



信息论与编码

徐冠军

gjxu@ee.ecnu.edu.cn

课程介绍

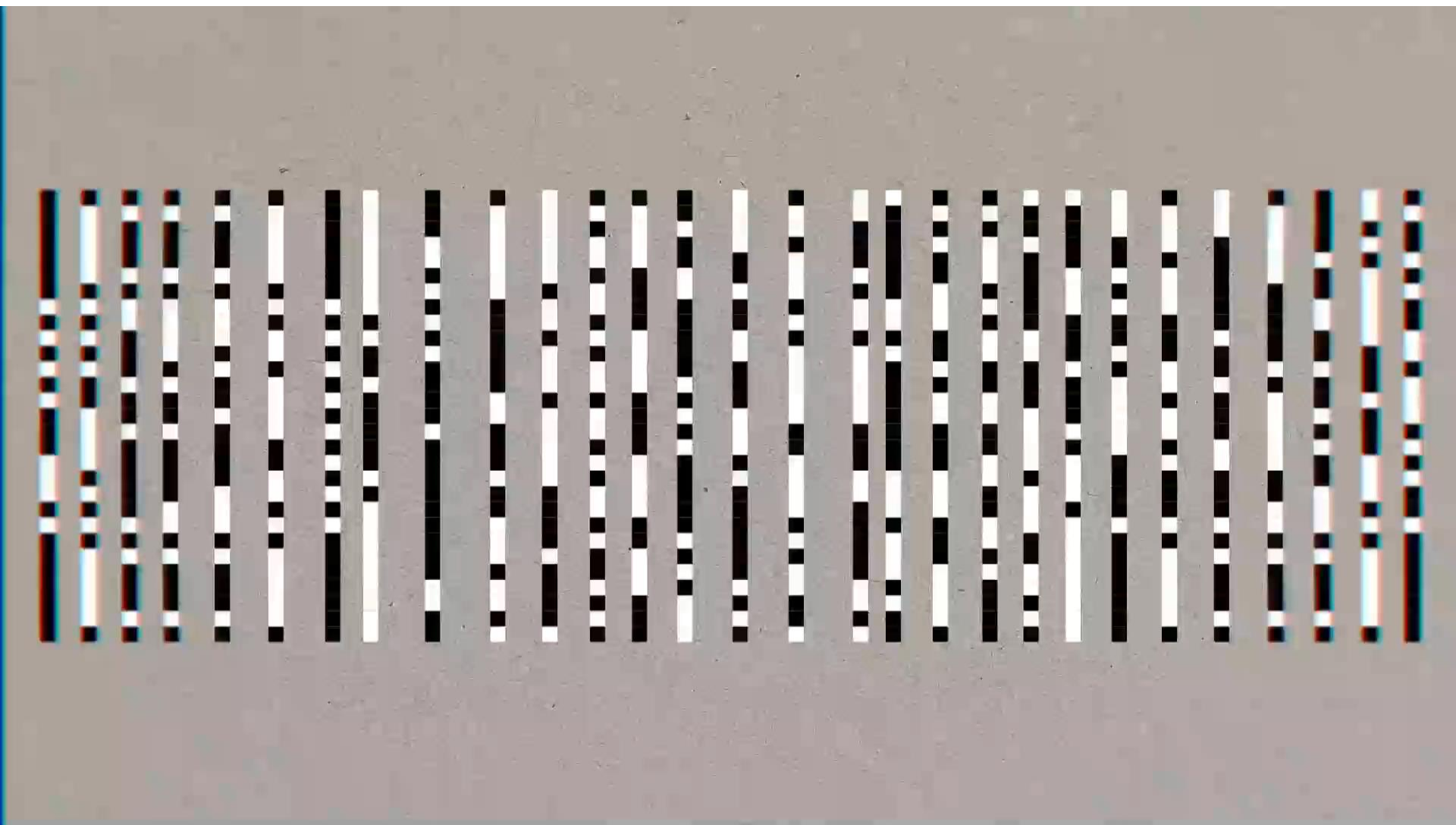
教材

- ◆ 曹雪虹、张宗橙, **《信息论与编码》**, 清华大学出版社



参考书

- 周荫清, **《信息理论基础 (第4版)》**, 北京航空航天大学出版社
- 焦瑞莉, **《信息论基础教程》**, 机械工业出版社
- 岳殿武, **《信息论与编码 (第2版)》**, 清华大学出版社



- 信息论的**基本概念**
- 数字通信系统的**模型**
- 信息论与编码**发展简史**
- 信息论与编码**主要内容和应用**

□ 信息论的基本概念

信息科学是研究信息**产生**、**获取**、**传输**、**存贮**以及信息**识别**和**利用**等等有关问题的一门科学。研究的范围极其广泛，包括了许多看来似乎极不相关的领域

信息论是信息科学的理论基础

一般，关于信息论的研究有三种理解：

狭义信息论

广义信息论

一般信息论

第一章 绪论

□ 信息论的基本概念

狭义信息论

➔ **香农信息论**，是应用近代数理统计方法来研究信息传输和处理的基本理论。主要讨论**信息的度量**，**信源和信道的特性**以及**有关编码**的问题

广义信息论

主要研究**通信理论问题**，包括噪声理论、信号检测和滤波理论、调制理论以及信息处理等理论问题

一般信息论

通称为信息科学，以**香农信息论**和**维纳控制理论**作为理论基础，运用信息概念和信息方法来研究自然界和人类社会有关的信息问题的理论

□ 信息论的基本概念

本课程所研究的范畴是**狭义信息论**，即**香农信息论**

香农信息论的最大特点是：

- ✓ 将概率统计的观点和方法引入通信理论的研究中，揭示通信系统中传输的对象是信息
- ✓ 并对信息给出了科学的、定量的描述
- ✓ 指出通信系统设计的中心问题是在随机噪声的干扰下如何有效而可靠的传输信息

□ 信息论的基本概念

- ◆ 花朵开放时的**色彩**是一种信息，它可以引来昆虫为其授粉 **色彩** ← 视觉
- ◆ 成熟的水果会产生**香味**，诱来动物，动物食后为其传播种子，果香也是一种信息 **果香** ← 嗅觉
- ◆ 药有苦味，让人难以吞咽，**药味**是一种信息 **药苦** ← 味觉
- ◆ 听老师讲课可以得到相应的知识，**知识**也是信息 **知识** ← 听视觉
- ◆ 听老师讲课可以得到相应的知识，**知识**也是信息…… **冷热** ← 触觉

信息处处存在，人们的眼、耳、鼻、舌、身体都能**感知**信息

□ 信息论的基本概念

关于信息的科学定义，国内外有许多流行的说法，它们从不同的侧面和不同的层次来揭示信息的本质

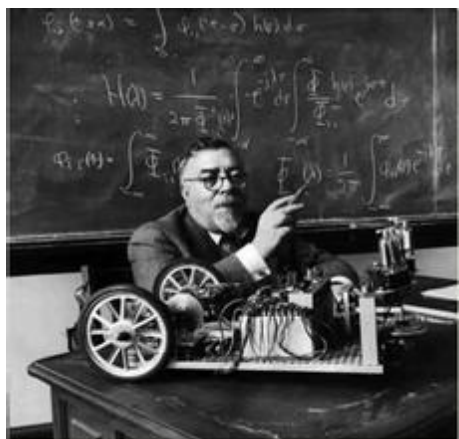
- 1928年,美国数学家**哈特莱 (Hartley)** 在《贝尔系统电话杂志》上发表了一篇题为《信息传输》的论文
“信息是选择的自由度”



□ 信息论的基本概念

关于信息的科学定义，国内外有许多流行的说法，它们从不同的侧面和不同的层次来揭示信息的本质

- 美国数学家、控制论的主要奠基人维纳(Wiener)在1950年出版的《控制论与社会》中写到：



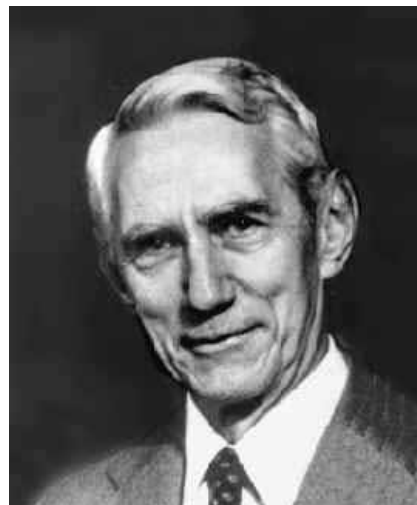
“信息既不是物质又不是能量，信息就是信息”
——揭示了信息的特质：即**信息**是独立于**物质**和**能量**之外存在于客观世界的**第三要素**

□ 信息论的基本概念

关于信息的科学定义，国内外有许多流行的说法，它们从不同的侧面和不同的层次来揭示信息的本质

- 1948年，香农(C.E. Shannon)发表了一篇著名的论文，“通信的数学理论”。他从研究通信系统传输的实质出发，对信息作了科学的定义，并进行了定性和定量的描述

信息是事物运动状态或存在方式的不确定性的描述



□ 信息论的基本概念

信息 ——事物运动状态或存在方式的不确定性的描述

消息 ——用文字、符号、数据、语言、音符、图片、图像等能够被人们感觉器官所感知的形式，把客观物质运动和主观思维活动的状态表达出来就成为消息

信号 ——把消息换成适合信道传输的物理量（如电信号、光信号、声信号、生物信号等），这种物理量称为信号

□ 信息论的基本概念

- “信息” 不同于消息，消息是表现形式，信息是实质
- “信息” 不同于情报，情报的含义较窄，一般只限于特殊的领域，是一类特殊的信息
- “信息” 不同于知识，知识是人们根据某种目的，从自然界收集得来的数据中整理、概括、提取得到的有价值的信息，是一种高层次的信息

□ 信息论的基本概念



- 微信内容首先被手机编码成无线电波发送出去, 该无线电波是内容的载体, 是实际存在的无线电“**信号**”
- 熊猫的手机接收到信号之后在屏幕上显示“Trump禁止TikTok与甲骨文交易”, 这是一条“**消息**”
- 种花家看到消息之后, 会在大脑中形成自己的理解, 有自己的感受, 这是“**信息**”

□ 信息论的基本概念

信息具有以下**特征**：可**识别**、可**存贮**、可**传递**、可**量度**、可**加工**、可**共享**、信息的载体是可以**转换**

信息的
独有性质

信息在信息化程度越来越高的社会中将起到越来越重要的作用，是比物质和能量更为宝贵的**资源**，全面掌握信息的概念，正确、及时、有效地利用信息，能够为人类创造更多的**财富**

□ 信息论的基本概念

信息的测量：信息量与不确定性消除的程度有关

从数学的角度，不确定就是随机性，可运用研究随机事件的数学工具——概率论和随机过程来测度不确定性的大小

某一事物状态的不确定性的大小，与该事物可能出现的不同状态数目以及各状态出现的概率大小有关

既然不确定性的大小能够度量，所以信息是可以测量的

□ 信息论的基本概念

✓ 信息的直观认识1

信道上传送的是随机变量的值

即，我们在收到消息之前，并不知道消息的内容
否则消息就没有必要发送

消息随机变量有一个概率分布

消息随机变量的一个可能取值就称为一个事件

□ 信息论的基本概念

✓ 信息的直观认识2

消息随机变量的随机性越大，此消息随机变量含有的信息量就越大

例：消息随机变量 X = “中国西部战区做好应战印军挑衅行为” 则消息随机变量 X 含有的信息量小

例：消息随机变量 Y = “中国西部战区痛击印军挑衅”
则消息随机变量 Y 含有的信息量大

□ 信息论的基本概念

✓ 信息的直观认识3

两个消息随机变量的相互依赖性越大，它们的互信息量就越大

例：假设 X =厦门十二月份日平均气温， Y =福州十二月份日平均气温， Z =北京十二月份日平均气温， W =纽约十二月份日平均气温

则： X 与 Y 互信息量大；
 X 与 Z 互信息量较小；
 X 与 W 互信息量更小。

□ 信息论的基本概念

✓ 信息的直观认识4

例1：甲袋红、白球各50个，乙袋红、白、蓝、黑球各25个。比较从甲袋中取出一个球是红球的事件和从乙袋中取出一个球是红球的事件发生的难易程度，也就是事件发生的不确定性

例2：本地十二月份可能出现的天气包括：晴、阴、雨、雪。比较天气预报为“晴”和天气预报为“雪”，给人们带来的信息量

□ 信息论的基本概念

样本空间：所有可能选择的消息的集合（某事物所有可能出现的状态）

概率空间：一个样本空间和它的概率测度

可用 $[X, P]$ 来表示
$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & \dots, & a_q \\ p(a_1), & p(a_2), & \dots, & p(a_q) \end{bmatrix}$$

□ 信息论的基本概念

先验概率：选择符号 a_i 作为消息的概率---- $p(a_i)$

自信息： a_i 本身携带的信息量（不确定性） $I(a_i) = \log \frac{1}{p(a_i)}$

后验概率：接收端收到消息(符号) b_j ，而发送端发的是 a_i 的概率 $p(a_i/b_j)$

互信息：收信者获得的信息量-----先验的不确定性减去尚存在的不确定性

$$I(a_i; b_j) = \log \frac{1}{p(a_i)} - \log \frac{1}{p(a_i / b_j)}$$

□ 信息论的基本概念

信息具有：

- **可靠性**：使信源发出的消息经过信道传输以后，尽可能准确、不失真的再现在接收端
- **有效性**：用尽可能短的时间和尽可能少的设备来传送一定数量的信息
- **保密性**：隐蔽和保护通信系统中传送的消息，使它只能被授权接收者接收，而不能被未授权接收者接收和理解
- **认证性**：接收者能正确判断所接收的消息的正确性，验证消息的完整性，而不是伪造和篡改的

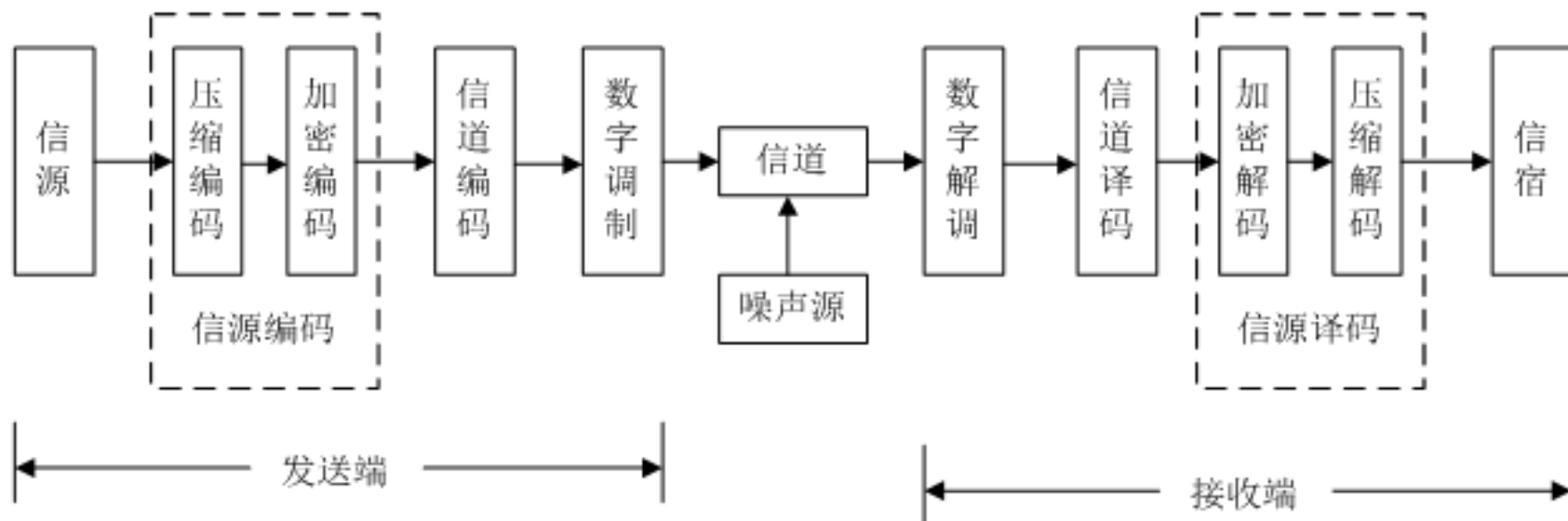
□ 信息论的基本概念

信息的特征：

- ✓ 信息是新知识、新内容
- ✓ 信息是能使认识主体对某一事物的未知性或不确定性减少的有用知识
- ✓ 信息可以产生，也可以消失，同时信息也可以被携带、储存及处理
- ✓ 信息可以度量

□ 数字通信系统的模型

数字通信系统是利用数字信号来传递信息的通信系统。数字通信涉及的技术问题很多，其中主要有信源编码/译码、信道编码/译码、数字调制/解调、数字复接、同步以及加密等



□ 数字通信系统的模型

在通信系统模型中

信源的核心问题是它所包含的信息有多少，如何定量表示

信宿问题是能收到或提取多少信息量

信道的问题，主要是它最多能传送多少信息的问题，也就是信道容量的问题

编译码的问题，主要目标是用尽可能小的平均码长来代表要传输的消息

□ 数字通信系统的模型

1. 信源编码与译码

信源编码的实质就是为了**去掉信源中的信息冗余**

信源编码的作用是数据压缩和模数转换

数据压缩——设法减少码元数目和降低码元速率。码元速率将直接影响传输所占的带宽，而传输带宽直接反映了通信的有效性

模数转换——当信息源给出的是模拟语音信号时，信源编码器将其转换成数字信号，以实现模拟信号的数字化传输

信源译码是信源编码的逆过程

□ 数字通信系统的模型

2. 信道编码与译码

数字信号在信道传输时，由于噪声、衰落以及人为干扰等，将会引起差错。为了减少差错，信道编码器对传输的信息码元按一定的规则加入保护成分（监督元），组成所谓“抗干扰编码”

接收端的信道译码器按一定规则进行解码，从解码过程中发现错误或纠正错误，从而提高通信系统抗干扰能力，实现可靠通信

□ 数字通信系统的模型

3. 加密与解密

在需要实现保密通信的场合，为了保证所传信息的安全，人为将被传输的数字序列扰乱，即加上密码，这种处理过程叫加密

在接收端利用与发送端相同的密码复制品对收到的数字序列进行解密，恢复原来信息，叫解密

□ 数字通信系统的模型

4. 数字调制与解调

数字调制就是把数字基带信号的频谱搬移到高频处，形成适合在信道中传输的频带信号。基本的数字调制方式有振幅键控ASK、频移键控FSK、绝对相移键控PSK、相对（差分）相移键控DPSK。对这些信号可以采用相干解调或非相干解调还原为数字基带信号。对高斯噪声下的信号检测，一般用相关器接收机或匹配滤波器实现。

□ 数字通信系统的模型

5. 同步与数字复接

同步是保证数字通信系统有序、准确、可靠工作的不可缺少的前提条件。同步是使收、发两端的信号在时间上保持步调一致。按照同步的功用不同，可分为载波同步、码元(位)同步、群同步和网同步，这些问题在《通信原理》课程中已讨论

数字复接就是依据时分复用基本原理把若干个低速数字信号合并成一个高速的数字信号，以扩大传输容量和提高传输效率

□ 数字通信系统的模型

数字通信的优点

- ✓ 抗干扰能力强，可消除噪声积累
- ✓ 差错可控，传输性能好。可以采用信道编码技术使误码率降低，提高传输的可靠性
- ✓ 易于与各种数字终端接口，用现代计算技术对信号进行处理、加工、变换、存储，从而形成智能网
- ✓ 易于集成化，从而使通信设备微型化
- ✓ 易于加密处理，且保密强度高

□ 数字通信系统的模型

数字通信的缺点

- ✓ 频带利用率不高——数字通信的许多优点都是用比模拟通信占据更宽的系统频带为代价而换取的。以电话为例，一路模拟电话通常只占据4 kHz带宽，但一路接近同样话音质量的数字电话可能要占据20~60 kHz的带宽，因此数字通信的频带利用率不高
- ✓ 对同步要求高，系统设备比较复杂

第一章 绪论

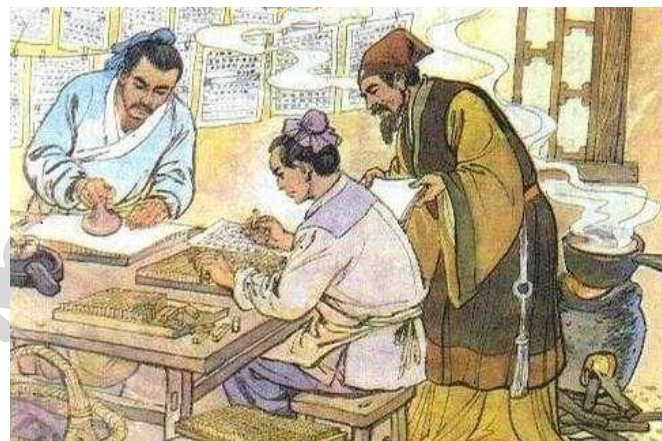
信息论与编码发展简史



最早的快速、远距离传递信息的方式



7. 到了西汉武帝时期，就已经有了纸。纸原料主要是大麻，掺有少量苧麻。纸在羊括切断植物纤维、煮煮、舂捣及抄造等熬纸的质地比较粗糙，表面也不够平。上最早的以植物纤维为原料的纸。



信息的表示和存储方式产生了一次重大的变化

没有物质的世界是虚无的世界，没有能源的世界是死寂的世界，没有信息的世界是混乱的世界

□信息论与编码发展简史

1. 信息论的提出

- 17世纪到19世纪，美国物理学家吉布斯(*Gibbs, Josiah Willard*)和奥地利物理学家波尔兹曼(*Boltzmann, Ludwig*)首先把统计学引入物理领域，为信息论的诞生做出了贡献。这种探究方法为信息理论的创立提供了方法论的前提
- 1832年莫尔斯电报系统中高效率编码方法对后来香农的编码理论是有启发的
- 1885年凯尔文(*L Kelvin*)曾经研究过一条电缆的极限传信串问题
- 1922年卡松 (*J.R.Carson*) 提出边带理论，指明信号在调制（编码）与传送过程中与频谱宽度的关系

□ 信息论与编码发展简史

1. 信息论的提出

- 1924年奈奎斯特(H.Nyquist)解释了信号带宽和信息速率之间的关系；1924年奈奎斯特发表了《影响电报速度的某些因素》一文，指出了电信信号的传输速率与信道频带宽度之间存在着确定的比例关系
- 1928年哈特莱(Ralph Vinton Lyon Hartley, 1888-1970)给出了信息度量方法；1928年，哈特莱在《信息传输》(Transmission of Information)一文中指出，信息是包含在消息（讯息）中的抽象量，消息是信息的载荷者

□信息论与编码发展简史

1. 信息论的提出

- 1936年Armstrong提出增大带宽可加强抗干扰能力
- 1939年H. Dudley发明了声码器(vocoder), 提出了通信所需要的带宽至少应与所传送的消息的带宽相同。达德利和莫尔斯都是研究信源编码的先驱者
- 1943年Norbert Wiener与别格罗和罗森勃吕特合写了《行为、目的和目的论》的论文, 从反馈角度研究了有目的性的行为, 找出了神经系统和自动机之间的一致性

□信息论与编码发展简史

1. 信息论的提出

- Shannon于1941~1944年对通信和密码进行了深入的研究，利用概率论和数理统计的方法研究通信系统，得出了几个重要具有普遍意义的结论，奠定了现代信息论的基础，通常称为香农信息论
- 1948年和1949年，香农在贝尔系统技术杂志上分别发表了两篇论文，即Mathematical Theory of Communication和“在噪声中的通信”两篇文章，他提出了信息量的概念和信息熵的计算方法，并因此被视为现代信息论的创始人

□信息论与编码发展简史

2. 信息论与编码的发展

- 1948年香农在论文中提出了无失真信源编码定理，给出了简单的编码方法——香农码
- 1949年，香农发表了论文“保密通信的信息理论”，首先用信息论的观点对信息保密问题作了全面论述
- 1950年，汉明码被提出
- 1951年，Institute of Radio Engineers成立信息论组，1955年正式出版了信息论汇刊
- 1952年R.M. Fano给出并证明了费诺不等式，并给出了关于香农信道编码逆定理的证明。同年，D.A. Huffman构造了著名的霍夫曼编码，并证明了它是最佳码

□信息论与编码发展简史

2. 信息论与编码的发展

- 1957年，沃尔夫维兹采用了类似典型序列方法证明了信道编码逆定理
- 1959年，香农发表 Coding Theorems for a Discrete Source with a Fidelity Criterion，系统地提出了信息率失真理论和限失真信源编码定理，为各种信源编码的研究奠定了基础
- 1960年，卷积码的概率译码被提出
- 1961年，香农的Two-way Communication Channels论文开拓了多用户信息理论（网络信息论）的研究

信息论与编码发展简史

2. 信息论与编码的发展

- 1961年，费诺描述了分组码中码率、码长和错误概率的关系，并提出了香农信道编码定理的充要性证明
- 1968年，P. Elias发展了香农-费诺码，提出了算术编码的初步思想
- 1969年，M.S.Pinsker提出具有反馈的非白噪声高斯信道容量
- 1971年，T.Berger给出更一般的信源率失真编码定理
- 1971年，R. Ahlswede和1972年H. Liao找出了多元接入信道的信道容量区

□信息论与编码发展简史

2. 信息论与编码的发展

- 1972年，Cover发表了有关广播信道的研究，以后陆续进行了有关多接入信道和广播信道模型和信道容量的研究
- 1972年，S. Arimoto和R. Blahut分别发展了信道容量的迭代算法
- 1976年，J. Rissanen给出和发展了算术编码
- 1976年，迪弗和海尔曼发表了论文“密码学的新方向”，提出了公钥密码体制
- 1983年，Cover和R. Ahlswede分别发表文章讨论相关信源在多元接入信道的传输问题

□信息论与编码发展简史

2. 信息论与编码的发展

- 1990年，贝尔等在L-Z算法的基础上又作出了一系列变化和改进。目前，L-Z编码（通用信源算法）已广泛用于计算机数据压缩中，如UNIX中的压缩算法就采用了L-Z编码算法
- 至此，香农理论已成为真正具有实用意义的科学理论
- 信息论形成与发展的三大定理：无失真信源编码定理（第一极限定理），信道编码定理（第二极限定理），限失真信源编码定理（第三极限定理）

□信息论与编码主要内容和应用

信息论研究的**目的**：**高效、可靠、安全**并且**自由交换**和**利用**各种各样的信息

信息论与编码研究的**主要内容**包括以下内容：

1. **通信的统计理论**的研究
2. **信源**的统计特性
3. **编码理论与技术**的研究
4. **提高信息传输效率**的研究
5. **抗干扰理论**与技术的**研究**
6. 噪声中**信号检测理论**与技术的**研究**

□信息论与编码主要内容和应用

本门课主要研究以下内容：

- (1) 信源与信息熵；
- (2) 信道与信道容量；
- (3) 信息率失真函数；
- (4) 信源编码；
- (5) 信道编码；
- (6) 加密编码；

□信息论与编码主要内容和应用

信息论与编码从它诞生的那时起就吸引了众多领域学者的注意，他们竞相应用信息论与编码的概念和方法去理解和解决本领域的问题。例如，信息论与编码在**生物学、医学、经济、管理、图书情报**等领域都有不同程度的应用

作业：

1. 教材思考题1~4题
2. 预习第二章2.1小节

第一章 绪论

考试安排：

平时成绩(课堂测试+期中考试) $\times$ 系数 ξ + **考试成绩** $\times$ 系数 $(1-\xi)$