



高等学校教材
Textbook for Higher Education

数据采集原理

肖忠祥 主编



西北工业大学出版社

“九五”中国石油天然气集团公司规划教材

数 据 采 集 原 理

主 编 肖忠祥
编 者 孟开元 尚海燕 金树波

西北工业大学出版社

【内容简介】 本书是中国石油天然气集团公司“九五”规划教材。全书分八章,较全面地介绍了数据采集系统的组成原理和 design 方法,内容包括理论基础、信号调理、浮点放大、A/D 转换、数据记录 and 系统设计方法等;对近年来兴起的若干新技术给予了重点介绍。

本书可作为测控技术与仪器、电子工程、自动化等专业的教材或参考书,也可供有关专业的技术人员参考。

图书在版编目(CIP)数据

数据采集原理/肖忠祥主编. —西安:西北工业大学出版社,2001.2
ISBN 7-5612-1334-4

I. 数… II. 肖… III. 数据采集 IV. TP274

中国版本图书馆 CIP 数据核字(2001)第 02975 号

出版发行:西北工业大学出版社

通信地址:西安市友谊西路 127 号 邮编:710072 电话:029—8493844

网 址: <http://www.nwpup.com>

印刷者:西北工业大学印刷厂

开 本:787 mm × 1 092 mm 1/16

印 张:15.25

字 数:367 千字

版 次:2001 年 2 月 第 1 版 2001 年 2 月 第 1 次印刷

书 号:ISBN 7-5612-1334-4/TP·199

印 数:1~3 000 册

定 价:19.00 元

前 言



数据采集技术是信息科学的一个重要组成部分。信息技术的核心是信息获取、通信和计算机技术,常被称为 3C 技术(即 Collection, Communication and Computer),其中信息获取是基础和前提。数据采集技术是信息获取的主要手段,它随着微电子技术和计算机技术的进步而得到迅速发展。目前各种各样的数据采集系统已得到广泛的应用,新型数据采集系统仍不断涌现。

“数据采集原理”是测控技术与仪器专业的主要专业课之一,一般安排在非电量电测技术(或传感器技术)课程之后,主要讲述从传感器输出的微弱电信号,经信号调理、模数转换到存储记录这一过程中所涉及的原理与技术。

本书的编写力求体现如下两个特色:

(1)系统性。尽可能全面地反映现代数据采集系统的组成原理和设计方法,书中包含了软件设计方法和系统标定等内容。

(2)先进性。力求反映数据采集系统近年来的最新进展,书中包含了 $\Sigma-\Delta$ 型高精度 A/D 转换技术、高速 A/D 转换技术、新型总线技术和虚拟仪器等内容。

全书共分八章。第一章是绪论,介绍了数据采集的基本概念、系统的基本组成和主要性能指标;第二章介绍与数据采集有关的理论基础,包括信息论基础知识、微弱信号检测的基本原理,重点介绍采样理论与技术;第三章讲述信号调理电路,重点是低噪声放大技术;第四章介绍 A/D 转换原理与技术,这是本书的重点部分,主要介绍了高速和高精度 A/D 转换原理及 ADC 的应用问题,还介绍了为扩大系统动态范围而设置的瞬时浮点放大器的原理;第五章简要介绍了数据记录原理;第六章讲述数据采集系统设计方法,包括系统结构、误差的合成与分配、总线及其接口、虚拟仪器等,最后还介绍了系统标定的有关概念与方法;第七章专门介绍软件设计方法;第八章以一种高级数据采集系统为例,介绍了实际数据采集系统的组成和关键部件的原理。

本书是编者们在教学、科研的基础上,参考近年来的最新文献编写的,在此编者对所引用的参考文献的作者们表示诚挚的谢意。

由于篇幅所限,书中对例题及某些常规内容做了必要的精简,在教学过程中,教师可根据实际情况增加某些内容及例题。

本书是中国石油天然气集团公司“九五”规划教材,由肖忠祥任主编,其中第一、二、四、五、六章由肖忠祥编写,第三章由孟开元编写,第七章由尚海燕编写,第八章由金树波编写。西安电

子科技大学张守宏教授仔细审阅了本书全部初稿,并提出许多宝贵意见,在此表示诚挚的感谢。

限于编者的学识水平和实际经验有限,书中定有不少缺点和错误,恳请读者批评指正。

编 者

2000 年 9 月

目 录



第一章 绪论	1
1.1 数据采集的基本概念	1
1.2 数据采集系统基本组成	2
1.3 数据采集系统的主要性能指标	3
第二章 数据采集的理论基础	5
2.1 信息论基础知识	5
2.1.1 有关信息的基本概念	5
2.1.2 信息的描述	7
2.2 微弱信号检测基础知识	13
2.2.1 微弱信号检测的概念	13
2.2.2 噪声及其抑制方法概述	13
2.2.3 接地的基本原则	15
2.2.4 微弱信号检测方法概述	18
2.2.5 相关函数和相关检测	19
2.2.6 取样积分器原理	25
2.3 采样理论	31
2.3.1 采样过程概述	31
2.3.2 采样定理	31
2.3.3 采样定理的几种基本形式	35
2.3.4 采样方式	36
2.3.5 量化与量化误差	37
2.3.6 编码	41
第三章 模拟通道电路	44
3.1 集成运算放大器的种类及作用	44
3.1.1 集成运算放大器的发展概况	44
3.1.2 集成运算放大器的分类	44
3.1.3 常用集成运算放大器型号对比	45
3.1.4 集成运算放大器使用选择原则	45
3.2 测量放大电路	46

3.2.1	测量放大器电路原理	46
3.2.2	测量放大器主要技术指标	48
3.2.3	测量放大器集成芯片简介	49
3.2.4	测量放大器集成芯片的使用示例	52
3.3	有源滤波器的设计	53
3.3.1	有源滤波器的分类和基本参数	53
3.3.2	组成典型二阶有源滤波器的基本方法	55
3.3.3	二阶有源滤波器传递函数的分析与设计	56
3.3.4	有源滤波器的设计	59
3.4	频率合成器电路	65
3.4.1	CC4046 单片锁相环电路	65
3.4.2	用 L562 单片锁相环构成的频率合成器	68
第四章	模数转换电路	70
4.1	模拟多路开关(Analog Multiplexer, MUX)	70
4.1.1	模拟多路开关的种类	70
4.1.2	多路开关工作原理	71
4.1.3	多路开关的主要技术指标	73
4.1.4	多路开关的选用	73
4.2	采样保持电路	74
4.2.1	采样/保持器的基本原理	74
4.2.2	采样/保持器的基本结构	75
4.2.3	采样/保持器的性能参数	77
4.3	A/D 转换器	78
4.3.1	A/D 转换器的主要性能指标	78
4.3.2	A/D 转换器分类	79
4.3.3	高速 A/D 转换器	81
4.3.4	高精度 A/D 转换器—— Σ - Δ 型 ADC	86
4.4	浮点放大器	94
4.4.1	基本概念	94
4.4.2	IFP 的结构	95
4.4.3	IFP 的增益比较器	98
4.4.4	循环浮点放大器原理	100
4.5	A/D 转换器的应用问题	106
4.5.1	A/D 转换器的选择原则	106
4.5.2	模拟信号的极性转换问题	106
4.5.3	A/D 转换器的接口方法概述	109
4.5.4	A/D 转换器接口设计实例	109

第五章 数据记录系统	119
5.1 磁表面存储器记录原理	119
5.1.1 磁记录的基本原理	120
5.1.2 评价记录品质的主要指标	121
5.1.3 数字磁记录编码方式	122
5.2 磁盘存储器	126
5.2.1 硬盘工作原理	126
5.2.2 硬盘驱动器的基本结构	127
5.3 数字磁带机	134
5.3.1 概述	134
5.3.2 基本组成	135
5.3.3 基本工作原理	137
5.3.4 磁带记录格式	138
5.3.5 磁带机的主要技术指标	140
5.4 光盘存储器	140
5.4.1 光盘存储器的种类及特点	141
5.4.2 光盘存储原理	142
5.4.3 光盘驱动器组成简介	147
第六章 数据采集系统设计	148
6.1 系统设计的基本原则	148
6.2 数据采集系统的基本结构	149
6.3 误差的合成与分配	152
6.3.1 误差的基本概念	152
6.3.2 测量误差的合成	152
6.3.3 测量误差的分配	154
6.3.4 最佳测量方案的选择	155
6.4 总线	155
6.4.1 总线的基本概念	155
6.4.2 总线的分类	157
6.4.3 总线控制与总线接口	157
6.4.4 VXI 总线	160
6.4.5 PXI 总线简介	163
6.5 虚拟仪器	164
6.5.1 基本概念	164
6.5.2 虚拟仪器的组成	165
6.5.3 仪器驱动器基本概念	166
6.6 数据采集系统的标定	169

6.6.1	标定的基本概念	169
6.6.2	静态标定	170
6.6.3	传感器的动态标定	171
第七章	数据采集系统软件设计方法	173
7.1	应用程序的设计步骤	173
7.1.1	设计任务书的编写	173
7.1.2	硬件电路设计	173
7.1.3	软件任务分析	174
7.1.4	数据类型和数据结构规划	175
7.1.5	资源分配	175
7.1.6	编程及调试	176
7.2	程序流程图与子程序设计	176
7.2.1	程序流程图	176
7.2.2	子程序设计	180
7.3	系统监控程序设计	186
7.3.1	监控程序的任务	186
7.3.2	监控程序的结构	187
7.3.3	监控程序的设计方法	190
7.4	抗干扰设计	199
7.4.1	干扰的作用机制及后果	199
7.4.2	数字信号输入输出中的软件抗干扰措施	200
7.4.3	数字滤波	203
第八章	数据采集系统实例	210
8.1	地震勘探对数据采集系统的要求	210
8.1.1	地震勘探的基本概念	210
8.1.2	地震波的基本特性	210
8.1.3	地震勘探对数据采集系统的要求	211
8.2	I/O SYSTEM 2000 地震数据采集系统	213
8.2.1	硬件组成	214
8.2.2	软件操作	215
8.3	数据采集站典型部件原理分析	215
8.3.1	Σ - Δ 型 24 位 A/D 转换器的应用	216
8.3.2	频谱整形滤波器	221
8.3.3	可预测陷波器	224
8.4	地震数据处理简介	227
8.4.1	地震资料处理的基本流程	227
8.4.2	预处理	229
8.4.3	基本处理	232

第一章 绪 论

1.1 数据采集的基本概念

随着计算机技术的发展与普及,数字设备正越来越多地取代模拟设备,在生产过程控制和科学研究等广泛的领域中,计算机测控技术正发挥着越来越重要的作用。然而,外部世界的大部分信息是以连续变化的物理量形式出现的,例如温度、压力、位移、速度等。要将这些信息送入计算机进行处理,就必须先将这些连续的物理量离散化,并进行量化编码,从而变成数字量,这个过程就是数据采集。它是计算机在监测、管理和控制一个系统的过程中,取得原始数据的主要手段。

数据采集就是将被测对象(外部世界、现场)的各种参量(可以是物理量,也可以是化学量、生物量等)通过各种传感元件做适当转换后,再经信号调理、采样、量化、编码、传输等步骤,最后送到控制器进行数据处理或存储记录的过程。控制器一般均由计算机承担,所以说计算机是数据采集系统的核心,它对整个系统进行控制,并对采集的数据进行加工处理。用于数据采集的成套设备称为数据采集系统(Data Acquisition System,DAS)。

数据采集系统是计算机与外部世界联系的桥梁,是获取信息的重要途径。数据采集技术是信息科学的重要组成部分,已广泛应用于国民经济和国防建设的各个领域,并且随着科学技术的发展,尤其是计算机技术的发展与普及,数据采集技术将有广阔的发展前景。

数据采集系统追求的最主要目标有两个:一是精度,二是速度。对任何量的测试都要有一定的精确度要求,否则将失去测试的意义;提高数据采集的速度不仅仅是提高了工作效率,更主要的是扩大数据采集系统的适用范围,便于实现动态测试。

现代数据采集系统具有如下主要特点:

(1) 现代数据采集系统一般都由计算机控制,使得数据采集的质量和效率等大为提高,也节省了硬件投资。

(2) 软件在数据采集系统的作用越来越大,这增加了系统设计的灵活性。

(3) 数据采集与数据处理相互结合得日益紧密,形成数据采集与处理系统,可实现从数据采集、处理到控制的全部工作。

(4) 数据采集过程一般都具有“实时”特性,实时的标准是能满足实际需要;对于通用数据采集系统一般希望有尽可能高的速度,以满足更多的应用环境。

(5) 随着微电子技术的发展,电路集成度的提高,数据采集系统的体积越来越小,可靠性越来越高,甚至出现了单片数据采集系统。

(6) 总线在数据采集系统中有着广泛的应用,总线技术它对数据采集系统结构的发展起着重要作用。

1.2 数据采集系统基本组成

数据采集系统包括硬件和软件两大部分,硬件部分又可分为模拟部分和数字部分。图1-1是硬件基本组成示意图。下面简单介绍一下数据采集系统的各个组成部分。

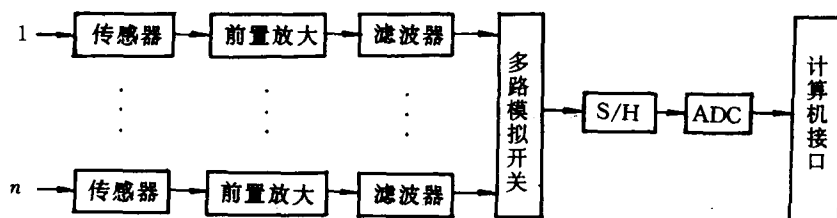


图 1-1 数据采集系统硬件基本组成

传感器的作用是把非电的物理量转变成模拟电量(如电压、电流或频率),例如使用热电偶、热电阻可以获得随温度变化的电压;转速传感器常把转速转换为电脉冲等。通常把传感器输出到 A/D 转换器输出的这一段信号通道称为模拟通道。

放大器用来放大和缓冲输入信号。由于传感器输出的信号较小,例如常用的热电偶输出变化,往往在几毫伏到几十毫伏之间;电阻应变片输出电压变化只有几个毫伏;人体生物电信号仅是微伏量级。因此,需要加以放大,以满足大多数 A/D 转换器的满量程输入 5~10 V 的要求。此外,某些传感器内阻比较大,输出功率较小,这样放大器还起到了阻抗变换器的作用来缓冲输入信号。由于各类传感器输出信号的情况各不相同,所以放大器的种类也很繁杂。例如为了减少输入信号的共模分量,就产生了各种差分放大器、仪用放大器和隔离放大器;为了使不同数量级的输入电压都具有最佳变换,就有量程可以变换的程控放大器;为了减少放大器输出的漂移,则有斩波稳零和激光修正的精密放大器。

传感器和电路中的器件常会产生噪声,人为的发射源也可以通过各种耦合渠道使信号通道感染上噪声,例如工频信号可以成为一种人为的干扰源。这种噪声可以用滤波器来衰减,以提高模拟输入信号的信噪比。

在数据采集系统中,往往要对多个物理量进行采集,即所谓多路巡回检测,这可通过多路模拟开关来实现。多路模拟开关可以分时选通来自多个输入通道的某一路信号,因此,在多路开关后的单元电路,如采样/保持电路、A/D 及处理器电路等,只需一套即可,这样节省成本和体积。但这仅仅在物理量变化比较缓慢、变化周期在数十至数百毫秒之间的情况下较合适。因为这时可以使用普通的数十微秒 A/D 转换器从容地分时处理这些信号。但当分时通道较多时,必须注意泄漏及逻辑安排等问题;当信号频率较高时,使用多路分路开关后,对 A/D 的转换速率要求也随之上升。在数据通过率超过 40~50 kHz 时,一般不再使用分时的多路开关技术。模拟多路开关有时也可以安排在放大器之前,但当输入的信号电平较低时,需注意选择多路模拟开关的类型。若选用集成电路的模拟多路开关,由于它比干簧或继电器组成的多路开关导通电阻大,泄漏电流大,因而有较大的误差产生。所以要根据具体情况来选择多路模拟开关

的类型。

模拟开关之后是模拟通道的转换部分,它包括采样/保持和 A/D 转换电路。采样/保持电路的作用是快速拾取模拟多路开关输出的子样脉冲,并保持幅值恒定,以提高 A/D 转换器的转换精度,如果把采样/保持电路放在模拟多路开关之前(每道一个),还可实现对瞬时信号进行同时采样。

采样/保持器输出的信号送至模数转换器,模数转换器是模拟输入通道的关键电路。由于输入信号变化速度不同,系统对分辨力、精度、转换速率及成本的要求也不同,所以 A/D 转换器的种类也较多。早期的采样/保持器和模数转换器需要数据采集系统设计人员自行设计,目前普遍采用单片集成电路,有的单片 A/D 转换器内部还包含有采样/保持电路、基准电源和接口电路,这为系统设计提供了较大方便。

A/D 转换的结果要送给计算机。有的则采用并行码输出,有的则采用串行码输出。使用串行输出结果的方式对长距离传输和需要光电隔离的场合较为有利。

1.3 数据采集系统的主要性能指标

对数据采集系统的性能要求和具体应用目的与应用环境有密切关系,对应不同的情况往往有不同的要求。以下给出的是比较主要和常用的几个指标的含义。

(1) 系统分辨率:是指数据采集系统可以分辨的输入信号最小变化量。通常用最低有效位值(LSB)占系统满度信号的百分比表示,或用系统可分辨的实际电压数值来表示,有时也用满度信号可以分的级数来表示。表 1-1 示出了满度值为 10 V 时数据采集系统的分辨率。

表 1-1 系统的分辨率(满度值为 10 V)

位 数	级 数	1 LSB (满度值的百分数)	1 LSB (10 V 满度)
8	256	0.391%	39.1 mV
12	4 096	0.024 4%	2.44 mV
16	65 536	0.001 5%	0.15 mV
20	1 048 576	0.000 095%	9.53 μ V
24	16 777 216	0.000 006 0%	0.60 μ V

(2) 系统精度:是指当系统工作在额定采集速率下,每个离散子样的转换精度。模数转换器的精度是系统精度的极限值。实际的情况是,系统精度往往达不到模数转换器的精度,这是因为系统精度取决于系统的各个环节(部件)的精度,如前置放大器、滤波器、模拟多路开关等等,只有这些部件的精度都明显优于 A/D 转换器精度时,系统精度才能达到 A/D 的精度。这里还应注意系统精度与系统分辨率的区别。系统精度是系统的实际输出值与理论输出值之差,它是系统各种误差的总和。通常表示为满度值的百分数。

(3) 采集速率:又称为系统通过速率、吞吐率等,是指在满足系统精度指标的前提下,系统对输入模拟信号在单位时间内所完成的采集次数,或者说系统是系统每个通道、每秒钟可采集的子

样数目。这里所说的“采集”包括对被测物理量进行采样、量化、编码、传输、存储等的全部过程。在时间域上,与采集速率对应的指标是采样周期,它是采样速率的倒数,它表征了系统每采集一个有效数据所需的时间。

(4) 动态范围:是指某个物理量的变化范围。信号的动态范围是指信号的最大幅值和最小幅值之比的分贝数。数据采集系统的动态范围通常定义为所允许输入的最大幅值 V_{imax} 与最小幅值 V_{imin} 之比的分贝数,即

$$I_i = 20 \lg \frac{V_{imax}}{V_{imin}}$$

式中最大允许输入幅值 V_{imax} 是指使数据采集系统的放大器发生饱和或者是使模数转换器发生溢出的最小输入幅值。最小允许输入幅值 V_{imin} 一般用等效输入噪声电平 V_{IN} 来代替。

对大动态范围信号的高精度采集时,还要用到“瞬时动态范围”这样一个概念。所谓瞬时动态范围是指某一时刻系统所能采集到的信号的不同频率分量幅值之比的最大值,即幅度最大频率分量的幅值 A_{fmax} 与幅度最小频率分量的幅值 A_{fmin} 之比的分贝数。若用 I 表示瞬时动态范围,则有

$$I = 20 \lg \frac{A_{fmax}}{A_{fmin}}$$

(5) 非线性失真:也称谐波失真。当给系统输入一个频率为 f 的正弦波时,其输出中出现很多频率为 kf (k 为正整数) 的新的频率分量的现象,称为非线性失真。谐波失真系数用来衡量系统产生非线性失真的程度,它通常用下式表示

$$H = \frac{\sqrt{A_2^2 + A_3^2 + \dots}}{\sqrt{A_1^2 + A_2^2 + A_3^2 + \dots}} \times 100\%$$

式中 A_1 —— 基波振幅;

A_k —— 第 k 次谐波(频率为 kf) 的振幅。

第二章 数据采集的理论基础

数据采集属于信息科学的一个重要分支,它是以传感器技术、信号检测与处理、电子学、计算机技术等高技术为基础而形成的一门综合应用技术学科,它所涉及的理论基础比较多。本章介绍几个主要方面,首先简要介绍信息论基础知识和微弱信号检测的基本原理,重点介绍采样理论与技术。

2.1 信息论基础知识

2.1.1 有关信息的基本概念

1. 信息的概念

人类从诞生开始就离不开信息活动。但是信息真正成为科学的研究对象并发展成为一门科学,却是始自 20 世纪 40 年代中后期。1948 年 N. 维纳(Norbert Wiener)在《控制论——动物与机器中的通信与控制问题》中指出:“信息既不是物质,也不是能量,信息就是信息”,即提出了“信息”是存在于客观世界的第三要素的著名论断。另一位美国学者 C. E. 仙依(Claude Elwood Shannon)第一次系统地给出了信息的定量描述,成功地用数学公式把信息传递过程中的物质、能量和信息之间的相互作用和依存关系统一起来。

如果我们用空间表示现实世界的话,那么可以说,维纳和仙依在人们已经熟识的物理性(物质和能量)空间之外,又把信息空间的存在揭示出来,这是人类认识史上的一次重大飞跃。

关于信息(Information)的定义至今众说不一,正式成文发表的就有 100 多种,例如:信息就是信息,既不是物质也不是能量;信息是事物之间的差异;信息是集合的变异度;信息是用以消除不定性的东西;信息是负熵;信息是消息,是情报,是知识;等等。

由于信息概念的复杂性,在定义信息的时候必须十分注意定义的条件。从广义上讲,信息是事物运动的状态和(状态改变的)方式。这里所说的“事物”,是广义的事物,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,包括物理的、化学的、生物的、思维的和社会的运动;“运动状态”是指事物运动在空间上所展示的性状和态势;“运动方式”则是指事物运动在时间上所呈现的过程和规律。从这个广义的信息概念出发,引入不同的约束条件,就可以得到不同的定义。例如:从受信者的角度出发,信息就是对受信者有用的知识。

尽管人类自诞生以来就开始利用信息,现代人们更加认识到信息对于人类生活与物质和能量同样重要。但是对它的认识,特别是作为重要信源的自然界如何产生信息等问题的认识还

很不够,还有待进一步研究。正如仙依自己所强调的那样,仙依信息论是“针对某些非常特殊的问题的(即通信问题),未必适合心理学、经济学和其它一些社会科学领域”。换言之,仙依信息论是建立在统计数学基础上用来解决统计通信问题的理论。按照通信工程的要求,这种理论只需要关心信号“波形的复制”(即只关心形式),不必关心它的内容和价值。要全面描述信息概念,还有待人们不断的努力。

2. 信息的基本特征

信息来源于物质,又不是物质本身;信息与能量息息相关,但又与能量有本质的区别。信息是具体的,并且可以被(动物、机器等)所感知、提取、识别,可以被传递、存储、变换、处理、显示、检索和利用。信息的传递要伴随能量的传递,信息在传递过程中的物理表现形式就是信号(Signal)。

从广义的信息概念来讲,信息传递过程(即广义的通信系统)可概括为图 2-1 所示的模型。

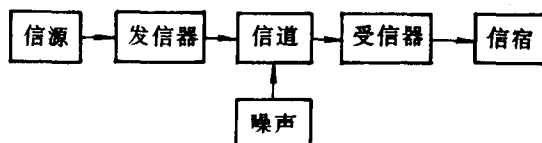


图 2-1 广义通信系统模型

这是一个单向通信系统,信息从信源(Information Source)传送到信宿(Destination)。发信器(Transmitter)又称编码器,是把信息变换成物理信号的装置;受信器(Receiver)又称译码器,是发信器的逆变器,是把物理信号转换为信宿能够感知的信息的装置;信道(Channel)是传送物理信号的媒介;信宿是信息传送的对象,即接收信息的人或机器。这是一个概括性很强的通信系统,它表达了信息传输过程的共同规律。

信息是普遍存在的,在整个宇宙时空中,信息是无限的,即使在有限的空间(时间有限或无限)中,信息也是无限的。由于信息可以脱离事物本身相对独立地存在,因此可以被无限制地进行复制、传播,即信息可以被众多用户所共享。

3. 信息技术

信息技术是指获取、传递、处理和利用信息的技术,是扩展人的信息器官(感觉器官传导神经网络、思维器官以及效应器官或执行器官)功能的一类技术。信息技术作为一个综合性的科学技术体系,它涉及下列许多学科与技术分支。

(1) 信息获取技术:这是感觉器官功能的延长,包括传感技术、测量技术和存储技术等。数据采集技术就是信息获取技术,它涉及自然信息、机器信息和社会信息的采集,特别是直接获取自然信息的技术。

(2) 信息传递技术:它是传导神经功能的延长,包括各种信息的发送、传输、接受、显示、记录技术,特别是“人—机”信息交换技术。这门技术的主体是通信技术。

(3) 信息处理技术:这是思维器官功能的延长,包括各种信息的变换、加工、放大、滤波、提取、压缩等技术,特别是数值信息处理与人工智能技术。这门技术的主体是计算机技术。

(4) 信息利用技术:这是效应器官功能的延长,包括各种利用信息进行控制、操纵、指挥、管理决策的技术,特别是“人—机”协调的智能控制与智能管理技术。这门技术特别广泛,涉及计算机技术与各种专业、学科、技术的多种多样的结合。例如,计算机技术与管理科学和有关专业相结合而形成了管理信息系统;办公自动化技术则是计算机技术、通信技术、管理科学和行为科学等结合而形成的一种信息利用技术;等等。

(5) 信息技术的支撑技术:是指信息技术的实现手段所涉及的技术。当前信息技术的支撑技术主要是电子技术,特别是微电子技术,还有激光技术、生物技术等等。信息技术的支撑技术还包括机械,特别是精密机械等工程技术。

2.1.2 信息的描述

1. 信息的数学模型

信源是信息的来源,但信息是较抽象的东西,所以要通过它的外在表现形式——消息(Message)来表达。在通讯系统中,受信者在未收到消息以前,对信源发出什么消息是不确定的,所以可用随机变量或随机矢量来描述信源输出的消息,或者说用概率空间来描述信源。

信源常常以一个个符号的形式发出消息,例如文字、字母等,而且这些符号的取值是有限的或可数的,所以,可用离散型随机变量来描述这些消息,这样的信源称为离散信源。它的数学模型就是离散型的概率空间

$$[X, P] = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ P(x_1) & P(x_2) & \cdots & P(x_N) \end{bmatrix}$$

概率空间 $[X, P]$ 在信息论中也常称为信源空间。

有些信源输出的消息的取值是连续的,即可能出现的消息数是不可数的无限值,这时可用连续型随机变量来描述这些消息。那么这些信源称为连续信源,其数学模型为连续型的概率空间

$$[X, p(x)] = \begin{bmatrix} (a, b) \\ p(x) \end{bmatrix} \quad \text{或} \quad \begin{bmatrix} R \\ p(x) \end{bmatrix}$$

并且满足

$$\int_a^b p(x) dx = 1 \quad \text{或} \quad \int_R p(x) dx = 1$$

其中, R 或 (a, b) 表示数集区间,而 $p(x)$ 是随机变量 x 的概率密度函数。

2. 自信息(Self Information)

前面已经叙述,信源发出的消息常常是随机的,所以在没有收到消息之前,受信者对信源发出什么消息是不确定的,这种不确定性是客观存在的。只有当信源发出的消息通过信道传输给受信者后,才能消除这种不确定性,也就是受信者获得了信息。因此,我们可以直观地把信息量定义为

收到某消息所获得的信息量 = 不确定性减少的量

事件发生的不确定性与事件发生的概率有关。事件发生的概率越小,不确定性就越大。这样的事件一旦发生,所能获得的信息量就越大。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函数

$$I(x_i) = f[P(x_i)]$$

式中 $P(x_i)$ —— 事件 x_i 发生的先验概率;

$I(x_i)$ —— 事件 x_i 发生所含有的信息量。

根据客观事实和人们的习惯概念,函数 $f[P(x_i)]$ 应满足以下条件:

(1) $f_i(P_i)$ 应是先验概率 $P(x_i)$ 的单调递减函数,即当 $P_1(x_1) > P_2(x_2)$ 时, $f_1(P_1) < f_2(P_2)$

(2) 当 $P(x_i) = 1$ 时, $f(P_i) = 0$

(3) 当 $P(x_i) = 0$ 时, $f(P_i) = \infty$

(4) 两个独立事件的联合信息量应等于它们各自信息量之和, 即统计独立信源的信息量等于它们各自的信息量之和。

根据上述条件, 可以从数学上证明这种函数形式是对数形式, 即

$$I(x_i) = \lg \frac{1}{P(x_i)} = -\lg P(x_i)$$

这就是事件 x_i 发生时, 该事件所含有的信息量。因为它描述的是事件 x_i 发生时的信息量, 故又称为自信息。 $I(x_i)$ 代表两种含义: 当事件 x_i 发生以前, 表示事件 x_i 发生的不确定性; 当事件 x_i 发生后, 表示事件 x_i 所含有(或所提供)的信息量。

自信息采用的测度单位取决于所取对数之底。如以 2 为底, 则所得信息量单位为比特(bit, binary unit 的缩写); 以 e 为底, 则称为奈特(nat, nature unit 的缩写); 以 10 为底, 则为哈特(Hart, hartley 的缩写)。一般都采用以 2 为底的对数, 因为当 $P(x_i) = 1/2$ 时, $I(x_i) = 1$ bit, 所以, 1 bit 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。

3. 信息熵(Information Entropy)

自信息 $I(x_i)$ 是指信源(物理系统)某一事件 x_i 发生时所包含的信息量, 物理系统内不同事件发生时, 其信息量不同, 所以自信息 $I(x_i)$ 是一个随机变量, 它不能用来作为整个系统的信息的度量。

仙依定义自信息的数学期望为信息熵, 即信源的平均信息量

$$H(x) = E[-\lg P(x_i)] = -\sum_{i=1}^N P(x_i) \lg P(x_i)$$

熵的单位是 bit/ 事件或 bit/ 符号。

信息熵表征了信源整体的统计特性, 是总体的平均不确定度的量度。对某一特定的信源, 其信息熵只有一个。信息熵具有以下基本性质。

(1) 对称性。当概率空间中 $P(x_1), P(x_2), \dots$, 顺序任意互换时, 熵函数的值不变, 例如下面两个信源空间

$$\begin{aligned} [X, P(x)] &= \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix} \\ [Y, P(y)] &= \begin{bmatrix} y_1 & y_2 & y_3 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{bmatrix} \end{aligned}$$

其信息熵 $H(X) = H(Y)$ 。该性质说明, 熵只与随机变量的总体结构有关, 与信源总体的统计特性有关。此点也说明了所定义的熵有其局限性, 它不能描述事件本身的主观意义。

(2) 确定性。如果信源输出只有一个状态是必然的, 即 $P(x_1) = 1, P(x_2) = P(x_3) = \dots = 0$, 则信源的熵

$$H(X) = -[1 \times \lg 1 + \sum_{i=2}^N 0 \times \lg 0] = 0$$

这个性质表明, 信源的输出虽有多种不同形态, 但其中一种是必然的, 这意味着其它状态不可能出现。那么, 这个信源是一个确知信源, 其熵为零。