 内 容 简 介

本书依据数字媒体技术特点、发展历程与趋势,主要从“理论—技术—应用—案例”四个维度进行阐述。全书共分为8章,从逻辑上分为三大部分。第一部分(第1章)主要介绍数字信息的一些基础理论;第二部分(第2章至第6章)分别介绍数字广播、数字电视、数字出版、互联网和移动互联网这些媒介形态的技术发展和应用案例;第三部分(第7章至第8章)则介绍了数字媒体未来发展所依托的云计算和大数据方面的知识。

本书从文科读者的认知和思维角度深入浅出地阐述,重点在于概念解释、技术应用和案例分析,力图使读者全面了解和正确理解数字媒体技术的基本知识,具有一定的系统性、先进性和借鉴性。本书不仅适合高等院校新闻与传播学院本科生和研究生作为学习教材使用,还可供从事数字媒体技术研究、开发和应用的工程技术人员以及数字媒体产业从业人员等学习参考。

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

数字媒体传播技术概论/胡兵编著. —北京:清华大学出版社,2015

ISBN 978-7-302-40886-4

I. ①数… II. ①胡… III. ①数字技术—应用—传播媒介—研究  
IV. ①G201.6\*239

中国版本图书馆CIP数据核字(2015)第165854号



责任编辑:朱敏悦

封面设计:汉风唐韵

责任校对:王荣静

责任印制:何 芊

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)

印 装 者:三河市金元印装有限公司

经 销:全国新华书店

开 本:170mm×240mm 印 张:21.5 字 数:429千字

版 次:2015年9月第1版 印 次:2015年9月第1次印刷

印 数:1~2000

定 价:59.00元

产品编号:060064-01

# 序 言

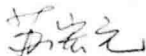
胡兵老师毕业于华南理工大学电子与信息学院，曾在广东省广电系统工作13年，于2008年调回华南理工大学，加盟新闻与传播学院。胡老师工科博士出身，有媒体技术工作经验，潜心专研教学和科研多年，再加勤奋和努力，助推其学术研究呈现厚积薄发的态势，成果丰硕。除了专业论文屡见专业学术期刊（包括在国外学术期刊发表的英语论文），2011年出版了《传播技术原理与应用》一书，如今新著《数字媒体传播技术概论》又顺利完工，真是可喜可贺。

相比《传播技术原理与应用》一书，这本书聚焦于数字媒体传播技术，尤其是移动“互联网技术与应用”、“云计算、云存储和云媒体”以及“大数据及其应用”这样一些信息与传播技术革命发展的前沿，是一本与时俱进、专业性很强的著作。此书不但内容新颖扎实，而且保持了胡兵老师作为工科博士思维严谨、叙述明晰的写作风格，其“理论—技术—应用—案例”交叉融合的写作方法也颇具特色，便于读者理解和掌握艰深的专业知识。

胡兵博士的成果，不但是其个人研究方向的产物，也是华南理工大学新闻与传播学院长期致力于打造“理科中的文科”所取得的实绩，甚至可以说是对新闻与传播学科变革所作的积极回应。毕竟，媒介与传播技术已是当今新闻与传播教学和研究的重要维度，理应成为广大新传学子专业知识结构的基本要素。

对胡兵老师的研究领域我不太熟悉，但仍欣然答应他写此短序，以此作为对自己的鞭策。

华南理工大学新闻与传播学院院长



2015年6月21日于广州小谷围

# 前 言

“媒介技术决定论”者马歇尔·麦克卢汉对现代传播媒介的分析深刻地改变了人们对 21 世纪生活的观念，特别是当代青年人的观念，他所预言的地球村已经变成了现实。虽然有学者批评麦克卢汉过多强调了传播技术的作用，而忽略了传播关系的支配作用，以及社会主体和社会环境的互动关系。但是，近年来以计算机为主体的互联网技术的发展和以手机为主体的移动互联网的兴起，正以前所未有的威力推动着人类经济和社会高速发展，将人类社会推到一个崭新的信息化时代——大数据时代。我们不可否认，近二十年来传播学的发展是由传播技术的发展引领的。

编者具有十多年的理工科教育背景，对数字媒体技术有着与文科教育背景的教师和学者不同的理解。本书依据数字媒体技术的特点、发展历程与趋势，主要从“理论—技术—应用—案例”四个维度进行阐述。本书的撰写结合了编者多年在教学中的经验与体会以及科研心得，并参考和汲取了与数字媒体技术相关领域的大量教材、文献及技术资料。

本书共分为 8 章，从逻辑上分为三大部分。第一部分（第 1 章）主要介绍数字信息的一些基础理论；第二部分（第 2 章至第 6 章）分别介绍数字广播、数字电视、数字出版、互联网和移动互联网这些媒介形态的技术发展和应用案例；第三部分（第 7 章至第 8 章）介绍了数字媒体未来发展所依托的云计算和大数据方面的知识。

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狭义的通信工程范畴，进入了信息科学这一更广阔、更新兴的领域，甚至包括心理学、语言学、神经心理学、语义学、社会学等学科领域。传播技术正是建立在以经典信息论为核心的通信理论基础上的，因此本书从介绍信息论的基础知识开篇。除此之外，在第 1 章中还介绍了数字信号的特点及模拟信号的数字化过程、数字通信的优点、数据压缩的原理和标准等方面的内容。

第 2 章介绍数字广播技术与应用。在简单介绍电磁波的性质、传播方式及其用途，模拟调制和数字调制的原理等基础知识基础上，重点介绍了数字音频广播和数字多媒体广播的技术发展概貌和发展趋势，网络广播的系统架构与制作，广播电台的数字广播实践案例等。

第3章，数字电视技术与形态，是第二部分的重点之一。首先介绍了数字电视的优点、数字电视系统的构成，以及数字电视的关键技术。其次阐述了广电系统作为“救命稻草”的高清互动数字电视发展所面临的机遇与挑战。在第2节还详细介绍了IPTV、OTTTV、手机电视等其他数字电视形态的特征、发展机遇和国内外的发展案例。

第4章介绍数字出版技术与产业发展。在这一章中，除了介绍数字出版的含义、范畴、特征和产品形态，跨媒体出版的工作流程和跨媒体出版系统外，还简明扼要地介绍了XML、数字版权保护等技术方面知识。本章最后总结了国内数字出版产业发展现状和归纳出国外数字出版产业发展的几种模式，以及报刊媒体成功转型案例的分析。

第5章，互联网新技术与应用，也是第二部分的重点之一。介绍了互联网的通信原理和IPv4协议，以及正在过渡的IPv6协议的新特性和地址结构。本章中还专门有一节详细介绍了互联网营销的概念、网络营销的推广方式、网络营销的衡量标准、网络营销效果的监测工具、互联网营销的发展趋势等内容。本章最后一节简单介绍了互联网的延伸——物联网的概念、体系结构、应用领域，以及车联网、可穿戴智能设备等物联网未来发展方向等方面的内容。

第6章，移动互联网技术与应用，也是第二部分的重点之一。本章从介绍移动通信的发展历程开始，并穿插简述了CDMA的关键技术。接着重点阐述了移动互联网的特征、业务应用和技术架构，以及移动互联网的终端软件技术、网络平台技术和应用平台技术。最后介绍了几种流行的Mobile2.0应用所依托的技术背景。

第7章，云计算、云存储和云媒体，属于第三部分。本章介绍了云计算的概念、特征、服务类型、体系结构和关键技术；云存储的概念、特点、服务类型和运营案例。本章还引入了“云媒体”的概念，并介绍了云媒体的两种具体形态——云电视和云报纸。

第8章，大数据及其应用，是第三部分的重点。本章简述了大数据的含义及其特征，以及大数据带来的思维变革、价值和挑战。在简单回顾数据库和数据挖掘知识之后，第2节介绍了大数据的技术体系及其关键技术。最后一节介绍了大数据在数据新闻、社会管理和网络营销三个行业的应用情况。

本书尽可能考虑到数字媒体技术与应用的发展和特点，以及所涉及的基本知识和成功案例，并且从文科读者的认知和思维角度深入浅出地进行阐述，具有一定的系统性、先进性和借鉴性。本书不仅适合高等院校新闻与传播学院本科生和研究生作为学习教材使用，还可供从事数字媒体技术研究、开发和应用的工程技术人员以及数字媒体产业从业人员等学习参考。

借此书出版之际，首先要感谢华南理工大学新闻与传播学院的两任院长——李幸教授和苏宏元教授。他们秉承着“工科学校中的文科，文科学院中的工科”的建院理念，坚持在学院对本科生开设《传播技术基础》这门专业基础必修课，并积极推动本书的编撰出版，希望有一本适合文科学生了解传播技术的参考书。2013年华南理工大学新闻与传播学院申报并成功建立了广东省新媒体与品牌传播创新应用重点实验室（项目编号：2013WSYS0002），将计算机软硬件技术、传播技术等理论方法应用于新媒体与品牌传播的模式之中，本书也是该重点实验室“大数据与新媒体传播效果研究”研究方向的研究成果。

这里还要感谢一位特殊的人，群邑一传立（中国）公司的互动营销主管林德仁先生，他曾是华南理工大学新闻与传播学院的学生，后去美国读了新媒体硕士，并在美国联合创办了 Soholute&HDL 公司，现回国从事互联网营销工作。他基于国内外教育的经历和在国外传媒行业创业工作的体会，认为传媒学生很有必要了解传播技术的相关知识，对本书的内容框架提出了许多建设性的意见，并撰写了“5.3 互联网营销”一节。

研究生何鑫为本书的撰写查阅了大量的参考书籍和资料，胡长建博士、叶昭高级工程师等人为本书提供了宝贵的建议和资料。本书的出版还得到了清华大学出版社编辑的大力支持。在此对提供支持的所有人表示深深的感谢！

鉴于数字媒体技术发展迅猛，涉及范围广泛，且对数字媒体的认识与界定也是仁者见仁、智者见智，加上编者水平有限，书中的不足之处恳请广大读者和同行批评指正。

编 者

2015年3月于广州大学城

# 目 录

<b>第 1 章 数字信息基础理论</b> .....	1
1.1 信息的概念及其测度 .....	1
1.1.1 信息的概念 .....	1
1.1.2 信息的测度 .....	3
1.1.3 信息论研究的对象和内容 .....	7
1.2 模拟信息与数字信息 .....	11
1.2.1 模拟与数字的区别 .....	11
1.2.2 数字信号与数字传输 .....	13
1.2.3 模拟信号的数字化 .....	15
1.3 数字信息的编码 .....	20
1.3.1 几种常用的编码方法 .....	20
1.3.2 数字信息的压缩 .....	24
1.4 分辨率与彩色显示 .....	29
1.4.1 数字信息的分辨率 .....	29
1.4.2 数字信息的彩色显示 .....	32
<b>第 2 章 数字广播技术与应用</b> .....	38
2.1 数字广播技术 .....	38
2.1.1 无线广播的基础理论 .....	38
2.1.2 数字广播技术概要 .....	45
2.1.3 数字多媒体广播应用——交通广播信息系统 .....	52
2.2 网络广播 .....	55
2.2.1 网络广播的分类与传播特点 .....	55
2.2.2 网络广播的架构与制作 .....	56
2.2.3 网络广播的创新案例 .....	58
<b>第 3 章 数字电视技术与形态</b> .....	64
3.1 数字电视系统及关键技术 .....	64



## 数字媒体传播技术概论

3.1.1	数字电视标准和优点	65
3.1.2	数字电视前端系统	67
3.1.3	数字电视关键技术	70
3.1.4	数字电视发展的挑战和机遇	80
3.1.5	高清互动数字电视	84
3.2	数字电视其他形态	91
3.2.1	IPTV 技术与业务	92
3.2.2	OTT TV 互联网电视	95
3.2.3	手机电视	101
<b>第 4 章</b>	<b>数字出版技术与产业发展</b>	<b>108</b>
4.1	数字出版	108
4.1.1	数字出版的含义与范畴	108
4.1.2	数字出版的特征与产品形态	110
4.2	数字出版的关键技术	111
4.2.1	XML 简介	112
4.2.2	数字版权保护技术	117
4.3	跨媒体出版	128
4.3.1	跨媒体出版的定义与特征	128
4.3.2	跨媒体出版工作流程	129
4.3.3	跨媒体出版系统	132
4.4	数字出版产业发展现状与模式创新	133
4.4.1	国内数字出版产业发展现状	134
4.4.2	国外数字出版产业的发展模式	138
4.4.3	报刊媒体的移动媒体运营策略分析——以现代传播集团为例	142
<b>第 5 章</b>	<b>互联网新技术与应用</b>	<b>148</b>
5.1	互联网通信原理	148
5.1.1	计算机网络的体系结构	148
5.1.2	Internet 与 TCP/IP 协议	153
5.1.3	互联网新协议——IPv6	163
5.2	Web 3.0 的含义与特征	171
5.2.1	Web 3.0 的含义	172
5.2.2	Web 3.0 的特征	172

5.2.3 Web 3.0 的核心——语义网 .....	173
5.3 互联网营销 .....	175
5.3.1 互联网营销概念与策划 .....	176
5.3.2 互联网营销常用的推广方式 .....	177
5.3.3 互联网营销衡量标准 .....	183
5.3.4 互联网营销监测 .....	184
5.3.5 互联网营销未来发展趋势 .....	187
5.4 互联网的延伸——物联网 .....	190
5.4.1 物联网的概念与内涵 .....	190
5.4.2 物联网体系架构 .....	192
5.4.3 物联网的应用领域与最新发展 .....	194
<b>第 6 章 移动互联网技术与应用 .....</b>	<b>200</b>
6.1 移动通信技术简述 .....	200
6.1.1 移动通信发展简史 .....	200
6.1.2 移动通信系统的分类 .....	206
6.1.3 CDMA 系统的关键技术 .....	207
6.2 移动互联网技术 .....	210
6.2.1 移动互联网的特征、业务应用和技术架构 .....	211
6.2.2 终端操作系统 .....	217
6.2.3 网络平台技术 .....	224
6.2.4 应用平台技术 .....	228
6.2.5 Mobile 2.0 技术的应用 .....	236
<b>第 7 章 云计算、云存储和云媒体 .....</b>	<b>246</b>
7.1 云计算 .....	246
7.1.1 云计算的概念与特点 .....	246
7.1.2 云计算的类型 .....	249
7.1.3 云计算的体系结构 .....	252
7.1.4 云计算的关键技术 .....	253
7.1.5 几种云计算平台介绍 .....	261
7.2 云存储 .....	263
7.2.1 云存储的起源 .....	264
7.2.2 云存储的概念、特点和分类 .....	264

# 数字媒体传播技术概论

7.2.3 云存储应用案例 .....	267
7.3 云媒体 .....	269
7.3.1 云媒体的概念和特征 .....	269
7.3.2 云媒体的具体形态 .....	269
<b>第8章 大数据及其应用 .....</b>	<b>274</b>
8.1 大数据及其特征 .....	274
8.1.1 大数据的定义 .....	274
8.1.2 大数据的特征 .....	276
8.1.3 大数据思维 .....	278
8.1.4 大数据的价值 .....	279
8.1.5 大数据的挑战 .....	281
8.2 大数据技术 .....	282
8.2.1 数据库与数据挖掘 .....	283
8.2.2 大数据技术体系 .....	291
8.2.3 大数据标准体系 .....	296
8.3 大数据在行业中的应用 .....	298
8.3.1 数据新闻学 .....	298
8.3.2 社会管理 .....	305
8.3.3 网络广告 .....	310
<b>主要参考文献 .....</b>	<b>315</b>
<b>主要缩写词索引表 .....</b>	<b>323</b>

# 第 1 章 数字信息基础理论



## 学习目标

- 理解信息的概念，并掌握信息量和信息熵的计算、通信系统的通用模型
- 理解模拟与数字的区别，掌握模拟信号数字化的过程
- 理解数字通信的优点，了解数字信号抗干扰的原理
- 了解数据压缩的原理和标准，掌握 Huffman 编码的过程
- 掌握各种分辨率和图像彩色模式的定义

源自香农和韦弗 (Shannon and Weaver, 1949 年) 发展的信息论被誉为引领媒介研究的主导范式中十分有影响的理论要素之一，该理论致力于传播渠道传递信息的技术功效研究。他们发展出一种分析信息传送的模式，这种“传播”模式并非直接与大众传播有关，但之后它作为一个万能的模式解释了众多人类传播的过程，尤其是信息传播的效果。因此，本书从介绍信息论的基础知识开篇。

## 1.1 信息的概念及其测度

在当今“信息社会”中，人们在各种生产、科学研究和社会活动中，无处不涉及信息的交换和利用。迅速获取信息、正确处理信息、充分利用信息、可以促进科学技术和国民经济的飞跃发展。可见，信息的重要性是不言而喻的。

那么，什么是信息呢？

### 1.1.1 信息的概念

“信息”一词在英文、法文、德文、西班牙文中均是“Information”，日文中为“情报”，我国台湾称之为“资讯”，我国古代用的是“音信”。作为科学术语，“信息”这一概念最早出现在哈特莱 (R. V. L. Hartley) 于 1928 年撰写的《信息传输》一文中，他认为，发信者所发出的信息，就是他在通信符号表中选择符

## 数字媒体传播技术概论

号的具体方式，并主张用所选择的自由度来度量信息。此后，许多研究者从各自的研究领域出发，给出了不同的定义，具有代表意义的表述如下：

信息论的奠基人香农（C. E. Shannon）在1948年发表的一篇著名论文《通信的数学理论》中，从研究通信系统传输的实质出发，给出了信息的科学定义，并进行了定性和定量的描述。他认为“信息是事物运动状态或存在方式的不确定性的描述”，这一定义被人们看作经典定义并加以引用。

控制论创始人维纳（Norbert Wiener）认为“信息是人们在适应外部世界，并使这种适应反作用于外部世界的过程中，同外部世界进行互相交换的内容和名称”，它也被作为经典定义加以引用。

美国信息管理专家霍顿（F. W. Horton）给信息下的定义是：“信息是为了满足用户决策的需要而经过加工处理的数据。”简单地说，信息是经过加工的数据，或者说，信息是数据处理的结果。

经济管理学家认为“信息是提供决策的有效数据”。

电子学家、计算机科学家认为“信息是电子线路中传输的信号”。

我国著名的信息学专家钟义信教授认为“信息是事物存在方式或运动状态，以这种方式或状态直接或间接的表述”。

根据对信息的研究成果，科学的信息概念可以概括如下：

“信息是对客观世界中各种事物的运动状态和变化的反映，是客观事物之间相互联系和相互作用的表征，表现的是客观事物运动状态和变化的实质内容。”<sup>①</sup>

信息一般具有如下一些特征：

- (1) 可识别。
- (2) 可转换。
- (3) 可传递。
- (4) 可加工处理。
- (5) 可多次利用（无损耗性）。
- (6) 在流通中扩充。

(7) 主客体二重性。信息是物质相互作用的一种属性，涉及主客体双方；信息表征信源客体存在方式和运动状态的特性，所以它具有客体性、绝对性；但接收者所获得的信息量和价值的大小，与信宿主体的背景有关表现了信息的主体性和相对性。

(8) 信息的能动性。信息的产生、存在和流通，依赖于物质和能量，没有物质和能量就没有能动作用。信息可以控制和支配物质与能量的流动。

---

<sup>①</sup>摘自百度百科

(9) 可共享性。这是信息与物质和能量的主要区别。

信息是信息论中最基本、最重要的概念，它是一个既抽象又复杂的概念。信息既是通信系统传输和处理的对象，又泛指人类传播的一切内容。人通过获得、识别自然界和社会的不同信息来区别不同事物，得以认识和改造世界。在一切通信和控制系统中，信息是一种普遍联系的形式。

### 1.1.2 信息的测度

根据香农有关信息的定义，信息如何测度呢？当人们收到一封电报，或听了广播，或看了电视，到底得到多少信息量呢？显然，信息量与不确定性消除的程度有关。消除多少不确定性，就获得多少信息量。那么，不确定性的数量能度量吗？

用数学的语言来讲，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——概率论和随机过程来测度不确定性的数量。直观来讲，不确定性的数量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。

举个例子，我们知道气象的状态有许多种。但是 1 月份广州的天气，经常出现的是“晴”、“晴间多云”或“多云”，其次是“多云转阴”、“阴”或“阴有小雨”等，而“小雪”这种天气状态出现的概率极小，“大雪”的可能性几乎为零。因此，在听天气预报前，我们大体上能猜测出天气的状况。由于出现“晴”、“晴间多云”或“多云”的可能性大，我们就比较能确定这些天气状况的出现。所以，当预报明天白天“晴”或“晴间多云”，我们并不觉得稀奇，因为和我们猜测的基本一致，所消除的不确定性要小，获得的信息量就不大。而出现“小雪”的概率很小，我们很难猜测它是否会出现，所以该事件的不确定性很大；如果预报是“阴有小雨”，我们就会大吃一惊，感到气候反常，这时就获得了很大的信息量。出现“大雪”的概率更小，它的不确定性更大；如果出现“大雪”的气象预报，我们会万分惊讶，这时将获得更大的信息量。由此可知，某一事物状态出现的概率越小，其不确定性越大；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零。

上面的例子告诉我们：某一事物状态的不确定性的数量，与该事物可能出现的不同状态数目以及各状态出现的概率大小有关。既然不确定性的数量能够度量，可见，信息是可以测度的。

### 1. 自信息

“自信息”一词由香农提出，用来衡量单一事件发生时所包含的信息量多寡。

由于种种原因（例如噪声太大），收信者接收到受干扰的消息后，对某消息发生的不确定性依然存在或者一点也未消除时，则收信者获得较少的信息量或者说一点也没获得信息。因此，获得信息量的大小，是与不确定性消除的多少有关。我们直观地把信息量定义为：

收到某消息获得的信息量 = 不确定性减少的量  
=（收到此消息前关于某事件发生的不确定性） -  
（收到此消息后关于某事件发生的不确定性）

在无噪声传输信道中，可以完全不失真地收到所发的信息，所以收到此消息后关于某事件发生的不确定性完全消除，此项为零。因此得

收到某消息获得的信息量 = 收到消息前关于某事件发生的不确定性  
= 信源输出的某消息中所含有的信息量

我们已经知道，事件发生的概率越小，我们猜测它有没有发生的困难程度就越大，不确定性也就越大；而事件发生的概率越大，我们猜测这事件发生的可能性就越大，不确定性也就越小。对于发生概率等于 1 的必然事件，就不存在不确定性。因此，某事件发生所含有的信息量应该是该事件发生的先验概率的倒数的某一函数。

根据概率函数的性质条件，我们可以从数学上证明这种函数形式是对数形式，因此，事件发生所含有的信息量  $I(a_i)$  为：

$$I(a_i) = \log \frac{1}{P(a_i)} \quad (1.1)$$

代表两种含义：当事件  $a_i$  发生以前，表示事件  $a_i$  发生的不确定性，称为该事件  $a_i$  的自信息；当事件  $a_i$  发生以后，表示事件  $a_i$  所含有（或所提供）的信息量。在无噪声信道中，事件  $a_i$  发生后，能正确无误地传输到收信者，所以  $I(a_i)$  可代表接收到消息  $a_i$  后所获得的信息量。

自信息采用的单位取决于对数所选取的底。由于  $P(a_i)$  是小于 1 的正数，又根据实际情况知道自信息  $I(a_i)$  也必然是正数，所以对数的底应选取大于 1 的任意数。如果取以 2 为底，则所得的信息量单位称为比特（Bit, Binary Unit 的缩写）；如果采用以 e 为底的自然对数，则所得的信息量单位称为奈特（Nat, Nature Unit 的缩写）；若采取以 10 为底的对数，则所得的信息量单位称为哈特（Hart, Hartley 的缩写）<sup>①</sup>。

一般情况，如果取以  $r$  为底的对数（ $r > 1$ ），则

$$I(a_i) = \log_r \frac{1}{P(a_i)} \quad (r \text{ 进制单位}) \quad (1.2)$$

<sup>①</sup>以纪念哈特莱首先提出用对数来度量信息。

通常都采用以 2 为底的对数,且为了书写简洁,把底数“2”略去不写。我们可以看到,如果  $P(a_i) = 1/2$ ,则  $I(a_i) = 1$  比特。所以 1 比特信息量就是两个互不相容的等可能事件之一发生时所提供的信息量。注意,这里的比特是抽象的信息量单位。与计算机术语中的“比特”的含义有所不同,它是代表二进制数字(Binary Digits)。这两种定义之间的关系是每个二进制数字所能提供的最大平均信息量为 1 比特。

**【例】**某离散信源<sup>①</sup>由 0, 1, 2, 3 四种符号组成,其概率场为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3/8 & 1/4 & 1/4 & 1/8 \end{bmatrix}$$

假定信源输出的每一个符号都是独立的,当信源输出的符号序列 S 为“113200”时,求其输出的信息量的大小。

解:根据式(1.1),信源输出的序列 S 的信息量为

$$\begin{aligned} I(S) &= \log \frac{1}{P(S)} = \log \frac{1}{P(1)} + \log \frac{1}{P(1)} + \log \frac{1}{P(3)} + \log \frac{1}{P(2)} + \log \frac{1}{P(0)} + \log \frac{1}{P(0)} \\ &= \log \frac{1}{1/4} + \log \frac{1}{1/4} + \log \frac{1}{1/8} + \log \frac{1}{1/4} + \log \frac{1}{1/4} + \log \frac{1}{3/8} + \log \frac{1}{3/8} \\ &= 2 + 2 + 3 + 2 + 1.415 + 1.415 = 11.83(\text{比特}) \end{aligned}$$

上式中第二个等号成立是因为符号的出现具有独立性。

## 2. 信息熵

自信息是指某一信源发出某一消息所含有的信息量,所发出的消息不同,它们所含有的信息量也就不同。所以自信息  $I(a_i)$  是一个随机变量,不能用它作为整个信源的信息测度。

我们定义自信息的数学期望(加权平均值)为信源的平均自信息量,即

$$H_r(X) = E\left[\log_r \frac{1}{P(a_i)}\right] = - \sum_{i=1}^q P(a_i) \log_r P(a_i) \quad (r \text{ 进制单位/符号}) \quad (1.3)$$

这个平均自信息量的表达式与统计物理学中热熵的表达式很相似。在统计物理学中,热熵是一个物理系统杂乱性(无序性)的度量,在概念上两者也有相似之处。因而借用“熵”这个词把  $H_r(X)$  称为信息熵。信息熵的单位由自信息的单位来决定,即取决于对数选取的底。一般选用以 2 为底时,信息熵写成  $H(X)$  形式,其中变量 X 是指某随机变量的整体。

信息论利用统计热力学中熵的概念,建立了对信息的量度方法。在统计热力

<sup>①</sup>离散是分散的意思,离散信源只能产生有限种符号,因此离散消息可以看作一种有限个状态的随机序列,如天气预报。



学中，熵是系统的无序状态的量度，即系统的不确定性的量度。一个系统越是有序，熵就越低；反之，一个系统越是混乱，熵就越高。所以，熵是系统有序化程度的一个度量。

信源的信息熵  $H$  是从整个信源的统计特性来考虑的。它是从平均意义上来说表征信源的总体信息测度的。对于某特定的信源，其信息熵是一个确定的数值。不同的信源因统计特性不同，其熵也不同。信息熵具有以下三种物理含义：

第一，信息熵  $H(X)$  是表示信源输出后，每个消息（或符号）所提供的平均信息量。

第二，信息熵  $H(X)$  是表示信源输出前，信源的平均不确定性。例如有两个信源，其概率空间分别为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0.99 & 0.01 \end{bmatrix}, \begin{bmatrix} Y \\ P(y) \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ 0.5 & 0.5 \end{bmatrix}$$

则信息熵分别为

$$H(X) = -0.99\log 0.99 - 0.01\log 0.01 = 0.08(\text{比特/符号})$$

$$H(Y) = -0.5\log 0.5 - 0.5\log 0.5 = 1(\text{比特/符号})$$

可见  $H(Y) > H(X)$ ，信源  $Y$  比信源  $X$  的平均不确定性要大。

第三，用信息熵  $H(X)$  来表征变量  $X$  的随机性。如上例中，变量  $Y$  中的两个事件是等概率的，所以其随机性大，而变量  $X$  中一个事件比另一个事件发生的概率大很多，这时变量  $X$  的随机性小。因此，信息熵  $H(X)$  正是描述随机变量  $X$  所需的比特数。

应该注意的是：信息熵是信源的平均不确定性的描述。一般情况下，它并不等于平均获得的信息量。只有在无噪声情况下，接收者才能正确无误地接收到信源所发出的消息，全部消除了  $H(X)$  大小的平均不确定性，所以获得的平均信息量就等于  $H(X)$ 。

**【例】** 离散信源与例 1.1 相同，(1) 计算该离散信源的熵；(2) 利用信源的熵估计序列  $S$  的信息量。

解：对于问题 (1)，根据式 (1.3)，该信源的熵

$$\begin{aligned} H(X) &= -\sum_{i=1}^4 P(x_i) \log P(x_i) = -\left[ \frac{3}{8} \log \frac{3}{8} + \frac{1}{4} \log \frac{1}{4} + \frac{1}{4} \log \frac{1}{4} + \frac{1}{8} \log \frac{1}{8} \right] \\ &= 1.906(\text{比特/符号}) \end{aligned}$$

该离散信源只有四种状态 (0, 1, 2, 3)。利用信源的熵，可以很方便地估算发出的符号序列的信息总量。对于问题 (2)，信源的熵为 1.906 比特/符号， $S$  序列共有 6 个符号，则  $S$  序列的总信息量为

$$I(S) \approx \left( \sum_{i=1}^4 n_i \right) H(X) = 6 \times 1.906 = 11.436(\text{比特})$$