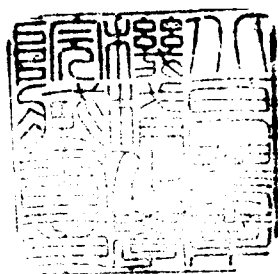


现代化信息丛书

信息论讲座

程明 朱筱 著



科学普及出版社

前 言

小小寰球，天涯咫尺。它每一处迸发的思想火花或实践着的变革行动，都在创造着全球居民们每日生活的信息需求。

现代的中国人，已经开始了“面向现代化，面向世界，面向未来”的新长征。面向现代化，需要现代化的知识；面向世界，需要了解世界潮流；面向未来，需要科学地预测未来，《现代化信息丛书》即为此目标而诞生。

《现代化信息丛书》将努力追踪世界科学技术和经济社会发展的主导趋向，及时介绍有益于社会主义建设事业的新成就、新知识和新观念，为新型干部队伍的知识化和专业化服务，并兼顾社会各界的一般需要。

这套丛书着眼于加速知识更新，在世界科技前沿和国家需要的结合点上精选最新题材。它将打破学科的局限，注重自然科学与社会科学的结合，注重科学技术发展对社会发展的影响，注重为全社会普遍关注的重大的综合性问题。

“声如千骑疾，气卷万山来”。世界新技术革命的严峻挑战和振兴中华的伟大变革，在我们面前汇成了势不可挡的时代潮流。我们将努力在这一潮流中创一家风格，新万人耳目，让丛书置于社会主义事业家和青年朋友的案头，在他们的思考、拼搏、奋进中助上一臂之力。

现代化信息丛书编委会

目 录

1. 通信与信息

- 1.1 信息技术和信息论的发展史..... (1)
- 1.2 物质、能量和信息..... (3)
- 1.3 通信系统..... (5)
- 1.4 编码和译码..... (8)
- 1.5 消息和信息的提取..... (9)

2. 信息源和信息量

- 2.1 信息与认识..... (11)
- 2.2 信息源..... (13)
- 2.3 信息量..... (15)
- 2.4 信息的组合..... (17)
- 2.5 信息源的熵和信码的平均信息量..... (19)
- 2.6 日常语言的熵和多余度..... (21)

3. 通信的可靠性

- 3.1 噪声与干扰..... (26)
- 3.2 抗噪声干扰的一些方法..... (29)
- 3.3 纠错码..... (31)

4. 信道容量和信道定理

- 4.1 传信率和码的效率..... (34)
- 4.2 信道容量..... (40)
- 4.3 无噪声信道的信道定理..... (41)
- 4.4 有噪声信道的信道容量和信道定理..... (42)

4.5 失真和保真度.....	(46)
5. 信号的分析 and 处理	
5.1 滤波.....	(49)
5.2 信号的检测.....	(50)
5.3 信号估计.....	(51)
6. 广义信息论	
6.1 广义信息.....	(53)
6.2 语义信息.....	(55)
6.3 信息加工的可逆性与不可逆性.....	(57)
6.4 信息化时代.....	(59)

1. 通信与信息

1.1 信息技术和信息论的发展史

通信的起源可以说和生物的出现同时开始。生物通信的最初形式也许是依靠嗅觉感受特殊种类化学物质的刺激。以后又发展了触觉和视觉。有趣的是,在进化的过程中,一些生物,如鱼和鳗等已经采用了原始的电通信方式。虽然许多生物都把声音作为通信的手段,但是只有人才在发音的基础上发展了语言的功能。遗憾的是我们现在很难知道早期的语言究竟是怎样的,因为有关的信息无法保存下来,供我们研究。我们只能从一些原始的民族作间接的研究。人们知道更多的是以文字为工具的通信方式。

最初人们直接使用图画作为符号,后来就演变成成为象形文字。汉字就是在象形文字基础上发展起来的。其他许多文字则走上了拼音化的道路。

文字的出现使人类成为真正意义上的整体。它不但使各地无法直接接触的人可以进行交流、通信,而且可以使不同时代的人也能“传话”。文字保留下来的大量信息是人类认识自己历史的重要手段。在这方面,中华民族尤其是值得自豪的。汉字历史文献的规模之大,历史之悠久和连续不断,其内容之可贵是举世无双的。例如我国的历史文献中就保留了关于新星爆发的记载,成为现代天文学研究的珍贵资料。

为了使书写速度与讲话速度同步,人们发明了速记方法。同时,人们也对使用密码感兴趣。伽利略、虎克等人都曾使用密码(字谜)的方式记载自己的研究成果。

人们除了用语言文字来传递信息外，还有许多其他方式。例如，每个民族都有自己独特的服饰、发型、礼节等。你很容易从某些民族女人的发型上判断她是否成年、已婚，有无孩子，甚至孩子的数目和性别。表情和手势也是人们常用的传递信息手段。值得一提的是，有许多民族发展了只采用两种不同符号的通信方式。例如，印第安人采用长短的烟信号，非洲人采用高音鼓和低音鼓进行通信。

中国人发明的造纸术和印刷术使人类的通信手段又一次发生了重大的变化。文字成为信息记录、贮存和传递的有效手段。与之相适应的是邮递系统和图书馆逐渐发达起来。

电报电话的发明是新的工业技术的产物。莫尔斯发明了由点划组成的电码系统。这种电码的有效性还在于发明者考虑到了字母出现的频率。电通信以其快速、便利、通信距离远，再次造成了通信手段的重大变革。

值得指出的是，在数学中使用符号代替文字也是一种提高通信效率的手段。于是，各种符号的形式系统发展起来了，成为一些学科的专业语言。

无线电通信技术和卫星通信技术的发展很快就把电报、电话、广播和电视的通信范围扩大到全球，把世界连成了一片。

电子计算机的迅速发展，特别是微型机的广泛应用大大改善了处理信息的能力和通信系统的管理，使人类社会逐步进入了信息化的时代。

同信息技术的古老历史相比，信息论是一门年轻的学科。关于信息论的研究工作可以说是从本世纪20年代卡森对调频的研究开始。他的研究纠正了当时一种流行的错误观

点，即认为可以通过调频将载频的频带扩大。实际上调频的优越性在于可以保持音质的优美，因而可以用作高质量的广播。

1924年奈奎斯特指出，对于确定的传信率需要最低限度的频带宽度。屈普夫米勒和哈特利也作了类似的研究。1928年哈特利指出，传递一定量的信息需要带宽与时间的乘积达到一定的值。加博尔发展了这种思想，并在哈特利的工作基础上进一步考虑了有噪声的情况。费歇尔强调了用统计方法处理信息论的有效性与充分性。

信息论作为一门严密的科学分支，主要归功于贝尔实验室的申农。他在1948年发表的论文《通信的数学理论》奠定了信息论的基础。控制论的创始人维纳也对信息论有不可忽视的贡献。申农和维纳的基本思想都是把通信作为统计过程来处理。他们采用的术语、方法也主要依靠统计理论。因此要想比较深入学习信息论的读者，如果先了解一些关于概率论和统计理论的知识，将是非常有益的。

1.2 物质、能量和信息

我们的生活一时一刻也离不开信息。清晨闹钟把我们从睡眠中唤醒，收音机播送出时事新闻。我们骑车上班要注意红绿灯信号，走进工厂机关要出示证件。这些都是在传递信息，也就是进行“通信”。如果有一位作家忽然心血来潮，想写一本科幻小说《没有信息的世界》，那么他的作品将是一无所有。因为没有信息与通信就失去了人类存在的基础，失掉了生命的基础。如果从最广泛的意义来理解信息，没有信息也就没有统一的世界。

统一的世界是由形形色色的事物构成的。这些事物之间有着密切而不可分割的联系。谁都知道，人无法脱离社会而独立存在。人类必须和动植物一起共存。生物圈与气圈、水圈、岩石圈保持着微妙的平衡。甚至在遥远星体发生的过程也对地球上的各种运动有着重要的影响。

但是，我们可以人为地把世界的某一部分与其他部分分割开来而称之为一个系统。一个系统可以是一个人、一个物体、一个社会，也可以是一组符号，如此等等。

我们可以把系统理解为由各个相互联系、相互作用的部分组成的统一体。这个统一体具有整体的功能和综合的行为。

系统的内部是协同有序、相互制约的。同时系统又作为一个整体与外部世界联系。为此，系统与外界、系统内部各部分之间就要相互发生作用。

系统之间的相互作用有三种基本方式，即物质、能量和信息。我们分别称它们为物质的传递、能量的传递和信息的传递。

物质的传递是很直观的。人员的流动、物资的输送、香味的扩散都是传递物质的过程。按照质量守恒定律，在一个系统中物质的减少必然引起其他系统中物质的增加。能量的传递也是大家所熟悉的。能量可以以机械能、热能、辐射能、化学能、生物能等形式从一个系统传递到另一个系统。按照能量守恒定律和各种形式能量相互转换的定量关系，对于能量在系统之间的传递也可以进行精确的测量和计算。

除了质量和能量的传递以外，系统之间相互作用的重要方式就是信息的传递。生物体要维持自己的生存，就要从外

界获得信息，并作出相应的反应。动物在进化的过程中发展了许多巧妙而有效的通信机制。蜜蜂会以各种花样的舞蹈互通情报，蝙蝠的超声波系统是它捕捉食物、逃避危险的主要手段。人类社会更是离不开语言、文字等各种通信手段。随着人类社会的发展，社会活动的规模越来越大，结构也越来越复杂，通信的手段也越来越先进。可以说，通信是人类社会生存和发展的基础。

应该指出的是，通信过程并不能离开物质和能量的传递过程而独立存在。我们在上海拍一份电报到旧金山，先要写下电文，然后译成电码，再由发报机通过发射装置转变成电磁波，通过卫星的中继再为旧金山的地面站所接受……。我们不想在这里列举整个过程的每个环节，只是强调每一个环节都离不开物质和能量的传递，即使“看不见、摸不着”的电磁波带着信息在空中传播也是如此。因为，电磁波是物质存在的一种特殊形式，即电磁场，而电磁场具有电磁能。

因此，系统之间的物质、能量和信息的传递往往是同时发生的，只是我们在研究通信过程时可以只考虑信息而不考虑物质和能量。

1.3 通信系统

我们可以把相互联系的系统看成一个更高一级的系统。这种系统结构层次的划分主要是根据我们分析问题的需要。当我们只考虑系统之间的信息交换时，我们就称之为信息系统，也叫通信系统。

最基本的通信系统由以下几部分组成：1. 信源；2. 发信机；3. 信道；4. 收信机；5. 收信者；6. 噪声

源。但是，一般把噪声源看成外界对通信过程的干扰。因而也可以把前 5 部分看成通信系统的基本组成部分。下面分别对这 5 部分作些说明。

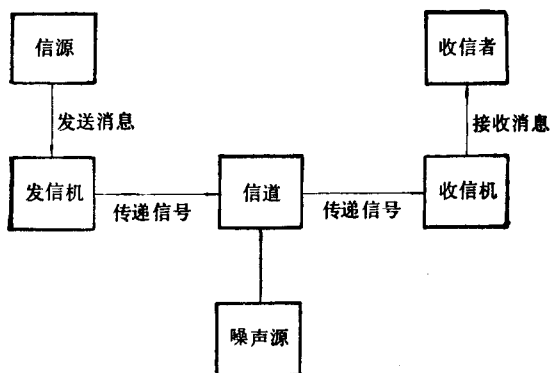


图 1.3.1

1. 信源，亦称为信息源，由此产生消息。消息有各种各样的形式：例如，由音素构成的语言，由字母构成的文本，由点、划和间隔构成的电报，由图象构成的电视节目。为了使问题简单起见，我们只考虑所谓离散型的信息源。也就是说这种信息源由不同的符号组成，并且是一个符号接着一个符号地产生消息。

2. 发信机。它把要传递的消息变换成为适合于在信道上传输的信号。例如，在电话中，话筒把声音变换成电流信号；在电报中，依靠电码系统把电文变换成点、划和间隔的序列，最后再由发报机变成电信号。

3. 信道。连接发信机与收信机的媒介。传递的信号通过这种媒介从一个系统进入另一个系统。导线、电缆、空

气、光导纤维就是一些典型的信道。

4. 收信机。执行与发信机相反的功能，把通过信道后的信号再变换成能被收信者所理解的消息。

5. 收信者。通信的目的地，接收信息的系统。通常指通过传递过来的消息获得信息的人。

电话系统是一个典型的通信系统。信源是言语的声音。话筒起发信机的作用，将声音（空气的振动）转换成电流强弱的变化。这种电流信号通过电话线传到听筒。听筒起收信机的作用，把电流信号还原成声音。这样，由声音携带的信息就从通信系统的信源传递到收信者那儿。

实际的通信系统要复杂得多。电话系统中就包括许多起中转作用的交换台和电话局。它们都可以包括在信道中。在许多现代通信系统中，需要不止一次地变换消息的形式，使用好几种不同的媒介作信道。卫星通信系统中，话筒将声音变为电流信号，在电话局载波机上它又变成高频电信号通过微波或电缆送到卫星地面站，由地面站变成极高频电磁波通过天线发送到卫星上。卫星接到信号后转发给另一个地面站，再经相反的程序变换回声音。信道包括电话线、电缆和空间。光纤通信也有类似情况。声音或图象先转换成电信号传送，然后再转换成光信号通过光纤传输。接收机将光信号转换成电信号送到用户，再复原成收信者能理解的消息。但是，不管通信系统多么复杂，我们还是可以从中区分出前述5个基本部分来。

通信系统可以根据它传递信号的性质分为三类，即离散的、连续的和混合的。电报是典型的离散式系统，而广播则是连续式系统。由于数字式离散系统的设备比较简单、通信

质量高，因而得到日益广泛的应用。为了便于读者理解，本书中只讨论离散的通信系统。它传递的信息都是离散符号的序列。但是，由此得出的结论对于其他情况也是适用的。

1.4 编码和译码

如上所述，通信系统是把信息从一个系统传递到另一个系统的传输系统，因而人们往往把通信系统称为通信手段。在人类社会发展的历史过程中，人们发展了语言、文字、电报、电话、广播、电视、传真、卫星通信等通信手段。从一定的意义上讲人类的历史也是发展和利用通信技术的历史。

我们可以把人和货物装在运输工具上，从一个地方运到另一个地方；我们也可以用输电网络把电能从一个地方输送到另一个地方。但是，怎样才能把信息“装载”在通信工具上呢？

这就要依靠编码。我们用语言、文字来表达一定的思想内容，这些都是编码。通过编码可以建立表达思想内容的语言与音素或字母之间的确定联系，即把思想内容用一段话或文字表现出来。话音可以以空气为信道来传递，而文字则可以借助报刊杂志和书籍来传递。

编码最初是指把文字变换成由点、划和间隔的空位组成的电码。例如，著名的莫尔斯编码就是其中的一种。点、划及它们之间的间隔很容易通过收发报机系统传递。以后人们把编码的概念加以推广。我们可以把信号理解为对于时间的物理变量，那么编码实际上就是把一种形式的信号变换成另一种形式的信号。电话话筒把声音转变成电信号，电视摄像机把图象转变成电信号都是进行编码。其目的是把信号变换成

为便于在信道中传递的形式。

通过与编码相反的过程可以把通过信道传来的信号变换成编码前的形式。这个过程就称为解码或者译码。

在一个通信系统中，发信机就起编码的作用，收信机则起译码的作用。两个系统必须具有共同的编码和译码机制才能进行通信。在描述间谍与反间谍活动的电影或小说中，使用密码的花样是层出不穷的。钢琴弹出的音符，身上的伤疤，一束鲜花，当然还有密码本，都可以成为编码的手段。而故事情节往往是沿着搜寻密码的线索展开的。其成功与否决定了双方较量的胜败。

1.5 消息和信息的提取

破译密码使得我们能够从一组信号中提取所需要的信息。我们把一句话、一段文字、一份电报或一幅图象称为一份消息，而把它们所要表达的内容称为信息。信息与消息就是内容与形式的关系。内容不能脱离形式而存在，但是在研究信息系统时，起决定性作用的是内容而不是形式，就是说信息是我们主要的考虑对象。

破译密码使我们能够从一份消息中提取它所包含的内容。如何做到这一点是考古学家、历史学家、人类学家、情报分析专家共同感兴趣的问题。因为在这些领域中，编码和译码的机制往往采取很隐蔽的方法，因而当人们看到一份很有价值的消息时，常常会“视而不见”，不知道有研究的必要；或者明明知道它包含着极有价值的内容，却苦于无法找到正确的译码方法。

这种提取信息的过程可以分为三个层次。首先是要判定

你所得到的是一份包含某种内容的消息。这在很大程度上取决于你的专业“嗅觉”。考古学家发现某些织物的残片上有一些图样。他根据自己的经验将其与装饰性的花纹区别开来，其依据很可能是因为有一些图样重复出现，但又没有一定的周期性。反间谍人员苦苦思索某些信号的含义，仅仅因为他所追踪的对象千方百计掩蔽这些信号，或者作出某些异常的反应。在确定这是一份消息后，便进入第二个层次，即找出解译这份消息的密码。例如由考古学家或语言学家判断这是采用哪一种文字写的，或者请密码专家来破译密码。最后一个层次才是运用密码将消息的内容翻译出来。

宇宙的其他星球上有没有高等生物一直是人们感兴趣的问题，也是科学家认真研究的课题。这种探索的重要方法之一就是认真地监测来自各种天体的电磁波。如果有一天能够接受到有规则的信号，那么肯定会引起世界的轰动，就会有成千上万的人兴致勃勃地去努力揭示这种信号的秘密。而一旦解开了这种秘密，就意味着我们和其他高等生物建立了通信联系。因为我们有了共同的语言，也就是共同的编码和译码机制。

由此可见，通信过程就是通过编码把要传递的信息变换成便于传递的形式，即把它“装上”运载工具，由运载工具运到目的地后再通过译码机制把信息“卸下”。于是，信息就从一个系统传到另一个系统。

2. 信息源和信息量

2.1 信息与认识

依靠编码我们可以从消息中提取信息。这对人类具有非常重要的意义。它是我们认识世界的基础。我们的认识过程是从不知到有知，从知之甚少到知之较多，就是一个不断获得信息的过程。

我们可以把认识的对象看成一个系统，所谓获得有关这个系统的知识就是了解该系统的状态。如果我们具有关于它的完备知识，那么我们就完全肯定它处于什么状态。反之，如果我们没有任何有关它的知识，那就对于它处在什么状态一无所知。如果我们只有关于一个系统的部分知识，那就只知道它可能处于哪些状态。每当我们获得新的知识，这个系统的状态对于我们来讲就更加确定了。因此我们可以把这种确定性作为我们关于该系统知识多少的量度，也就是作为我们具有关于该系统信息多少的量度。从这种意义上讲，我们获得有关系统的信息就是我们所认识的系统变得更加确定。

举一个例子来讲。一个剧场有20排座椅，每排有30个座位。如果想知道一个人在什么座位上就需要获得一定的信息。这里我们把一个人看成一个系统，而把他在什么座位上看成是该系统所处的状态。也就是说，这个系统有600种可能出现的状态。

按照这些约定，并不是有关他的任何消息都包含我们所需的信息。例如“他和妻子坐在一起”，如果只有这条消息

就没有提供任何信息。但“他坐在妻子右边”就提供了一些信息，因为这些信息排除了他坐在最左边一行座位的可能性。也就是说，这个系统的可能状态从600种减少到580种。如果有一条消息告诉我们“他坐在单号”，那么系统的可能状态就减少到300种。显然第二条消息提供的信息比第一条消息要多。

我们常常会遇到一些有趣的智力游戏题。要寻找的答案实质上是确定一个系统的状态。在智力题所给出的条件里实际上已经包含了找出答案所需要的全部信息。不过这些信息以巧妙的方式被掩藏起来，需要你用自己的机智把这些信息提取出来。

一个最简单的例子就是小学生的一则智力题：“树上停着7只鸟，一枪打落一只鸟，树上还有几只鸟？”“打枪有声响，声响会惊走所有鸟”，这些信息隐含在原题的“一枪”这两个字里。只要会提取这个信息，问题就迎刃而解了。

各种形式的谜语也是如此。谜面中包含着可以确定谜底的足够信息。谜语越难，从谜面提取信息的机制就越复杂，所需要的技巧也就越高。

从一条消息中可以获得的信息与一个人的智力和知识背景有着很大的关系。“心有灵犀一点通”可以用来比喻一个人善于把新消息带来的知识与其他知识贯通起来。例如，迈克尔逊等人的实验证明“光速是不变的”。这对于普通人来讲只是叙述了一种物理现象。他可以记住这条结论，获得了一点物理知识。但是对于一个有素养的理论物理学家来讲，就可以在这块基石上建立新的物理学理论体系。他从这个消息中获得的信息无疑要大得多。情报分析专家往往是从不被人

们注意的蛛丝马迹中获得大量重要的情报（即信息）。

2.2 信息源

顾名思义，编码需要一组码。这些码是构成一段消息的建筑材料，是传递信息的工具。我们称之为信码或者符号。歌唱家用歌词传递各种信息，作曲家用更加抽象的符号——音符创作乐曲，表达自己的喜怒哀乐。数学家也许会使用一般人莫名其妙的数学符号来描述自然美妙和谐的旋律。他们都是在以自己熟悉的方式传递各种信息。

我们已经说过消息与信息是形式与内容的关系。我们最熟悉的消息是一段文字的描述。只要有文化的人都能把自己的思想用文字表达出来，也能够从别人写下的文字中理解其内容。以英语为例，一段文字由一些单词组成，每个单词又是由一些字母组成。众所周知，英语的字母只有26个，这些字母是构造一段英语消息的基本材料。

把我们的思想写成文字是编码的过程。相反，从一段文字理解它所表达的内容就是解码的过程。一个人学习文化就是建立上述编码与解码的机制。在这种机制建立起来之后，文字就成为有效的通信手段。

用文字构造消息，字母就是信码。用音符作曲，音符就是信码。同样我们也可以说数学符号也是信码。用文字通信，所有字母的集合称为信息源。用26个英语字母构成的信息源产生了拜伦和雪莱热情澎湃的诗篇，产生了莎士比亚不朽的名著。它的确是神奇而万能的源泉。

但是你是否想到过，在英语中这些字母的出现有没有什么规律性？显然是有的。只要你打开一本英语词典，马上会