电信技术普及丛书

信 息 论 入 门

梁传甲 编著

人民邮电出版社

内 容 提 要

本书对通信的最基本问题：提高通信有效度和提高通信可靠性的问题作了较有系统的叙述。对于各种重要的编码也作了介绍。全书着重于物理概念，深入浅出，是一本关于信息论和编码的入门书。

本书可供在电信方面，以及在遥测、遥控方面工作的人员作参考之用。

电信技术普及丛书

信 息 论 入 门

梁传甲 编著

人民邮电出版社出版

北京东长安街27号

河北省邮电印刷厂印刷

新华书店北京发行所发行

各地新华书店经售

开本：787×1092 1/32　1988年2月第一版

印张：3 16/32　页数：72　1988年2月河北第1次印刷

字数：101千字　印数：1—4 100册

ISBN7115-03522-9/TN

定价：1.00元

出版者的话

为了普及电信技术知识，特别是电信新技术知识，为我国的通信现代化服务，我们组织编写了一套“电信技术普及丛书”，陆续出版。这套丛书的主要读者对象是具有中学文化水平、有一些电信基本知识的工人、管理干部和初级技术人员。在编写

概括的了解，作为进一步学习的入门向导。我们殷切希望广大读者对这套丛书提出意见和建议，帮助我们做好这一工作。

目　录

概　　说

信息、通信与社会

花朵开放时的色彩是一种信息，它可以引来昆虫为其授粉。许多成熟的水果会产生香味，诱来动物，动物食后为其传播种子。有的杨树在遭到病虫害侵袭时会产生一种化学物质来抵御它们所受到的危害，同时“通知”未受这种病虫害侵袭的杨树产生该物质以作好预防。

人类的信息交换更为普遍。群落愈是有组织，其成员间的信息交换就愈是频繁和迫切：集体捕猎时进行的联络和协调、警戒时发布的各种情况、分配食物时所作的安排等，都需要进行信息交换。在这种交换中，语言是最主要的信息载体，它们可用语言在有限的范围内进行信息交流。

人类从蒙昧时代分散的部落，逐步统一和发展成为中央集权统治的国家，从茹毛饮血的原始社会发展到今日的工业化社会，通信联络是行使政令、组织生产和实施军事指挥的重要保障。这种联络，开始时主要藉助于语言，后来也用书信，战斗时还用鼓角、旌旗、烟火等各种声、光信号进行。十六至十九世纪对电磁进行了研究，在此基础上发明的电报、电话标志着通信方式的一场革命，并已成为当今发达国家的一种主要通信方式。

为了使信息交换不但可以超越空间进行，而且能在时间上进行“延伸”．人类从最早利用结绳记事，后来逐渐发展，创

造了文字。文字可以认为是信息的第二个载体。来往的书信、历史事件的记载、小说和戏剧的流传等都主要藉助于文字才得以实现。文字或者刻在“龙骨”、竹简、木牍、石碑上，或者写在帛和纸张上。现在，由于计算技术的发展，信息还可以用符号表示，存储在磁带、磁盘和光盘等介质上。令人惊叹的是，一张如唱片大小的光盘可以存储十年人民日报的内容。它比竹简、纸张的存储密度有了数量级的增加，是继发明纸张后的又一重大进展。

随着生产力的发展，人类互相联系变得愈来愈频繁。如果以前的工业革命主要发生在机械制造、化学工业等基础工业中，那么现在我们面临的则主要是一场“信息革命”。世界正在向信息化社会前进。实现信息化社会就是指信息能及时、准确和可靠地处理、传输和存储。这里技术的关键在于通信和计算机的发展。卫星和光纤通信的发展使得我们可以在全球范围内进行通信。把各种终端接入计算机网后，许多工作就可以在家进行：工程设计、计算机模拟、数值计算、情报检索、报表统计、资料分析、指挥生产、召开电视电话会议、进行学术讨论和了解世界最新消息等。这一切发展使得地球上任意两地间的联系如同在同一个“村落”中那样方便和紧密。信息化的实现将不断地影响和改变我们的生活。

通信与计算机都牵涉到信息的存储、处理、传输与接收。我们知道，许多信息，如图象、语言、文章等中都有许多多余的成分。存储时希望尽可能地去掉这些多余成分以节省容量，这就是提高有效性的问题。信息传输时会遭受各种干扰，如何消除这些干扰的影响，尽可能恢复原来的信息则是抗干扰或提高可靠性的问题。一般情况下，提高可靠性就要增加多余度，提高有效性将使抗干扰的能力下降。我们要研究的问题正是

如何统一解决这一对矛盾，使之既有效又能抗干扰。

我们前面一再提到了“信息”，那么“信息”到底指什么呢？就字意来说，信息与消息、情报、知识等没有大的区别。确切些说，信息是事物间互相交换的所处状态的内容，即指的是事物运动的状态、规律。它们可以用文字、图形、图象、声音等形式表示。然而，作为一门与技术密切有关的科学，我们不满足于这种过分一般的含义。这种情况正如同我们不满足“力”的一般含义一样，因为力可以指权力、力量、能力、体力、智力等等。但物理学中只是将改变物体运动状态的作用称为力，它有大小、方向和作用点三要素。为了给出信息的恰当含义，建立起有效和可靠等概念的确切数量关系，我们需要能对信息定量地加以度量，只有这样，信息论才能作为一门技术科学加以发展。

科学家仙农在贝尔系统技术杂志上发表的“通信的数学理论”论文中将信息定义为解除的不确定性的多少，从而明确了“有效”及“可靠”能够达到的目标及实现该目标的原则途径。这篇文章以其思想的深刻、方法的新颖，引起了广泛的注意，被公认为标志着信息论作为一门学科的诞生。仙农强调信息论是通信的数学理论。有人称之为仙农信息论、古典信息论、狭义信息论，甚至就称为通信理论，以区别于更一般的含义。本书介绍的信息论就是这一意义下的内容，不作一般讨论。

通信系统的构成

通信系统包含发、收两端，它的一般构成如图一所示。信息由信源发出，它经发送设备送往信道，信号经过信道到达接

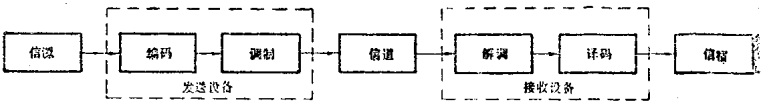

图 一　通信系统的构成

收端，在接收端由接收设备还原成发送的信息送给信宿。

信源可以是文字、数据、话音或图象。根据包含的符号数为有限或无穷（不可数），可分为离散或连续信源。载荷信息的载体称为信号，它们可以是声波、光波等。为了将这些信号送往对方，我们常常将它们变成电磁信号。因为电磁信号的传输速度快，消耗的能量很少，目前主要用它作用通信中传送信息的载体。

为了将信源送出的信息变换成电磁信号，就需要有能起这种变换作用的发送设备。发送设备常常包含编码及调制两个部分。前者将信息变为易于传输的形式，有时还使之有一定的抗干扰能力。如果信息是文字，我们难以直接用电磁信号传输它，但我们可以用四个阿拉伯数字来表示一个汉字，每一数字又用四个二进制数字表示，如3用0011，9用1001表示等。我们将这种用离散信号表示文字或信息的过程叫做编码。

为了将编码符号变换成电信号送往信道，常常还需要进行“调制”。或者说，调制是用信源信息或编码符号改变电磁信号，使它能准确反映信息的过程。例如，我们可用电信号幅度的有无或正负反映0和1，也可用电信号的十种幅度代表十个阿拉伯数字等。选择调制方式的主要着眼点是使信号便于在信道中传输，即尽量使信号与信道“匹配”。好的调制制式占有较小的频带，也易于消除传输中受到的干扰。

发送设备有时可以没有编码，有时可以不用调制。例如传

输语言等模拟信号时，可以不必编码，这时调制设备将声波变成易于传输的电信号，且电信号的变化规律和声波的变化规律“一致”，以便在接收端还原成原来的信号。若我们传输的是文字等离散电信号，在最简单的情况下，不必再作调制。图一画的是通信系统最一般的情况。

信息经发送设备变换成信号后送往信道。信道是传输信号的媒质或通道，它常常会使传输的信号失真且存在各种加性干扰。干扰有时又叫做杂波，它是我们不需要的成分，常常与信号无关且叠加在信号上。信道可以是有线信道也可以是无线信道。前者如明线、电缆、光缆等各种传输媒质；后者如中波、短波、微波等信道。

信号经过信道到达接收端后，接收设备或者将它们还原成原来的信号送给接收者，或者还原成接收者所需要的形式。如果信源是语言，受信者是人，我们就希望接收设备能将信号尽可能不失真地还原成讲话者的声音。若受信者是计算机，则语言常是一些指令、数据、程序等，接收设备应将该语言准确地变换成机器的符号。由于接收者可以是人，也可以是机器，因此今后就称之为信息的归宿，简称为信宿。

参看图二。发送设备与接收设备实际上是一变换设备，它将输入函数（信号）变换成另一函数（信号）。我们以$f\langle\cdot\rangle$表示发送端编码与调制的作用，即：

$$s_i(t)=f\langle c_i(t)\rangle$$

以$g\langle\cdot\rangle$表示接收端解调与译码的作用：

$$c_0(t)=g\langle s_0(t)\rangle$$

若干扰具有相加性，即$s_0(t)=s_i(t)+n(t)$

则 $$c_0(t)=g\langle s_i(t)+n(t)\rangle$$

当信道中没有干扰，即$n(t)=0$时，我们只要选择$g\langle\cdot\rangle=$

$f^{-1}(\cdot)$，即可不失真地恢复原发送信号；否则选择$g(\cdot)=f^{-1}(\cdot)$不一定能使干扰的影响为最小。因此，设计通信系统时应当针对干扰的性质来选择$f(\cdot)$及$g(\cdot)$，使信号能较好地在信道上传输，实现起来比较容易，且在有高效率的前提下经过解调与译码，使干扰的影响减至最小。

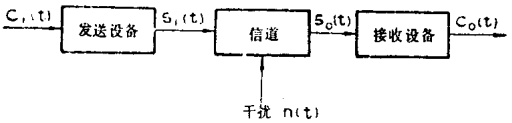

图 二　干扰的影响

由于信源符号种类只有有限多种时最为简单，因此我们主要考虑这种情况时的离散信源。对于有无限多种可能的连续信源，我们只作简单的讨论，因为它们的分析要牵涉到积分，且理论上远不如离散的成熟。离散信源的传输只要采用编码就可使性能满足要求，因此关于调制问题本书将不再涉及。

通信系统的上述模型并不局限用于电信传输这一狭窄的范围，实际上情报检索、事物观察、感觉传递等领域均与通信系统中传递信息的情况类似。例如，情报检索相当于从极为广泛的议题中寻找我们所要的情报，一切无关的题目均是易于消除的“干扰”。名称接近但实际内容并非我们所需的题目排除起来比较困难。情报检索中也有编码（编目）、存储、处理等问题，需要将情报、数据作适当的压缩，增加存储的内容。同时，存储的数据也应当准确、可靠、不易篡改与丢失等。

仙农信息论强调的是通信，由于人们愈来愈认识到物质、信息与能量是客观世界的三项要素，因而受到广泛的重视。信息论发展至今已有三十余年，作为一门学科有其成熟的一面；

它研究的对立双方比较明确，有较为统一的处理方法，获得一些较基础的结果。另一方面不能不说信息论还是一门发展着的学科，许多问题还没有完全解决，甚至还有一些争论。

第一章　信息如何表示——信息的编码

§1.1　怎样表示文字才便于传输
——离散信源的等长编码

十九世纪英国科学家法拉第发现电磁感应现象后，美国的莫尔斯、英国的惠斯顿等相继发明了电报。电报的出现标志着人们主要依赖“信使”传递信息的时代即将结束。新一代的通信方式一电信通信开始呈现在人们的面前。

电报通信不能像书信那样将信息整体送出，而是将文字用符号表示后顺序传输。一般情况下，电码常由电的有无构成，我们以0表示无电，1表示有电。0和1既可以代表电的有无，也可以表示开关的断和通、晶体管的截止和导通、磁性材料的磁化方向等。一般地说，它们可以看作是某一器件的两种状态，计算时把它们看作是二进制数的两个基本符号。

二进制数的表示与运算和十进制类似。十进制中数字自右至左代表个位、十位、百位、千位等等。二进制数自右至左代表个位、2位、$2^2=4$位、$2^3=8$位、…、等。例如

$$10110$$

相当于十进制的

$$1\times2^4+0\times2^3+1\times2^2+1\times2^1+0\times2^0=22$$

反之，十进制数也可化成二进制数，它们之间有一一对应关系。

十进制的加法规则是“逢十进一”，二进制则为“逢二进一”（表1.1.1(a)）。如

$$\begin{array}{r} 11 \\ +)\ 1 \\ \hline 100 \end{array}$$

个位数 1 + 1 = 2，“逢二进一”为10，进上去的 1 和被加数在二位数的 1 相加又逢二进一，最后得100，它相当于十进制的4。二进制的乘法也与十进制类似，但显得简单得多（表1.1.1(b)）。

表 1.1.1　　二进制的加法表及乘法表

+	0	1
0	0	1
1	1	0

(a)

×	0	1
0	0	0
1	0	1

(b)

一般情况下为了便于传输，我们常常用有限个符号的组合表示文字、数字、指令、…，这就是**编码**。由于二状态的器件容易实现，编码时常用二进制或二元数字表示，这样的编码叫二元编码或二进制编码。信息的这种表示叫**码字**，码字中包含的符号数叫**码长**。若码字数为M，用的是二元表示，则码长n应是满足

$$2^n \geqslant M$$

的最小整数。例如M＝32，n＝5；M＝60，n＝6等。

如果我们仅考虑26个拼音字母、十个阿拉伯数字、以及十、一、×、÷等符号，那么共有六十多个符号。打印时，若一个键分上下两部分，故共有三十几个键即可，它们的表示方式如表1.1.2所示。三十几个键只要用五位二进制数就可表示，

表 1.1.2　　五 单 位 电 码 表

八进制表示	字符	八进制表示	字符	八进制表示	字符	八进制表示	字符
		10	换行	20	3 E	30	- A
01	5 T	11	7 L	21	+ Z	31	2 W
02	回车	12	4 R	22	* D	32	@ J
03	9 0	13	> G	23	# B	33	
04	空格	14	8 I	24	' S	34	7 U
05	↑ H	15	0 P	25	6 Y	35	1 Q
06	, N	16	· C	26	< F	36	(K
07	·· M	17	= V	27	/ X	37	

这就是经常使用的五单元电码。表中的32用五单元电码表示为11010，15表示为01101等，即后一位数用三个二进制数表示，前一位数因不超过3，可用两位二进制数表示，故每一信息符号用五位二进制数表示。由于五单元电码中每一信息符号均用五位二元数字表示，即长度均为5，故是一种等长编码。等长码对区分整个电文的信息符号及设计电报机等均比较方便，是目前各种机械和电子设备中用得最多的一种。

若我们还希望区分大小写字母及增加一些控制符号，应用五单位电码就有困难。为此美国国家标准局建议使用一种电码—ASCⅡ码，它由七位二进制数表示，共可代表$2^7=128$个符号。它已被作为国际通用的一种标准电码，并已在电传电报机、计算机终端等设备上被广泛采用。ASCⅡ码的编码如表11..3所示。实际上，ASCⅡ码常由八位二进制数表示，这是因为计算机中常用8位二元符号作为一个字节。多余的一位有时选1用作同步；有时取0或1以使八位中1的总数为偶数（或奇数）；或者有时就空着不用。

表1.1.2和表1.1.3中的电码若用二元数字表示，分别需用

表 1.1.3　　ASCII　　码

八进制表示	字符	八进制表示	字符	八进制表示	字符	八进制表示	字符
000	NUL	040	SP	100	@	140	、
001	SOH	041	!	101	A	141	a
002	STX	042	"	102	B	142	b
003	ETX	043	#	103	C	143	c
004	EOT	044	$	104	D	144	d
005	ENQ	045	%	105	E	145	e
006	ACK	046	&	106	F	146	f
007	BEL	047	'	107	G	147	g
010	BS	050	(	110	H	150	h
011	HT	051	)	111	I	151	i
012	LF	052	*	112	J	152	j
103	VT	053	+	113	K	153	k
014	FF	054	,	114	L	154	l
015	CR	055	-	115	M	155	m
016	SO	056	.	116	N	156	n
017	SI	057	/	117	O	157	o
020	DLE	060	0	120	P	160	p
021	DC1	061	1	121	Q	161	q
022	DC2	062	2	122	R	162	r
023	DC3	063	3	123	S	163	s
024	DC4	064	4	124	T	164	t
025	NAK	065	5	125	U	165	u
026	SYN	066	6	126	V	166	v
027	ETB	067	7	127	W	167	w
030	CAN	070	8	130	X	170	x
031	EM	071	9	131	Y	171	y
032	SUB	072	:	132	Z	172	z
033	ESC	073	;	133	[	173	{
034	FS	074	<	134	\	174	|
035	GS	075	=	135	]	175	}
036	RS	076	>	136	^	176	~
037	US	077	?	137	_	177	DEL

五位和七位，这样书写起来较长，不太方便。现在表中这样的表示方法为八进制表示。八进制数与二进制数之间的关系如表1.1.4所示。计算机中由于每个字节通常为八位，故用十六进制表示更方便一些（如表1.1.5所示）。十六进制数与我国旧的16两一斤的进位方式相同。为了避免采用新的符号和增加改装现用设备的困难而用字母A～F表示10～15。一般情况下，可根据使用场合判断A～F的含义而不致引起混淆。

表 1.1.4

八进制数	0	1	2	3	4	5	6	7
二进制数	000	001	010	011	100	101	110	111

表 1.1.5

十六进制数	二元表示	十六进制数	二元表示	十六进制数	二元表示	十六进制数	二元表示
0	0000	4	0100	8	1000	C	1100
1	0001	5	0101	9	1001	D	1101
2	0010	6	0110	A	1010	E	1110
3	0011	7	0111	B	1011	F	1111

当用二进制数代表汉字时，由于可用的汉字共约六万个*，

* 历史上汉字字数的概况如下

汉	杨 雄	训纂编	5340字
	许 慎	说文解字	9353字
晋	吕 忱	字林	12824字
魏	张 揖	广 雅	18151字
宋	丁 度	集 韵	52525字
	司马光	类 编	31000字
明	梅膺祚	字 汇	33179字
清	张玉书	康熙字典	47035字