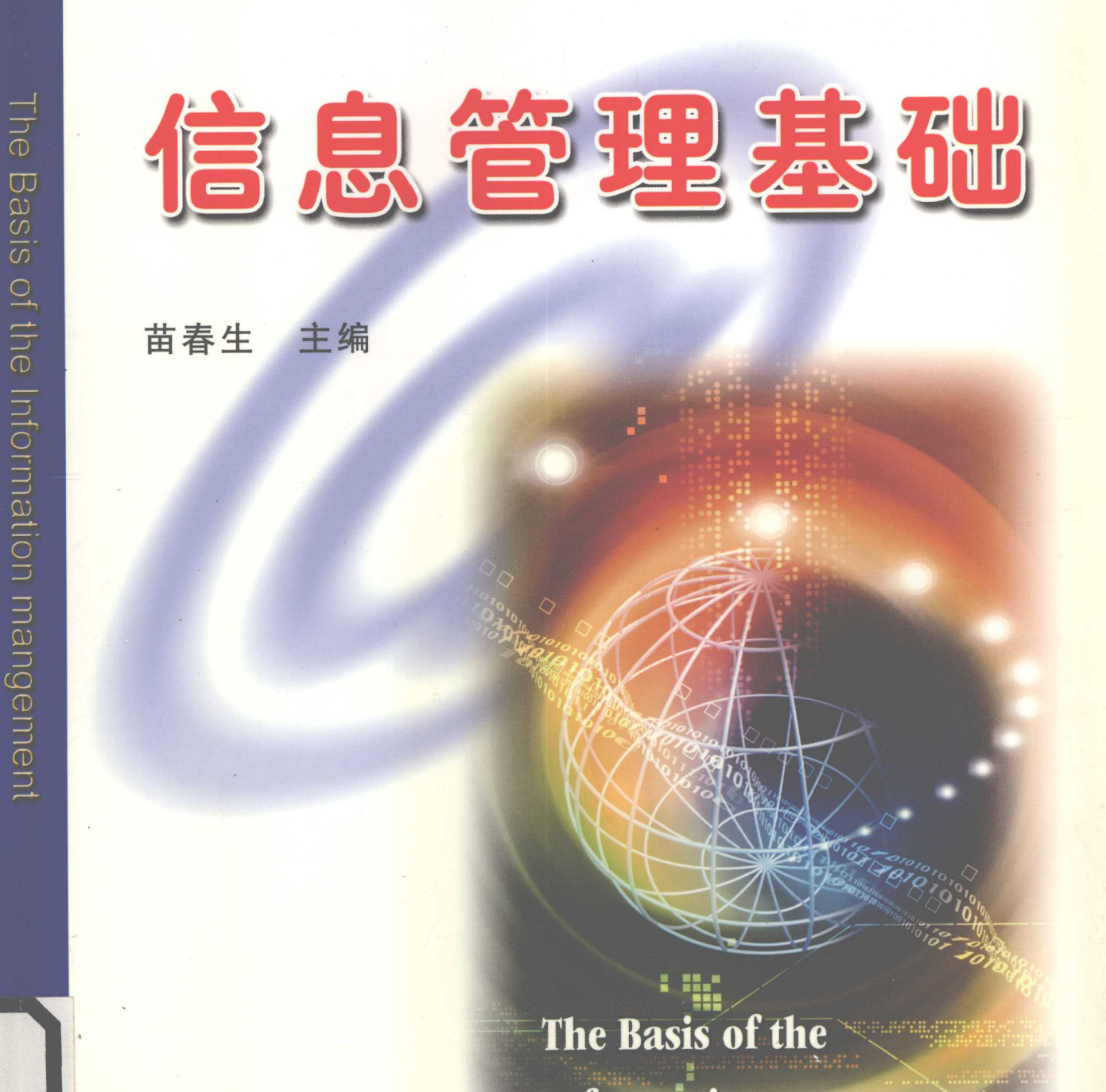




信息管理与信息系统系列教材

信息管理基础

苗春生 主编



**The Basis of the
Information management**

中国矿业大学出版社

China University of Mining and Technology Press

The Basis of the Information management

543

内 容 提 要

信息管理基础课程科技含量高、交叉学科多、涉及范围广、应用性强,因此,作者力图不仅要培养学生具有扎实的、多层次的信息理论和信息管理理论基础,还要培养学生的较强的信息处理和信息管理技术能力,培养出面向与信息处理和信息管理有关各行业通用的、高质量信息管理技术人才。本书全面介绍了信息管理的信息科学基础、管理科学基础,重点讲述了与现代信息处理和信息管理关系密切的信息处理技术、通信网络技术、数据资源管理技术和信息管理系统开发流程等内容。

本书可作为高等院校信息管理类专业本、专科学生专业基础教材,也可供广大信息管理技术人员参考。

图书在版编目(CIP)数据

信息管理基础/苗春生主编. —徐州:中国矿业大学出版社,2004.7

ISBN 7 - 81070 - 153 - 3

I. 信... II. 苗... III. 信息管理 IV. G202

中国版本图书馆CIP数据核字(2004)第065871号

书 名 信息管理基础
主 编 苗春生
责任编辑 孙建波
责任校对 张海平
网 址 <http://www.cumtp.com> E-mail: cumtpvip@cumtp.com
出版发行 中国矿业大学出版社
(江苏省徐州市中国矿业大学内 邮编 221008)
排 版 中国矿业大学出版社排版中心
印 刷 淮阴新华印刷厂
经 销 新华书店
开 本 787×1092 1/16 印张 17.25 字数 417 千字
版次印次 2004年7月第1版 2004年7月第1次印刷
定 价 32.80元

(图书出现印装质量问题,本社负责调换)

前 言

2001年9月我们编写了同名讲义,在南京气象学院信息管理和信息系本、专科教学中使用,3年来我们根据教学的需要,不断修改和完善,现在正式出版此书。

信息管理是一门理论性和实践性都很强的学科。理论性很强,是因为实践中很多信息管理系统开发及信息管理中所出现的问题都与理论上不成熟或缺乏理论指导有直接的关系;实践性很强,是因为该学科源于实践,在实践中形成并提出进一步发展要求。同时,信息管理是一门既古老又新兴的复合学科,其学科内容是随着计算机技术特别是网络技术和数据库技术的发展不断完善的。基于此,我们这本教材的特点是重基础、重技术,全面介绍了信息管理的科学基础、管理科学基础,重点讲述了与现代信息处理和信息管理关系密切的信息处理技术、通信网络技术、数据资源管理技术和信息管理系统开发流程等内容。

我们编写该教材,主要是基于教学和实践的需要,在强调理论和方法重要性的同时,突出可操作性和实践性。本书附有复习思考题,教学中可根据讲课内容,结合书中的案例、复习思考题进行讨论。希望通过感性认识和理性思考,提高同学们认识问题、解决问题的能力 and 水平。

将理论、技术、方法和应用科学地结合在一起,是我们在编写教材时的努力目标。我们希望在这方面的探索和努力通过实践的检验,不断完善和发展。编写一本好教材是件不容易的事情,无论从内容还是体系上均是基于前人在诸多方面进行了深入研究的基础上完成的,我们必须向参考书的作者们表示敬意和感谢。本教材编写过程中,我们还得到了刘宣飞博士、闵锦忠博士、丁治英副教授和张智祥先生、韦龙先生的大力支持和帮助,在此一并表示感谢。

本书共11章。第1、2章由王咏青老师编写,第3、4章由李丽萍老师编写;第5、6、7、8、9、10、11章(案例分析)由苗春生老师编写。由于水平有限,时间仓促,错误和不妥之处,敬请读者批评指正。

编 者

2004.2

目 录

第一章 信息管理的信息学基础与管理学基础	1
第一节 信息论基本原理	1
一、信息的概念	1
二、信息分类	3
三、信息的特点与价值	5
第二节 信息的定量表示	6
一、信息的定量表示	6
二、信息熵	8
三、语法、语用、语义信息的度量	14
第三节 信息论与信息科学	16
一、现代通信理论	16
二、信息科学的基本内容	18
第四节 管理学基础	20
一、管理的概念	20
二、现代管理的基本原理	21
第五节 信息管理科学	25
一、信息管理的意义	25
二、信息管理的发展	26
三、信息管理学的研究对象和内容	28
第二章 信息行为论	31
第一节 信息需要与信息动机	31
一、信息需要	31
二、信息需要的结构分析	32
三、信息动机的形成与转化	34
第二节 信息行为	36
一、信息查寻行为	36
二、信息选择行为	37
三、信息利用行为	39
第三章 信息交流论	42
第一节 信息的表述	42
一、符号	42

二、编码	44
第二节 社会信息流通的基本模型与方式	49
一、社会信息流通模型	49
二、人际信息传播与流通	50
三、组织信息传播与流通	51
四、社会信息的大众传播与流通	51
第三节 网络新闻组概述	53
一、新闻组概述	53
二、新闻组结构	53
第四章 信息产品的采集与分析	55
第一节 信息采集	55
一、信息源	55
二、信息采集的原则	58
三、信息采集的途径	59
四、信息采集的方法	61
第二节 信息整序	64
一、信息整序的目的与要求	64
二、信息整序的方法	65
第三节 信息分析	72
一、信息分析的工作程序	73
二、信息分析的方法	74
第五章 信息产品的流通	78
第一节 信息服务	78
一、信息服务简述	78
二、信息服务的原则	79
三、信息服务的主要类型	80
第二节 信息市场	87
一、信息市场的基本情况	88
二、信息市场的形成与发展	90
三、信息市场的结构与运行机制	91
四、信息市场的营销	96
第六章 信息管理技术基础.....	102
第一节 信息技术概论.....	102
一、信息技术的概念范围	102
二、信息技术的社会作用	103
三、信息技术的发展规律	105

第二节	信息处理技术	107
一、	计算机技术的发展	108
二、	计算机系统结构	110
三、	计算机系统软件	111
四、	程序设计语言简介	115
五、	计算机应用技术	115
第七章	通信与网络技术简介	123
第一节	通信技术	123
一、	通信网络定义	123
二、	计算机通信网络的演变与发展	123
三、	远程通信模型	125
四、	通信通道的特征	126
五、	通信介质	127
六、	通信处理机及其功能	130
第二节	Internet 技术	131
一、	Internet 的起源与发展	132
二、	TCP/IP 协议简介	133
三、	Web 及 Web 服务系统	135
第三节	电子商务用的网络技术	138
一、	电子商务用通信网络	138
二、	电子商务安全技术	140
第四节	电子结算技术	143
第八章	Web 编程语言介绍	146
第一节	HTML 语言介绍	146
一、	HTML 文档基本结构与标签	146
二、	HTML 的列表表格	158
三、	HTML 超级链接	165
第二节	JavaScript 语言简介	168
一、	JavaScript 语言及其历史	168
二、	JavaScript 与 HTML 的关系	169
第三节	Java 语言	176
一、	Java 语言的特点	176
二、	Java 语言在 Web 中的应用	180
第九章	数据资源管理技术	184
第一节	文件组织	184
一、	数据组织的层次	184

二、文件分类	185
三、数据的物理储存与传送	186
四、文件组织方式	187
第二节 数据库技术基础	191
一、传统的文件处理和数据库处理	191
二、数据库管理系统	193
三、数据库的类型	194
四、数据库模型与数据库组织结构	196
五、数据库设计	198
六、数据库设计举例——“教学管理”数据库模型的设计	202
第十章 信息管理系统开发简介	204
第一节 软件工程的基本概念	204
一、软件生命周期模型	204
二、软件生命周期各阶段的主要任务	206
三、Web 信息管理系统开发	207
第二节 系统分析	208
一、Web 信息管理的主题分析	208
二、信息内容分析	209
三、用户信息分析	210
四、软、硬件基础分析	210
五、开发成本分析	212
第三节 系统设计	213
一、设计方法	213
二、信息内容设计	214
三、链接设计与导航设计	217
四、Web 站点文件目录结构设计	221
五、页面布局与外观设计	222
六、主页设计	227
第四节 系统实现	227
一、系统实现小组的分工	228
二、Web 信息管理系统测试	229
三、Web 信息管理系统发布	230
第五节 Web 站点的运行与维护	231
第十一章 现代信息管理案例介绍	233
第一节 现代管理信息系统案例	233
一、案例简介	233
二、系统调查	233

三、系统分析.....	235
四、系统设计.....	238
五、系统实施.....	244
六、讨论.....	247
第二节 现代企业信息管理系统案例.....	249
一、案例简介.....	250
二、系统调查.....	250
三、需求分析.....	251
四、系统设计目标.....	252
五、业务功能分析.....	253
六、系统性能要求分析.....	253
七、系统设计.....	253
八、系统详细设计.....	254
九、系统源代码编写.....	259
十、系统.....	263
参考文献.....	264

例如,一块化石只有为考古工作者发现时才具有信息的形式和意义,因为考古工作者能够解读它所标示的古生物状况,而对于其他人而言,其不过是一块普通石头而已,并不具有信息的形式和意义,至少不会成为古生物学方面的信息并具有其意义。又如,1976年7月28日唐山大地震前,曾出现很多异常现象,诸如鸡上树、驴马不进棚、狗叫不停、井水浑浊等等,其实这都是一些地震信息,但由于未引起人们注意和认识,因而这些现象只停留在自然信息范围内,而未成为地震信息而为人们所认识。

从信息形成的上述两个条件可以看出,信息是一个关系概念,而非实体概念。第一,只有能够标示他事物存在方式或属性的事物才是信息;第二,只有为特定主体所感知的事物才成为信息。离开了这两个关系,信息就不能存在了。

(三) 信息主体包括人类、生物、机械三种形态

1. 人类主体

人类是具有无限的感知能力的。人类可以通过一个客体认识众多的客体,如通过现在的自然界认识过去的自然界和将来的自然界,通过现在的社会认识以往的社会和将来的社会,通过表面现象看到事物的本质等等。因此,对于人类而言,信息的范围是无限的。

2. 生物主体

实验表明,生物也是具有感知能力的,因而,很多事物对于生物来说也具有信息意义。例如,据华盛顿大学研究,柳树在遇到毛虫侵害后会放出一种物质,其他柳树在接收到这种物质以后,就能很快做出反映,增加体内的碳酸含量,使毛虫不敢前来侵袭。对于其他柳树而言,该树放出的这种物质就是信息,它预示着毛虫的侵袭。另据报道,洋槐树、槭树等植物也具有类似的能力。

动物比植物的感知能力要高级得多,因而对于动物来说,信息的范围也更大。例如,动物可以通过声音来辨别对方是同伴、幼仔还是敌人或猎物,还可以通过气味找到配偶、食物或道路,还有的动物能通过形象来辨别目前处境是否危险(美国科研人员利用这一点研制出猫头鹰活擒乌鸦的模型放在乌鸦为害的菜田里,据说颇有“杀一儆百”的作用;日本科学家则在田野上每隔十几米架设彩带,随风发声飞舞,再加上阳光反射效果,亦使鸟类感到危险而不敢靠近)。在上述情况下,声音、气味、颜色、形象……对于动物来说就都是信息了。

3. 机械主体(主要指智能机)

随着电子科技的发展,现代智能机的功能愈来愈多,例如有的智能机能够接收和处理文字信息,有的智能机能接收处理声音信息、图像信息等。美国于20世纪70年代末曾研制出一种会料理家务的机器人(智能机),能按人的指令打扫卫生、操作洗衣机、开关门户、挂取衣帽、端送咖啡、准备餐具等,很像一个能干的家庭主妇。日本东京警视厅亦曾研制成一个机器人警察,身高一米七,头戴大沿帽,身穿深色警服,宛若真人一样指挥交通,凡有不听指挥的人次日就会接到传票,受到处理。

(四) 信息客体包括自然现象、社会现象和思维现象

各种自然现象无疑是可以做为信息客体而存在的,例如,为人和生物所感知的声音、颜色、形状、硬度、气味等等就都是信息(自然信息)。而为人所感知的各种社会事件的发生,诸如文艺演出、电影放映等等,则属于社会信息了。思维现象也可以作为信息,如对于搞哲学史研究的人来说,古代哲学家的思想体系就是信息了,对于搞社会调查的人来讲,群众的心理状态也无疑是一种重要的信息。

(五) 关于信息的哲学属性问题

信息是物质的,还是精神的?抑或是物质与精神的统一?这是近年来学术界讨论的重要问题之一。如果把所有的信息都归结为物质的东西,那显然是片面的,因为人的感觉、思维等信息明明是主观的东西(尽管它包含着客观的内容),怎么说它是物质性的呢?那不就同哲学上的庸俗唯物主义关于意识的观点划不清界限了吗?但把信息理解为纯粹精神的东西是否就准确了呢?也未见得。所谓精神,就是指客观事物在人脑中的反映,某些植物对外界事物的反映便不属于精神范畴,而只能称之为物质性的了。因为物质就是存在于人脑之外的一切事物。至于实物信息也不能说成是精神的东西。因此,如果准确地回答信息的属性问题,则应该说:有一类信息(如人的感觉、思维等等)是精神的,而此外的信息都是物质的。

二、信息分类

信息可以做各种各样的分类,如:

从信息主体的性质来划分,有生物信息,人类信息,机械信息;从信息客体的性质来划分,有自然信息,社会信息,思维信息;从信息的功能来划分,有有用信息,无用信息,干扰信息;从信息的逻辑意义来划分,有真实信息,虚假信息,不定信息;还有其他分类。

而每类信息又可以做进一步的细小划分。

这里介绍一种根据信息本身特性而进行的分类。因为在目前信息科学的发展中,这种信息分类方法受到人们普遍的重视。

(一) 语法信息

所谓语法信息就是指单纯表示符号或符号关系(符号的数目、符号持续的时间以及符号的结构等)的信息。如由莫尔斯电码所构成的信息,数字排列构成的信息,由各种图形、字母、文字和实物系统的组合所构成的信息等等。当人们考察这些信息内部的符号或符号关系时,它们就做为一种语法信息而存在了。如邮电局的报务员所处理的都是语法信息,因为他们只注意电码符号间的关系,而不管各组电码所表达的意义。打字员所处理的也是语法信息,因为他在打字的时候只注意和原稿的字序排列一致,而不管文字的意义如何。电子计算机所处理的信息绝大多数都是语法信息。

(二) 语义信息

所谓语义信息就是指表示一定意义的信息。例如,在交叉路口的交通信息中,红灯出现表示停止前进,绿灯出现表示放行,黄灯则表示准备前进或停止。这些信号灯所提供的信息就是语义信息。此外,旧时店铺前的各种招幌,公路铁路两旁的行车标志,表示剧毒物品的图形——骷髅,以及信件、文章、论著等等也都属于语义信息或具有语义信息的特征。由于信息的意义往往具有多样性(如一词多意),因而把握和处理语义信息要比把握和处理语法信息困难得多。现在只有极少数高级智能机能处理某些语义信息,如按人的简单话语去动作等。

(三) 语用信息

所谓语用信息就是指具有一定价值或效用的信息。例如,天气预报可以起到保证安全生产和生活的作用,商品广告可以起到指导人们购物的作用,科学理论可以起到指导实践的作用,等等。总之,当我们只是强调某一信息的价值或效用时,该信息就是语用信息或具有语用信息的意义了。

一个信息的价值和效用往往是随着收信主体的不同、时间的差异而改变的。例如,商品信息对欲购物者就有用,而对非欲购物者则无意义,有时还会起到干扰收信者接收其他信息

的作用。此外,新闻信息、军事信息、经济信息等等的时间性也都是很强的,有时一条消息迟发了一小时就会丧失其重大意义。正因为如此,对语用信息的确定和把握就更加困难,现代智能机尚未达到处理语用信息的水平。

语法信息、语义信息、语用信息实际上是同一信息的3个不同方面的属性,因而它们又是相互联系着的,而不是各自孤立的。同一个信息既有一定的结构,又有一定的含义和功能,而且这3方面往往具有一致性,就是说,一定的结构具有一定的意义并预示着一一定的效用。因而我们欲全面把握信息,就必须依次注意信息的这3方面特征。当然,由于实践要求的不同,人们有时只侧重把握信息的某一方面特性也就够了。

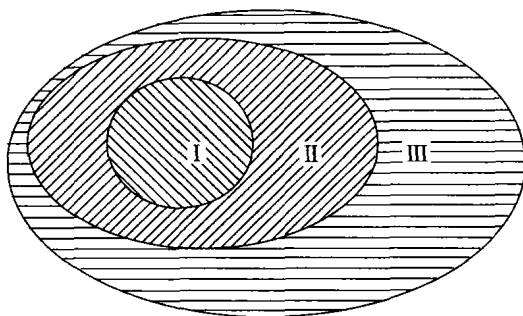


图 1-1 语法信息(I)、语义信息(II)、语用信息(III)的关系

我们再通过一个简单的例子来说明语法、语义和语用信息的含义。Einstein 的著名物理学公式 $E=mc^2$ 给出了能量 E 与质量 m 之间转换关系的信息。若从语法信息的角度看,这个公式规定了这些英文字母和数字的一种特定的排列方式。在没有看到这个公式之前,我们不知道 26 个英文字母中哪几个字母会出现,也不知道 10 个阿拉伯数字中哪个或哪几个会出现,而且也不知道这些字母和数字会以什么方式来排列,这就是我们在观察(阅读)之前所具有的不定性。当我们观察到这个公式的时候,我们就知道了它们的“运动状态和方式”,获得了语法信息,消除了原先所具有的不定性。

但是,如果我们仅仅观察到了这个公式,不知道这些字母和数字以及它们的这种排列方式的含义,那么,我们对于这个公式的语义就具有不定性。乃至我们知道了 E 代表能量、 m 代表质量、 c 代表速度、 2 代表平方关系,才知道了这一公式的真正含义,获得了相应的语义信息。

那么,这一公式以及它所代表的物理含义对于观察者具有什么效用呢?在弄清了这一公式的含义、并且确实证明了这一公式在逻辑上的真实性之后,观察者就懂得,利用这一公式所表明关系,就可以通过改变原子核的质量状态来获得巨大的原子核能。当然,这个信息对某个具体的观察者是否有价值,则与观察者的目的和其他条件有关。以上的分析和例子都表明,在语法、语义和语用信息三者之间,语法信息是最简单、最基本的层次,语用信息则是最复杂最实用的层次。在信息理论发展的初期,人们故意排除语义信息和语用信息的因素,先从语法信息入手来解决问题,这既是迫不得已的事情,同时又是很明智的选择。问题在于我们不能总是停留在语法信息这个相对简单的层次,应当继续深入地研究和解决语义信息和语用信息的问题。因为,语法信息只能解决通信工程这样一类传递信息的问题,而凡是有

智能、有“目的”行为的系统，都必然要涉及到语义信息和语用信息的问题。信息科学技术要有效地扩展人类信息器官的功能(特别是扩展人的智力功能)，就不能不利用语义和语用信息。

三、信息的特点与价值

(一) 信息的特点

信息存在许多有趣的特点或属性。

信息没有质量，正因为这一点它区别于物质。一部孙子兵法，可载于一车竹简，也可载于一本书，还可微缩于几平方厘米的胶片或存于一张光盘。载体或媒体的改变一点也不影响信息。

信息容易拷贝，正因为这一点它区别于能量。能量虽然守恒，但一定形式的能量，只会使用一点少一点。信息则不然，信息拷贝或传播给别人，其本身一点也不丧失，信息也不改变。信息可以共享，所以维纳说：“信息就是信息，既不是物质，也不是能量。”

信息取之不竭，信息是事物运动的方式和状态。世上事物的运动是永恒不息的，故信息永不枯竭，不会出现类似于材料和能源的短缺现象。

信息需要载体 任何信息都必须依附在其载体上，才能存储和传播。如前所述，数据是信息的逻辑载体。除了一般常用的数据外，还有像人类语言、文化习惯等高度浓缩的数据，在它们上面记载和沉积着人类或民族发展进步的历史信息。客观物质是信息(或数据)的物理载体。存储信息需要物理载体，如绳结、石子、竹简、陶器、丝绸、纸张、磁盘、光盘、脱氧核糖核酸以及人脑等；传播信息需要物理载体，如声波、电缆及光纤等。

信息超越时空 信息可自由地超越时间和空间进行传播。在地球范围内乃至浩瀚宇宙间的通信，是信息在超越空间。阅读历史，古人教诲今人，这是信息在超越时间。信息在传播过程中，有的会因“噪声”而变质，有的被存储，有的自行消亡。

(二) 信息的价值

众所周知，没有材料(或物质流)，制造机器如同空中楼阁；没有动力(或能量流)，机器如同废铁；没有信息(或信息流)，则机器如同野马失去控制，达不到目的。物质、能量、信息三位一体，共同构成了客观世界。人类的一切生产、生活的活动都是物质、能量与信息组合的过程。可见信息是一种极为重要的资源，因而也具有极为重大的价值。著名的量子力学家薛定谔就说过：“人并非以能量为生的，而是以负熵(信息)为生的。”

信息支持决策。所谓决策，就是把收集到的信息与要求的目标信息进行比较分析，选择最合理的对策进行实施，并随时监督实施，依据实施反馈的新信息调整对策。简单地说，决策就是为了达到行为目标而采取某种对策的过程。管理需要决策，指挥需要决策，任何有目的的行动都需要决策，而任何决策都必须先有信息。否则，如果没有及时与适用的信息，决策必然是瞎指挥、乱拍板，后果不堪设想。一方面收集和分析信息需要付出社会劳动，另一方面准确、及时与适用的信息会导致正确的决策，进而取得效益，因此信息是具有价值的。

信息构成知识。知识是人类在不断认识自然与改造自然的过程和进步中形成的。随着时代的前进和科学技术的发展，我们面临的许多问题往往都比过去更复杂，而问题的复杂度又往往与该问题相关的知识难度成正比，所以如果具备了解决该问题的有关信息及知识，就为最终解决该问题奠定了基础。从寻找和收集信息的角度看，可以说，哪里有知识，哪里有决策，哪里就有信息。而如果从使用信息的角度看，又可以说，哪里需要知识，哪里需要决策，哪

里就需要信息。

信息增进有序。信息普遍存在于自然界、人类社会和思维领域。它是生物界、人类社会和人造系统赖以生存和发展的重要资源。任何系统物质和能量的变化、运动和交换一般都以信息为先导,并受信息的控制。如生物个体的发育受脱氧核糖核酸(DNA)中遗传信息的控制和影响,候鸟的迁徙、鱼类的洄游、爬虫的冬眠和惊蛰等无不受信息的控制。人类社会更是这样,一切生产和生活的活动都是在人类发出的信息的控制下进行的,只不过人类对信息不像生物那样简单被动地受控,而具有记忆和处理的能力,能进行逻辑推理和形象思维,建立新概念,发现新规律,并进而改造客观环境,使之适合于人类生活的需要。

总之,由熵增原理可知:一个孤立封闭的系统总是在或快或慢地走向无序,要改变这种状况或促进系统走向有序,就必须使系统向环境开放,并从外部环境中输入负熵,进而使整个系统的熵减少,系统朝确定的方向演进。所以,信息反映着系统的确定程度,信息增进系统的有序发展。人利用信息的基本过程如图 1-2 所示。

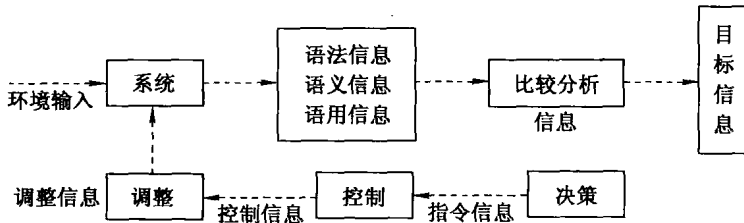


图 1-2 人利用信息的基本过程

第二节 信息的定量表示

一、信息的定量表示

(一) 定量表示信息的必要性

自然科学通常都应该作定量的描述。定性描述只能解释一些现象,即使认为能作直观理解也是悬而未决的,尚未上升到学术水平。对某个量作定量表示时,往往将它与某一适当的标准量进行比较。例如,长度、重量、时间等都是与国际规定的标准作比较来表示的。这是物理学的一个基本问题。功、电量、克当量等量与时间、长度等相比,较难理解,因为对于比较的标准我们可能没有切身的体验。

由于我们要从理论上研究信息,所以必须对信息的大小作定理的描述。但是,看来信息也不能像重量、长度那样定出标准进行比较。究竟怎样才能对信息作定量描述呢?这是一个长期没有解决的问题,因而信息论也在很长一段时间内没有作为一门科学建立起来。

(二) 信息量的性质

下面请看两句一对的几个例子。这些句子所具有的信息量究竟哪一个更多呢?即使不知道信息量是什么也可以根据常识直观地感觉出哪个的信息更多。

(1) A 客机坠落

B 小型卡车掉到了马路外面

(2) A 读者的话:“这本书极好”

- B 出版社的广告：“这本书极好”
- (3) A 6 月份的天气预报“明日有雨”
B 11 月份的天气预报“明日有雨”
- (4) A 民用电源的电压表指示为 220 伏
B 民用电源的电压表指示为 160 伏
- (5) A 高才的王某考取了
B 高才的李某的落第了

(1)的 A 是新闻速报栏上应该报导的大事件,故信息量多。

(2)的 B 的广告宣传方式,怎么也被认为是些奉承话,因此 A 的信息量多。

(3)不太明显,不过 B 信息量多些,其原因是 11 月份晴天很多,雨天非常少见。

(4)电力公司有保证的供电通常是 220 伏,出现 160 伏几乎不可能,因此出现 160 伏成了一大新闻,故 B 的信息量大。

(5)当然是 B 的信息量多。

由以上各例各见,(1)、(2)、(3)、(4)、(5)各自对应的 A、A、B、B、B 情况都是少见的,表明信息量多些。如果改用统计学的术语来描述,则出现概率小的事件信息量多,即对应有

概率减小→信息量增大

因此,信息量应该是概率的单调减函数。

信息量应具有必要的特性——可加性,现举例说明。O 公司以某大楼的第 5 层左起第 3 间借作办公室。这个大楼第一层的房间编号为 11,12,⋯,18;二层编号为 21,22,⋯,28;三层编号为 31,32,⋯,38 等。每层都有 8 个房间。这时告知 O 公司的“办公室在 53 号房间”这一消息应与告知“办公室在第 5 层”和“从左数起第 3 间”这两个消息是一致的。这时后面两个消息具有的信息量合起来应该等于前一消息的信息量才合理。为此应该取怎样一种函数才行呢?按此例,前一消息所示事件出现的概率应等于后面两个事件出现的概率之积。能将“乘积”以“求和”方式表示的单调函数是什么函数呢?显然,对数函数是恰当的。因此作为信息量的定义被确定为事件出现概率的倒数的对数。为了方便,对数的底取为 2,这时信息量的单位是比特。即信息量定义为

$$\text{信息量} = -\log_2 P$$

式中 P 是消息所示事件出现的概率。按上例全部共有 48 个房间,指定其中一间的概率为 $1/48$ 。因此

“O 公司的办公室是 53 号房间”→ $\log_2 48 = 5.58$ 比特。

指定 6 层中的某一层的概率为 $1/6$,故

“O 公司的办公室在 5 层楼”→ $\log_2 6 = 2.58$ 比特。

指定 8 间中的一间的概率为 $1/8$,因此

“O 公司的办公室在第 3 间”→ $\log_2 8 = 3$ 比特。故有

$$\log_2 48 = \log_2 6 + \log_2 8$$

即

$$5.58 = 2.58 + 3.00$$

(三) 概率与信息量

信息量的定义为 $-\log_2 P$,而其中的概率并不限于上例的单纯情况。但是考虑方法的基础是相同的,因此只要具有有关概率的知识,即使十分复杂的消息也能求出它的信息量。本节

就来作这方面的讨论。

首先来考虑独立事件的信息量。例如，“今天的天气极好”，“今天是休息日”这两个消息，它们是相互独立的事件。“天气极好”如占一年 365 天中的 $\frac{1}{4}$ ，则具有信息量为

$$-\log_2 \frac{1}{4} = 2 \text{ 比特}$$

又设学校一年中有 $\frac{1}{4}$ 为假日，则“今天是休息日”的信息量也是 2 比特。

同时给予这两个消息：“今天是天气极好的休息日”，则具有 4 比特信息量。因此，对于独立事件来说信息量以“和”的形式出现。独立事件同时发生的概率等于各事件发生的概率之积。从这一点来看也是很清楚的。

相反，对于不独立事件，信息量就不是它们之和。例如，某工厂得到“今天停电”的消息，同时又得到“今天是休息日”的消息，信息量就不一定是两者信息量之和。因为如已决定停电时作为停电休息日，那么后一个消息的信息量必为零。

练习一 某机器的故障率为 6%，其原因分为：机械方面的占 32%；电气方面的占 12%；材料方面的占 56%。考察“是机械故障”这一消息的信息量以及对于预先知道有故障的人得到“是机械故障”这一消息的信息量之差别。

练习二 给出消息“现在的气压是 940 毫巴且是满潮”，试求具有多少比特信息量。已知气压降到 940 毫巴一年只有一天，满潮在一天中有 2 小时。

相互独立的事件的出现概率是“乘积”，而信息量是“求和”。很明显，概率是“和”的情况，信息量一定不是“和”。互斥事件的概率以“和”来表示，而将这些事件的消息结合起来，则信息量反而会变小。

“骰子的点数是偶数”这一消息具有 1 比特信息量，它是由骰子的点数是 2 或 4 或 6 三个互斥事件结合而成的，即“骰子的点数是 2、4、6”，如果分别只看其中一个消息，则具有信息量为 $-\log_2 \frac{1}{6} = 2.58$ 比特，然而结合在一起时却是 1 比特，可见并不是它们之和。

练习三 某汽车工厂按同一流程生产 U、S、D 和 B 型 4 种产品。U 占 10%，S 占 30%，D 占 35%，B 占 25%。工程 A 把 U 与 S 作为同一种产品装配，而 D、B 则作为不同的产品装配。试求对于这个流程中的人来说，“拿到一个 U 型产品”这一消息的信息量是多少？

二、信息熵

(一) 熵

所谓信息量，应该理解为由某个消息描述的事件出现的概率所决定的量，大多数情况下，研究一个单独的消息的信息量是不够的，应该知道消息系列的整体性质。画出香农线图也是一种观察整体性质的方法。这里将说明称为熵的整体平均信息量。

为了直观地理解平均信息量，先看下面的例子。公司的董事长每天必须听取营业部长的报告，而不是每天听取警卫人员的报告。警卫人员的报告往往信息量较少，不必每天听，虽然它也可能包含有关盗窃事故等具有很大信息量的重要报告。即使如此，还是只听营业部长的报告，这是因为长时间的信息量要大得多。这种按较长时间考虑的平均信息量称为熵。

其次来说明熵的定量表示方法。设 n 个消息， n 相当大。为简单计，设三种消息按不规则的次序出现，总共有 n 个。设三种消息分别出现的概率为 p_1 、 p_2 、 p_3 ， n 个信息中含有的消息数分别为 np_1 、 np_2 、 np_3 个，各消息所含的信息量分别为 $-\log p_1$ 、 $-\log p_2$ 、 $-\log p_3$ 。则有

消息出现的概率	1 个消息的信息量	n 个中的消息数	它的信息量
p_1	$-\log p_1$	np_1	$-np_1 \log p_1$
p_2	$-\log p_2$	np_2	$-np_2 \log p_2$
p_3	$-\log p_3$	np_3	$-np_3 \log p_3$

全部信息量 $-np_1 \log p_1 - np_2 \log p_2 - np_3 \log p_3$

平均信息量(熵) = 全部信息量/ n = $-p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3$

按一般形式写出熵为

$$H = - \sum_i p_i \log p_i \quad (1-1)$$

练习 设某地方的天气预报晴(记为 F) 占 $4/8$, 阴(记为 C) 占 $2/8$, 小到中雨(记为 r) 占 $1/8$, 大雨(记为 R) 占 $1/8$, 又设另一地方 F 占 $7/8$, r 占 $1/8$ 。试求各自的熵。极端的情况, 如 F 的出现的概率 P_F 为 1 而其余为 0, 以及 $P_F = P_C = P_r = P_R = 1/4$ 时, 试计算这两种极端情况下的熵。

熵是单位消息的信息量。如按比重和比热等的叫法, 也可以称为比信息量。将消息数 n 乘以熵 H , 其值 nH 则是全信息系列的信息量。这一点与比重乘以体积算出重量是相似的。不过, 因为信息量具有统计的性质, 按大数法则 n 必须是很大的值。

当求两类信息系列组成的系列的熵时, 正好与求两种比重的物质所构成的物体的比重相同。

即各比重乘以其对应的那部分体积, 然后加起来与总体积之比。总熵也必须由两种熵各乘上出现的概率加权相加起来。例如熵为 H_1, H_2 的信息系列, 分别以概率 p_1, p_2 出现, 系列全体的熵为

$$H = p_1 H_1 + p_2 H_2 \quad (1-2)$$

又如图 1-3 所示, 出现概率为

$$\frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}$$

的消息构成的系列的熵为

$$H = \left(\frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8} \right) = \frac{1}{8} \log 8 + \frac{1}{8} \log 4 + \frac{2}{8} \log 4 + \frac{3}{8} \log \frac{8}{3}$$

如将它按图 1-3(b) 的形式分解, 原来的熵分成两部分之和, 即以 $5/8, 3/8$ 出现的第一段的熵 $H(3/8, 5/8)$ 与它右边第二段的熵之和。第二段的熵又分为两部分, 上面的 $H(1/3, 2/3)$ 以它的出现概率加权, 下面的 $H(2/5, 3/5)$ 以 $2/8 + 3/8 = 5/8$ 加权, 然后相加起来。即

$$H = \left(\frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8} \right) = H\left(\frac{3}{8}, \frac{5}{8}\right) + \frac{3}{8} H\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{5}{8} H\left(\frac{2}{5}, \frac{3}{5}\right)$$

如图 1-3 这种表示概率选择方式的图称为概率树图, 树的分枝方法一般形式记为

$$H(p_1 \cdots p_n, p_{n+1} \cdots p_N) = H(p_1 \cdots p_n + p_{n+1} \cdots p_N) + (p_n + p_{n+1}) H\left(\frac{p_n}{p_n + p_{n+1}}, \frac{p_{n+1}}{p_n + p_{n+1}}\right) \quad (1-3)$$

其次求马尔可夫过程的熵。马尔可夫过程的熵与按前述单纯的出现概率定义的信息量不同, 它是对于按迁移概率定义的信息量的平均值。例如以数字 1、2、3 表示三个状态, 作为

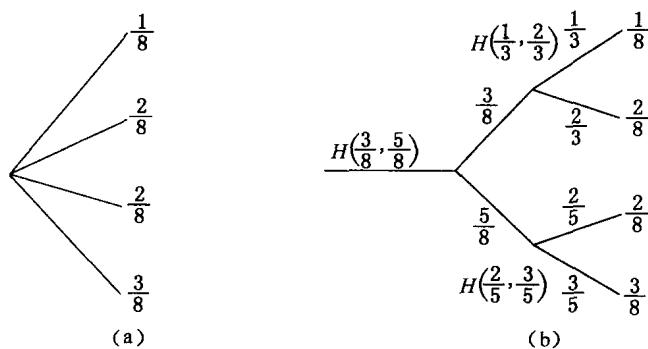


图 1-3 熵的分解与合成

单纯的概率过程时,其熵是信息量:

$$-\log p(1), -\log p(2), -\log p(3)$$

的平均值。如按一阶马尔可夫过程来考虑,则是则迁移概率 $p(1|1), p(2|1), p(3|1)$ 等定义的信息量:

$$-\log p(1|1), -\log p(2|1), -\log p(3|1)$$

的平均。这时总的熵 H 表为

$$H = p(1)H_1 + p(2)H_2 + p(3)H_3$$

式中, H_1, H_2, H_3 分别是 1, 2, 3 迁移的那部分的熵。例如 H_1 具有下式形式

$$H_1 = -p(1|1)\log p(1|1) - p(2|1)\log p(2|1) - p(3|1)\log p(3|1)$$

一阶马尔可夫过程的熵的一般形式为

$$H = - \sum_{i,j} p(i)p(j|i)\log p(j|i) \quad (1-4)$$

(二) 剩余度与熵

所谓熵大的信息源实际究竟是怎样一种信源呢?各消息的信息量全部相同时熵为最大。一年中只送一次具有非常之大的信息量的消息,由于其余 364 天几乎不送什么,平均信息量就会很小。相反,每天都不断送某个大小程度相等的信息量,整个说来,可以送出更多的信息量,即熵也变大。为什么是这样,下面加以简单说明。首先来考虑信息量为 $-\log p_1, -\log p_2, -\log p_3$ 的三种消息组成的系列的熵,求能使此熵为最大的 p_1, p_2, p_3 。且总满足

$$p_1 + p_2 + p_3 = 1$$

这里只以三种消息的情况来作讲解,当有任意多个消息时也完全相同。现在先固定 p_3 改变 p_1, p_2 使 H 为最大。为此必须令

$$\frac{dH}{dp_1} = \frac{\partial H}{\partial p_1} + \frac{\partial H}{\partial p_2} \frac{\partial p_2}{\partial p_1} = 0$$

又由于有 $H = -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3$

故有

$$\frac{\partial H}{\partial p_1} = -\frac{\partial}{\partial p_1}(p_1 \log p_1), \frac{\partial H}{\partial p_2} = -\frac{\partial}{\partial p_2}(p_2 \log p_2)$$

这里由于 $p_1 + p_2 = 1 - p_3$ 恒成立,故有

$$dp_1 + dp_2 = 0, \frac{\partial p_2}{\partial p_1} = -1$$