

高等学校交通信息工程系列教材



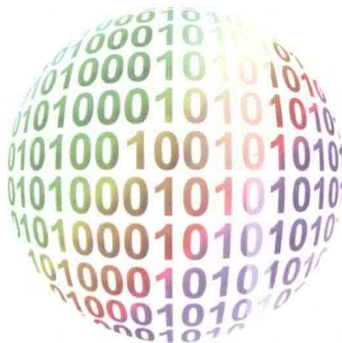
普通高等教育“十一五”国家级规划教材

Information Transmission Principles

张树京 欧冬秀 编著

信息传输原理

(第2版)



同济大学出版社
TONGJI UNIVERSITY PRESS



普通高等教育“十一五”国家级规划教材

Information Transmission Principles

张树京 欧冬秀 编著

信息传输原理 (第2版)



同济大学出版社
TONGJI UNIVERSITY PRESS

内 容 提 要

本书第1版在2004年出版,主要是为交通信息工程专业编写的教学用书。本次第2版中删除了目前已经逐步淘汰的模拟信息传输内容,新增了广泛应用的光纤信息传输原理一章,并且补充了在现代数字信息传输中的新技术,使得全书内容更加全面、充实,也更加切合交通信息工程专业的实际需求。全书共由6章组成:第1章信息与信息系统,第2章信息编码技术原理,第3章数字调制技术原理,第4章移动信息传输原理,第5章光纤信息传输原理,第6章信息传输网络。

本书既可以用作交通信息工程专业本科生教材以及交通信息工程与控制学科研究生教学参考书,也适用于从事交通信息工程专业技术人员的培训及自学。

图书在版编目(CIP)数据

信息传输原理/张树京,欧东秀编著.--2版.--上海:
同济大学出版社,2011.5

普通高等教育“十一五”国家级规划教材.高等学校信
息工程系列教材

ISBN 978-7-5608-4543-2

I. ①信… II. ①张…②欧… III. ①信息传输—通
信理论—高等学校—教材 IV. ①TN919.1

中国版本图书馆 CIP 数据核字(2011)第 062172 号

普通高等教育“十一五”国家级规划教材

信息传输原理(第2版)

张树京 欧冬秀 编 著

责任编辑 高晓辉 责任校对 徐春莲 封面设计 陈益平

出版发行 同济大学出版社 www.tongjipress.com.cn
(地址:上海市四平路1239号 邮编:200092 电话:021-65985622)

经 销 全国各地新华书店

印 刷 同济大学印刷厂

开 本 787mm×1092mm 1/16

印 张 18.75

印 数 1—3100

字 数 468 000

版 次 2011年5月第2版 2011年5月第1次印刷

书 号 ISBN 978-7-5608-4543-2

定 价 36.00 元

本书若有印装质量问题,请向本社发行部调换 版权所有 侵权必究

第2版前言

经过几年来的教学实践,笔者感到《信息传输原理》第1版的内容已不能满足交通信息工程专业的教学需求和科研工作,因此有必要对第1版进行改版。第2版的内容不仅从概念上加强了对信息系统和各种信息传输技术的原理性介绍,削减了许多繁琐的公式推导,而且在编排章节上也进行了较大调整,在内容上充实和补充了现代信息传输技术的最新成果。因此第2版不仅更适用于作为交通信息工程专业本科生和研究生的教学用书,而且也适用于从事交通信息工程专业设计和科研人员的培训及自学。

交通信息工程专业是电子信息工程与交通运输工程这两大学科门类的全新结合,它是将信息科学技术融合到现代交通运输系统中去,充分发挥信息技术在交通运输工程中的作用。因此交通信息工程专业的培养目标应该是运用先进的信息技术更好地在交通运输工程中得到推广应用,以便更快地推动和促进交通运输事业的现代化和信息化。根据交通信息工程专业的教学安排,“信息传输原理”这门课程是唯一属于电子信息类的专业基础课,它的后续课程都是交通运输类的专业课,因此“信息传输原理”的教学内容比较广泛,不仅要介绍电子信息方面的基础知识,而且还要包括通信控制和信息处理等方面的专业知识,这是与电子通信工程专业的教学计划不同之处。由此可见,《信息传输原理》教材不能简单地与通信原理等同视之,因为通信原理的后续课程是各种分门别类的通信专业课。

信息传输是所有一切应用信息系统的核心,它不仅肩负着将海量信息有效准确且安全可靠地有序流动和传递的任务,而且在一定程度上决定了信息系统的服务质量和应用价值。另外,信息传输技术也是在信息技术中发展得最快、应用得最广泛的热点课题之一,特别是在交通运输系统步入智能化时代之后,更是离不开信息传输技术的研究和应用。

根据上述指导原则,本书第2版在第1版基础上删除了已经逐步淘汰的模拟信息传输一章,新增了光纤信息传输一章,并对其他各章内容也都作了增减和调整,第2版的内容更加全面充实,并且更加切合交通信息工程专业教学用书的实际需求。

本书总共由6章组成,各章主要内容如下:

第1章是信息与信息系统。该章着重介绍信息的属性和度量方法,信息量和信息熵的计算,不同信源的信息熵,不同信道的传输特性,以及以通信信息系统与交通信息系统为代表的信息应用案例。在修订版中进一步加强了对信息概念和其属性的理解,摒弃了部分信息理论的内容,突出了信源和信道的物理特征。另外,首次将通信信息系统纳入应用信息系统的

范畴之内,保留了交通信息系统的内容,删去了社会信息系统这一小节。

第2章是信息编码技术原理。该章着重介绍字符类信源编码算法、音频类信源编码算法,以及视频类编码算法的不同特点以及多媒体信息压缩编码的国际标准。还详细介绍了线性分组码的编译码方法,特别是用得较多的循环码,以及同时能对抗随机噪声和突发干扰的卷积码,最后是简单介绍目前最先进的级联码。由此可见,在第2版中增加补充了信源编码的内容,同时保留了原来信道编码的主要章节,这样可以更加全面完整地理解信息编码技术在信息传输系统内的地位与作用。

第3章是数字调制技术原理。该章原名数字信息传输,第2版改名为数字调制技术原理,这样可以更加名符其实。该章首先介绍基带数字信号的传输性能,然后是三种最基本的二元数字调制方式,即幅移键控(ASK),频移键控(FSK)和相移键控(PSK),并对它们的传输性能进行详细比较。在此基础上,进一步介绍相对应的三种多元数字调制方式。最后是重点介绍频带利用率最高的最小频移键控(MSK)和抗干扰能力较强的正交幅度调制(QAM)技术。该章内容数学公式较多,在第2版中尽量做到简明扼要,突出重点结论。

第4章是移动信息传输原理。该章原名无线信息传输,第2版改名为移动信息传输原理,这是因为无线通信的工作频段很宽,而移动通信只是使用其中的特高频频段,或称超短波通信,因此改名后更加突出移动通信的本质和特色。该章中着重介绍电磁波辐射原理,适用于移动通信的线状天线发射和接收、超短波传播特性、移动通信的关键技术(包括多址技术、扩频技术、正交频分复用技术),以及数字移动通信系统和移动自组织信息传输网络。

第5章是光纤信息传输原理。着重介绍光纤信息传输原理、光纤通信系统、光同步数字传输网络(SDH)以及光波分复用技术(DWDM)。

第6章是信息传输网络。简要介绍信息传输网络基本概念及分类,重点介绍通信网中的交换技术;电信网中的数据通信网,包括分组数据网、数字数据网和综合业务数字网(ISDN)等;计算机网中的局域网(LAN)、互联网、无线传感器网等;通信网的接入网。

本书由同济大学资深教授张树京负责全书编审和撰写第1,2,3章,同济大学副教授欧冬秀撰写第4,5,6章。

本书难免有不妥,甚至错误之处,谨请读者批评指正。本书能够出版面世,得到同济大学出版社的大力支持,并获得同济大学教材/学术著作出版基金资助,在此表示衷心的感谢。

编者
2011年1月

第 1 版前言

从信息论的观点来研究信息传输技术是撰写本书的宗旨。随着信息时代的脚步越来越临近,全世界信息产业蓬勃发展,已将信息传输这门学科推向最新的科技前沿。信息传输是所有信息系统不可缺少的重要内容,而且在一定程度上决定了信息系统的服务质量和应用价值。从历史的沿革出发,信息传输技术是从传输话音信息开始,习惯上称为模拟通信时代。经过百年来的更新改造,目前信息传输的对象已经不限于话音信息,而是非话业务信息越来越占有很大的比例。特别是集话音、文字、图像、数据、音频视频业务为一体的综合业务数字信息(也称多媒体信息)得到了广泛应用。包括文体传媒在内的各种应用信息领域都要求建立宽带高速信息传输系统;在金融保险、证券交易等场所又要求建立可靠和安全的信息传输系统。所有这些都说明信息传输技术在信息时代的地位和作用越来越重要,各种各样的信息资源都要依靠信息传输系统来开发应用。与此同时,信息产业的飞快发展,也推动了信息传输技术自身的创新和改造,从模拟通信到数字通信,目前正朝着个体化、网络化、智能化的发展方向前进。这里有必要澄清一个基本概念,即信息传输与通信技术是否有区别?回答是:既相同,又有区别。在通信发展史上,曾经历过模拟通信和数字通信两个不同的时代,但是从信息传输的观点来理解,它们是统一的,只不过传输信息的对象和特征有所区别而已。另外,在模拟通信时代是以传输话音信息为主,但自从模拟信源可以改造成为数字信源后,话音和图像信息都可以成为数字通信的对象。再加上大量涌现的数据信息和音频视频业务,因此信息传输的对象已从原来的模拟信息转向包括数字话音在内的多媒体数字信息,而且后者的比例急速上升,成为信息传输的主要对象。可以这样来理解,信息传输已经突破了原来通信的概念,它包括传输所有的话音和非话业务信息,因此信息传输已经包含了通信的全部内容,并且极大地拓展了原有通信的业务功能。

循着上述思路,本书不同于作者原来撰写过的通信系统原理教材。本书的特点是在信息论的基础上强调信息传输的一致性和多样性,主要内容包括各类信息的传输原理,并着重分析各类信息系统的传输性能。具体内容除绪论外包括 7 章,其中第 1 章是信息和信息系统,着重介绍信息论的基本概念和有关信源与信道的物理特性和数学模型,还介绍两个应用信息系统。第 2 章是模拟信息传输,在对随机模拟信号和噪声分析的基础上,着重分析各种模拟信号调制

的传输带宽和抗噪声性能,并指出其中的矛盾所在。第3章是数字信息传输,在分析基带传输信号的基础上着重讨论各种数字信号调制方式以及它们的传输性能,其中包括信息传输速率和误码率。第4章是数字信息编码,其中包括属于信源编码的模拟信号数字化方法和性能分析,重点是介绍属于信道编码的差错控制编码方法,其中包括线性分组码和卷积码。第5章是多媒体信息传输,在介绍不同媒体特征的基础上着重分析传输多媒体通信业务的特点和多媒体通信系统的实现方案。第6章是无线信息传输,该章简要介绍天线和电波传播方面的基本概念和计算方法,着重讨论无线信息传输的某些专门技术,包括多址技术和扩频技术。第7章是信息传输网络,其中除了电信网以外还包括计算机网和国际互联网,着重介绍各类网络的通信协议,特别是TCP/IP协议,最后还介绍了有线和无线接入网的多种方案。各章均配有小结和习题,便于教学。

本书按照交通信息工程专业本科教学计划的安排,属于专业基础课程教材或教学参考用书,因此着重于基本原理和方法的介绍。不涉及到专业课程所要讨论的内容,但又与专业课密切相关。本书针对交通信息工程专业的特点偏重于应用方面,因此除了必要的数学分析外尽量减少繁琐的数学推导,突出论述物理概念。另外,本书篇幅较大,内容较多,在使用时可根据教学时数灵活掌握。

本书也适用于其他各类应用信息工程专业、通信工程及相关专业的学生和工程技术人员参考使用。

编著者
2003年6月

目 录

第 2 版前言

第 1 版前言

第 1 章 信息与信息系统	1
1.1 信息与信息量	1
1.1.1 信息的度量与属性	1
1.1.2 信息量的计算方法	4
1.2 信源与信息熵	6
1.2.1 信息熵的计算	6
1.2.2 离散字符信源的信息熵	10
1.2.3 连续模拟信源的信息熵	14
1.3 信道与传输性能	17
1.3.1 模拟调制信道的传输性能	17
1.3.2 数字编码信道的传输性能	20
1.4 信息系统	23
1.4.1 信息系统的组成与分类	23
1.4.2 通信信息系统	24
1.4.3 交通信息系统	31
本章小结	34
思考题	36
第 2 章 信息编码技术原理	37
2.1 信源编码	37
2.1.1 信源编码器模型	37
2.1.2 信源编码效率	39
2.1.3 信源编码方法	40
2.2 信源编码技术	41
2.2.1 熵编码	41
2.2.2 算术编码	44
2.2.3 量化编码	45
2.2.4 差值编码	47
2.2.5 正交变换编码	49

2.2.6 多媒体信息压缩编码	51
2.3 差错控制编码	55
2.3.1 差错控制方式	55
2.3.2 差错控制编码分类	56
2.3.3 差错控制码的属性	57
2.4 信道编码技术	60
2.4.1 线性分组编码	60
2.4.2 循环编码	65
2.4.3 卷积编码	69
2.4.4 级连码	73
本章小结	75
思考题	77
第3章 数字调制技术原理	78
3.1 基带数字信号传输	78
3.1.1 基带数字信号的码型	78
3.1.2 基带数字信号的传输性能	81
3.1.3 基带数字信号的误码性能	86
3.2 二元数字调制	89
3.2.1 ASK 信号的传输性能	89
3.2.2 FSK 信号的传输性能	94
3.2.3 PSK 信号的传输性能	100
3.3 多元数字调制	104
3.3.1 MASK 信号的传输性能	105
3.3.2 MFSK 信号的传输性能	106
3.3.3 MPSK 信号的传输性能	107
3.4 最小频移键控	112
3.4.1 MSK 信号的特点	112
3.4.2 MSK 信号的形成	114
3.4.3 MSK 信号的功率谱	115
3.5 多元正交调幅	116
3.5.1 QAM 信号的形成	116
3.5.2 MQAM 信号图	117
本章小结	118
思考题	120
第4章 移动信息传输原理	121
4.1 无线电波	121
4.1.1 电波辐射	121
4.1.2 高频传输线	123

4.2	线状天线	125
4.2.1	对称振子	126
4.2.2	天线特性	130
4.3	超短波传播	137
4.3.1	自由空间传播	137
4.3.2	地面传播	139
4.3.3	无线衰落信道的传播特征	141
4.4	移动信息传输关键技术	145
4.4.1	多址技术	145
4.4.2	扩频技术	150
4.4.3	正交频分复用技术	156
4.5	移动信息传输系统	159
4.5.1	蜂窝式移动通信系统	159
4.5.2	公众移动通信系统	163
4.5.3	专用移动通信系统	169
4.5.4	移动自组织信息网络	175
	本章小结	180
	思考题	183

第5章	光纤信息传输原理	184
5.1	光纤信息传输	184
5.1.1	光纤信息传输系统的组成	184
5.1.2	光纤信息传输的特点	184
5.2	光纤与光缆	186
5.2.1	光纤结构与分类	186
5.2.2	光纤导光原理	188
5.2.3	光纤的传输特性	190
5.2.4	光缆结构与分类	193
5.3	光纤通信系统	194
5.3.1	光源和光发射机	195
5.3.2	光放大器	198
5.3.3	光检测器和光接收机	201
5.3.4	光耦合器与光连接器	203
5.4	光同步数字传输网络	205
5.4.1	准同步数字体系	205
5.4.2	同步数字体系	207
5.4.3	多业务传输平台	213
5.5	光波分复用	217
5.5.1	光波分复用原理	217
5.5.2	光波分复用和解复用技术	219

5.5.3 光波分复用网络	221
本章小结	224
思考题	225
第6章 信息传输网络	226
6.1 信息传输网的基本构成	226
6.1.1 信息传输网的拓扑结构及分类	227
6.1.2 信息传输网的质量要求	230
6.2 信息传输网的交换技术	231
6.2.1 电路交换	232
6.2.2 分组交换	232
6.2.3 ATM交换	234
6.2.4 IP交换	235
6.3 数据通信网	237
6.3.1 分组交换数据网	237
6.3.2 数字数据网	243
6.3.3 综合业务数字网	247
6.4 计算机通信网	251
6.4.1 局域网	252
6.4.2 高速局域网	255
6.4.3 无线局域网	258
6.4.4 无线传感器网络	259
6.5 互联网	265
6.5.1 互联网的互连设备	265
6.5.2 宽带互联网	267
6.5.3 互联网协议	268
6.5.4 互联网地址结构	270
6.6 接入网	272
6.6.1 接入网概述	272
6.6.2 有线宽带接入网	275
6.6.3 无线宽带接入网	278
本章小结	284
思考题	286
参考文献	287

第 1 章 信息与信息系统

本章在简要介绍信息与信息量的基础上,着重讨论信源的信息熵和信道的传输性能,最后以通信信息系统和交通信息系统为例,说明信息系统的组成。

1.1 信息与信息量

1.1.1 信息的度量与属性

自从 1928 年哈特莱(Hartly)首次提出信息的概念和信息的量度方法以来,信息这个名词就与人类产生了解之缘。无论是人际交往或者社会生活,还是文化经济发展和军事国防建设,都离不开信息的流动和变化。例如从原来的电报电话通信与广播电视传媒,发展到现在的电子商务和电子金融信息服务,无不都是利用电子信息在各种信息系统内的有序流动和充分发挥信息系统的各种功能。

信息的物理概念应该是客观存在的,信息和物质、能量一样是人类生存和社会发展的三大要素之一,因此信息资源的不断开发和利用是人类共同的任务。信息虽然没有像物质那样的形态和重量,但是信息的产生却来源于物质的运动和形态的变化。因此信息是一切事物运行状态和运动方式的表征,也可以说信息是物质存在的一种固有属性。虽然信息不等同于物质,但是没有物质和没有物质的运动就不会有信息的产生和信息的生存条件。与此同时,信息与能量也有着密切的关系,因为没有能量,一切事物就不能发生运动和变化,自然就不会有信息产生和存在的前提,因此信息虽然也不等同于能量,但是能量却是事物运动和发生变化的起因,而信息则是事物运动和变化的结果。

另外,物质和能量要受到空间和时间的限制,但是信息原则上可以延伸和开拓到无限的空间和时间。物质和能量只存在于客观世界,但是信息除了客观存在外,还接受主观世界的影响,例如文学艺术信息的内涵就与感知者的理解与思维能力,以及爱好和欣赏水平等因素有关系。有人将信息分为客观信息和主观信息两大类,前者是以统计概率作为基础来分析研究信息的量度、信息的容量以及信息的速度等,这是现代信息技术的主要内容。主观信息则是在统计分析的基础上,进一步研究信息的逻辑结构、内容意义以及实用效果等,这在信息科学及信息系统中应用比较广泛。

按照信息的表现特征可以分为模拟信息和数字信息两大类。其中模拟信息所反映的事物状态变化是模拟量,它是一类在数值轴上不可细分,或者说分层级差可以达到无限小的数量,例如语音幅度、图像亮度、温度湿度、距离长度以及高度宽度等物理参数都是模拟量。而数字信息所反映的事物状态变化则是数字量,它在数值轴上可以分级,并且具有有限的量化分级数量,例如各种字符、指令或数据等。但是有限的模拟量也可以通过量化分级后,在给定的量化误差范围内转变为多进制的数字量。如果再通过数字编码过程之后,可将多进制数字信息转变成二进制码元序列,其结果就是将一个模拟信息转换成等值的一连串二进制数码序列,它就体现了原来的模拟信息的存在和信息量大小。

同时,按照信息在时间轴上的分布还可以分为连续信息和离散信息两大类。其中连续信息是指事物状态在时间轴上的变化是渐进和连续的,例如各种周期信号、不同频率的电波、随机噪声等。离散信息则在时间轴上的变化是锐变和离散的,例如各种脉冲波形、数据信号、天电干扰等。但是连续信息通过在时间轴上的采样过程之后,只要在保持足够采样频率的条件下也可以保留原来连续信息所拥有的信息量,实现连续信息的离散化,将原来的连续信息等值地转换成相应的离散信息。

根据原先哈特莱对信息量所定义的量度方法,它是在概率统计的基础上以随机信息出现概率的倒数为基础来计算的。其物理意义就是出现概率越小的随机信息,亦即比较罕见的信息,它就应该拥有越多的信息量;反之,出现概率较大的随机信息,亦即常用信息,则它所拥有的信息量就比较少。同时,根据信息的累加性原则,当多个随机信息同时出现时,它的合计信息量应该是各单个随机信息所拥有信息量的累加和。另外,哈特莱还采用对数的计量方法,使概率的相乘转化为对数概率的相加,这时出现概率很小的信息也能计算出比较精确的信息量,同时对数概率相加正好符合信息累加性原则。因此哈特莱对信息的科学量度方法确定为

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

式中, P 就是信息的出现概率,它也可称为信息的先验概率。

因为概率 P 总是小于 1 的,故概率 P 的对数总是负值,按式(1-1)计算可知,信息量 I 应该是正值。

如果将公式(1-1)中的对数采取以 2 为底来度量,则其计量单位为比特(bit),并可将公式(1-1)改写成 $I = -\lg(P)$ 。如果采用自然对数,即以 e 为底来度量,则其计量单位为奈特(net),并可将公式(1-1)改写成 $I = -\ln P$ 。奈特与比特的对换关系为 1 奈特 = $\lg(e)$ 比特。例如一个以等概率出现的二进制码元(0 或 1),它的出现概率相等,即为 $P(0) = P(1) = 1/2$,故一个等概率码元的信息量应该等于 $-\lg(1/2) = 1$ 比特。

在日常生活中经常会接触到消息这个名词,其实消息与信息的物理概念和它们的属性都是一致的。可以简单地说,消息就是由许多信息有序地组合而成的信息集或者排列而成的信息序列,而信息则是消息中最简单的组成单元。例如一句话是语音消息,一段文字是文字消息,它们都是由若干个音符信息或者字符信息所构成的信息序列,因此由消息中信息序列的长度就可以衡量出一个消息所拥有的信息量,而单个音符或单个字符就是其中一个最简单的信息。但是消息不是若干信息的随机组合,无论是语音消息还是文字消息都是要按照一定规则才能构成有序的信息序列,这就是消息的有序性。例如组成一个单词要符合词法规则,组成一个句子更要符合句法规则,因此有意义的消息一定是许多信息有规则的组合,它所拥有的信息量应该就是在组成该消息中所有单个信息的信息量的累加和。例如在数据消息中是以二进制码元所组成的数码信息序列,因此数据消息的信息量就是该数码序列的信息量,它应该等于每个码元信息的累加和。前面已经知道,单个二进制码元(即 1 与 0)是个等概率事件,它的信息量就是 1 比特。因此如果一个数据消息是由 n 个二进制码元所组成的数码序列,则该数据消息的信息量就等于 n 比特。通常在数据通信系统中都是采用比特数作为衡量数码序列的信息量大小,因而比特这个名词已经成为计算二进制数据消息的基本单位。如果是一个由十进制数字序列所组成的数据消息,则首先要经过二进制或十进制数字编码转换后,再将多位十进制数字转换成一个由多位码元(二进制数字)所组成的等值信息序列,然后再按该序列中的二进

制码元数目来计算数据消息的信息量(即比特数)。同样地,在数字通信系统中还经常用每单位时间内传输的信息量来表示信息传输速率,它的计量单位通常采用比特/秒(bit/s),即每秒内传输的信息量。

另外,在通信信息系统中还经常使用信号这个名词来替代信息的概念。简单地说,信号就是电子信息,这是因为信号是信息的载体,一个电信号就承载着全部信息序列在通信信息系统内的流动和传输。特别是采用现代数字信号处理技术可以实现模拟信号数字化(即模数转换),因此不管原来是数据信息还是数字化后的模拟信息,都可以用数码信号序列在数字通信系统内传输,此时的数码信息序列就可以代表原来的数据消息,也可以代表数字化后的模拟消息。在传统的模拟通信系统中常用有效传输频带作为一项重要的技术指标,它的计量单位是赫兹(Hz);而在先进的数字通信系统中却是采用信息传输速率作为重要技术指标,其计量单位是每秒比特数(bit/s)。

信息的属性主要有:

(1) 随机性。从计算信息量的定义中可以看出,出现信息的概率是决定信息量大小的唯一依据,这就表明信息具有随机性或不确定性。如果该信息的出现是个完全确定的事物,即它的出现概率 P 等于 1,则它的信息量 I 就等于零;相反地,如果该信息是完全不确定的,即其出现概率 P 为零,则相应的信息量 I 将为无限大。由此可见,越是意料之外的信息,其信息量就越大;而经常会发生的信息,其信息量就越小。通常随机信息的出现概率是在 1 与 0 之间,因此信息量应该是个有限值。

(2) 动态性。从信息的物理概念中得知,信息是反映一切事物发生和发展的状态变化,因此信息具有动态性。对于固定不变的事物状态,它就不会产生新的信息,因此静态信息的信息量也应该等于零,只有动态信息才具有一定的信息量和信息价值。同时,人们在研究发现信息出现概率分布规律的基础上,可以利用信息的动态性来实现自动控制,甚至是智能控制功能。

(3) 时效性。信息的产生与变化具有很强的时效性,亦即在不同的时间或者在不同的场合,同一事物的状态变化可以产生完全不同的效果。在通信信息系统中要求既快速又准确地传输大量信息,因此特别要强调其实时性,以便最大限度地发挥传输信息的作用和效果。相反地,一切过时的信息就会失去其信息价值和效用。

(4) 真实性。只有信息的真实性才能保证其有效的信息价值。如果信息失去真实性,则这些假信息肯定会产生负面影响。同时,同样的信息在不同的场合下对不同的人群可能会产生不同的信息价值和效果。因此人们在利用信息的功能和价值之前,务必要识别信息的真伪,剔除一切假信息,以防以假乱真,造成不良后果。

(5) 安全性。保证信息的安全非常重要,特别是在信息传输过程中要求信息安全可靠。例如在金融经济信息系统中要保证金融经济数据的绝对安全、有序可靠地进行金融经济服务,防止造成不必要的经济损失;在交通信息系统内对信息的安全性和可靠性都要求很高,特别是要防止恶性交通事故发生;在国防军事及社会治安部门的信息系统内对信息的安全保密性要求就更为严格,通过数字编码技术可以实现信息加密。

(6) 累加性。如果同样的信息重复多次,或者同样的信息来自多处,都可以给接受者累加信息量和信息价值,避免假信息和传输干扰的影响。例如在通信信息系统中,如果同样的数据信息重复传输多次,则可以减少在信道中产生的传输错误,提高信息传输的可靠性,在技术上可以采用差错控制机制来实现。

(7) 关联性。通常在有实际意义的消息中,信息之间具有关联性,它反映前后信息的依赖

程度。关联性可用条件概率来度量,即该信息是以前面的信息已经存在作为条件的出现概率,可称之为条件概率。如果信息出现的条件概率越大,则说明它们之间的依赖度越高,即关联性越强;反之,则关联性越低。在极端情况下如果在消息中各信息均彼此统计独立,则表明它们之间没有任何依赖关系,因此关联性就等于零。

1.1.2 信息量的计算方法

根据哈特莱的信息度量方法,香农(Shanon)再针对不同的情况给出了不同信息量的计算公式。常用的信息量主要有自信息量、条件自信息量、互信息量和条件互信息量几种。

1. 自信息量

按照信息量的计算公式(1-1),若单个随机字符 X 的出现概率为 $P(X)$,则其自信息量 $I(X)$ 可定义为

$$I(X) = -\log P(X) \quad (1-2)$$

如果在一个信源内同时有两个随机字符 X 与 Y 发生,它们的联合出现概率为 $P(XY)$,则 X 与 Y 的联合自信息量就可定义为

$$I(XY) = -\log P(XY) \quad (1-3)$$

当 X 和 Y 是两个独立的随机字符时, X 与 Y 的联合出现概率 $P(XY)$ 应该等于它们各自出现概率之乘积,即 $P(XY) = P(X)P(Y)$ 。因此,它们的联合自信息量 $I(XY)$ 就等于

$$I(XY) = -\log P(X) - \log P(Y) = I(X) + I(Y) \quad (1-4)$$

即为它们各自的自信息量之和。式中 $P(X)$ 和 $P(Y)$ 分别是字符 X 与 Y 的出现概率。

依此类推,我们可以推广到计算由多个随机字符所组成的字符序列的联合自信息量。如果这些字符信息都是彼此统计独立的,则该字符序列的联合自信息量就应该等于各个字符的自信息量总和。

2. 条件自信息量

条件自信息量的定义为由条件出现概率来确定的自信息量。若随机字符 X 是在另外一个字符 Y 之后出现,则字符 X 的条件出现概率就是 $P(X|Y)$,故字符 X 的条件自信息量应为

$$I(X|Y) = -\log P(X|Y) \quad (1-5)$$

它说明随机字符的条件自信息量是在限定条件下唯一能确定该字符所拥有的自信息量。如果随机字符 X 和 Y 是彼此统计独立的,那么条件出现概率就等于独立无条件的出现概率,即 $P(X|Y) = P(X)$,此时条件自信息量 $I(X|Y)$ 应该等于随机字符 X 独自出现的自信息量 $I(X)$ 。

3. 互信息量

假定随机字符 X 与 Y 来自两个不同的信源,则在 X 与 Y 字符对之间存在互信息量 $I(X, Y)$,它的定义为

$$I(X, Y) = \log\{P(X|Y)/P(X)\} \quad (1-6)$$

式中, $P(X|Y)$ 是在字符 Y 给定后,出现 X 的条件概率; $P(X)$ 则是字符 X 的出现概率。由概率论可知,联合概率 $P(XY) = P(X)P(Y|X) = P(Y)P(X|Y)$,因此 X 与 Y 字符对之间所拥有的互信息量也可以写成

$$I(Y, X) = \log\{P(Y|X)/P(Y)\} \quad (1-7)$$

其中 $P(Y)$ 是字符 Y 的出现概率;而 $P(Y|X)$ 则是在字符 X 给定后出现 Y 的条件概率。人们通常将条件出现概率 $P(X|Y)$ 或者 $P(Y|X)$ 称为后验概率,而将字符 X 或者 Y 的出现概率 $P(X)$ 或者 $P(Y)$ 称为先验概率。由此可见, X 与 Y 字符对之间所拥有的互信息量实际上是各自的后验概率与先验概率之比的对数。

4. 条件互信息量

条件互信息量是针对三个随机字符以上的字符集来计算的。它的定义是指在给定随机字符 Z 的条件下,出现字符对 X 与 Y 之间的互信息量 $I(X, Y|Z)$ 为

$$I(X, Y|Z) = \log\{P(X, Y|Z)/P(X|Z)\} \quad (1-8)$$

式中, $P(X, Y|Z)$ 是在给定字符 Z 已经出现的条件下,同时出现 X 与 Y 字符对的条件概率; $P(X|Z)$ 是在给定 Z 已经出现的条件下,仅出现字符 X 的条件概率。

如果 X, Y, Z 这三个随机字符都保持彼此统计独立,则原来的条件出现概率都变换成为无条件出现概率,因此字符集的条件互信息量也就等于字符 X, Y, Z 各自的自信息量之和。

自信息量具有以下属性:

(1) 非负性。由于每个随机字符的出现概率通常都是在 0 与 1 之间,即 $P(X) < 1$, 故 $P(X)$ 的对数就一定是一个负值,而随机字符 X 的自信息量 $I(X)$ 就一定是一个正值。如果 $P(X) = 1$, 即这里 X 是代表固定不变的确定性事件,则此时事件 X 的自信息量就等于零,这就说明一成不变的事物或状态早已被人们所熟知,所以它是不拥有自信息量的。同样,在 X 与 Y 字符对之间的条件自信息量也具有非负性,只有当字符 X 与 Y 相互统计独立的情况下, X 与 Y 之间的条件自信息量就等于独立随机字符 X 或者 Y 的自信息量。

(2) 相加性。在同时出现两个以上的随机字符时,其联合自信息量应该决定于它们的联合出现概率 $P(XY)$,即在对 $P(XY)$ 取对数的基础上就可获得联合自信息量 $I(XY)$ 。在一般情况下,字符对 X 与 Y 的联合自信息量 $I(XY)$ 应该是字符 X 的自信息量 $I(X)$ 与在 X 之后再出现字符 Y 的条件自信息量 $I(Y|X)$ 之和,即 $I(XY) = I(X) + I(Y|X)$ 。但是在字符 X 与 Y 之间保持相互统计独立的情况下,它们的联合自信息量就应该等于各自的自信息量之和,即 $I(XY) = I(X) + I(Y)$,此时条件自信息量 $I(Y|X)$ 就等于字符 Y 的自信息量。

互信息量的主要属性有:

(1) 互易性。 X 与 Y 字符对之间的互信息量具有互易性,即 $I(X, Y) = I(Y, X)$ 。这说明字符 X 与字符 Y 的位置可以互换,而仍保持它们之间的互信息量不变,这在通信信息系统中就表明双方通信的可行性。在双向通信中当甲方向乙方传递信息时字符 X 代表在发送端发出的信息,而字符 Y 就是在接收端收到的信息。反过来,如果是乙方向甲方反馈信息,那时字符 Y 就是发端发送的信息,而字符 X 则是接收端收到的信息,这就表明双向通信的互易性。

(2) 相关性。由概率论可知,联合概率 $P(XY)$ 就等于先验概率 $P(X)$ 与条件概率 $P(Y|X)$ 的乘积,即 $P(XY) = P(X)P(Y|X) = P(Y)P(X|Y)$,因此在 X 与 Y 字符对之间的互信息量就是 $I(X, Y) = \log\{P(XY)/P(X)P(Y)\}$ 。如果字符 X 与 Y 是相互统计独立的,则因 $P(XY) = P(X)P(Y)$,此时在 X 与 Y 字符对之间的互信息量 $I(X, Y)$ 等于零,意即字符 X 与 Y 之间不存在任何相关性。但是在一般情况下只要在随机字符之间彼此相关,它们的条件概率 $P(X|Y)$ 或 $P(Y|X)$ 就不能用先验概率 $P(X)$ 和 $P(Y)$ 来代替,因此 X 与 Y 之间的互信息

量就不可能等于零。这说明互信息量的大小与它们之间的相关性有着密切关系。

(3) 互信息量可正可负。当 X 的后验概率 $P(X|Y)$ 大于它的先验概率 $P(X)$, 即 $P(X|Y) > P(X)$ 时, 在字符 X 与 Y 之间的互信息量 $I(X, Y)$ 为正值; 反之, 当 X 的后验概率小于先验概率时, 它们之间的互信息量就是负值。这点说明, 互信息量 $I(X, Y)$ 与联合自信息量 $I(XY)$ 的属性有所不同, 前者不具有非负性, 而后者则具有非负性。

1.2 信源与信息熵

顾名思义, 信源就是信息的发源地。根据信源所发出的信息在时间域内的表现形式不同, 信源大致可以分为两大类, 即信源在时间轴上发出离散信息和连续信息的离散信源和连续信源。通常字符型或者数据型信源都属于离散信源, 因为无论是字母、符号、数字, 或者文字都是在时间轴上分隔开的信息, 即信源所发出的信息都是属于离散信息, 所以说它们是离散信源。而像语言音频类或者图像视频类信源就属于连续信源, 这是因为无论语音、音乐、广播或者影视节目等, 它们在时间轴上所播放的信号都是连续信息, 所以说它们是连续信源。

另外, 如果根据信源所发出的信息在数值域内的表现形式不同, 还可以分为模拟信源和数字信源两大类。模拟信源所发出的信息在数值轴上表现为无限细分的模拟量, 例如语音和音响等音频类信源所发出的音量与音调等信息, 在数值轴上都是属于模拟信息; 或者图像和影视等视频类信源所发出的亮度与色彩等信息, 也都是属于模拟信息, 所以这类信源都是模拟信源。而字符与数据等类信源所发出的信息在数值轴上则表现为可以量化分级的多进制数字量, 例如数据、指令或者文字符号等字符类信源都属于二进制数字信源。模拟信源在自然信源中占绝大多数, 但是在现代信息社会中数字信源却飞速发展, 这要归功于模拟信源数字化技术的结果, 此时可以将原来的模拟信源改造成为等值的数字信源。

1.2.1 信息熵的计算

香农是在信息量的基础上引入更为实用的信息熵概念及其度量方法。在热力学中熵是代表分子紊乱度的概念, 在农业土壤中水熵则代表土壤中含水量随机分布的概念, 因此信息熵也是衡量信息出现随机性的测度。

下面以字符型离散信源为例, 介绍几种最常用的信息熵。

1. 信息熵

在实际文字信源所发出的消息中总是包含有许多个随机字符, 它们所组成的一串有序的随机字符序列就代表该消息的内容。例如在英语中每一个单字就含有若干个字母, 一个单词则由若干个单字按语法规则组成, 一句话和一段文字加上标点符号则是由若干个单词或词组按句法来构成, 因此一段文字消息就是由许多个字符所组成的有规则的随机字符序列, 其中每个字符(代表字母、单字、单词、词组或标点符号)的出现都具有随机性。如果将消息中每个随机字符的自信息量按概率加权平均的方法来计算该随机字符序列的平均自信息量, 则它就代表该文字消息所拥有的信息熵。

假定某文字信源在所发出的文字消息 X 中包含着许多个字符信息 x , 而且各个字符信息的出现概率 $P(x)$ 往往都是不同的。因此该消息的信息熵可以定义为