



XI'AN JIAOTONG UNIVERSITY

# 信息论与编码

---

## 第1章 信息与信息的度量

张建国





XI'AN JIAOTONG UNIVERSITY

## 主要内容与基本要求

- > 主要内容
  - 信息的定义、特点;
  - 信息的度量: 信息量与信息熵。
- > 基本要求
  - 了解信息论的产生及发展;
  - 理解信息的概念和含义;
  - 掌握信息的度量方法——信息量, 信息熵。

2013-3-9 《信息论与编码》——信息与信息的度量 2





XI'AN JIAOTONG UNIVERSITY

## 本章目录

- 1.1 引言
  - 1.1.1 信息论与信息简介
  - 1.1.2 通信系统模型
- 1.2 信息的度量
  - 1.2.1 信息量
  - 1.2.2 信息熵
- 1.3 本章小结

2013-3-9 《信息论与编码》——信息与信息的度量 3

XI'AN JIAOTONG UNIVERSITY

## 1.1 引言

### 1.1.1 信息论与信息

人类社会的生存和发展, 无时无刻都离不开对信息的获取、传递、处理、控制和利用。

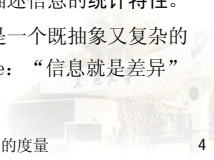
信息论是一门应用概率论、随机过程和数理统计等方法来研究信息的存储、传输和处理中一般规律的学科。它主要研究如何提高信息系统的可靠性、有效性、保密性和认证性, 以使信息系统最优化。从通信的角度研究信息。


信息论是研究信息的科学和理论, 它定量地描述信息的统计特性。

信息是信息论中最基本、最重要的概念, 它是一个既抽象又复杂的概念。Hartley: “信息是选择的自由度”; Longe: “信息就是差异”

从信息的近亲词“消息”出发来讨论信息。

2013-3-9 《信息论与编码》——信息与信息的度量 4




  
 XI'AN JIAOTONG UNIVERSITY

## 1.1 引言

信息是对**认识**在传输过程中的**特性**的度量。


认识（或认知）是人脑对客观事物的反映，它通过消息的形式进行传递。或者换句话说，消息是认知在传递过程中的表现形式，是传递中的知识。常见的三种形式：声音、文字、图像。

消息传递的特点：

- 消息知识量——固定；
- 目的——使接收者对消息完全知晓；
- 传递信息量——随接收者而异。

消息与信息区别：信息不能等于消息。消息中包含信息，是信息的载体。同一则信息可用不同的消息形式来载荷。**消息的知识量是固定的，但消息中所含的信息量与接收者有关。**

2013-3-9
《信息论与编码》——信息与信息的度量
5


  
 XI'AN JIAOTONG UNIVERSITY

## 1.1 引言

实际需要传输的，只是为使接收者达到确知水平所需的那部分知识差额。如何对这种差额进行度量，对它进行高效、可靠地传递是香农信息论中重点要研究的。而信息正是对这一差额的定量描述。

信息是对在传消息中接收者尚未了解的（不知道的、不掌握的）那部分知识的度量。


信息是对传输过程前后接收者知识变化的度量。

信息是差的概念，无法定义绝对值。

信息的特点：

- 与接收者对这一消息的先验知识水平有关；
- 与接收者对这一消息的主观感受及兴趣无关；
- 与消息的具体含义和重要性无关。

2013-3-9
《信息论与编码》——信息与信息的度量
6


  
 XI'AN JIAOTONG UNIVERSITY

## 1.1 引言

### 1.1.2 通信系统模型


通信中常见传输模型

消息传递过程

产生消息 → 提取信息 → 传输信息 → 恢复信息 → 吸收消息

信息论研究内容

2013-3-9
《信息论与编码》——信息与信息的度量
7


  
 XI'AN JIAOTONG UNIVERSITY

## 1.1 引言

### 信源编码

目的是用二进制（或多进制）数字序列表示信源输出，所关注的主要问题之一是对给定信源模型的输出，确定在单位时间需要用多少二进制位表示。**高效地提取信息。**


常见的信源编码

- 文字 ASCII, GB2312, GBK
- 声音 PCM编码, MP3 (MPEG-1 Audio Layer 3)
- 图像 BMP, JPG, AVI (DivX, Xvid, H.264), RMVB

### 信道编码

信道编解码器的目的是在信道解码器的输出端可靠地重现二进制数字序列，关注的主要问题之一是能否以及如何实现这一点。**可靠地传输。**

2013-3-9
《信息论与编码》——信息与信息的度量
8




XI'AN JIAOTONG UNIVERSITY

## 1.1 引言

- > 将编解码分为信源编解码和信道编解码
  - 这是否对通信系统施加了基本限制?
  - 从实用观点: 非常方便, 利于不同信源使用相同信道或相同的信源使用不同的信道。
- > 信息论与编码的关系
 

信息论为编码提供了理论基础, 编码是信息论的应用
- > 编码历史上的几次大突破
  - 1950 Hamming的分组码
  - 1967 Viterbi的卷积码译码算法
  - 1993 Turbo码的发现和Turbo译码算法

2013-3-9
《信息论与编码》——信息与信息的度量
9



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

### 1.2.1 信息量的定义

信息只有在传输过程中才有意义。(狭义信息论)

通信过程是一种消除(收方)不确定性的过程。信息量与不确定性消除的程度有关, 消除了多少不确定性, 就获得了多少信息量。信息同接收者的先验知识有关。同样的消息, 可能对应着不同的信息量。如何来衡量这种不确定性呢?


概率表示了某一事物出现的几率或可能性, 它反映了人们对某事件的认识水平。因此可以用概率度量这种不确定性。

信息量(用  $I_x$  表示)是一个与概率有关的函数。

$$I_x \triangleq f(P_x)$$

信息论被认为是概率论的分支。

2013-3-9
《信息论与编码》——信息与信息的度量
10



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

根据前面对信息的讨论, 该函数  $f(\bullet)$  应满足以下三个条件:

- $f(1)=0$  (收方) 已知事件的信息量为0;
- $f(P_x)$  是  $P_x$  的减函数;
- 如果消息  $z$  由两条独立消息  $x$  和  $y$  组成, 则  $I_z = I_x + I_y$ 。

条件三要求:

$$f(P_z) = f(P_x) + f(P_y)$$

又因为  $x$  和  $y$  独立, 则:


$$P_z = P_x P_y$$

所以函数必须满足:

$$f(P_x P_y) = f(P_x) + f(P_y)$$

满足乘积的函数等于函数之和的初等函数只有对数函数。

2013-3-9
《信息论与编码》——信息与信息的度量
11



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

再考虑到前两个条件, 得到如下信息量的定义:

$$I_x \triangleq \log_a \frac{1}{P_x} = -\log_a P_x$$


底  $a$  可以取任意大于1的数, 并不影响使用, 通常取以下两种:

- $a = 2$ , 信息量的单位为比特 (bit)。
- $a = e$ , 信息量的单位为奈特 (nat)。

信息量的性质:

- $I_x \geq 0$ , 当且仅当  $P_x = 1$  时等号成立;
- $I_x > I_y$ , 当  $P_x < P_y$  时;
- $I_{x+y} = I_x + I_y$ , 当  $x$ 、 $y$  独立时。

2013-3-9
《信息论与编码》——信息与信息的度量
12



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

一条消息所含的信息量仅与它的先验概率有关，定义为  $-\log_a P_x$ 。


该信息量定义（技术定义）的优点：（傅，信息论：基础理论与应用）

- 它是一个科学的定义，有明确的数学模型和定量计算；
- 它与日常用语中的信息的含义是一致的；
- 定义排除了对信息一词某些主观上的含意。因此，信息的概念是纯粹的形式化的概念。

该定义的限制性

- 定义的出发点是假定事物状态可以用一个以经典集合论为基础的概率模型来描述；
- 定义和度量没有考虑收信者的主观特性和主观意义，也撇开了信息的具体含义、具体用途、重要程度和引起后果等因素。

2013-3-9
《信息论与编码》——信息与信息的度量
13



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

### 1.2.2 信息熵


信息量是随机事件的概率的函数。一个事件（信源的输出）有多种可能的结果，每种可能的结果都有自己的信息量。因此，信息量是一个随机变量。可以用其一阶统计量（均值）作为整个信源的信息测度。通常把这个均值称为**信息熵**，其本质是**平均信息量**。

设一消息源发出消息 $X$ ，以概率 $P_{x_i}$ 分别取 $x_i, (i=1, \dots, N)$ ，则 $X$ 所含的信息熵（平均信息量）定义为：

$$H(X) \triangleq E[I_{x_i}] = \sum_{i=1}^N p_{x_i} \log \frac{1}{p_{x_i}}$$

$$= \sum_x p(x) \log \frac{1}{p(x)}$$

2013-3-9
《信息论与编码》——信息与信息的度量
14



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

对于特定的信源，给定其概率空间，信息熵（平均信息量）就是一个确定的数值。不同的信源因统计特性不同，其熵也不同。

如：有两个信源，其概率空间分别为：

$$\begin{bmatrix} x \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0.99 & 0.01 \end{bmatrix}, \quad \begin{bmatrix} y \\ p(y) \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ 0.5 & 0.5 \end{bmatrix}$$

根据定义计算可得：

$$H(X) = -0.99 \log 0.99 - 0.01 \log 0.01 = 0.08(\text{bit/符号})$$


$$H(Y) = -0.5 \log 0.5 - 0.5 \log 0.5 = 1(\text{bit/符号})$$

$$H(Y) > H(X)$$

所以信源 $Y$ 比信源 $X$ 的平均不确定性要大。

熵描述了一个系统的混乱程度。

2013-3-9
《信息论与编码》——信息与信息的度量
15



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

**信息熵的性质1:** 设一消息源 $X$ 可取 $K$ 个状态或值，当各个状态或值有相同的概率（即等概）时，信源的信息熵最大。等价于下述定理。

**定理1.2.1:** 令 $X$ 为一集合，是含有 $K$ 个元素的样本空间，则 $H(X) \leq \log K$ 。当且仅当所有元素等概时，上式取等号。

证明：先证明一个引理，

$$\ln z \leq z - 1 \quad (z > 0), \text{ 等号当且仅当 } z = 1 \text{ 时成立。}$$

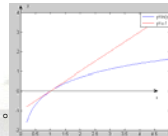
构造函数： $f(z) = \ln z - (z - 1)$ ，显然只需证明 $f(z) \leq 0$ 即可。

求其一、二阶导数，有：


$$f'(z) = 1/z - 1 \quad f''(z) = -1/z^2 < 0$$

由微积分知识可知：函数在 $z = 1$ 处取得最大值。

代入函数得最大值为0。所以： $f(z) \leq 0$ ，引理得证。



2013-3-9
《信息论与编码》——信息与信息的度量
16



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

再证:  $H(X) \leq \log K$  或  $H(X) - \log K \leq 0$       $\log_a x = \log_a e \cdot \ln x$

$$H(X) - \log K = \sum_x p(x) \log \frac{1}{p(x)} - \sum_x p(x) \log K$$


$$= \sum_x p(x) \log \frac{1}{Kp(x)} = (\log e) \sum_x p(x) \ln \frac{1}{Kp(x)}$$

$$\leq (\log e) \sum_x p(x) \left[ \frac{1}{Kp(x)} - 1 \right] \quad p(x) > 0$$

$$= (\log e) \left[ \sum_x \frac{1}{K} - \sum_x p(x) \right] = (\log e)(1-1) = 0$$

当且仅当  $Kp(x) = 1$  时取等号, 即  $p(x) = 1/K$ , 即所有元素等概。 #

2013-3-9     《信息论与编码》——信息与信息的度量     17



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

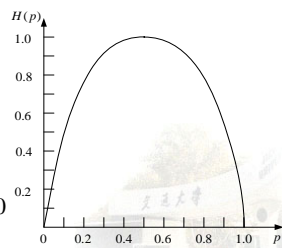
显然,  $K$  至少为2。当  $K$  取2时, 可以得到二元变量信息熵。

设  $P(0) = p, P(1) = 1 - p$


$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \leq \log 2$$

可以看出它是  $p$  的一元函数, 常被称为“马蹄”函数, 记为  $\Omega(p)$ 。

$$\lim_{p \rightarrow 0} \Omega(p) = \lim_{p \rightarrow 1} \Omega(p) = 0$$

$$\lim_{p \rightarrow 0} \Omega(p) = \lim_{p \rightarrow 0} p \log \frac{1}{p} = \log e \lim_{p \rightarrow 0} p \ln \frac{1}{p} = 0$$


2013-3-9     《信息论与编码》——信息与信息的度量     18



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

**信息的单位:** 定义一个二进制符号所携带的最大信息熵为一个信息单位, 即  $\log 2$ 。

对数底取2时,  $\log_2 2 = 1$  比特, bit (binary digit);

对数底取e时,  $\log_e 2 = \ln 2 = 0.693$  奈特, nat (natural digit)。

对数底取10时,  $\log_{10} 2 = 0.301$  哈特, hart (Hartley)


单位换算: 1bit=0.693nat; 1nat=1.443bit; 1hart=3.322bit。

**信息熵的性质2:** 一个随机变量加上一个常量所得到的新变量, 所含的信息熵与原变量所含的信息熵相同。

数学表述:  $H(V+C) = H(V)$

$$H(VC) = H(V) + H(C) = H(V) + 0 = H(V)$$

2013-3-9     《信息论与编码》——信息与信息的度量     19



XI'AN JIAOTONG UNIVERSITY

## 1.2 信息的度量

**非负性,** (离散信源)

**对称性,** 仅与信源的统计特性有关, 与事件本身的具体含义无关。

$$\begin{bmatrix} x \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ 1/2 & 1/3 & 1/6 \end{bmatrix}, \begin{bmatrix} y \\ p(y) \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 \\ 1/6 & 1/2 & 1/3 \end{bmatrix}$$

**可加性,** 统计独立信源  $X$  和  $Y$  的联合信源熵等于两个信源各自的熵之和。

$$H(XY) = H(X) + H(Y)$$

**强可加性,** 两个互相关联的信源  $X$  和  $Y$  的联合信源熵

$$H(XY) = H(X) + H(Y|X)$$

**扩展性**

$$\lim_{\varepsilon \rightarrow 0} H_{M+1}(p_1, p_2, \dots, p_M - \varepsilon, \varepsilon) = H_M(p_1, p_2, \dots, p_M)$$

2013-3-9     《信息论与编码》——信息与信息的度量     20



## 1.2 信息的度量

递增性  $H_{n+m-1}(p_1, p_2, \dots, p_{n-1}, q_1, q_2, \dots, q_m)$

$$= H_n(p_1, p_2, \dots, p_{n-1}, p_n) + p_n H_m\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \dots, \frac{q_m}{p_n}\right)$$

where,  $\sum_{i=1}^n p_i = 1, \sum_{j=1}^m q_j = p_n$

运用递增性可以简化熵的计算。

例:

$$\begin{aligned} H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}\right) &= H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) + \frac{1}{3} H\left(\frac{1}{2}, \frac{1}{2}\right) \\ &= \log 3 + \frac{1}{3} \log 2 = 1.918 \text{ bit/符号} \end{aligned}$$



## 1.3 本章小结

- 信息量的定义、意义及计算

$$I_x \triangleq \log_a \frac{1}{p_x} = -\log_a p_x$$

- 信息熵的定义、意义及计算

$$H(X) = \sum_x p(x) \log \frac{1}{p(x)}$$

- 信息量的单位的定义、意义及单位转换

$$\text{信息的单位} \triangleq \log 2$$

- 作业: 1, 2, 3, 5, 7, 9