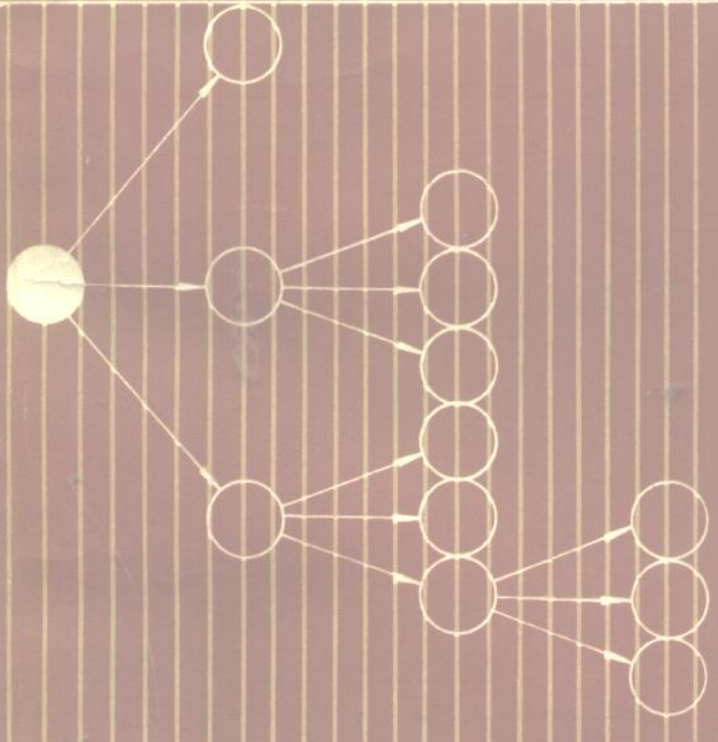
〔美〕R. W. 汉明 著

编码和信息理论



科学出版社

编码和信息理论

〔美〕R. W. 汉明 著

朱雪龙 译

科学出版社

内 容 简 介

本书用浅显易懂的语言，对信息和编码理论的研究对象、方法和所要解决的问题作了透彻而明确的阐述．全书共分十一章，包括检错码、纠错码、变长信源编码、熵、信道与互信息、信道容量等．本书把编码理论和信息理论有机地融合在一起，分析深入浅出、材料选取得当、物理概念清楚、理论联系实际，是编码和信息理论方面一本很好的入门书．

本书可供从事通信和计算机科学的科技人员和大学生阅读，也可供涉及信息的产生、传送和处理的其他有关学科的工作者参考．

R. W. Hamming

CODING AND INFORMATION THEORY

Prentice-Hall, 1980

编码和信息理论

〔美〕R. W. 汉明 著

朱雪龙 译

责任编辑 刘兴民

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1984 年 2 月第 一 版 开本：787×1092 1/32

1984 年 2 月第一次印刷 印张：8 1/8

印数：0001—7,800 字数：180,000

统一书号：15031·553

本社书号：3415·15—7

定价：1.30 元

译 者 的 话

编码和信息理论是在四十年代末发展起来的．由于这一理论抓住了信息的部分本质，同时也由于当时部分人的误解，使该理论在早期受到了相当普遍的欢迎和重视．但是几年以后，这一理论在实际应用上的进展却不很理想，部分人的过高的期望就逐渐转为失望，有关的科研和教学工作一度也曾有所减少．七十年代以来，计算机科学和数字通信的迅猛发展给这一理论增添了活力，从而使这一理论在较前更为广阔的基础上再次受到普遍的重视．正是在这样的背景下，R. W. Hamming 写了《编码和信息理论》这本书．

R. W. Hamming 是编码理论方面众所周知的先驱，他曾在 Bell 实验室作过长期的研究工作，后来又在美国海军研究院进行多年的教学工作，在教学和科研两方面都有比较丰富的经验．在本书中，他一方面对编码和信息理论作了通俗的理论联系实际的介绍，另一方面又对这一理论的作用和价值给予了恰如其分的估价．这两个方面，作者的阐述都是很出色的．

本书对于那些想了解和应用这一理论的计算机科学工作者和通信工程师将是一本很好的入门书．由于阅读本书不需要专门的数学知识和专业知识，所以他们将不费很大力气就可读完这本书并获得一个比较全面和准确的了解，在此以后，他们就可以知道这一理论对于他们有何意义以及如何用这一理论去开展他们的工作．

吴佑寿同志对本书的翻译给予了很大帮助，在此表示感

谢．由于译者水平有限，译文中的错误和不妥之处在所难免，欢迎读者批评指正．

译　者

1981年9月

原　　序

编码理论和信息理论这两个领域所研究的问题都是关于抽象符号的表示问题，在这本书中我们把它们自然地融合在一起了．这两个领域现在都已非常广阔，因此在这样一本小册子中我们只能简要地予以介绍．

通常认为信息理论研究信息在空间的传送（信息传输），但实际上信息在时间上的传送(信息存储)和在空间的传送完全一样．这两种情况在处理信息时都是经常发生的．很明显，如何对信息进行编码以便有效地进行存储以及在有干扰时可靠地进行恢复是计算机科学中一个很基本的问题．

和在计算机科学中的情况一样，信息的表示、传输和变换在其它很多学科中也是一个很基本的问题．不论何时何地，只要有信息的产生、存储和处理的问题发生，就总需要知道如何对冗长的材料进行压缩以及如何防止可能的差错．因此，现在该是使这一理论通俗易懂和便于应用的时候了．在已知的很多编码方法中，本书将只能介绍最重要的一些方法，但是我们希望书中提及的一些例子能使读者联想到其它的一些方法．

本书的目的是要阐述这两个领域的基本知识并给出应用这些概念的一些实例．书中所用的数学和电气工程知识已被减少到最低的程度，我们只利用了简单的数学分析和概率论的一些知识，任何超出这一范围的数学知识都在需要时作了推导．书中很多结果的证明和表示方法都用近年来计算机科学中发展起来的一些技术作了简化．这些技术在我们用到它

时也都作了解释，所以并不要求读者有计算机科学方面的专门知识．还有一些其它的证明也得到了很大的简化，同时在必要的地方又增加了一些新的材料以满足当前技术的需要．我们还在内容的编排上，特别是关于 Shannon 基本定理的证明上作了努力，以便使读者能够明白为什么这些定理是正确的而不仅仅是在数学上作出证明．

第十一章是关于代数编码的，在这一章中我们介绍了必要的有限域方面的知识．由于数学上的原因，这一章没有按应有的次序编排而是放在最后．如果需要的话，这一章也可以放在第三章的后面．书中有一些重复的地方是经过考虑有意这样做的，重要的概念一般至少要提到两次以保证读者真正理解这些概念．

本书舍弃了这些领域中的大量内容，我们相信真正掌握一点知识要比一知半解地知道一大堆东西好得多．这样，根据教师的判断，如果认为需要增加一些内容，那么增加起来也是很容易的．

我在这书中提到 Hamming 码和 Hamming 距时完全是按照通常的惯例．不这样叫，反而会造成学生的误解，这种谦虚是不必要的．

R. W. 汉明

目　录

第一章 绪 论

1.1 一个抽象而又概括的理论

编码和信息理论是一个相当抽象的理论．虽然在本书中我们也使用了一些具体生动的词汇，如“信息”、“传输”和“编码”等等，但是深加推究就会发现我们在这里所作的全部假设只是一个由符号 $s_1, s_2, s_3, \cdots, s_q$ 所组成的信息源．对于这些符号本身以及它们可能的含意，在这里一点也没有谈及，只是假定它们都是可以唯一地识别的．

随后就引入这些符号的发生概率 $p_1, p_2, \cdots, p_q$．但是如何去确定 p_i 的值不属于这一抽象理论的范畴．对于任意一个离散的概率分布有一个对应的函数 H：

$$H = \sum_{i=1}^{q} p_i \log\left(\frac{1}{p_i}\right)$$

这一函数称为熵函数，它是这一分布 p_i 所含不确定性、意外程度或信息的度量．这一函数在我们这一抽象理论中起着决定性的作用并给出平均码长的下限．以后我们还要考虑符号 s_i 的概率结构更加复杂的熵函数．

用另一种符号系统（一般讲是用只含 0 和 1 这两个符号的二元系统）来表示信源字母表的符号是本书的主要课题．在这一表示中两个主要的问题是：

1. 用什么方法可以使这些用新符号表示的信源符号相互间能有某种意义上的最大距离．这样，在它们的表示

有一些微小的变化(由于干扰)时，就可以发现其中是否有差错，甚至有可能纠正这些差错.

2. 从效率的观点出发，用什么方法表示信源符号可以获得最小的码长. 这就是使平均码长

$$L = \sum_{i=1}^{q} p_i l_i$$

达到最小，其中 l_i 是第 i 个符号 s_i 的码字长度. 熵函数给出了 L 的下限值.

这样，从原理上讲，这一理论仅仅是研究如何用一种确定的字母表(一般是二元系统)去表示某种信源符号的抽象的数学理论. 在这一抽象的数学理论中没有信息的传输、存储，也没有什么"干扰加在信号上". 这些含义具体的词汇在本书中只是为了推动这一理论才加以使用. 我们将一直使用它们，但是读者不要因此上当；一定要记住，从根本上讲这仅仅是一个有关符号的表示方法的理论.

1.2 历　史

编码和信息理论的历史可以追溯到很久很久以前. 还在1948年(也就是这两个理论第一次在坚实的基础上建立起来那一年)以前它的很多基本的概念就已为人们所了解. 1948年 Claude E. Shannon 在 Bell 系统技术杂志上发表了两篇有关"通信的数学理论"的文章(重印在参考文献[S]上). 这两篇文章的发表使信息理论这一领域得到了迅速普及，有关论文很快就出现在各种期刊上，在各大学的电气工程系和其它系里也先后开设了这一方面的课程.

如同其它很多被骤然打开的领域一样，信息论的很多早期的应用都是考虑欠佳的，但是这对于一个刚被发现的新的

领域来讲总是会有这种缺陷的．由于对信息论能做些什么开始时期望过高，所以渐渐地就产生失望，这门课程的授课时间也接着减少下来．现在，我们或许可以来作一个比较恰如其份的估价，一个介于初期的狂热和随后的失望之间的恰当的估价．

信息论给出了可能达到的性能的极限，但对于如何设计一个具体的系统却没有多大帮助．然而，因此就认为信息理论没有什么用这种想法也是错误的，以下的类比有助于说明上面的这一看法．我们来考虑一下生物学中讲授的进化论．虽然很少有学生在生活中直接应用到它，但它却仍然是概念上的一个重大里程碑．而且尽管缺少直接的应用，但只要掌握这样两个概念，即

1．物种的微小变化(变异)；

2．适者生存(选择)，

就可以把这些概念适当地应用到很多其它的、而且经常是远离生物学的领域中去*．例如，看一下社会组织，诸如计算机科学系、大学、军事机构、银行系统、政府以至于家庭，人们会问："目前这种状况是怎样造成的？""是什么力量使得这种状况得以生存下来？"

对这一理论作用的稍进一步的认识还会提出这样的问题："在目前这些力量的作用下，社会组织和习俗的哪些方面会发生变化(它适应于这些力量的能力)？"以及"它将要如何变化(什么东西将会生存下来)？"这样，进化论的概念就能够在很多远离生物学的领域中得到应用．

---

* 把进化论的观点简单地应用于社会生活，从辩证唯物论和历史唯物论的观点来看是不科学的，但作者对信息论用途的看法基本上是正确的．为使读者对作者的看法有所了解，我们仍把这一不恰当的类比全文译出．——译者注

类似地，信息论的概念也同样能在远离通信的很多领域中得到广泛的应用．在这些领域中这些概念的应用并非总是确切和恰当的——它经常只是启发性的——但这些概念仍然是非常有用的．

大约在信息论创建的同时，而且差不多在同一个地方，编码理论也建立起来了．然而，有关它的主要论文由于专利的缘故而一直拖延到 1950 年 4 月才出现在 Bell 系统技术杂志上（重印在参考文献［B2］和［B1a］中）．在最初的时候编码理论的数学基础并没有信息理论那样完善，而且在很长的时间内也没有得到理论家的重视．然而，随着时间的流逝，各种数学工具，诸如群论、有限域(Galois 域)理论以至于线性规划都已用于编码理论，因而编码理论成为数学研究中一个活跃的分支(参考文献［B1］，［Gu］，［MS］，［Mc］和［W］)．

自然科学中的多数学科都把误差放在次要的地位，只是在随后的设计阶段才考虑到它．但是编码和信息理论却把误差作为它的核心问题，由于误差(干扰)在现实生活中无处不在，所以人们对这样的理论便有特殊的兴趣了．

从逻辑上讲是编码理论导至信息理论，而信息理论又给出编码所可能达到的极限．所以尽管在过去这些年中这两个理论在很大程度上是独立地得到发展的，但这两个理论却是密切相关的．本书的主要目的之一就是要说明它们之间的相互关系．至于编码理论历史的更详细的情况可看参考文献［B2］．

1.3 通信系统的模型

一个一般的通信系统的模型是这样的：

1. 信息源．

2. 信源的编码.

3. 信道,信息通过它进行传输.

4. 噪声(误差)源,它作用在信道的信号上.

5. 译码,希望从受干扰的接收信号中恢复原始的信息.

6. 信宿.

表示在图 1.3-1 中的这一模型是我们将要使用的通信系

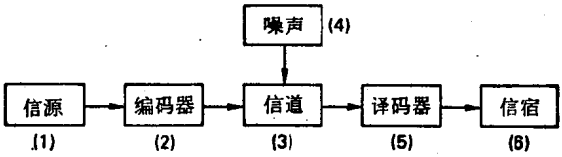


图 1.3-1 标准通信系统

统模型. 它具有现今通信系统的许多特征. 通常, 编码器被分成二级, 第一级是信源编码而第二级是为了适配信道而进行的编码. 译码器当然也相应地要分成二级. 下面我们将从系统的两端开始讨论,然后逐步讨论到中间.

1.4 信 息 源

我们从信息源开始讨论. 编码和信息理论的威力, 在很大程度上是由于我们没有去定义信息是什么——我们只假定有一个信息源,一个取自信源字母表$s_1, s_2 \cdots, s_q$的符号序列.到第六章时我们将看到信息论使用熵函数作为信息的度量, 而这就意味着对"信息的数量"下了定义. 但这是一个抽象的定义,它和一般人们所接受的信息的概念只是部份地有所吻合. 在信息论发展的早期,由于它貌似讨论一般所指的信息,所以很快地就得到广泛的传播. 人们在当时还没有注意这种差别, 还认为这一理论在任何场合下都能提供正确的含义. 由