信息論與編碼

[美] 诺曼·阿布拉姆逊 著
郭宝兰 译　　张彩录 校

XINXILUNYUBIANMA

河北大学教材科

信息论与编码

（校内教学参考书）

河北大学教材科

保定市郊区兴华印刷厂印

￥1元

校内教学
参 考 书

信 息 论 与 编 码

〔美〕诺曼·阿布拉姆逊　著
郭宝兰　译　　张彩录　校

河 北 大 学 教 材 科
八三·九

目　　录

前　　言

本书是在 ITT Federal Laboratories 讲授信息论与二元编码的一系列讲义的基础之上形成的。将此讲义又增加一些内容以后在 Stanford 大学的电气工学部的研究生院作为一个学期的教材使用，此后又在 IBM Research Laboratories, San Jose, California 作为一个学期的信息论的讲义讲授，最后才形成此书的初稿。

本书以交待清楚有关信息论的基本概念及其考虑方法为目的，尽量不拘泥于表现这些考虑方法的数学公式。可以把信息论认为是，用数学来研究由概率测度定义的抽象量的性质的科学。然而我们感兴趣的是理论与现实世界的关系，我们所研究的量是与在各种领域中起重要作用的自然产生的种种概念相对应的。为了充分明确地表示这种对应关系，我们使用作为表现信息论手段的数学，用数学可以严密地表示和证明信息论的各种重要定理。本书中对于定理和系均给出完全的证明，而对其结果则给出直观的解释。

阅读本书所要求的数学预备知识是不多的，只要具备对数的知识並能直观理解概率及平均值等的意义便足够了。本书未使用微积分学，与此相应地数学预备知识不太多的读者，即对极简单的讨论也不甚了解的读者，希望能通过花足够的时间达到理解本书所讲述的一些证明及其意义的程度。而理解书中证明的本身及其意义是並不需要特别的数学预备知识的。

本书之所以记述简明，是因为仅限于使用一般的数学知

识，书中所讨论的信源为具有有限个输出符号，其记忆也是有限的信源。同样所讨论的信道也是具有有限个输入与输出符号记忆为零的信道。在这样的限制条件下是可以展开信息论的所有重要概念的。因为有了这样的限制，故使其失去了数学的优雅性，然而对此优雅性有兴趣的读者可以把本书给出的证明移到 Borel 域上其结果将很容易地得到推广。

本书的内容是经过在大学和工场的教学实践充分检验过的，其内容足够对电子工程系、管理系、数理系、计算机系的学生进行一个学期的教学。而对于有更深数学预备知识的学生，或对特殊学科有兴趣的学生希望利用各章后面的注解探讨更高一级的概念。在这些注解中给出了信息论中一些有趣的研究领域。本书的各章后面还附有习题。习题中加有“*”号的是其解答需要数学分析知识的。

最后，在本书的形成过程中得到许多人的帮助，在此表示感谢。在本书的初期阶段 Richard Hamming 博士对原稿进行了详细的校阅，David Braverman 教授，Thomas Kailath 教授和 Wesley Peterson 教授提出了种种宝贵建议。本书还收录了 Thomas Cover，Arthur Geoffrion 和David Howell等位的许多启示。正如以上所述，本书是在 Stanford 大学的两个班和 IBM 的一个班的教学讲义的基础上写成的，对于曾指出本讲义中的错误和印刷错误的各位同学表示欠意：“实际上在你们的成绩中我未将你们提出的更正一览表考虑在内”。

Norman Abramson

第一章 序 论

1—1 什么是信息论（一）

信息论作为某一科学体系的名字是相当具有魅力的，本书的内容在某些地方使用信息论这一名字多少有些不妥。一般的公认信息论起源于仙侬（C·E.Shannon）1948年在贝尔系统技术杂志上发表的论文。仙侬给自己的论文命题为“通信的数学理论”。他大概认为把自己的论文定名为信息论是不太恰当的。仙侬的论文並未论及通常我们称之为信息的东西。而论及的是信息的载体，即关于符号的问题，因此，他的论文讨论通信及通信的方法，而没有讨论作为通信的结果在最终得到的抽象的东西——信息及其理论。

我们将要讨论的一些概念是极其重要的。在第二章我们将导出传送信息时所采用的符号的一些基本性质，及为使符号传递信息所必须遵从的一些规则，並表明符号的性质定然与其传送的信息量有关。现在考虑一个符号，此符号实际上是否传递了信息呢？一般地说，在本书所论及的范围内並不能决定。比如“le soleil brille”这句话，仅给本书读者中的一部分人带来信息。使用共同的语言传递信息是相当容易的，然而在研究信息时心理学因素虽不象语言那样重要，但却是不能忽视的。比如对“皓月当空”这句话，精神病患者听后大概有时会不只从气象学方面来领会其含义。从这种意义上考虑，对于同样一些单词的集合，听的人不同会有不

同的理解。仙依在论文中写道："关于通信中传递的意思这一侧面是与工程技术上的问题无关的"（1948年），然而，威沃（Weaver）在1949年指出，与其相反的结论必定是不正确的，也就是说有许多事情从工程技术方面讲与思想，心理学和语言的侧面是相关联的。本书2—8节中谈到在语言学上如何应用书中所述的理论。然而，除2—8节及各章的注解部分外，本书並未涉及信息论在其它领域的种种更专门的应用。

本书以信息论的各种重要概念，特别是信息的量度及其解释为重点。一些读者可能会希望多学一些信息论在其他领域中的应用，然而，这种要求是无界限的，例如本书所讨论的概念与由统计学上的实验得到的信息是相关的（林德里 Lindley 1956，库尔贝克 Kullback 1959，格雷顿伯格 Grettenberg 1962）。而进一步引出的信息论中心概念熵与热力学上的熵至少在形式上是一致的（布赖沦 Brillouin 1956，杰尹内斯 Jaynes 1959）。此外，奎斯特利尔（Quastler 1956）研讨了信息论在心理学方面的应用，皮尔斯（Pierce 1961）讨论了信息论在艺术方面的应用，巴—希列尔（Bar—Hillel）、卡那普（Carnap 1952）谈到信息论在动机方面的应用。而埃利斯（Elias 1958）对信息论在神学方面的应用做了有趣的探讨。

1—2 什么是信息论（二）

在研究信息时，首当其冲的是定义信息的量度，並应充分弄清该量度的性质。这样可进一步确认所定义的信息量度的正确性，並进一步明确作为数学理论出发点的物理模型与

理论本身的关系。然而，我们定义的信息量度是否正确，当然不能从以此定义为基础的理论体系完全导出。此后在本书中所讲述的信息论的体系，其自身当然是前后一致且完全正确的。然而，即使象这样的体系，如不进一步从体系之外给以证明，那也就仅仅只会成为数学理论而已。此后所述之体系是否合理是通过查清与该体系完全不同的一些量和该体系本身的关系来明确的。因此，本书首先给出信息的定义，並在此定义的基础上推导出一些关系，这些关系本身无疑是完全正确的，然而其出发点即信息的定义本身是否正确呢？这仅仅从以其为基础导出的各种关系相互间无矛盾这一点是不能证明的，只有把这些关系实际地应用于该信息体系之外的各个量来看才能证明有关信息的定义是否正确。明确此后所述之数学模型与物理世界如何对应这一点是必要且非常重要的。在本书的第一章中阐述一些与具体信息的量度定义完全无关的重要课题，这仅仅是强调以上所述之种种概念的重要性。而在定义信息的基础上如何给这些课题以定量的或数学的解释则是第二、第三、第四章的内容。

1—3　信息的编码

以下叙述一些信息传输的例子，並由此导出信息论的基本概念。在此首先考虑极其简单然而却非常重要的信息形态——二元信息的情况。要给出二元信息的例子是很容易的，比如穿孔卡片存储的信息，由电传打字机的通、断发出的信息，电子计算机中的双稳态器件所存储的信息都是二元信息的例子。为使论点明确，本章仅限于讨论这种形态的信息。

一般来说，对于信息的二元表示人们会觉得比较新奇，

其实这是自古以来就已经使用了的。在马太（Matthew）的福音书中就谈到二元消息的重要性，在那本书的第5章37节中有这样的话："你们的话仅仅是，是，是；非，非，超出此则是从恶处来的东西。"这种考虑方法也许有些极端。关于既包含有二元信息又包含有非二元信息的理论将在第二章以后论及。

表1—1给出了用二元数字"0"和"1"表示非二元信息的例子。此表为十进制数与二元序列的对应关系。是一个简单代码的例子。表1—1中的十个二元序列分别被称之

表1—1　十进制数的二元编码

十进制数	二元表示
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

为码字，而与其相应的十个十进制数被称之为消息符号。码与码字的严格定义将在3—1节给出。希望注意此处所述是稍欠严密的。若使用表1—1给出的码，则无论什么样的十进

制数序列均可用二元数字序列来表示。反之对由此代码得到的二元数字序列也可得到相应的十进制数字序列。

然而，由此並不能简单地下结论说从二元码字序列一定可以求得与其相应的消息符号。试考虑表1—2的代码。

表1—2　二元码

消息符号	代　码
S_1	0
S_2	01
S_3	001
S_4	111

当用表1—2的代码给出码字序列时，有时与此相应的消息符号序列並非只有一个，例如：给出

111001　…………（1—1）

序列，直观地可以看出与此相应的消息符号序列有

S_4S_3　…………（1—2）

或　$S_4S_1S_2$　…………（1—3）

两种可能性。

大概此时读者会提出，在代码中间插入“，”号或空不就可以防止这种混乱了吗？不必说这是正确的，然而使用了逗号或空，就增加了使用的符号，从而已不是二元代码了。事实上为使代码互相隔开而使用逗号，其结果是在码字中使用了0，1和逗号等三个符号。

其实能很简单地找出没有表1—2中代码之缺点的代码。例如表1—3中的代码，无论什么样的码字序列均只有

一个与其相应的消息符号序列。本章中仅使用这种码。

表1—3 二元码

消息符号	代 码
S_1	0
S_2	10
S_3	110
S_4	1110

1—4 信息传输中的一个问题

为明确编码的基本考虑方法与信息之间的关系，试考虑以下问题。首先在桑福兰西斯克与纽约之间建立一通信系统。其目的是每隔一定的时间把桑地的气候状况传送给纽约。假定此系统是只用通、断两种状态的二元信息传送装置，为使事情简单，在此将桑地的气候状态分为晴、阴、雨、雾四种，任何时候的天气均是此四种状态中的一种。可以把四种状态分别以消息符号表示，在此假定此四种状态的发生概率如表1—4所示。

表1—4 桑地气候状况

消 息	概 率
晴	¼
阴	¼
雨	¼
雾	¼

对于这样的四种状态用二元符号编码的方法是有多种的。以下给出编码的例子，以α代码表示。

α 代码	
晴……………………	0 0
阴……………………	0 1
雨……………………	1 0
雾……………………	1 1

………（1—4）

使用代码α则晴、雾、雾、阴可以编为0 0 1 1 1 1 0 1。

当使用代码α时，可以看出对于任何码字序列与其相应的消息符号序列只有一种，在这种意义上可以认为代码α对于传送这样的消息是恰当的。

在代码α中传送各个消息要用二个二元数字*。在这种情况下，如所使用的代码与消息相对应的码字不满足二位二元数字，则不能保证在任何情况下均能找到唯一的与其相应的原消息序列。

现在考虑一同样的问题，即在楼斯昂载尔斯与纽约之间建立同样的通信系统，以把楼地的天气状态传送到纽约。众所周知楼地的天气与桑地的天气在气象学上是有很大差异的，通常楼地的天气按晴、阴、雨、烟雾分成四类，这里所说的雾与烟雾的区别，对于当地居民来说也许是重要的。然而对于通信系统的设计却无任何影响，仅仅是将四种状态分别变换成二元序列。对于某一特定的序列代表何种意思，从

* 以下将二元数字（二进制数）简称为binit，区别二进制数（binit）与比特（bit）是很重要的，（比特是信息单位将在第二章定义）通常并不是一个二进制数就具有一比特的信息，只有在某种特定的情况下才具有一比特的信息。

通信的观点来说是没有什么关系的。

有一点，从通信的观点来看是不能忘掉的，即楼地的天气四种状态产生的概率是不相同的，表1—5给出了各状态的概率。

表1—5　楼地气候状况

消　息	概　率
晴	$\frac{1}{4}$
阴	$\frac{1}{8}$
雨	$\frac{1}{8}$
烟雾	$\frac{1}{2}$

如果使用α代码传送这些消息，是与桑地的通信状况完全相同的。使用α代码就是不问天气状态怎样，相应于一个消息使用二个二元数字，除此之外别无其它。下面试考虑使用β代码。

β代　码

晴……………………10
阴……………………110
雨……………………1110　　………（1—5）
烟雾…………………0

依此代码对于晴、烟雾、烟雾、阴，这样的消息序列进行编码，得到的二元序列为：1000110。

对于由β代码得到的二元序列，通常仅有唯一的消息序列与其相应，这由于此代码中的码字都以0结束，因此在接收端只要收到一个0，就可判定一个码字已完。β代码的平均长度（单位为二进数）可计算如下：

$$
\begin{aligned}
L &= 2P_r(\text{晴}) + 3P_r(\text{阴}) + 4P_r(\text{雨}) \\
&\quad + 1P_r(\text{烟雾}) \\
&= 2\cdot(\tfrac{1}{4}) + 3\cdot(\tfrac{1}{8}) + 4\cdot(\tfrac{1}{8}) + 1\cdot(\tfrac{1}{2}) \\
&= 1\tfrac{7}{8}\,\text{binits/消息} \qquad \cdots\cdots\cdots\cdots (1-6)
\end{aligned}
$$

这就是说使用β代码，在楼地与纽约之间传送一个天气状态平均只需使用$1\frac{7}{8}$个二进制数，而不再需要二个二进制数，然而若使用β代码传送桑地的天气状态（见表1—4）则相当于一个消息的平均长度为$L=2\frac{1}{2}$二进制数，由此可见，对楼地来说，欲传送同样的消息使用β代码时要比使用α代码少用6％的二进制数。在实际通信系统中所传送的二元数字少用6％是会得到很大好处的。在这里仅仅改变传送消息的表现形式就能获得好处。这一点有必要给与注意。

1—5　一些问题

由前节的例子得到一些需要研究的基本问题。首先，在上节的讨论中很简单地就获得了6％的利益，难道得到的利益不能比此更大些吗？也就是说在如何用二元序列表示消息上悉心研究就不会得到更多的好处吗？其次如果这样的改进是可能的（实际上前节的例子是可能的）那么会有一个界限吗？这也就是，消息的平均所需最低限度二进制数的个数的问题。如果决定了最小值为L，那么就又产生了如何构造这个最小值的代码的问题，也就是在实际上用什么方法构成代码的问题。

下面叙述与前节的例子有关的最后一个问题，就是“为什么”的问题，由上节的讨论已知，传送楼地的天气状态用比较少的二进制数就可以了，楼地与桑地的天气有什么样的