

第1章

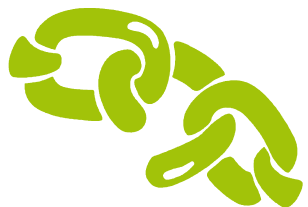
绪论



关于我



- 吴嘉婧 中山大学 数据科学与计算机学院 副教授
- 主要研究方向：区块链，图挖掘，网络科学，网络表示学习。
- 联系方式：wujiajing@mail.sysu.edu.cn
- 助教：林凯欣
- 办公室：微电子学院楼406



教材和参考资料

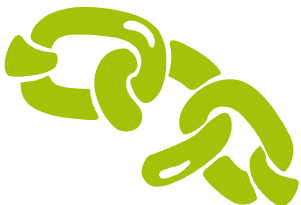


✓ 课堂教材

- 田宝玉等，《信息论基础》，人民邮电出版社。

✓ 参考教材

- Thomas M. Cover and Joy A. Thomas, “Elements of Information Theory” , Jone Wiley & Sons, Inc. 1991.
- 曹雪虹，张宗橙，《信息论与编码》，清华大学出版社。



主要内容



1.1 信息的基本概念

1.1.1 信息论的产生

1.1.2 信息的基本概念

1.2 香农信息论研究的内容

1.2.1 通信系统模型

1.2.2 香农信息论的主要内容





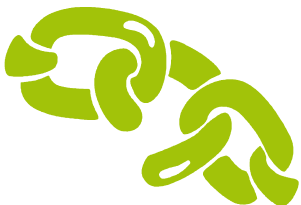
1.3 香农信息论研究的进展与应用

1.3.1 香农信息论创立的背景

1.3.2 香农的主要贡献

1.3.3 香农信息论研究进展

1.3.4 香农信息论的应用



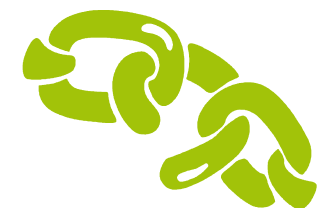
信息论定义



信息论即香农信息论，也称经典信息论

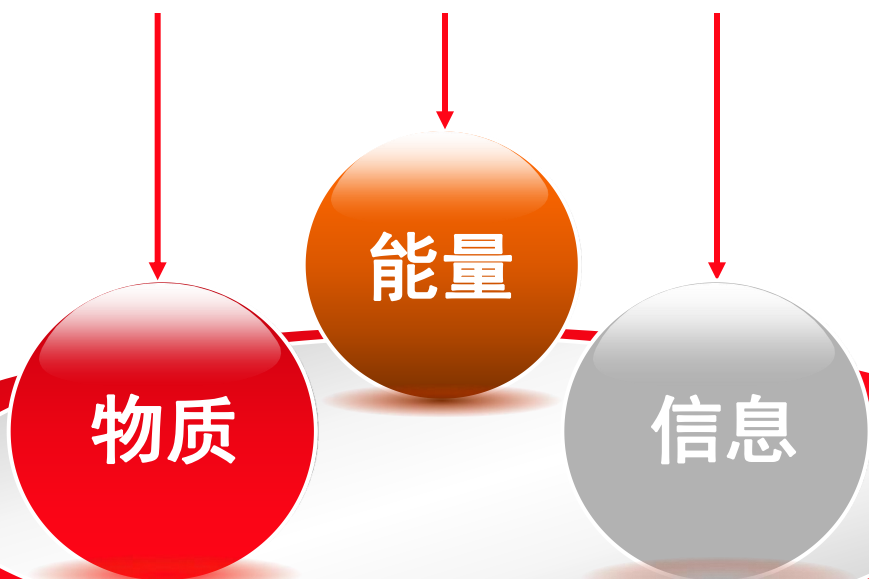
信息论是研究通信系统极限性能的理论

香农(Claude Edwood Shannon, 1916—2001)



§ 1.1 信息的基本概念

客观世界三大基本要素



信息时代社会的发展离不开物质(材料)、能量(能源)和信息资源。

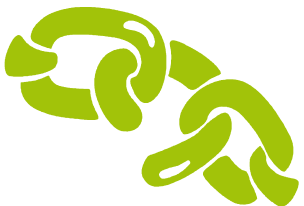
美国学者欧廷格：

没有物质什么都不存在，没有能量什么都不发生，没有信息什么都没意义。

§ 1.1.1 信息论的产生



按理说，关于信息的课程应该象物理、化学、生物等课程一样，是基础课。但是，由于信息的抽象性以及当前人们且对信息的认识并不完全清楚，所以在当前只能是专业课。

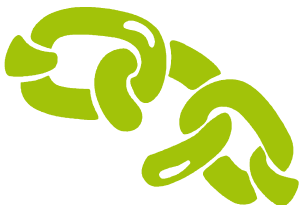


§ 1.1.1 信息论的产生



1948年，美国工程师和数学家Claude Edwood Shannon (1916-2001)发表《通信的数学理论》，标志着信息论的产生。

(A Mathematical Theory of Communications, Bell Sys. Tech. Journal, July and October 1948)



§ 1.1.1 信息论的产生

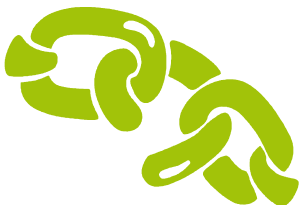


通信的基本问题

在一点精确地或近似地恢复
另一点所选择的消息

通信的三项 性能指标

传输的有效性
传输的可靠性
传输的安全性



§ 1.1.1 信息论的产生



有效性

对于离散信源，信源符号平均码长度应尽量短；
信息传输应尽量快，即高的传信率；
信息传送应该有高的频谱利用率。

可靠性

传输差错要尽量少，对数字传输就是要求低的误码率。

安全性

传输的信息不能泄露给未授权人。



§ 1.1.1 信息论的产生



三项指标所对应的三项基本技术：

数据压缩

数据纠错

数据加密

香农建立了通信中的三项基本技术的理论基础

信息论是前两项技术的理论基础。



§ 1.1.1 信息论的产生



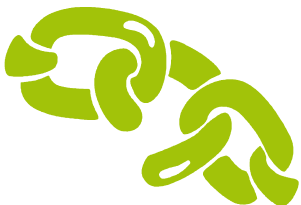
香农信息论解决了前两项技术的理论问题

- 提高有效性可通过信源编码来实现

即信源压缩编码，并给出了压缩编码的极限

- 提高可靠性通过信道编码来实现

并给出实现可靠传输的信息传输速率的极限



§ 1.1.1 信息论的产生



传输安全性的问题也是香农解决的。

- 《密码学的数学理论》

(A Mathematical Theory of Cryptography)

- 《保密系统的通信理论》

(Communication Theory of Secrecy System)

香农揭示了数据压缩和传输的基本定律，并建立了通信中的三项基本技术的理论基础。



§ 1.1.2 信息的基本概念

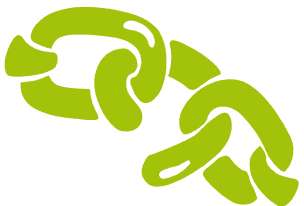


信息论的产生引起了很多专家学者对信息研究的兴趣，他们从不同的角度和侧面研究和定义信息。

例如：

- 信息是事物之间的差异
- 信息是物质与能量在时间与空间分布的不均匀性
- 信息是收信者事先不知道的东西

通俗概念：信息是一种消息



§ 1.1.2 信息的基本概念



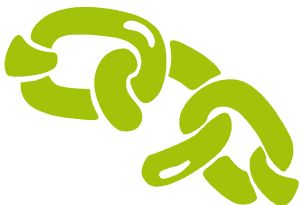
信息的基本特征

- 信息有许多与物质、能量相同的特征，例如信息可以产生、消失、携带、处理和量度。
- 信息也有与物质、能量不同的特征，例如，信息可以共享，可以无限制地复制等。

实际上，信息可以划分为两个大的层次：

- 本体论层次：信息是客观的，即它是独立于人或其他有感知的事物而存在的
- 认识论层次：信息是通过认识主体的感受而体现出来的

现在我们所说的信息实际上是指认识论层次的信息。



§ 1.1.2 信息的基本概念

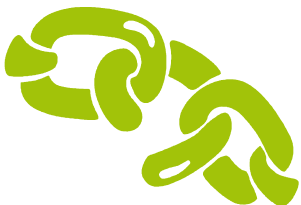


1949年，Weaver在《通信的数学》中解释香农的工作时，把通信问题分成三个层次：

第一层：通信符号如何精确传输？（技术问题）

第二层：传输的符号如何精确携带所需要的含义？（语义问题）

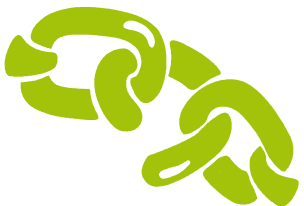
第三层：所接收的含义如何以所需要的方式有效地影响行为？
（效用问题）



§ 1.1.2 信息的基本概念



- 广义信息的描述：
信息是认识主体（人、生物、机器）所感受的和所表达的事物运动的状态和运动状态变化的方式。
- 以这种定义为基础，信息分成三个基本层次，即语法信息，语义信息，语用信息，分别反映事物运动状态及其变化方式的外在形式、内在含义和效用价值。
- 与Weaver对信息的解释基本一致。



§ 1.1.2 信息的基本概念

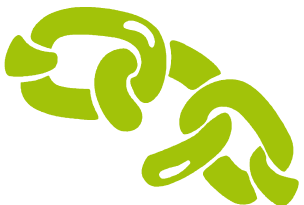


信息的三个基本层次：

语法(Syntactic)信息

语义(Semantic)信息

语用(Pragmatic)信息

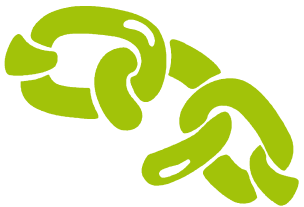


§ 1.1.2 信息的基本概念



- 语法(Syntactic)信息

事物运动的状态和变化方式的外在形式，不涉及状态的含义和效用。像语言学领域的“词与词的结合方式”，而不考虑词的含义与效用。在语言学中称为语法学。语法信息还可细分为，概率信息、偶发信息、确定信息、模糊信息等。



§ 1.1.2 信息的基本概念

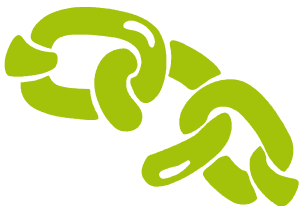


- 语义(Semantic) 信息

事物运动的状态和变化方式的含义。在语言学里，研究“词与词结合方式的含义”的学科称为语义学。

- 语用(Pragmatic)信息

事物运动状态及其状态改变方式的效用。



§ 1.1.2 信息的基本概念



举例说明信息三个层次的含义

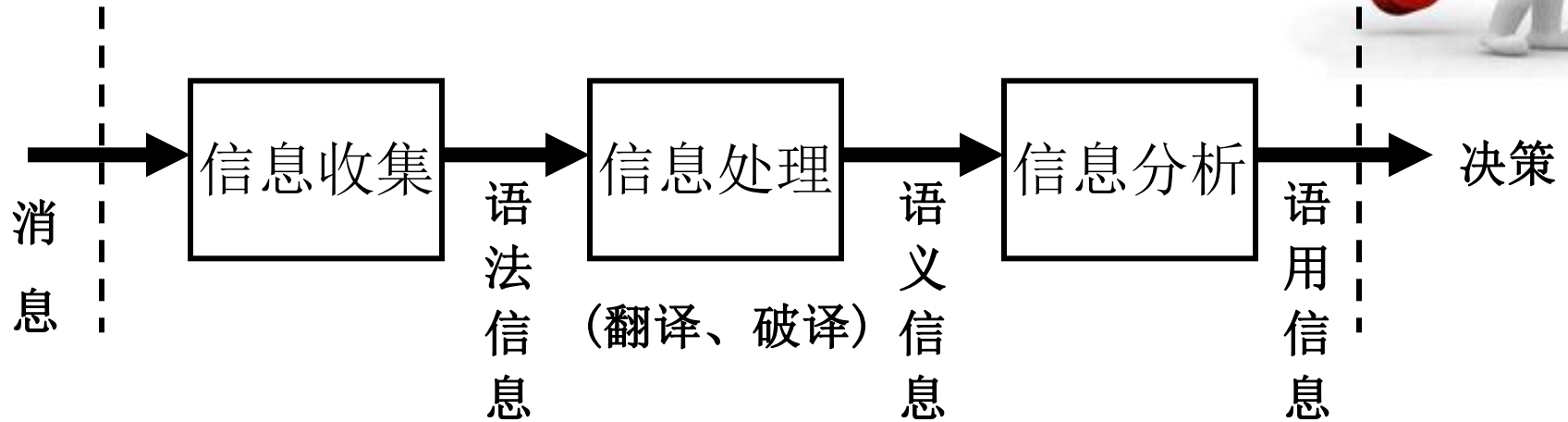
有一个情报部门，主要对经济情报进行收集、整理与分析，以提供给决策机构。

- 信息收集组：将收集到的资料按中文、英文或其他文字、明文、密文进行分类，不管这些资料的含义如何都交到信息处理组。
- 信息处理组：翻译或破译得到这些资料的含义，然后交到信息分析组。
- 信息分析组：挑选出有价值的情报提交给决策机构。

§ 1.1.2 信息的基本概念



举例说明信息三个层次的含义

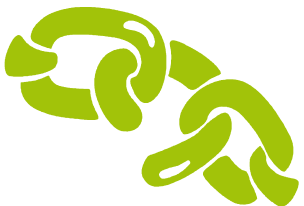


- 语法信息：英文，中文，密文等
- 语义信息：含义
- 语用信息：用途

§ 1.1.2 信息的基本概念



- 研究语义信息要以语法信息为基础
研究语用信息要以语义信息和语法信息为基础。
- 语法信息是最简单、最基本的层次
语用信息则是最复杂、最实用的层次。



§ 1.1.2 信息的基本概念



香农在1948年指出：

“通信的基本问题是在一点精确地或近似地恢复另一点所选择的消息。通常，这些消息是有含义的，即它对于某系统指的是某些物理的或概念的实体。这些通信的语义方面与通信问题无关，而重要的方面是实际消息是从一个可能消息集合中选择出的一条消息。”

§ 1.1.2 信息的基本概念



香农先从语法信息入手解决当时最重要的通信工程一类的信息传递问题。

香农信息论研究的事物运动的状态和变化方式的外在形式是遵循某种概率分布的随机事件。

香农信息论或经典信息论所研究的信息是：
语法信息中的概率信息

§ 1.1.2 信息的基本概念



- 香农对信息的定义：
 - 是对事物运动状态或存在方式不确定性的描述。
 - 概率空间 $[X, P]$ ， X 为样本空间， P 为概率函数。
 - 离散概率空间

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{bmatrix} \quad \sum_{i=1}^n p_i = 1$$

- 自信息和信息熵

§ 1.1.2 信息的基本概念



通信信息的三个层次



- 消息是信息的携带者；
信息包含于消息之中
- 信号是消息的载体；
消息是信号的具体内容

§ 1.1.2 信息的基本概念

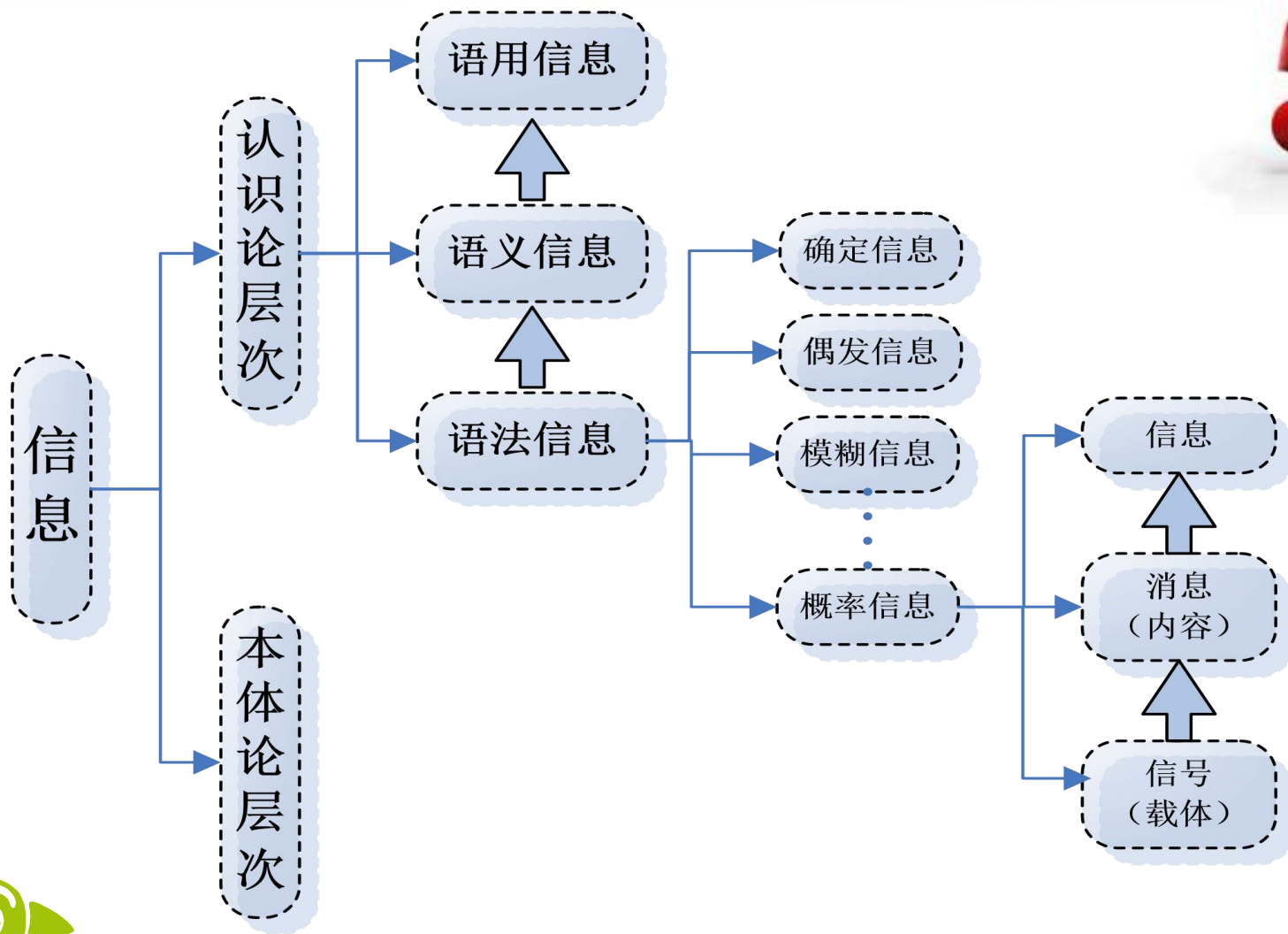
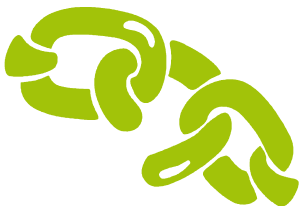


图1.1 信息各层次之间的关系图

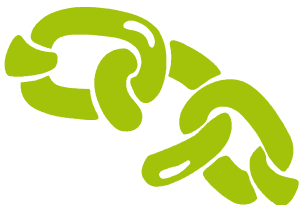


§ 1.2 香农信息论研究的内容



1.2.1 通信系统模型

1.2.2 香农信息论的主要内容



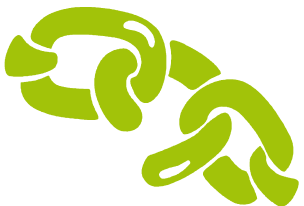
§ 1.2.1 通信系统模型



点对点的通信的通信系统模型是指“从一个地方向另一个地方传送信息的系统”

通信系统是从空间的一点到另一点传送信息的

存储系统是从时间的一点到另一点传送信息的



§ 1.2.1 通信系统模型

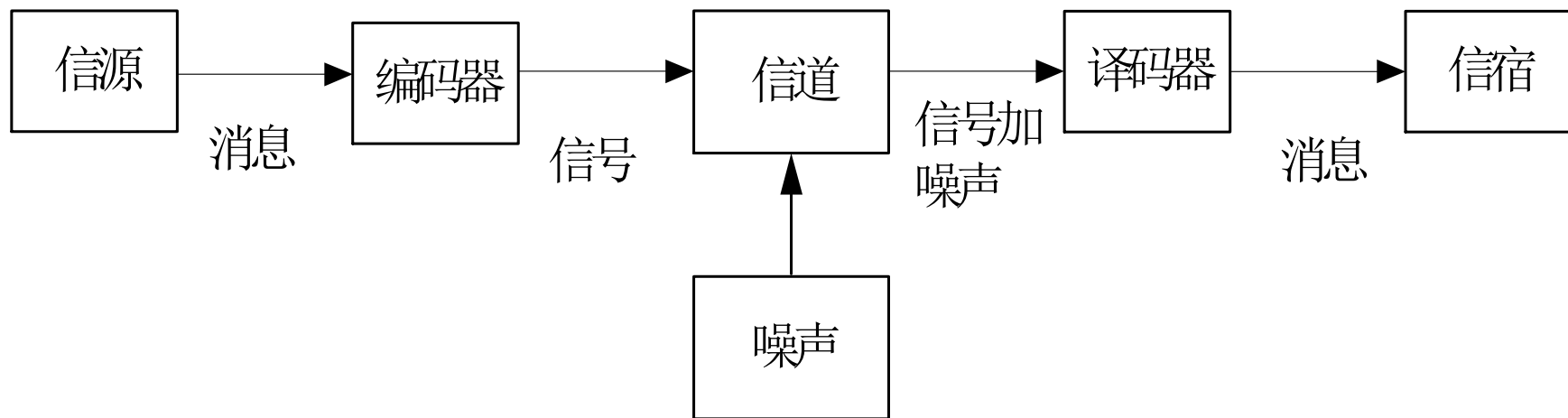
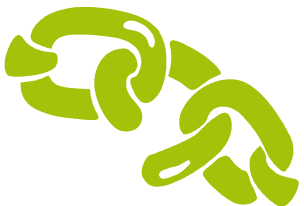


图1.2 通信系统模型



§ 1.2.1 通信系统模型



信源(information source)

功能：直接产生消息或消息序列

按输出符号的取值分类：离散信源、连续信源

按输出符号之间的依赖关系分类：无记忆信源、有记忆信源

信源的消息中所包含的信息量
以及信息如何量度

核心
问题



§ 1.2.1 通信系统模型



编码器(Encoder)

编码器的功能是将消息变成适合于信道传输的信号
编码器包括：

信源编码器 (source encoder)

信道编码器 (channel encoder)

调制器 (modulator)



图1.3 编码器的组成



§ 1.2.1 通信系统模型

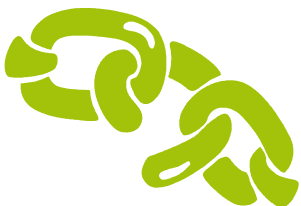


信源编码器

功能：将信源消息变成符号

目的：提高传输有效性

即压缩每个信源符号传输所需代码（通常为二进制代码）的数目（对二进制代码称比特数）。



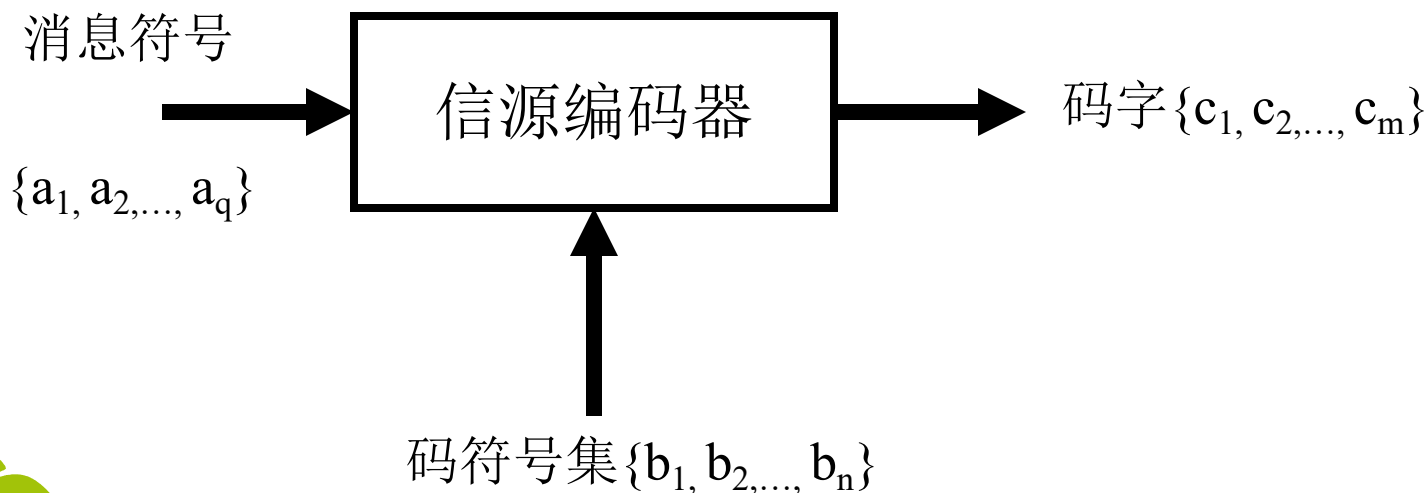
§ 1.2.1 通信系统模型



例如，一个信源含4个符号{a,b,c,d}，概率分别为1/2，1/4，1/8，1/8。

(1)如果不采用信源编码，每个信源符号至少需要用2个二进制代码传输。

(2)如果采用信源编码，分别将a,b,c,d编码成为：0，10，110，111，那么平均每信源符号只需1.75个二进制代码传输。



§ 1.2.1 通信系统模型



信道编码器

功能：给信源编码符号增加冗余符号

目的：提高传输可靠性

信源编码输出直接传送，不能保证传输可靠性。

信道编码对信源编码器的输出符号增加一些冗余符号，并让这些符号满足一定的数学规律，使传输具有纠错或检错能力。



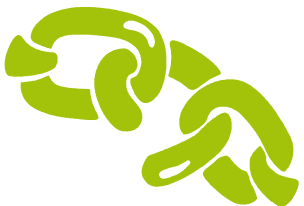
§ 1.2.1 通信系统模型



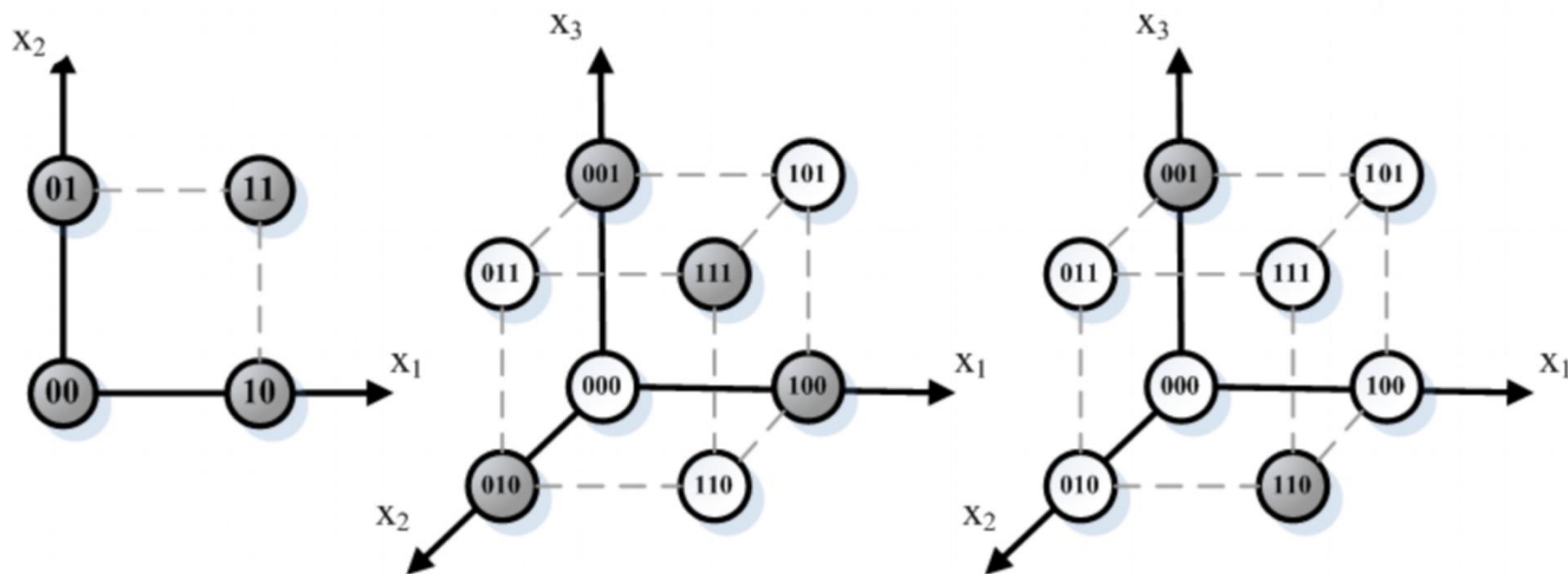
例如，奇偶纠错

将信源编码输出的每个码组的尾补一个1或0

当传输发生奇数差错，打乱了“1”数目的奇偶性，就可以检测出错误。

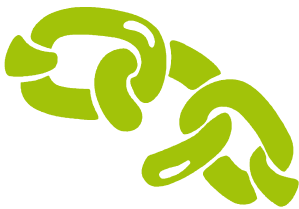


§ 1.2.1 通信系统模型



(a) 无检错 (b) 可检错 (奇校验) (c) 可纠错 (纠一个错)

图1.4 增加冗余符号增加可靠性示意图



§ 1.2.1 通信系统模型

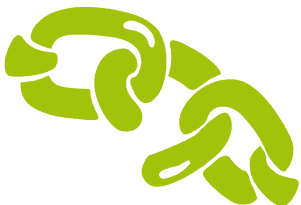


调制器

功能：将编码器的输出符号变成适合信道传输的信号

目的：提高传输效率

信道编码符号不能直接通过信道输出，要将编码器的输出符号变成适合信道传输的信号，例如，0、1符号变成两个电平，为远距离传输，还需载波调制，例如，ASK，FSK，PSK等。



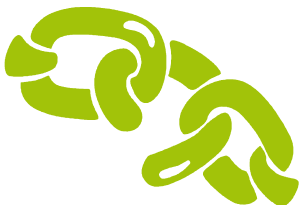
§ 1.2.1 通信系统模型



信道(channel)

信道是信号从编码器传输到译码器的中间媒介。

- 信道
- 狭义信道：某些物理通信信道，也可以是物理的存储介质
 - 广义信道：一种逻辑信道，它和信息所通过的介质无关，只反映信源与信宿的连接关系



§ 1.2.1 通信系统模型



信道(channel)

信道 { 无噪声信道
有噪声信道

通信系统中主要有两种噪声：加性噪声和乘性噪声。
一般，背景噪声为加性，而衰落为乘性。

在信息论中研究最多的是理想加性高斯白噪声（AWGN）信道。



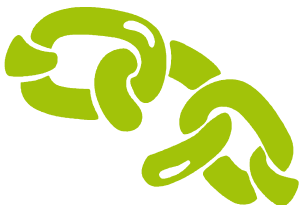
§ 1.2.1 通信系统模型



信道(channel)

高斯分布的普遍性主要基于两种原因：

- (1) 根据中心极限定理，无数独立随机变量的和的分布趋近高斯分布，因此高斯噪声普遍存在
- (2) 在限功率条件下产生最大熵的信源分布为高斯分布，而最大熵分布是最容易被观察到的分布



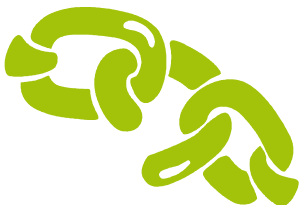
§ 1.2.1 通信系统模型



信道(channel)

信道 { 离散信道
离散时间连续信道
波形信道（或模拟信道）

- 离散信道和离散时间连续信道输入与输出都是符号序列，只不过符号取值不同，前者取离散值，而后者取连续值
- 波形信道的输入与输出均为时间的连续波形。

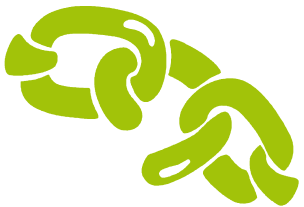


§ 1.2.1 通信系统模型



信道(channel)

- 信道也可有无记忆信道和有记忆信道的区分
- 离散信道和离散时间连续信道可以是无记忆的，也可以是有记忆的
- 波形信道是通常有记忆的。



§ 1.2.1 通信系统模型



译码器(decoder)

译码器的功能：从信号中恢复消息

- 解调器
- 信道译码器
- 信源译码器



图1.5 译码器的组成

§ 1.2.1 通信系统模型



信宿(destination)

信宿的功能是接收信息，包括人或设备。

信宿的研究和压缩编码结合可以在不影响视听效果的前提下，显著压缩码率。

§ 1.2.1 通信系统模型



通信系统性能指标的评价

有效性用频谱利用率来衡量。

采用信源编码以压缩码率，采用频谱利用率高的调制减小传输带宽

可靠性用传输错误率来衡量。

采用信道编码以降低错误率

安全性可通过信息的加密强度来衡量。

采用强度高的加密和伪装技术



§ 1.2.2 香农信息论



狭义信息论

即香农信息论或经典信息论。所研究的基本问题是：信源、信道及编码问题，核心是三个编码定理。

广义信息论

香农信息论（经典信息论）、信源编码、信道编码、近代信息论、统计通信理论、通信网理论、信号与信息处理、保密通信等。

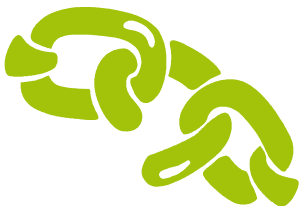


§ 1.2.2 香农信息论



1. 关于信息的度量

- 信源限制为具有某一先验概率的随机过程，所以在通信系统中各节点产生消息的实体也是随机过程
- 包括信源在内的通信系统中所有过程的输出所含信息的度量称为信息的度量
- 信息的度量也可视为是对随机变量的信息度量。

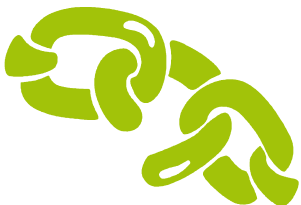


§ 1.2.2 香农信息论



1. 关于信息的度量

有两种含义不同但又密切相关的信息度量方式，一种是随机变量本身所含信息量多少的量度，另一种是随机变量之间相互提供信息量多少的量度。前者用信息熵来描述，后者用互信息来描述。



§ 1.2.2 香农信息论



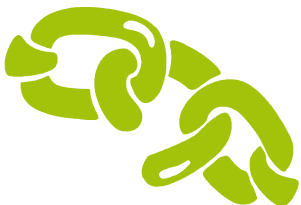
2. 无失真信源编码定理（香农第一定理）

如果编码后的信源序列信息传输速率不小于信源的熵，则可实现无失真编码，反之，不存在无失真编码。

$$R \geq H \Leftrightarrow \text{存在无失真信源编码}$$

例：ASCII码，摩尔斯电码

最优编码： Huffman编码（DATA）， L-Z通用编码算法



§ 1.2.2 香农信息论

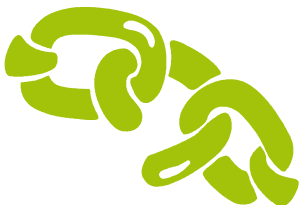


3. 关于信道容量与信息的可靠传输（香农第二定理）

如果信息传输速率小于信道容量，则总可找到一种编码方式使得当编码序列足够长时传输差错任意小，反之不存在使差错任意小的编码。

$$R \leq C \Leftrightarrow \text{存在译码差错任意小的编码}$$

其中， R 为信息传输速率，也称信道编码码率， C 为信道容量。

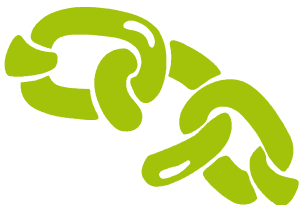


§ 1.2.2 香农信息论



3. 关于信道容量与信息的可靠传输（香农第二定理）

根据实际研究的问题建立信道模型，分析特性，计算容量，以及在这些信道条件下各类通信系统的性能。寻找接近或达到信源信道容量的信道编码技术、分析其性能并有效实现。



§ 1.2.2 香农信息论



4. 信息率失真理论（有损数据压缩的理论基础）

限失真信源编码定理（香农第三定理）

对任何失真测度 $D \geq 0$ ，只要码字足够长，总可找到一种编码，使得当编码后信息传输速率 $R \geq R(D)$ 时，码的平均失真 $d \leq D$ 。 $R(D)$ 称为信息率失真函数。

$R \geq R(D) \Leftrightarrow$ 存在平均失真 $\leq D$ 信源编码

其中， R 为信源编码码率， $R(D)$ 称为信息率失真函数，是满足平均失真 D 下，每信源符号所需最小编码比特数。

§ 1.2.2 香农信息论



香农信息论特点

- 以概率论、随机过程为基本研究工具；
- 研究通信系统的整个过程，而不是单个环节，并以编、译码器为重点；
- 关心的是最优系统的性能和怎样到达这个性能，并不具体设计系统；
- 语法信息中的概率信息，要求信源为随机过程。



§ 1.3 香农信息论研究的进展与应用



1.3.1 香农信息论创立的背景

1.3.2 香农的主要贡献

1.3.3 香农信息论研究进展

1.3.4 香农信息论的应用



§ 1.3.1 香农信息论创立的背景



1. 技术背景

通信理论与技术已有较大发展

2. 理论背景

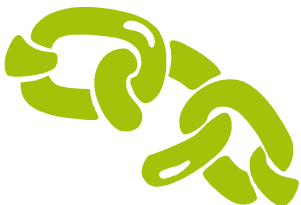
Nyquist, Hartley, Wiener, Rice做了许多有影响的
相关工作。为信息理论的建立产生了重要影响。



§ 1.3.2 香农的主要贡献



- 1948年，发表《通信的数学理论》，奠定了信息论的理论基础；
- 1956年，发表*The Zero-Error Capacity of a Noisy Channel* 《噪声信道的零差错容量》，开创了零差错容量的研究领域；
- 1959年，发表*Coding Theorem for a Discrete Source with a Fidelity Criterion* 《在保真度准则下的离散信源编码定理》，推动了信息率失真理论研究；
- 1961年，发表*Two-Way Communication Channels* （《双路通信信道》），从而开拓了多用户理论的研究。

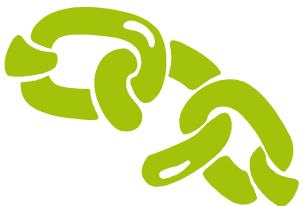


§ 1.3.3 香农信息论的研究进展



1.信息的量度

- 以香农熵为基础，研究信源信息熵的估计方法
- 将信息熵用做推断工具的最大熵原理，另外就是提出新的信息量度方法，以适应新的应用。

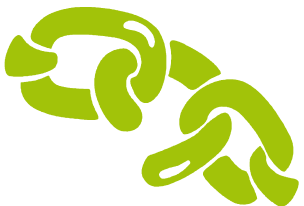


§ 1.3.3 香农信息论的研究进展



1.信息的量度

- 信息熵的估计
- 最大熵原理
- Kolmogorov熵
- Renyi熵
- Tsallis熵

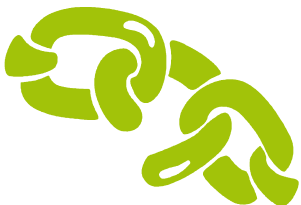


§ 1.3.3 香农信息论的研究进展



2. 无损数据压缩

- 无损数据压缩的理论基础是香农第一定理和通用信源编码的理论。通用信源编码就是对某种宽泛的一类可能信源中的每一个都渐近达到最佳性能的信源编码。



上节回顾

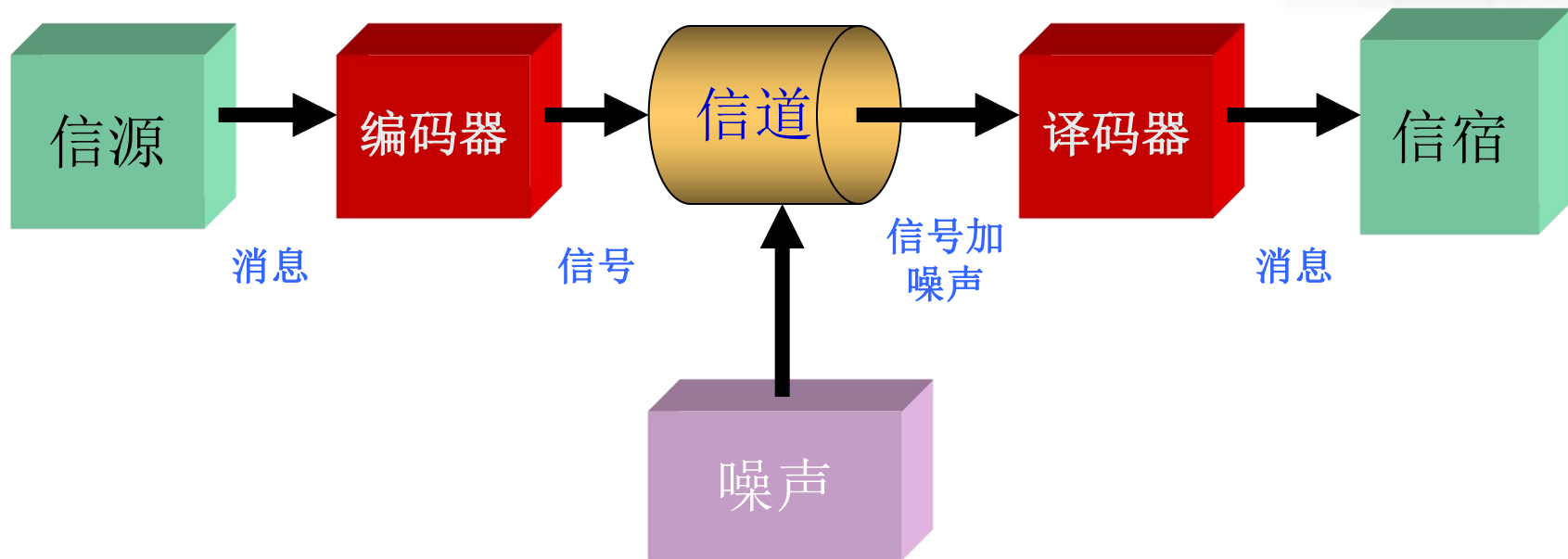
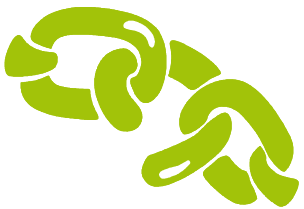


图1.2.1 通信系统模型



§ 1.3.3 香农信息论的研究进展



2.无损数据压缩

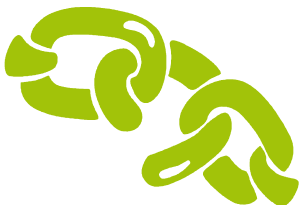
- 熵编码

- ① Huffman编码

- ② Tunstall编码

- ③ Golomb编码

- ④ 算术编码



§ 1.3.3 香农信息论的研究进展



2.无损数据压缩

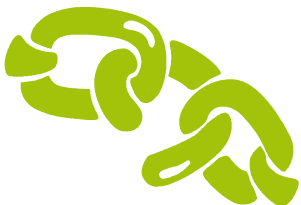
● 通用编码

① LZ算法

② Broose-Wheeler变换的压缩编码

③ 部分匹配预测（Prediction by Partial Matching, PPM）压缩编码

④ 上下文树加权编码方法



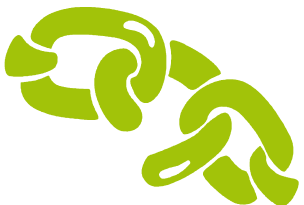
§ 1.3.3 香农信息论的研究进展



2.无损数据压缩

● 分布信源编码

D.Slepian和J.Wolf提出相关信源编码是数据压缩理论最重要的进展之一。他们证明，即使对两个相关信源分别编码但联合译码，所需总的码率也可以小于两个信源熵的和。此外具有边信息的信源编码和数据压缩的研究还在开展。



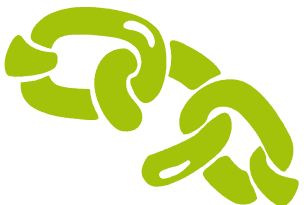
§ 1.3.3 香农信息论的研究进展



3. 可靠通信

- 多用户信道的研究
- 各种特殊信道容量与有关编译码技术研究

主要包括：多接入信道、无记忆高斯多接入信道、干扰信道、广播信道、未知参数信道、窃听信道等，现已构成信息论中一重要分支网络信息论。



§ 1.3.3 香农信息论的研究进展



3.可靠通信

- 各种特殊信道容量与有关编译码技术研究

主要包括：高斯信道、衰落信道、反馈信道、迭代译码、有约束信道和零差错信道等

- 多入多出（MIMO）信道与时空编码的研究

Paulraj、Telatar、Foschini等人的工作奠定了时空信道研究的理论基础

Tarokh提出了时空格码技术，推动了时空编码的研究



§ 1.3.3 香农信息论的研究进展



3.可靠通信

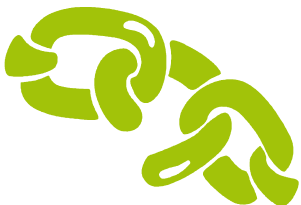
- 编译码理论的研究

- ①代数编码理论

- ②BCH、RS码、卷积码、级联码以及TCM编码

- ③Turbo码LDPC码

- 信道编码定理的研究

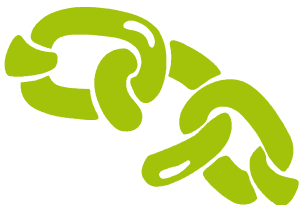


§ 1.3.3 香农信息论的研究进展



4.有损数据压缩

- ①Berger给出了更一般的率失真编码定理
- ②Wyner和Ziv研究了在译码器具有边信息的有损信源编码问题，推动了多端有损压缩编码的研究
- ③Hajek和Berger建立了随机场率失真编码理论的基础

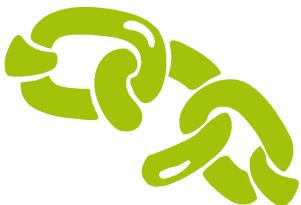


§ 1.3.3 香农信息论的研究进展



4.有损数据压缩

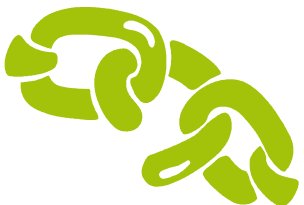
- ④ Yang和Kieffer证明，对于信源统计特性和失真测度都通用的有损压缩编码是存在的
- ⑤ 信源和信道联合编码
- ⑥ 实现有损压缩的技术得到很大发展



§ 1.3.4 香农信息论的应用



- ① 信息论提出了评估通信系统性能优劣的基本标准
- ② 信息论提供了从全局的观点观察和设计通信系统的理论方法
- ③ 信息论是从事与信息处理有关的研究、开发与管理人员的必备知识
- ④ 信息论在其他领域的应用已获得很大的成功



谢谢大家!

