



信息理论与编码

武汉理工大学 信息工程学院

教材:

吕锋、王虹. 信息理论与编码(第2版). 北京:
人民邮电出版社, 2010

参考书:

- 1) 王虹、刘雪冬. 信息理论与编码学习指导. 北京:
人民邮电出版社, 2011
- 2) 傅祖芸. 信息论——基础理论与应用(第3版).
北京: 电子工业出版社, 2011
- 3) Robert J. McEliece The Theory of Information and
Coding (第2版). 北京: 电子工业出版社, 2006

课程网站: <http://jpkc.whut.edu.cn/xxllybm1>



第1章 绪论

信息论是研究**信息及其运动规律**的科学。

人类社会的生存和发展，无时无刻都离不开发送、传递、接收、处理和利用信息。

最早对信息问题进行系统理论研究的是通信工作者。



第1章 绪论

信息论诞生的标志：**香农**发表的“通信的数学理论”。信息论是信息科学的**经典理论**。

科学的各个领域都存在大量的信息问题。（生物学、医学、管理学、经济学等），信息科学与其它学科**交叉**。



第1章 绪论

本节课主要内容：

1.1 信息的概念

1.2 信息论的主要研究内容

1.3 信息论的发展历程

1.1 信息的概念

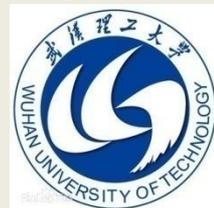
信息是信息论**最基本和最重要**的概念。

信息是信息论的**出发点**：认识信息的本质和运动规律。

信息是信息论的**归宿**：利用信息来达到某种具体的目的。

到目前为止，国内外已有不下百余种关于信息的定义。

从不同的侧面和不同的层次揭示了信息的本质。



1.1.1 信息概念的复杂性

从《钟义信》先生的《信息科学原理》选取部分关于对“信息”概念的描述。

用人们熟知的概念来定义信息：

信息是消息（消息是信息的载体）
信息是数据（数据是记录信息的形式之一）

用某些学科专门术语来定义信息：

信息是集合之间的变异度、信息是场、信息是信号。该类定义难以理解。



1.1.1 信息概念的复杂性

从**哲学意义**上来定义信息：

信息是信息，既不是物质也不是能量。

信息是事物之间的差异。

信息是事物相互作用的表现形式。

信息是事物联系的普遍形式。

信息是物质的普遍属性。

该类定义**难以**引出信息的**度量**方法。



1.1.1 信息概念的复杂性

从**控制论**和**系统论**的角度定义信息：

- 信息是与控制系统相联系的功能现象。
- 信息是控制的指令。
- 信息是系统组织程度的度量。
- 信息是有序性的度量。
- 信息是负熵。
- 该类定义**难以**引出信息的**度量**方法。
- 后三种与香农对信息的理解接近。

1.1.1 信息概念的复杂性

从**随机不确定性**的角度定义信息：

- 信息是收信者事先不知道的内容。
- 信息是用以消除不确定性的东西。

下此类定义者对**通信**有所了解，或者本来就是通信领域的专家，定义信息时考虑到了通信的实际情况。

例：一场巴西和阿根廷的足球赛。

赛前，有两队势均力敌的先验知识。

上半场得到“巴西队**3:0**”领先。



1.1.2 信息的定义

牛津英文字典里对**information**给出的解释是“某人被通知或告知的内容、情报、消息”。该解释不是科学名词或技术术语。

哈特莱（R.V.L.Hartley）在1928年发表的《信息传输》中提出“发信者所发出的信息，就是他在通信符号表中**选择符号**的具体方式”。该定义的缺陷是未考虑各种选择方式的统计特性。

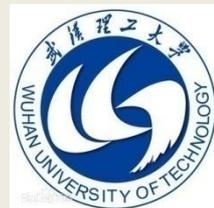


1.1.2 信息的定义

1948年，美国科学家**维纳**（N.Wiener）在《控制论——动物和机器中通信与控制问题》一书中指出：“信息是信息，不是物质，也不是能量”。

后来，维纳在《人有人的用处》一书中提出：“信息是人们适应外部世界，并且使这种适应反作用于外部世界的过程中，同外部世界进行互相交换的内容的名称”。

维纳的定义不够确切。



1.1.2 信息的定义

香农 (C.E.Shannon) 在 1948 年发表的著名论文“通信的数学理论”中对信息的定义，被认为是科学的定义，并对信息进行了定性和定量的描述。

香农对信息的定义：

信息是事物运动状态或存在方式的**不确**
定性的描述。



1.1.2 信息的定义

香农1916年4月30日生于美国密歇根州。

1936年在密歇根大学电子工程和数学系获得理学学士学位。

1940年在麻省理工学院获博士学位。

之后加入贝尔电话实验室。

1941年，开始了对通信问题的深入研究。

1948年凭借发表的论文，被认为是信息论之父。



1.1.2 信息的定义

香农的以下三大论点，以新颖的思想和方法，跨出了用数学方法**定量**描述信息的关键一步，开创了通信领域信息理论新局面。

① 形式化假说

它使利用数学工具定量度量信息成为可能。

② 非决定论

它从原则上解决了用什么样的数学工具解决信息度量问题。

1.1.2 信息的定义

③ 不确定性

通信的过程是一种**消除不确定性**的过程。
不确定性消除得越多，获得的信息就越多。

“不确定性”是与“可能性”相联系的。

“可能性”在数学上是用概率来表示。

概率大 \longrightarrow 可能性大 \longrightarrow 不确定性小。

信源符号 a_i ($i = 1, 2, \dots, r$)

先验概率 $P(a_i)$

自信息量（不确定性） $I(a_i) = f[P(a_i)]$

1.1.2 信息的定义

从客观事实和人们的习惯概念出发，函数 $I(a_i)$ 必须满足以下四个公理性条件：

1) $I(a_i) = f[P(a_i)]$ 是先验概率 $P(a_i)$ 的单调递减函数。

若 $0 < P(a_i), P(a_j) < 1$,且 $P(a_i) > P(a_j)$,则 $I(a_i) < I(a_j)$ 。

2) 若 $P(a_i) = 0$, 则 $I(a_i) \rightarrow \infty$ 。

3) 若 $P(a_i) = 1$, 则 $I(a_i) = 0$ 。

4) 统计独立事件的联合信息量，等于它们各自信息量之和。





1.1.2 信息的定义

在数学上可以证明，满足以上四个公理条件的**自信息**函数为

$$I(a_i) = \log \frac{1}{P(a_i)} = -\log P(a_i)$$

单位由对数的底决定：

以2为底，单位为bit（比特）

以e为底，单位为nat（奈特）

以10为底，单位为hart（哈特）





1.1.2 信息的定义

事实上信息的概念远远超出原来通信领域的范畴。

哲学家和科学家普遍认为，**物质、能量和信息**，是物质世界的三大**支柱**。

世界是物质的，信息是物质的一种普遍属性。

信息的性质：

普遍存在；寄存性；可传递；可共享；可变换；可转化；可伪；可量度；客观的；无限的；相对的；抽象性；时效性；异步性



1.1.2 信息的定义

我国学者钟信义教授在**本体论**（最高、最普遍，无约束条件的层次）的层次，对信息的定义：信息是事物运动的状态和方式，就是关于事物**运动的千差万别的状态和方式**的知识。

在本体信息的基础上，引入认识主体的一些约束条件，从而又提出了信息的不同定义。

语法信息：是事物运动状态和状态改变的方式本身，是信息最抽象和最基本的层次。香农信息的定义属于这个层次。

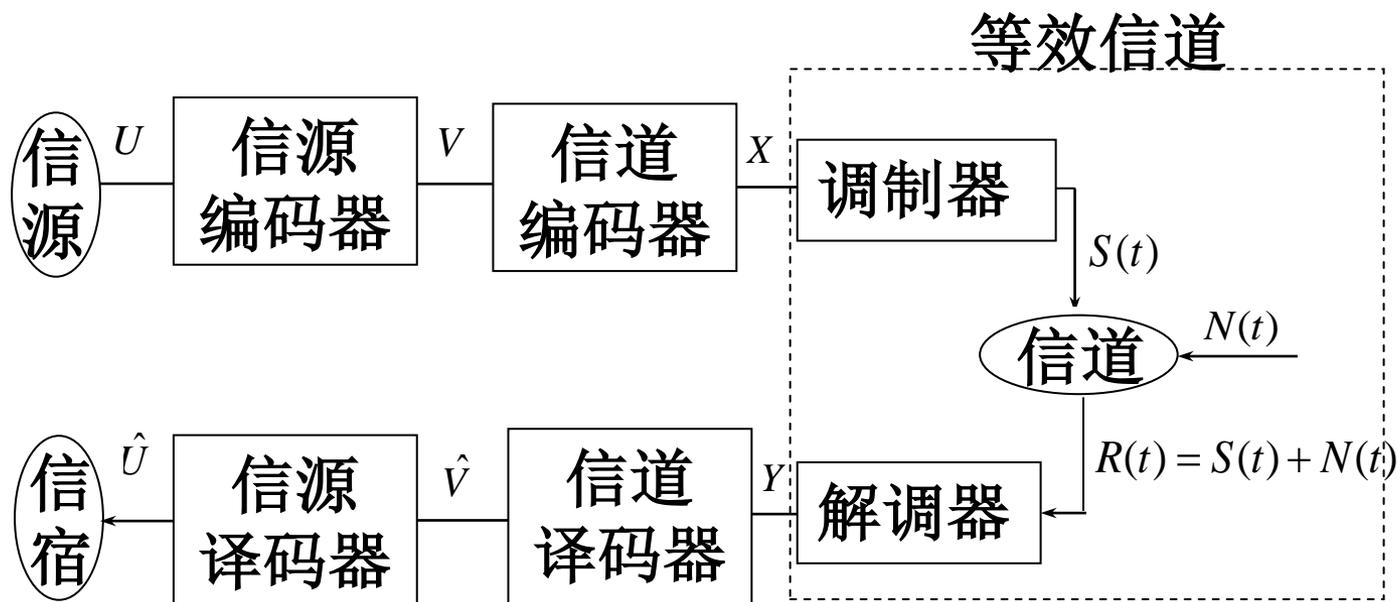
语义信息：是事物运动状态和方式的具体含义。

语用信息：是事物运动状态和方式及其含义对观察者的效用，是信息的主观价值。



1.2 信息论的主要研究内容

研究对象：统一的通信系统模型



信源的度量：第2章；

信道的数学模型和信道容量：第3章；

信源编码：第4、6章；信道编码：第5章。



1.2 信息论的主要研究内容

研究目的：

找到信息传输过程的**共同规律**，以提高信息传输的可靠性、有效性、保密性和认证性，以达到信息传输系统最优化。

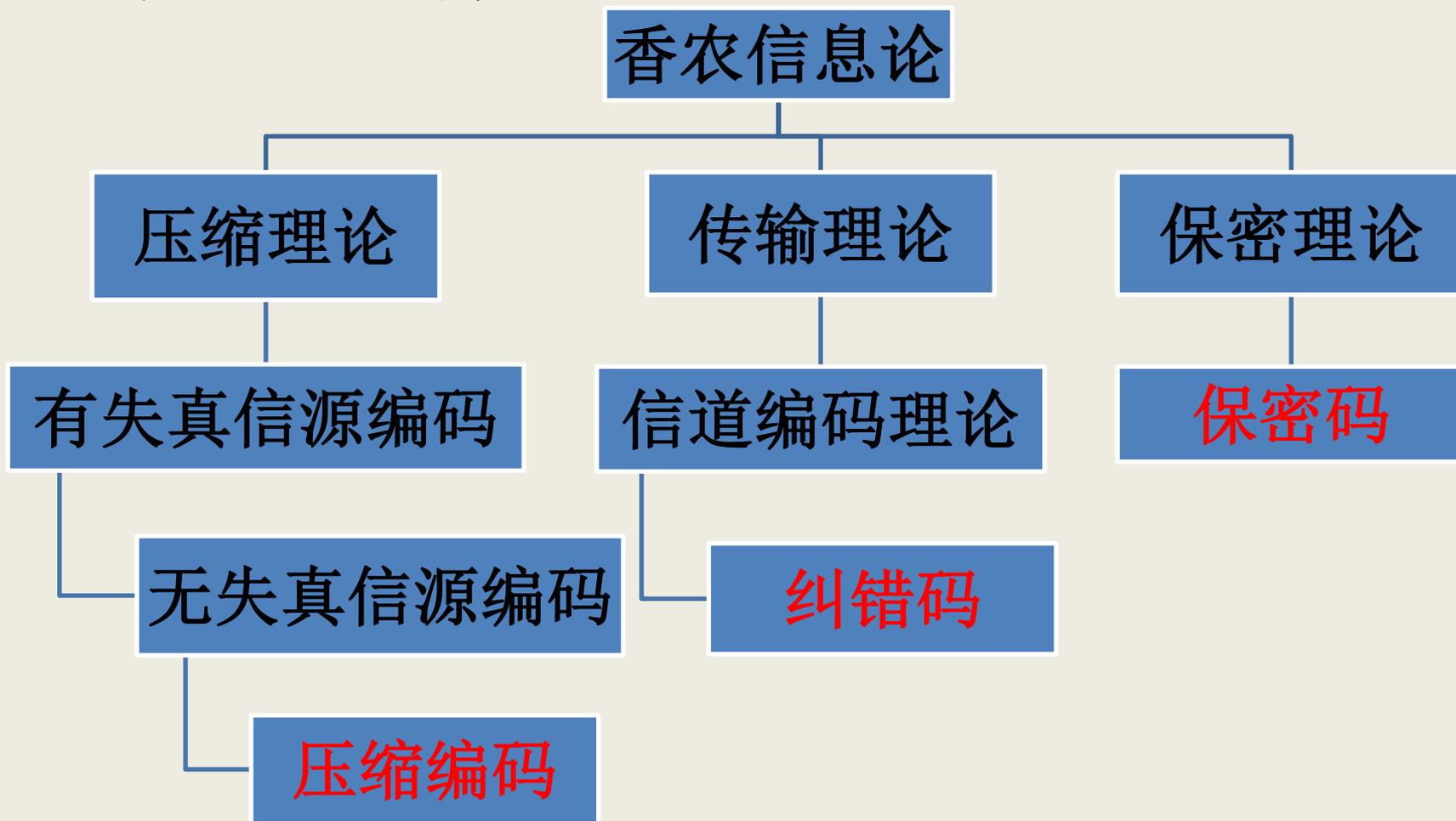
研究内容：一般有三种理解

1) 狭义信息论（经典信息论）

又称香农基本理论。主要研究信息的测度、信道容量、信源编码理论、信道编码理论。

1.2 信息论的主要研究内容

香农信息论的科学体系：



1.2 信息论的主要研究内容

2) 一般信息论

香农理论、噪声理论、信息处理理论、统计检测与估计理论等。

3) 广义信息论

广义信息论是一门综合、交叉的新兴学科，包括上述两方面内容，以及所有与信息有关的自然和社会领域。

新兴交叉学科：**量子信息论、生物信息学、医学信息学、经济信息学。**





1.3 信息论的发展历程及应用

1) 信息论的发展历程

信息论的早期酝酿 20世纪初~ 20世纪40年代

① 早期编码问题: Morse码

② 通信的有效性与可靠性:

奈奎斯特 (H.Nyquist)

③ 香农熵的其它产生途径

维纳 (N.Wierner)

④ 纠错与纠错码的产生





1.3 信息论的发展历程及应用

香农信息论的确立 1948年~1960年代

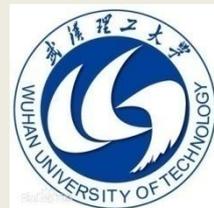
1948年，《贝尔系统技术》“通信的数学理论”为奠基性论文。无失真信源编码定理、无失真信源编码方法（香农编码）

1949年，“保密通信的信息理论”。用信息论的观点对信息保密问题进行了全面论述。

1959年，“保真度准则下的离散信源编码定理”。提出了率失真编码定理。率失真编码理论是信源编码的核心问题，是数据压缩的理论基础。

1961年，“双路通信信道”。开拓了网络信息论的研究。





1.3 信息论的发展历程及应用

主要特点:

对通信系统数学模型与基本问题的讨论

对信息量、香农熵的讨论

对信源、信道的讨论

对信源、信道编码定理的讨论与证明

信源、信道编码的实现与应用

主要标志:

对以上问题实现了严格的数学描述与证明、一系列专著地完成。





1.3 信息论的发展历程及应用

2) 信息论的发展与应用

信息论是理论基础、思想先导、技术关键性。

信源编码:

1948年, Shannon码; 1952年, Fano码、Huffman码 (最佳码)。

20世纪70年代以后, 人们转向与实际应用有关的信源编码问题。1968年, P.Elias提出了算术编码的初步思想, 1976年, J.Rissanen给出和发展了算术编码。用于数据压缩、视频压缩等领域, 部分技术受美国专利的保护。

1977年, L-Z编码, 之后L-Z78、LZW等变化和改进, 广泛应用于计算机数据压缩中。



1.3 信息论的发展历程及应用

数据压缩**国际标准**:

语音 G.711、G.721、G.722、G.726、G.727等

会议电视、视频 H.261、H.264

静止图像 JPEG

常规电视 MPEG-1

高清晰度电视 MPEG-2、

低比特率音视频 MPEG-4、

多媒体信息描述标准 MPEG-7

多媒体框架 MPEG-21



1.3 信息论的发展历程及应用

信道编码（纠错码）：

1950年，汉明码；同时期其它代数编码还有循环码。

BCH码，纠正多个错误的循环码。RS码，多进制BCH码。

1955年，Elias提出卷积码，1960年，以Viterbi译码为代表的译码方法被美国卫星通信系统采用。

交织码、Turbo码、TCM码、LDPC码等。

香农信道编码理论是真正具有实用意义的科学理论。



第2章 信息的度量