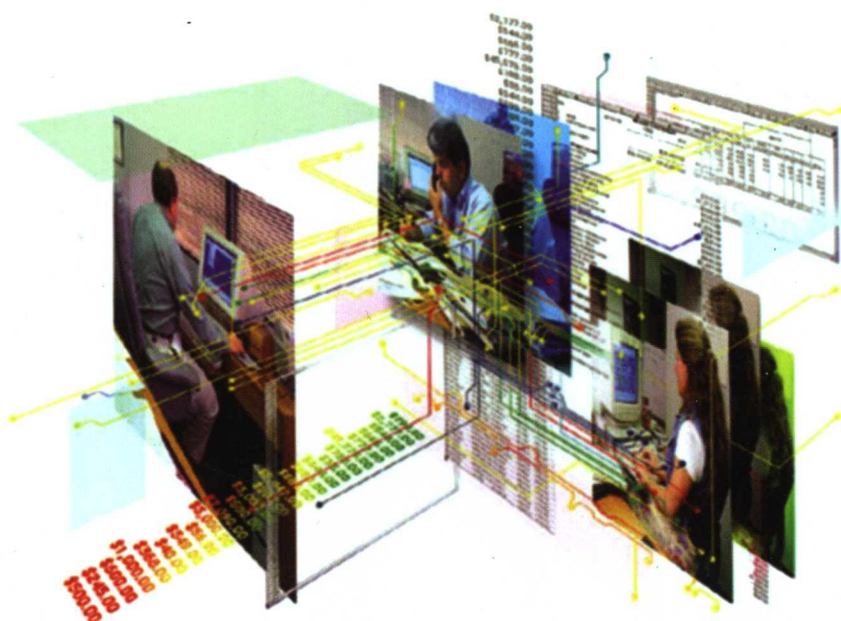


信息方法概论

蔡筱英 金新政 陈 氢 编著



 科学出版社
www.sciencep.com

信息方法概论

蔡筱英 金新政 陈 氢 编著

科学出版社

北 京

内 容 简 介

信息科学是一门正在形成和发展中的学科,其研究对象是构成人类社会发展的三种基本资源[物质(材料)、能量(能源)和信息]的信息。信息科学具有其特殊的研究方法,即以信息方法为核心而构成的信息科学方法论体系。本教材是为了适应信息时代的需要,为培养实务型的信息分析和应用人员,在吸收国内外先进的信息分析理论与方法的基础上编写而成。本书在系统论述信息学各种方法的基本定义、原理、特征、功能和作用的基础上,对其进行了深入的研究,并列举一些应用实例,以此说明这些方法在信息学研究中的重要性和实用性。

本书图文并茂,结构合理,内容翔实,适合卫生管理、临床医学等专业学生作为教材使用,也可作为卫生管理人员和医院信息设计开发等人员的参考书。

图书在版编目(CIP)数据

信息方法概论/蔡筱英,金新政,陈氢编著. —北京:科学出版社,2004

ISBN 7-03-013068-5

I. 信… II. ①蔡… ②金… ③陈… III. 信息学—科学方法论 IV. G201

中国版本图书馆 CIP 数据核字 (2004) 第 022046 号

责任编辑:李国红 / 责任校对:柏连海

责任印制:刘士平 / 封面设计:卢秋红

版权所有,违者必究。未经本社许可,数字图书馆不得使用。

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2004年8月第 一 版 开本: 787×1092 1/16

2004年8月第一次印刷 印张: 21 1/2

印数: 1—3 000 字数: 506 000

定价: 39.00 元

(如有印装质量问题,我社负责调换(环伟))

前 言

人类已经过机器隆隆的机械社会进入电光闪闪的信息社会,在信息社会的任何时间空间都充满信息,但对海量信息进行深入而有效的分析需要借助一定的科学方法。所谓方法即是实现目的的途径、完成任务的手段、解决问题的方式。人类的自我实现总要依赖一定的方法,任何物质文明与精神文明的创造与提高总是同相应的方法联系着的。广义方法的进步会促进社会的进步,任何宏伟目标都要借助方法去实现。可以这样说:真理诚可贵,方法价更高。

本教材的出版是为了适应信息时代的要求,培养实务型的信息分析和应用人员,作者在同仁们的鼓励与关怀下编著出版这本书旨在既适合我国国情又吸收国内外先进的信息分析理论与方法,以期训练培养和提高读者信息学方法的科学素养。本教材具有实用性、技术性、先进性、系统性、前瞻性、资料性等特点。

参加编写本教材的教师均长期从事本学科及相关学科的教学、研究和实践工作,积累了丰富的教学、科研和实践经验。本教材在原有讲义及研究成果的基础上,通过不断的补充和修改,将零散的资料系统化,融合为一个整体,并大量吸收了国内外有关文献和最新研究成果。本书共 16 章,第一、三、四、五、六、七、十四、十五和十六章由蔡筱英副教授编写,第二章由金新政教授编写,第八、九、十、十一、十二和第十三章由陈氢博士编写。我们在撰写本书的过程中,参阅和引用了国内外许多学者的著作(书末参考文献和书中注释列出了这些作者和著作的名称),我们向这些作者表示由衷的感谢,他们的成果为本书提供了丰富的素材。同时,本书的出版得到了科学出版社的大力支持,在此一并表示感谢!

由于我们水平有限、时间仓促,书中缺点错误在所难免,敬请广大读者和专家批评指正。

编著者

2003 年 7 月于武汉

目 录

前言

第 1 章 绪论	(1)
1.1 信息	(1)
1.2 信息学	(9)
1.3 信息方法	(14)
第 2 章 新老三论	(28)
2.1 系统论	(28)
2.2 信息论	(38)
2.3 控制论	(42)
2.4 耗散结构理论	(46)
2.5 协同学	(50)
2.6 突变论	(52)
第 3 章 信息方法论概述	(54)
3.1 信息方法论的基本原理	(54)
3.2 信息方法论体系结构	(58)
第 4 章 经验科学方法	(63)
4.1 专家调查法	(63)
4.2 信息咨询方法	(69)
4.3 观察方法	(72)
4.4 预测方法	(78)
4.5 德尔菲法	(86)
第 5 章 逻辑思维方法	(97)
5.1 比较法与分类法	(98)
5.2 归纳法与演绎法	(108)
5.3 分析法与综合法	(120)
5.4 证明与反驳法	(131)
5.5 抽象与概括法	(135)
第 6 章 信息计量学方法	(140)
6.1 信息计量学概述	(140)
6.2 信息计量学的基本原理	(145)
6.3 网络信息计量学	(148)

第 7 章 引文分析方法	(158)
7.1 引文分析方法概况	(158)
7.2 引文分析方法的应用	(166)
7.3 《科学引文索引》	(176)
第 8 章 内容分析法	(183)
8.1 内容分析法概述	(183)
8.2 内容分析的应用	(188)
8.3 主题词词频分析	(191)
8.4 指示词词频分析	(193)
8.5 篇幅分析	(198)
第 9 章 多元分析法	(203)
9.1 回归分析	(205)
9.2 聚类分析	(209)
9.3 判别分析	(214)
9.4 主成分分析	(217)
9.5 因子分析	(224)
第 10 章 趋势外推法	(231)
10.1 趋势外推法概述	(231)
10.2 线性趋势外推法	(232)
10.3 指数曲线法	(234)
10.4 生长曲线法	(235)
10.5 包络曲线法	(237)
第 11 章 时间序列法	(239)
11.1 时间序列法概述	(239)
11.2 移动平均法	(242)
11.3 指数平滑法	(246)
11.4 季节变化预测法	(253)
第 12 章 模型模拟法	(258)
12.1 模型方法	(258)
12.2 模拟方法	(274)
12.3 应用实例:远程教学系统模型研究	(281)
第 13 章 最优化方法	(284)
13.1 最优化方法概述	(284)
13.2 线性规划	(290)
13.3 目标规划	(294)
13.4 动态规划	(298)
第 14 章 相关分析方法	(304)
14.1 相关关系的基本概念	(304)

14.2	相关分析的基本内容·····	(306)
14.3	相关分析的类型和参数·····	(308)
14.4	相关分析的应用·····	(310)
第 15 章	系统分析方法 ·····	(313)
15.1	系统分析的含义·····	(313)
15.2	系统分析的内容·····	(314)
15.3	系统分析的步骤·····	(314)
15.4	系统分析中的注意事项·····	(317)
15.5	应用实例·····	(318)
第 16 章	移植法 ·····	(321)
16.1	移植法的基本概况·····	(321)
16.2	科学研究中的移植法·····	(324)
参考文献	·····	(333)

第 1 章 绪 论

1.1 信 息

“信息”(information)是一个既古老又年轻的词汇。我国古人很早就已采用了“信息”的概念,例如在“梦断美人沉信息,目穿长路倚楼台”(南唐李中《暮春怀古人》诗)和“欲传春信息,不怕雪埋藏”(南宋陈亮《梅花》诗)的优美诗句中就出现了“信息”这个词汇,而且用得非常恰当:前者有“音信”、“消息”之意,后者则已经给人以一种现代化的信息意境。

然而“信息”一词大量、广泛的应用却是最近二三十年的事情。今天在报刊、广播、电视、计算机网络等传媒以及人们日常的交谈中,都经常可以看到或听到诸如“商品信息”、“信息服务”、“信息时代”、“信息传播”一类名词。可见,“信息”这个词已经悄然进入我们的生活,成为我们这个时代的鲜明特征。

信息同物质、能源一样重要,是人类生存和社会发展的三大基本资源之一。物质相对直观,信息相对抽象,能源介于两者之间。这三者显示出的阶段性的历史进步,标志着人类认识、利用自然,在经历了数万年的游牧、几千年的农耕、几百年的工商后,人类社会迈进了辉煌的信息时代。信息作为一种富有生命力的资源,将如同阳光、空气和水一样,成为人类生活必不可少的要素。

1.1.1 信息的概念(the concept of the information)

信息是一个内容丰富、运用普遍、含义又相当模糊的概念,要对信息一词做出确切的定义是很困难的。现在各种有关信息的定义多达几十种。正如 19 世纪的印度哲学家辨喜所说,“世界上最伟大的东西是最简单的东西,它和你自己存在一样简单”。信息概念广泛地渗透到各门学科之中,人们可以根据各学科自身的特点为信息做出各种各样的定义。

一般认为,信息是关于客观事实的可通信的知识。这是因为:

第一,信息是客观世界各种事物的特征的反映。这些特征包括事物的有关属性状态,如时间、地点、程度和方式等等。

第二,信息是可以通信的。大量的信息需要通过各种仪器设备获得。

第三,信息形成知识。人们正是通过人类社会留下的各种形式的信息来认识事物、区别事物和改造世界的。

对于“信息”的理解可以有“广义的”和“狭义的”两个层次。“广义的”从纯客观的立场来定义信息,因此又叫“本体论信息”。本体论信息认为,任何一个事物的运动状态以及状态变化的方式,都叫信息。例如,花落、鸟啼、气温变化、股市的起落等等,都是信息。信息是一种客观存在,与我们主观是否感觉到它的存在没有关系。

“狭义的”从认识主体的立场来定义信息,因此又叫“认识论信息”。认识论信息认为,只

有那些认识主体(譬如说人、动物或机器)所能感受得到的“某个事物状态及状态的变化方式”才是信息;那些认识主体感觉不出来的,或者感觉到了但不能理解的东西,都不叫信息。例如,在中国古代,以点燃烽火或击鼓来通报军情,时明时灭的烽火以及有节奏的鼓声都是人能感觉到的,而且事先的约定又使特定的接收者能理解它所代表的意思,因此,从认识主体的角度上看,烽火和鼓声都是“信息”。相反,一些至今仍未被破译的古代文字和符号,由于我们尚不能理解它,所以还不能说它是“信息”。

在我们日常生活中,信息是指具有新内容、新知识的消息。但是,信息的含义要比通常的消息广泛得多。它包括消息、情报、指令、代码以及包含有一定内容的信号等。我们的感觉器官随时在接收各种信息,如语言、文字、图像、颜色、声音、自然景物等等。人们在获得信息之后,就能够消除某种知识上的不确定性,改变原有的知识状态,由不知到知,或由知之不多到知之较多的状态。例如:你已知道有一位远方朋友要来做客,但不知何时到达。当你接到他的电报后,知道了他到来的时间。电报中的信息就使你消除了原来关于这件事的不确定性。

在通信和控制系统中,信息是系统之间普遍联系的特殊形式,它与组织结构密切相关。如,遗传信息与核苷酸的排列顺序有关;计算机的技术信息与所给指令和程序有关;单词的信息与字母的排列顺序有关;大自然景色给人们的信息则与阳光的强弱、山脉的蜿蜒、瀑布的喧腾、树叶的颜色、小鸟的鸣声等有关。

由于各学科领域的专家学者们对信息意义的理解不同,所以,很难给信息下一个被各学科普遍接受的、统一的定义。我们认为:信息是物质的一种属性,它必须通过主体的主观认识才能被反映和揭示。信息是用来表现事物特征的一种普遍形式,是事物存在的方式或运动状态,以及这种方式、运动状态的直接或间接的表述。在此,“事物”泛指一切可能的研究对象,既包括外部世界的物质客体,也包括主观世界的精神现象;“运动”泛指一切意义上的变化,包括机械运动、物理运动、化学运动、生物运动、思维运动和社会运动等;“运动方式”是指事物运动在时间上所呈现的过程和规律;“运动状态”则是指事物运动在空间上所展示的形状与态势。

1.1.2 信息的类型(the type of the information)

由于信息与通信有着密切的联系,因此随着对通信问题的深入探讨,就产生了三种不同的信息类型。

通信可分为三个不同阶段:第一阶段是语法的,这阶段的主要目的是保证信号尽可能正确地接收即复制出来;第二阶段是语义的,它和语言的意义有关,保证理解所接收到的信号、语义,这就必须借助于记忆;第三阶段是语用的,把信号变为行动,行动也包括“代码化了的和符号化了的行动系统”,即“谈话和书写的语言”。

第一种信息为语法信息,即信息符号之间的关系。语法是从符号学借用过来的术语,在符号学中把符号之间的关系称为语法。在信息论中,把只研究信息符号之间统计关系的部分称为语法信息。按照语法信息设计的系统,只传递信息,而不考虑该信息是否真实、可靠、有用。这类信息系统的服务质量不高,属最低层次的认识论信息。

第二种信息为语义信息,即信息符号与其实体意义之间的关系。语义是从符号学借用

过来的一个术语。在符号学中,把符号与实体间的关系称为语义。在信息论中,信息的语义涉及到信息本身的含义及它在逻辑上的真实性和精确性。它是从意义和内容的角度反映信息特性。有语义的信息,即语义信息的描述,需用状态描述代替基本事件,用句子代替随机变量,用逻辑概率代替统计概率。如果原先有 N 个可能情况,即 N 件状态描述,而句子 I 的信息使可能情况的数目减少到 N_r ,则句子 I 给出的语义信息为:

$$I = \log_2(N/N_r) \text{ 比特}$$

按照语义设计的信息系统要比语法信息系统的服务质量高得多,它将排除那些逻辑不真实的信息,并用更精确的信息来代替那些语义不精确的信息。这是较高层次的认识论信息。

第三种信息为语用信息,即信息符号与其使用者之间的关系。语用是从符号学借用过来的术语。在符号学中,把符号与其使用者之间的关系称为语用。语义反映符号的功能含义,语用反映符号的操作含义。在信息论中,任何信息的语用特性都与信息的发出者及接收者的过去的经验、现在的环境、思想状态有关。通常说某种信息是“有用的”、“无用的”、“有价值的”等等,都是对信息的语用性判断。如果按照信息的语用设计信息系统,不仅包含语法与语义的信息系统的全部内容,而且还将包含分析处理大量历史资料,信息使用者个人的计划、习惯、爱好、环境、心理等方面的大量资料,能从多方面分析比较,因而具有做出最佳决策的能力。这是最高层次的认识论信息。

一般说来,语法信息、语义信息和语用信息是密切相关的,不可能撇开其中的一个方面而孤立地研究其他方面。当认识和研究一个事物时,人们多是遵循从语法、语义到语用的认识顺序。这三种信息都有它自己的信息理论。目前比较成熟的是语法信息理论,即以仙农信息论、维纳控制论为代表的信息理论。它在许多实际应用方面取得了卓越成果,解决了许多当代的重大理论和技术问题,如自动化技术、人工智能、宇宙空间探索等等。语义信息理论和语用信息理论尚处于初始发展阶段。

在日常生活中,人们常把信息与消息和信号等同起来,其实这样的认识并不确切。信息、消息和信号之间有密切联系,信息常以消息形式表现出来,并通过信号来传递,但是三者之间是有区别的。消息有可能包含甚为丰富的信息,但也可能信息甚少,若这种信息并未给人们带来新知识的话,那么这种消息所包含的信息实际上等于零。所以信息是给人们带来新知识的消息。信息与信号也是有区别的,信号是携带信息的载体,信息则是这个载体所携带的内容,同一种信息可用多种信号来表示,一种信号也可能用来传递多种信息。

此外,信息和数据(data)也是有区别的,数据是对某种情况的记录,包括数值数据(例如各种统计资料数据)以及非数值数据两种,后者如各种图像、表格、文字和特殊符号等;而信息则是经过加工处理后对管理决策和实现管理目标或任务具有参考价值的数据,它是一种资源。作为资源的信息具有如下特点:①信息是和决策密切相关的,正确的决策有赖于足够可靠的信息,信息又是通过决策来体现其自身价值的。②信息可影响甚至决定组织的生存,能够给组织带来收益。③获取和利用信息时往往要花费一定的费用成本,正因为如此,信息利用者就必然会考虑到他们的花费对改进管理带来的功效是否合算,来决定要否获取和利用该信息。④信息往往具有很强的时效性,延迟的信息可使其功效减少或全部消失,甚至可能起到截然相反的作用。

我们把信息与消息、信号、数据、情报、物质和能量的关系描述如下:

(1) 消息是信息的外壳,信息则是消息的内核。比如,同是一分钟的消息,有的包含信息量很大,有的则可能很小。

(2) 信号只是信息的载体,信息则是信号所载荷的内容。同一个信息既可以用这种信号载荷,也可以用那种信号载荷。比如,用“0,1”或用“正、负”来载荷。

(3) 数据只是记录信息的一种形式,而且不是惟一的形式。

(4) 情报在汉语中只是一类专门的信息(情报在日语中就是信息)。

(5) 物质是信息的载体,物质的运动是信息的源泉,而信息是事物运动的状态和方式,它并不就是事物本身。

(6) 信息与能量也是既相联系又有区别的两类范畴。传递信息需要能量,驾驭或控制能量则需要信息。

1.1.3 信息的本质(the innate character of the information)

第一,信息与物质密切相关联,但不是物质本身。信息定义中的“事物”可以是物质客体,因此,信息可以是物质客体的运动状态和状态变化的方式。例如“车辆在匀速前进”,就是与车辆这种物质客体相关联的信息。可见,信息与物质有密切的联系。没有物质和物质的运动,就没有与物质相联系的信息。但是,信息又不等同于物质本身。“车辆在匀速前进”不等同于“匀速前进的车辆”。

第二,信息与精神密切相关,但又限于精神。信息定义中的“事物”也可以是精神现象,因此,信息可以是精神现象的运动状态和状态变化的方式。例如“他很悲伤”,就是与他的精神现象相关联的信息。可见,信息也与精神现象有密切联系。没有精神与精神的运动,也就没有与精神相关联的信息。但是,信息又限于精神本身。

第三,信息与能量相辅相依,但不是能量。由信息的定义可知,没有能量就没有事物的运动,也就没有运动状态和状态变化的方式,于是就没有信息。但是两者又有本质的区别:信息是事物的运动状态和状态变化的方式,能量是事物做功的本领。

第四,信息具有知识的秉性。信息论的创始人仙农(C.E.Shannon)在他的代表作《通信的数学理论》中写道:“信息是能够用来消除不定性的东西”。例如,我们在收听(或收看)天气预报之前,不知道明天的天气情况如何(即存在“不确定性”),而在听了(或看了)天气预报后,知道明天的天气是晴、是阴还是雨,还知道大致的气温情况(这叫“消除”了某些“不确定性”)。在看完一本有关环境保护的科普读物后,我们了解了许多过去不了解的环境保护知识,增强了环境保护意识,以至进一步将它应用于指导宏观决策。所有这一切,都说明信息具有知识的秉性。它能改变人们的知识状态,使人们对某些事物从不知到知,从知之甚少到知之甚多。

第五,信息普遍存在,而且永不枯竭。事物(物质和精神)无处不在,而且总在不停地运动,不停地按照某种方式改变着自己运动的状态,也就是不停地产生信息。因此,信息普遍存在,而且不断更新,永不枯竭。了解信息的这一特点十分重要。它要求任何一个人,特别是领导者要不断掌握新的信息,并据此做出及时、准确的判断和正确的决策。如果把信息看成是一成不变的,就会做出错误的判断,使政策、决定成为“明日黄花”。

第六,信息可被感知、传递、处理和共享。信息可以被主体直接或间接所感知(通过技术措施,如传感技术),被感知的运动状态和状态变化方式可以被记忆(存储)、被变换(重新表示)和被加工处理。另一方面,由于信息只是事物的运动状态和状态变化的方式,不是事物本身,因而就可以通过一定的技术措施使它脱离原来的事物,附着于新的事物(称为载体)。这是信息可以被摄取、记录、复制、传输以及被共享的根据。而且,这种复制和共享可以重复进行。

1.1.4 信息的属性

物质在使用中是消耗的,能量就其个体而言在使用中也是消耗的,就其整体而言则是守恒的,信息虽然也是一种资源,但它是不同于物质和能量的一种特殊的资源。这种资源具有许多属性,其中最主要的有以下几点。

1. 价值性 信息是经过加工并对生产经营活动产生影响的数据,是一种资源,因而是有价值的。其价值在于它的知识性和技术性。事实是信息的核心价值,不符合事实的信息不仅没有价值,而且可能价值为负。信息的价值 = 使用信息所获得的收益 - 获取信息所用成本。

2. 可处理性 信息是可以处理加工的,如信息的有序化、压缩、提取、再生等。信息通过标引、分类、组织等有序化处理后,便于检索利用;通过概括、归纳、总结等方法可使信息更加精练;通过录放机、计算机等设备可以提取存储在磁带上的信息;收集到的信息经过处理后可以用语言、文字、图像等形式再生成。

3. 可识别性 由于信息反映了物质和能量在空间结构和时间顺序上分布的不均匀的状态,这样人们就可以对信息进行识别。通过感官进行的,是直接识别;通过各种探测手段的,为间接识别。不同的信息源可以用不同的识别方式,在识别中要特别注意选择。选择不当,就可能导致识别不佳、一无所获,甚至失误。

4. 可转换性 信息可以从一种形态转换为另一种形态。如物质信息可以根据需要转换为语言、文字、图表、图像等信息方式,也可以转换为电子计算机的代码、电信、广播、电视信号,而代码和信号又可以转换为语言、文字、图表、图像等等。认识了这个特征,我们接受知识和传播信息的渠道和范围就扩展了。

5. 可存储性 如果信息无法存储,人类至今仍会处于蒙昧时代。结绳记录是人类创造的最原始的存储信息的方法。以后纸和笔的利用,印刷术的发明,留声机的产生,照相术的出现,磁记录技术的突破,激光应用技术的推广,半导体规模集成电路的发展,使得存储信息的物质载体大大增加,存储形式更加多样。动物的信息存储叫记忆,人脑的记忆系统由长期记忆和短期记忆两部分组成,电脑也由内存储器 and 外存储器两部分存储信息。信息的可存储性为信息的积累、加工和在不同场合下的应用提供了可能。

6. 可替代性 信息已经成为现代人类社会能够进行交换和创造价值的东西,是生产力、竞争力的一个重要因素。信息的正确利用,可以替代资金、劳动力和物质材料。如果在生产领域中掌握了信息,就可以减少资金、劳动力和原材料的消耗,降低成本,提高产品质量,提高劳动生产率。如果在流通领域中及时获得了市场信息,就可以合理组织货源,打开销路,既不积压(商品)又不“断档”,最佳地搞好购、销、调、存,提高经济效益和社会效益。

7. 可传输性 凡是信息都可以通过一定的信道和载体进行传输。从信源(发信者)到信宿(收信者)是信息的空间传输,从过去到将来是信息的时间传输。信息存储实质是时间传输的延续。不同的材料、能量构成的信道可以传输同一个内容的信息,不同的信息可以通过同一种信道传输。个人间的信息传输主要靠语音、表情、动作等,社会性活动的信息则主要通过报刊、告示、广播、电视以及其他通信工具(如吹军号、打旗语、放信号弹等)进行传输。又如:在因特网上可收发邮件,手机可收发短消息,把文字记在笔记本上、放在磁盘、磁带、光盘上可传递使用。传输的速度和效率取决于传输手段和通信工具,目前最先进的通信技术是光导纤维通信和人造卫星通信,这两种通信技术通信容量大,传输速度快,效果可靠。

8. 无限性 在人类生存和接触到的一切领域都随时产生着信息。随着时间的推移,信息又在无限地产生和发展。物质世界是无限的,对物质世界的认识是无限的,人类的创造力是无限的,因而信息也是无限的。信息的无限性还表现在它的可扩充性。用于某种目的的某种信息可能会随着时间的推移而失去作用,但对于另一个过程而言又会成为一种有用的信息。例如,昨天的天气预报到了明天就只有历史意义了,但对研究一个时期的天气变化规律来说又是重要的信息来源。从总体而言,信息是随着它的利用而扩充的。

9. 共享性 信息是资源,但又有不同于一般资源的属性。其他资源在交换过程中遵循等值交换原则,失去一物才能得到另一物。而信息则不同,在交换信息的双方,不但不会失去原有的信息,而且还会增加新的信息。在信息进行单方面转让过程中也是这样,转让者不会因转让信息而失去信息,相反会使自己掌握的信息得到巩固。如教师向学生传授知识,科研单位转让专利成果等。这些都说明,信息是作为一种共享性资源而存在的,是可以利用扩散的方式而共享的。例如:对于同一个学术报告,一个人去听与一百个人去听,各人所得的信息可以完全相同,绝非一个人听时所得信息为百分之百,百人听时为百分之一。有的信息涉及到商业的、政治的、军事的秘密,扩大对这类信息的享用时可能影响某些享用者对这类信息的利用,但不会改变信息的内容。

除以上列出的信息的九种基本属性外,还有普遍性、客观性、相对性、抽象性、动态性、异步性等。在信息的实际利用过程中,还要注意信息的滞后性(即任何信息总是产生、传递在事实之后的,先有了事实,而后才可能有信息)以及不完全性(任何关于客观事实的知识都不可能包揽无余,为此,在收集、提供信息时,要力求真实、快速、全面、准确。)

1.1.5 信息量(information quantity)

一门科学,只有当它从定性研究进入定量研究后,才能成为“精确科学”。信息学要成为一门精确科学,就必须对信息进行度量。信息量是信息论中量度信息多少的一种物理量。它从量上反映具有确定概率的事件发生时所传递的信息。信息的量度与它所代表的事件的随机性或各个事件发生的概率有关,当事件发生的概率大时,事先容易判断;有关此事件的消息排除事件发生的不确定程度小,则包含的信息量就小,反之则大。

信息作为事物运动中“可能出现”的一般形式,它是可以测量的。在我们日常经验中常常采用重复多次通信或传递消息的方法(如多次向远方喊话、发出信号等等),使对方能够在噪音和背景干扰大的情况下收到信息,能够听清楚、看清楚和搞清楚它的内容。就是在这种实践经验的基础上,找到了用重复(发)和积累(收)的方法接收掩蔽在强噪音中的微弱信号。

由此可见,信息的数量大小是可以测定的,并且符合数学上累加法则。从这一点出发,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。统计热力学中,熵是系统的无序状态的量度,即系统的不确定性的量度。假设用 n 表示 n 个消息代表的可能事件,各个事件的概率分别是 $P_1, P_2 \cdots P_n$, 那么每个消息的平均信息量是:

$$H(x) = -K \sum_{i=1}^n P(x_i) \log P(x_i)$$

式中 $H(x)$ 是以 x 为某事物(事件集)的信息量, K 为系数,与不同的单位制有关, $P(x_i)$ 为 x 事物中某事件可能出现的概率, $i=1, 2, 3, \cdots n$; n 为 x 事物可能出现的事件总数。

当对数底取 2, 且 $n=2, P(x_1)=P(x_2)=\frac{1}{2}$ 时, 则:

$$H(x) = -K \sum_{i=1}^2 P(x_i) \log_2 P(x_i) = 1$$

以此为信息单位,称为比特(bit), 则有 $K=1$ 。1 比特信息量就是含有两个独立等概率可能状态的事件所具有的不确定性被全部消除所需要的信息。在这种单位制下,信息量的公式就成为:

$$H(x) = -K \sum_{i=1}^n P(x_i) \log(x_i)$$

从分子运动论的观点来看,一个系统的热运动是和它的无序程度有关的,在没有外界干预条件下,一个系统总是自发地从有序到无序的方向发展。在这过程中,系统的熵的变化总是增加的。但是,信息量和熵所反映的系统运动过程 and 方向相反。系统的信息量的增加总是表明不确定性的减少,有序化程度的增加。其减少的量的多少反映了该系统被消除的不确定度的多少,即有序化程度增加的多少。因此,信息在系统运动过程中可以看做是负熵。信息量愈大(即负熵愈大),熵值愈小,反映了该系统的无序度愈小,有序化程度愈高。在一个系统中,组织化和无组织化即有序和无序是对立统一的。两者互相联系、不可分割,并且在一定条件下可以互相转化。当我们知道某系统的无序度时,同时也知道了该系统的有序度。

信息量的大小取决于消息所讲事件发生的可能性大小。如果所讲事件是比较少见的,发生的可能性小,则这条消息消除的不确定性要多些,其信息量也就大些。例如,如果你的子女或朋友今年要参加高考,当有人告诉你,今年高考于暑期进行,你大概会认为这话等于没说。因为高考在正常情况下都放在暑期,这条消息对你来说,信息量极小。如果有人告诉你,今年高考要放在春季了,你肯定会大吃一惊。因为这种情况发生较少,你从这条消息中获得的信息量就大些。用信息论的观点来描述,可表达为概率小的事件,信息量大。

信息的计量有两方面的要求:一方面,要求提高通信系统传输信息的能力(即通信系统的工作效率),就是要解决在同一个信道中怎样加大信息的容量,或加快信息传输速度,从而提高通信的经济性,减少能量消耗;另一方面,又要求提高信息传输的可靠性,力求减少噪声的干扰,提高通信的质量。信息方法暂时撇开信息的语义方面,单从技术方面研究信息量的计算问题,要运用数学理论寻找合适的数学模式。

信息度量表述了系统的有序化过程,由此我们可以给出更广泛的信息含义:信息是任何一个系统的组织性、复杂性的度量,是有序化程度的标志。

1.1.6 信息传递(information transfer)

信息之所以称为信息,就是它的可传递性。信息传递就是信息由发生源经过一定的媒介体输送到接收器的过程。信息的传递有时间传递和空间传递之分。时间传递是指信息的储存(如用写字、打字、印刷、照相、唱片、磁带、磁盘等介质储存信息),使信息随时间的流逝而传递下去。空间传递是指由通信传输系统把信息从一方传到另一方,通信传输系统是由发送信息和接收信息的电子设备以及通信线路等硬设备与通信软件组成的,它把信息产生单位和信息用户连接起来,实现信息的空间传递。现代通信传输系统(能够把电子计算机与各种不同类型的用户终端联结成交时的信息传递系统)可以有不同的传输速率和传输方式,

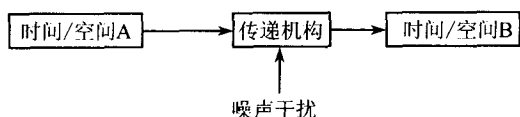


图 1-1 信息传递基本模型

在通信控制部件的控制下,实现信息的自动传递。传递信息有几个重要的环节,一个称为信息源,一个称为信息的接受者。在信息源和接受者之间还必须存在传递信息的通道,信息通道把信息源“发生了什么事件”传递给接受者。其基本模型可以表示如图 1-1 所示。

那么,信息传递过程中的传递是什么意思呢?是传递物质还是传递能量呢?都不是。信息的传递是指可能性空间缩小过程的传递。信息源发生的确定性事件使它的可能性空间缩小了,经过传递,这种缩小最终导致了信息接受者可能性空间的缩小。因此所谓信息的传递也就是可能性空间变化的传递。从这点来看,传递信息与控制有密切的关系,所谓控制也是一种使可能性空间缩小的过程。实际上,信息的传递离不开控制,控制也离不开信息的传递。

我们常说,DNA 携带着遗传信息,这是什么意思呢?大家知道,DNA 是一个双螺旋结构,它含有四种不同的碱基:腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶。这四个碱基可以组成不同结构的 DNA。不同的碱基排列的形式控制着不同种氨基酸,不同氨基酸可以结合成不同的蛋白质,组成不同的酶,不同的酶控制着不同的细胞形态,不同的细胞形态决定了生物不同的特征,决定了这一种生物是狗、是猫,还是耗子,这里我们看到一个控制链。DNA 的结构一旦确定(可能性空间缩小到某一状态),就引起一系列可能性空间的缩小,最后决定了遗传特征可能性空间的缩小,决定了某一生物的形态。所谓 DNA 携带遗传信息,实际上是指 DNA 结构可以控制物种的形态和特征。

传递信息需要我们实行某种控制,反过来,控制过程又必须依赖信息的传递。很多时候,我们不能实现有效的控制,是没有获得足够的信息之故,生物反馈在这方面提供了一个很好的例子。

我们知道,一个人不能控制自己的心跳速度,也不能控制自己的血压高低。所谓不能控制,人们认为这是人的意志不能对它们施加影响,因此这一类内脏器官的活动通常被称做“不随意运动”,跟人的意志能够自由控制的骨骼肌的随意运动相区别。实际上,人的意志之所以不能随意控制这些器官的活动,很大程度上,是没有获得它们活动情况的足够的信息。一般说来,位于内脏部位的各种内感受器接受刺激并将冲动传入中枢后,显然有时也可引起一些较模糊的感觉(如饱感、尿意等),但常常不引起明晰的主观感觉,而主要引起某种内脏

或躯体反射,使体内各器官、系统的活动自动达到平衡与协调,如使内环境理化功能相对稳定,心率、血压保持相对恒定等。而肌肉、关节的运动和位置的感觉被称为本体感觉,它们跟其他外感受器的感觉(如皮肤、视觉、听觉、嗅觉和味觉)一样,通过特异性传入系统传到大脑皮质相应的特定部位,能够引起清晰的特定的主观感觉,人的大脑每时每刻都可以清楚地接受到自己四肢位置和姿势的信息,因而也就可以有意识地控制它们。解剖学表明,人体大脑皮质的躯体感觉区中相应于手指和口唇部的区域最大,说明大脑对这些部位的信息最敏感,因此人对手指和口唇的控制也最自如,这显然跟人类的劳动和语言有关。根据以上原理,发展了生物反馈疗法。这种方法认为,只要人能够时时刻刻清晰地获得自己内脏活动的信息,比如能够像感觉到手上拿了什么东西那样感到自己的血压是多少,再经过适当的训练,就可以控制自己的血压了。生物反馈治疗仪的原理就非常简单,它是把心率、血压、痛觉等内脏信息转换为数字、灯光、音响等显示出来,利用眼睛、耳朵等外感受器输入大脑。由于条件反射的本领,大多数病人经过一段时间的训练之后,离开反馈治疗仪也能够获得一定的控制能力。这本质上就是利用反馈放大了自己的控制能力,从而控制了自己原来没有能力控制的行为。信息论的研究指出,这种反馈之所以能够做到,关键在于构成了信息传递的新通道。

1.2 信 息 学

信息学是研究信息现象及其规律、认识信息和利用信息的科学。当然,这里所说的信息是指“全信息”。下面我们从信息学的研究对象,研究内容、研究方法、研究目标四个方面来进一步理解信息科学的概念。

1.2.1 信息学的研究对象

传统的自然科学包括物理学、化学、天文学、地学、生物学等等,它们的研究对象都可以归结为物质与能量。它们的研究内容主要是各种各样的物质结构和形形色色的能量转换规律。传统自然科学的基础性和重要性是毋庸置疑的。问题在于,随着科学技术的发展,越来越多的复杂信息系统,如生物系统、人体系统、思维系统、社会系统等等,日益成为科学研究的重要对象。显然,对于这些以信息流程为主导特征的高级、复杂的信息系统来说,仅仅依靠物质结构和能量转换规律去分析是远远不够的,即不能以此去揭示它们工作机制的奥秘。比如,大脑是由生物物质构成的,也包含着丰富的生物—化学和生物—电等的能量转换,但是,对于大脑来说,更为重要的却是这些物质和能量所支持的信息活动。

因此,信息学的崛起,是科学技术研究深入的必然结果。信息学的发展,将使原先以物质和能量两者为中心观念的传统自然科学转变为以物质、能量、信息三者为中心观念的现代自然科学,也就是使传统自然科学的二元结构转变成为现代自然科学的三元结构,这将在很大的程度上改变自然科学的发展方式和思维方式。

信息作为信息学研究的基本对象,与传统科学相比,在人类科学史和认识史上,它也是一种新颖的、独立的研究对象。这可以从两个相互联系着的方面来加以理解。

一方面,从本体论意义来说,信息是事物运动的状态和方式,它与物质既相联系又互有区别。物质是信息的载体。物质的运动是信息的源泉,而信息是事物运动的状态和方式,它

并不就是事物本身。信息与能量也是既相联系又有区别的两类范畴。传递信息需要能量,驾驭能量则需要信息。但是,信息毕竟只是事物运动的状态和方式,而能量只是物体做功的本领。因此,也不能把信息与能量等同起来。所以在研究信息时,应当把信息与物质、能量联系起来,从它们交互作用的动态发展过程中来研究信息的本质。

另一方面,从认识论意义来理解,信息是认识主体所感受或所表述的事物运动的状态和方式。人们要认识事物(这里所说的事物,不仅包括外部世界的实在客体,也包括主观世界的精神现象),必须从客体获取信息,即将本体论意义上的信息转变为认识论意义上的第一类信息。随着实践和认识矛盾运动的发展,这类信息经过高度复杂优化的信息处理(控制中心——大脑系统的处理)加工,形成认识论意义上的第二类信息,即形成判断,作出决策,然后通过效应器官作用于外部世界。这样看来,认识论意义上的信息过程,既包括客观主观化的过程,即从本体论意义上的信息转化为认识论意义上的第一类信息,也包括主观客观化的过程,即将认识论意义上的第二类信息作用于客观世界。由此可以看出,无论是认识世界还是改造世界,都贯穿着信息运动的过程,都体现着认识信息和利用信息的过程。

1.2.2 信息学的研究内容

当前,根据国内外学者特别是根据中国著名的情报学家——北京邮电学院副院长钟义信教授的研究成果,可以把信息学的内容概括为下列五个方面:

- (1) 探讨信息的本质并创立信息的基本概念。
- (2) 建立信息的数值度量方法,包括语法信息、语义信息和语用信息的度量方法。
- (3) 研究信息运动的一般规律,包括信息的提取、识别、交换、传递、存储、检索、处理、再生、表示、施效(控制)等过程的原理和方法。
- (4) 揭示利用信息进行有效控制的手段和开发利用信息资源实现系统优化的方法。
- (5) 寻求通过加工信息来生成智能和发展智能的动态机制与具体途径。

上述信息学五个方面的内容,既涉及到现代科学的广阔领域,如信息学广泛地渗透到系统科学、控制论科学、人工智能科学、认知科学、思维科学等等领域,也可以促进对信息学的哲学问题的思考,如信息的哲学本质,智能的哲学本质,信息与反映、意识的关系,人工智能与人类智能的关系等等。信息学所涉及到的上述诸多方面,人们都可以做深入的、开创性的研究。但是,如果从认识论角度考虑问题,则可以把信息学上述五个方面的内容做一次思维行程中的抽象,那么,就可以看出,关于信息的本质和度量方法的研究,目的是要从质和量两个方面来把握信息的实质,这是对信息的认识问题。而关于信息的处理、加工和利用以生成智能和发展智能等问题的研究,则是要揭示信息的运动规律,解决如何利用信息来为人类服务的问题。

什么是信息运动?什么是信息运动规律?分析表明,典型的信息运动的全过程包含信息的产生、信息的获取、信息的传递、信息的处理、信息的再生、信息的施用(“施用”是一个专门术语,有“实施应用”的含义)等过程,如图 1-2 所示。

图中还标出了研究每个过程规律的理论,它们分别是信息论(信息的概念与度量方法)、识别论(信息获取的规律)、通信论(信息传递的规律)、认知论—智能论—决策论(信息处理