以饱满的热情
优良的学风
明显的进步

迎接教育部对我校进行本科教学评估

信息论与编码

Information Theory & Coding

- 主讲教师： 于 工
- 通信工程教研室： B区410
- **Email**： qdyugong@163.com

0 绪 论

（第1讲 2007.9.4.）

- 本课学什么？

  0.1《信息与编码》的主要内容

- 有什么用？

  0.2 学习《信息与编码》的意义与重要性

0.1《信息与编码》的主要内容

三个编码和一个理论：

编码--------信源编码

信道编码

保密编码

理论--------香农信息论

1. 编码的定义与作用

- **编码**，是用符号(或数字)表达信息的一种方案，是表达信息的符号组合。广义地说，文字也算是一种编码。
- 为了对信息传输、存储与处理，现代通信与计算机技术中，需要把信息符号通过设定的数学关系，用另一套代码来替换原来的代码，使更加适合于传输，使通信更加高效、可靠、安全。因此，说到底是编码两套符号之间数学映射。

例如ASCII码用1字节(8bit)表示256个基本字符和一些控制符以便数字通信与计算机处理。

下表列出部分英语字母与数字的ASCII码：

高4位 \ 低4位	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001
0011	0	1	2	3	4	5	6	7	8	9
0100	@	A	B	C	D	E	F	G	H	I
0101	P	Q	R	S	T	U	V	W	X	Y
0110	'	a	b	c	d	e	f	g	h	i
0111	p	q	r	s	t	u	v	w	x	y

国标GB2312规定将6763个常用汉字存储在94行(区)94列(位)的表中，每个汉字用2字节(16bit)表示：

区＼位	0010 0001	0010 0010	0010 0011	0010 0100	0010 0101	0010 0110	0010 0111	0010 1000
1011 0000	阿	啊	挨	埃	哎	唉	哀	皑
1011 0001	薄	雹	保	堡	饱	宝	抱	报
1011 0010	病	并	玻	菠	播	拨	钵	波
1011 0011	场	尝	常	长	偿	肠	厂	
1011 0100	敞							
	础	储	矗	搐	触	处	揣	

2. 三大类编码

ASCII码与BG2312码是把文字转换成二进制代码的编码。通信技术中要讨论的三大类编码是：

- **信源编码**（Information Source Coding）

  压缩代码长度的编码，它使通信更加有效。

- **信道编码**（Information Channel Coding）

  检错、纠错的编码，它使通信更加可靠。

- **保密编码**（Cryptography）

  对信息加密和认证的编码，它使通信更加安全。

3. Shannon信息论

- **信息论**（**Shannon’s Coding Theorems**）关于信息传输与编码的理论。它使通信有了科学指导。
- **1948**年香农（**Shannon**）创立信息论，给信息下了科学的定义，并给出了无失真信源编码定理和有噪信道编码定理。
- **1958**年香农又补充了限失真信源编码和保密编码的信息理论，形成了完整的理论体系。

4. 信息论与编码的关系

- 香农理论是高度概括性的理论，它揭示了信息传输的普遍原理和基本原则，给出了编码的存在定理和理论极限，对于通信系统的设计和编码方案的构建具有指导意义。
- 有了信息理论，编码就不会盲目，工作就会目标明确，原理清晰，心中有数，事半功倍。

- 但是，信息论不能代替编码，正如通信原理不能代替通信电路一样。实际问题千差万别，仅有原则性的理论指导是远远不够的。
- 编码是为解决各种实际问题而设计的算法和搭建的系统。每个编码都是具体的，必须具体问题具体分析，对症下药个别解决。

0.2 学习本课的意义和重要性

1. 信源编码、信道编码和保密编码这三大类编码是实现高效、可靠、安全通信的重要保障。

- 人类进行信息交互的基本方式不外乎语言、文字和图像。
- 1832年Morse发明电报开始了用电子技术传输文字的时代 。从那时起，文字和符号的编码问题就已被提上了日程 。
- **1876年Bell**发明的电话标志着现代语音通信的诞生。
- 二十世纪上半叶发展起来的电视则开创了图像传输的先河。

- 随着数字语音通信和数字图像传输的兴起，语音编码与图像编码的课题就成了研究的热点。
- 直到今天，有线和无线的电话网、有线和无线的电视网仍然是语音和图像的主要传输途径，因特网则是目前除报刊书籍外最主要的文字传输媒体。
- 随着数字技术的发展，三大信息网络在数字通信的平台上融合为一的趋势日益加速，集语言、文字和图像为一体的，世界范围的信息实时交互已不是梦想。
- 一方面是通信技术一日千里地发展，技术更新的周期越来越短；另一方面，是人们对信息数量和质量的需求不断增长；

- 如何更加有效、更加可靠、更加安全地传输信息，成了人们非常关注的问题。
- 为了用较短的代码表达更多的信息，人们提出了压缩代码长度的问题，并发明了多种压缩方法和实施方案，它们被称为信源编码。
- 为了及时发现并纠正信息传输中出现的错误，人们采用了各种检错和纠错技术，由此发展起来了信道编码技术，它使通信更加可靠。
- 为了信息传输的安全，人们采用了多种保密编码和认证措施，密码学在网络时代焕发出新的生命
- 三大类编码是实现高效、可靠、安全通信的重要保障。

2. “信息论”是指导现代通信技术的基础理论。

- “信息论”是通信的数学理论，对通信的发展起了积极的指导作用。
- 它揭示的带宽与信噪比的Shannon公式直接解释了模拟通信中的调制解调过程中的数量与质量的辨证关系。
- Shannon公式还从本质上说明了CDMA无线数字通信技术高灵敏、低噪声、低功耗的原因。

3. “信息论与编码”是现代信息技术通信技术不可缺少的组成部分。

- 没有信源编码，IP电话将不可能，视频技术（VCD、DVD、可视电话以及视频聊天等都不存在。
- 没有信道编码，计算机网络与移动通信的部分协议将无法进行，数据通信将错误频频，难保质量。
- 没有保密编码，电子公务和电子商务将无法运行，，个人通信将失去安全。

4. “信息论与编码”是相关课程《通信原理》课的补充，是后续课程《通信网》、《无线通信》、《图像处理》等课的基础。

- 本课是从《通信原理》中分割出来又加以扩展形成的，加强了信息论，补充了信源编码与保密编码。
- 《通信网》、《无线通信》、《图像处理》等课中常常涉及编码问题。

0.3 本课的要求和安排

- 要求：掌握基本理论，学会典型编码。
- 安排：先扼要地介绍香农信息熵的基本理论，然后分别对信源编码、信道编码和保密编码基本原理和方法进行讨论。
- 讲法：把第1章的相关理论分散到相应

授课内容与课时安排

绪论　　（0.5学时）

第一章　信息论基础　　（3.5学时）

- 1.1通信与信息
- 1.2离散信源

第二章 无失真信源编码　（12学时）

- 2.1 信源编码概述
- 2.2 霍夫曼编码
- 2.3 游程编码
- 2.4 算术编码
- 2.6 通用编码

- **1.3离散信道　　（4学时）**

第三章 信道编码　　（14学时）

- 3.1 检错、纠错原理
- 3.2 差错控制理论
- 3.3 线性分组码
- 3.4 循环码
- 3.5循环码的扩展
- 3.6 卷积码
- 3.7 纠正突发错误的编码

- **1.4 连续信源和波形信道　　（2学时）**

第四章 限失真信源编码 （4学时）

- 4.1 信源的有损压缩
- 4.2 率失真函数
- 4.3保真度准则下的信源编码

第五章 密码 （8学时）

- 5.1 密码学的基本概念
- 5.2 序列（流）密码
- 5.3 分组（块）密码
- 5.4 保密编码的信息理论
- 5.5 公开密钥系统
- 总结、复习 （2学时）

第1章 信息论基础

- 通信与信息
- 离散信源
- 离散信道
- 连续信源与波形信道

教学目的与要求

1. 深刻理解信息的定义与概念。

2. 牢固掌握离散有记忆信源信息熵的计算方法与规律。

3. 熟练掌握信道有关的信息熵及平均互信息的计算方法，深刻理解信道容量概念。

4. 掌握连续信源信息熵的计算，深刻理解Shannon公式的意义。

参考文献

1.傅祖芸：**信息论—基础理论与应用**

电子工业出版社（2001年8月第一版）

2.傅祖芸：**信息理论与与编码学习辅导及精选题解**

电子工业出版社（2004年7月第一版）

3.仇佩亮：**信息论与编码**

高等教育出版社（2003年12月第一版）

4.曹雪虹：**信息论与编码**

北京邮电大学出版社 (2001年8月第一版)

第1章 信息论基础

1.1 通信与信息

- 计划学时：1学时
- 要求掌握的主要内容：

  1. 深刻理解Shannon关于信息的定义。

  2. 熟练掌握自信息熵的计算公式与单位。

- 重点难点：

  重点----信息的定义

- 外语关键词：

信源：**Information Source**

信道：**Channel**

信宿：**Information Sink**

消息：**Message**

信号：**Signal**

信息：**Information**

信息的定义

信息（**Information**）作为一个专业名词，其科学定义来自通信。

通信系统的基本组成

发出信息　　传输信息　　接收信息

- 消息（Message）：

信源发出的语言、文字、公式、数据、声音、图像等等。每个消息都是具体的，其内容千千万万，形式多种多样。消息中包含着信息，但消息却不等于信息。

- 信号（**Signal**）：

替代消息并适合于在信道中传输的是连续或脉冲的电压、电流、电磁波及光波，它们叫做信号，信号是信息的载体，信号也不等于信息。

- 信宿获得的是知识、情报、机密、商情、情感交流和视听享受等等，它们仍不能作为信息的确切定义。

- 信息（**Information**）：

什么是信息？

- 信息是消息的内涵，是信号的价值，信息是能使信宿得以获知解惑的东西。
- 它应当是从千千万万不同形式不同内容的消息中抽象出来的、具有共性的、可定量测度的一个量，应该有它的单位和数学表达。

[几个实例]

❖天气预报。假如某地晴间多云天气大约占75%，降水天气约占5%，偶尔有（大约0.1%）台风。

❖彩票（抽奖）假如中头奖万分之一，二奖千分之一，末奖百分之一。

❖短信。假设长度50个字。

共同点是：通信使信宿获得了原来不明确的事情；不确定性越大的消息，所含的信息越多。

- 分析：通信过程发生前，信宿并不知道信源将发出什么消息，就是说，信源发出的消息存在着某种不确定性；借助通信，信宿明确了原来不明确的一些事情，增加了对信源所述事情的了解，减少甚至消除了原来的疑问，于是认为他获得了一些信息。

- 结论1：**通信是消除不确定性的过程。**

- 结论2：**信息的多少可以用通信所消除掉的不确定性来度量。**

信息的定义：

- 消息的不确定性又来自何处呢？
- 信源消息的不确定性归根结底来自客观事物的多样性和随机性。从这个意义上讲，**信息是客观事物存在方式和运动状态的多样性和不确定性的度量。**

信息的度量：

- 不确定性就是随机性，随机事物是用概率统计方法描述的。
- 概率大的事物，出现可能性大，不确定性就小；概率最大为1，概率百分之百就是肯定要发生的事物，其不确定性为0。
- 概率小的事物，出现可能性小，不确定性就大；概率为0，就是不会出现的事物，其不确定性无穷大。

- 因此，如果某事物出现概率为p，则可定义：　　$I = \log(1/p) = -\log p$

为该事物的不确定度，也称为该事件的自信息。

- 自信息是不确定度的代名词。
- 信息量=通信所消除掉的不确定度

=通信前的不确定度--通信后的不确定度

- 之所以取对数，是考虑到我们的定义应当符合以下事实：互不相关事件同时出现的概率是各事件单独出现概率之积，而总的自信息却应当是各事件自信息之和：

$$I = \log(1/p_1p_2) = \log(1/p_1) + \log(1/p_2)$$

$$= -\log p_1 - \log p_2 = I_1 + I_2$$

- 对数的底取2时，自信息的单位叫比特(*bit*)。
- 对数的底取10时，自信息的单位叫哈特（*hart*）
- 对数的底取e时，自信息的单位叫奈特（*nat*）
- 若不加声明，对数的底均取为2。

第1章 信息论基础

1.2 离散信源

- 计划学时：2学时
- 要求掌握的主要内容：

  1. 深刻理解离散信源信息熵有关概念。

  2. 熟练掌握无记忆信源与有记忆信源的消息序列的信息熵的计算方法与变化规律。

  3. 学会马尔科夫信源信息熵的计算。
- 重点难点：

  重点----有记忆信源的消息序列的信息熵的计算

  难点----马尔科夫信源

- 外语关键词：

离散信源： Discrete Information Source

信息熵： Information Entropy

有记忆信源： Information Source with memory

马尔科夫信源： Markov Information source

概率空间：Probability Space

相对信息率：Relative Entropy rates

冗余度： Redundancy

1.2.1 信息熵

- 离散信源：

信源X从m个符号 $\{a_1、a_2、a_3、……、a_m\}$ 组成的字符集中随机选取字符来发送信息，这样的信源属于离散信源。

- 概率空间：

$$\begin{pmatrix} X \\ p(\mathrm{X}) \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \cdots\cdots a_m \\ p_1 & p_2 \cdots\cdots p_m \end{pmatrix}$$

- 信源发出各字符的自信息分别为

$$I_i = -\log p_i \quad (i=1,2,\ldots\ldots,m);$$

- 所以信源发出单个字符的平均自信息为：

$$H(X) = \sum_{i=1}^{m} p_i I_i = \sum_{i=1}^{m} p_i \log \frac{1}{p_i} = -\sum_{i=1}^{m} p_i \log p_i$$

$H(X)$叫做信源的(单符号)信息熵，也叫先验熵，它反映信源平均发送每个符号的不确定度。

[例1]信源发出A、B、C、D四个符号，其概率分别为3/8、1/4、1/4、1/8；　(1)求各符号的自信息和信源信息熵；

(2)求信源发出符号序列ABAABCDBADCADCBDABCABA BAACAAADCABCDACBABCABCABAACBDDCAAABC 时序列总自信息和平均每符号自信息。

解:(1)∵ p(A)=3/8、p(B)=1/4、p(C)=1/4、p(D)=1/8；

$\therefore I(A)=\log(8/3)=1.415\text{bit},$

$I(B)=I(C)=\log 4=2\ \text{bit},$

$I(D)=\log 8=3\ \text{bit};$

$$H(X)=\frac{3}{8}I(A)+\frac{1}{4}I(B)+\frac{1}{4}I(C)+\frac{1}{8}I(D)=1.906\ bit/\text{符号}$$

(2)序列总自信息为：

$$I=23I(\mathrm{A})+14I(\mathrm{B})+13I(\mathrm{C})+7I(\mathrm{D})=107.546\ \mathrm{bit}$$

平均每符号自信息为：

$$H=107.546\ /57=1.89\ \mathrm{bit/符号};$$

讨论： 信息熵$H(X)$是信源发送大量符号的统计平均结果，反映信源的性质；而给定序列的平均自信息H只代表某次通信的结果，它只对这个样本序列有意义。随机样本偏离统计平均值是正常现象。

信息熵的性质

- 非负性：$H(X) \geqslant 0$
- 对称性：$H(p_1 p_2 \cdots\cdots) = H(p_2 p_1 \cdots\cdots)$
- 极值性：等概信源具有最大信息熵：

$$H_{max} = \log m$$

等概信源指：$p_1 = p_2 = \cdots\cdots = p_m$

[例2]二元信源的概率矢量可写为$p=(x, 1-x)$，画出信息熵随概率x的变化曲线$H(x)$。

解：根据 $H(x)=-x\log x-(1-x)\log(1-x)$ 得到下表

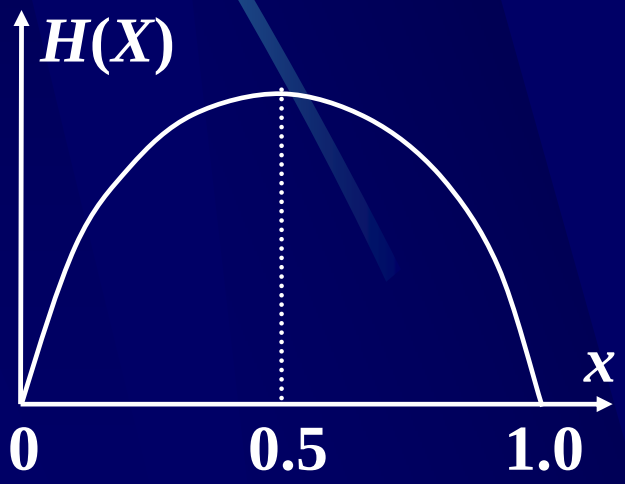
x	0	0.1	0.2	0.3	0.4	0.5
$H(x)$	0	0.47	0.72	0.88	0.97	1.0

特点：

（1）左右对称

（2）极大值在$x=0.5$

（3）$x\to 0$ 时$x\ log\ x\to 0$



小结：

- **自信息的定义：** $I = \log(1/p) = -\log p$
- **信息的单位：** 对数的底取2时，自信息的单位叫比特(***bit***)。
- **信息熵：** $H(X) = -\sum_{i=1}^{m} p_i \log p_i$
- **等概信源信息熵最大：** $H_{max} = \log m$

课后复习题

- 思考题：

  1. 为什么用不确定性来定义信息？

  2. 为什么等概信源信息熵最大？

- 作业题：

  教材第31页习题一第3、5题；

[温旧引新]

第2讲2007.9.6.

- 信息的定义、计算与单位；
- 什么是离散信源？什么是概率空间？
- 信息熵的定义、计算与单位；
- 信息熵的性质；

当信源发出字符序列……

①随着序列的伸延，信源选取字符的概率是否随着时间改变；

- 平稳信源——即信源选取字符的概率不随时间改变。

②序列前后字符之间是否统计相关。

- 字符之间不存在统计关联的信源叫做无记忆信源；
- 字符之间存在统计关联的信源叫做有记忆信源。

有关数学知识

- 概率：P(A)，P(B)
- 条件概率：P(A|B), P(B|A)
- 联合概率：P(AB)=P(A)P(B|A)

  或：P(AB)=P(B)p(A|B)
- 统计无关（统计独立）：

  若P(A)=P(A|B)，则P(AB)= P(A) P(B)
- 归一化条件：Σ_A P(A)=1；Σ_A P(A|B)=1；

  Σ_AP(AB)= P(B)；Σ_BP(AB)= P(A)；

  Σ_A Σ_B P(AB)=1；

1.2.2 无记忆信源的信息熵

- 无记忆信源发出序列的概率：

$$p(x_1x_2\ldots\ldots x_N) = p(x_1)\,p(x_2)\,p(x_3)\,\ldots\ldots p(x_N)$$

- 无记忆信源序列的信息熵（联合熵）：

$$H(X_1X_2\ldots\ldots X_N) = -\sum p(x_1x_2\ldots\ldots x_N)\log p(x_1x_2\ldots\ldots x_N)$$

$$= H(X_1)+H(X_2)+H(X_3)+\ldots\ldots+H(X_N)$$

$$= NH(X_1)$$

无记忆信源

因此平均每符号的信息熵：

$$H_N = H(X_1X_2\ldots\ldots X_N) / N = H(X_1)$$

1.2.3 有记忆信源的信息熵

- 有记忆信源发出序列的概率：

$$p(x_1x_2\ldots\ldots x_N)$$

$$= p(x_1)\,p(x_2|x_1)\,p(x_3|x_1x_2)\,\ldots\ldots p(x_N|x_1x_2\ldots\ldots x_{N-1})$$

- 有记忆信源序列的信息熵：

$$H(X_1X_2\ldots\ldots X_N) = -\Sigma\, p(x_1x_2\ldots\ldots x_N)\, log\, p(x_1x_2\ldots\ldots x_N)$$

$$=-\sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} p(x_1x_2\cdots x_N)\cdot[\log p(x_1)+\log p(x_2\,|\,x_1)+$$

$$\log p(x_3\,|\,x_1x_2)+\cdots+\log p(x_N\,|\,x_1x_2\cdots x_{N-1})\,]$$

利用以下关系：

$$\sum_{x_2}\sum_{x_3}\cdots\sum_{x_N} p(x_1x_2\cdots x_N) = p(x_1)$$

$$\sum_{x_3}\sum_{x_4}\cdots\sum_{x_N} p(x_1x_2\cdots x_N) = p(x_1x_2)$$

$$\cdots\cdots$$

$$\sum_{x_N} p(x_1x_2\cdots x_N) = p(x_1x_2\cdots x_{N-1})$$

有记忆信源

- 有记忆信源序列的信息熵：

$$H(X_1X_2\ldots\ldots X_N) = H(X_1) + H(X_2|X_1) + H(X_3|X_1X_2) + \ldots\ldots + H(X_N|X_1X_2\ldots\ldots X_{N-1})$$

式中各级条件熵分别为：

$$H(X_1) = -\sum_{x_1} p(x_1)\log p(x_1)$$

$$H(X_2 \mid X_1) = -\sum_{x_1}\sum_{x_2} p(x_1x_2)\log p(x_2 \mid x_1)$$

$$H(X_3 \mid X_1X_2) = -\sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1x_2x_3)\log p(x_3 \mid x_1x_2)$$

$$\cdots\cdots$$

$$H(X_N \mid X_1X_2\cdots\cdots X_N) =$$

$$= -\sum_{x_1}\sum_{x_2}\cdots\sum_{x_{N-1}} p(x_1x_2\cdots x_N)\log p(x_N \mid x_1x_2\cdots x_{N-1})$$

- **性质1：** 条件熵不大于无条件熵，强条件熵不大于弱条件熵。

$$H(X_1) \geqslant H(X_2|X_1) \geqslant H(X_3|X_1X_2) \geqslant \cdots$$

$$\cdots\cdots \geqslant H(X_N|X_1X_2\cdots\cdots X_{N-1})$$

(对无记忆信源，取等号)

- 理由：有条件约束相当于自由度减少，即不确定性的减少，熵是平均不确定性的度量。

- 性质2：条件熵不大于同阶的平均符号熵：

$$H_1 = H(X)$$

$$H_2 \geqslant H(X_2|X_1)$$

$$H_3 \geqslant H(X_3|X_1X_2)$$

$$\cdots \geqslant \cdots$$

$$H_N \geqslant H(X_N|X_1X_2\cdots\cdots X_{N-1})$$

- 性质3：序列越长，平均每个符号的信息熵就越小。

$$H_1 \geqslant H_2 \geqslant H_3 \geqslant \cdots\cdots \geqslant H_N$$

- 三条性质合起来，基本规律是：

$$\begin{array}{ccccccccc} H(X_1) & \geqslant & H(X_2|X_1) & \geqslant & H(X_3|X_1X_2) & \geqslant \cdots\cdots & \geqslant & H(X_N|X_1X_2\cdots\cdots X_{N-1}) \\ \| & & \wedge\!\vee & & \wedge\!\vee & & & \wedge\!\vee \\ H_1 & \geqslant & H_2 & \geqslant & H_3 & \geqslant \cdots\cdots & \geqslant & H_N \end{array}$$

- 直到$N \to \infty$，二个极限就一样大了：

$$H_\infty = \lim_{N\to\infty} H_N = \lim_{N\to\infty} H(X_N \mid X_1X_2\cdots X_{N-1})$$

- H_∞叫做极限熵，代表实际信源的熵。
- 连同等概信源最大熵$H_0 = \log m$就有：

$$H_0 \geqslant H_1 \geqslant H_2 \geqslant H_3 \geqslant \cdots\cdots \geqslant H_N \geqslant H_\infty$$

- 等概信源信息熵最大；不等概意味着信源选取符号有倾向，导致确定性增加，信息熵减小；有记忆信源符号间的相互关联导致确定性进一步增加，信息熵进一步减小；并且关联越长，确定性越大。

- 对于有记忆信源，只有极限熵才是严格意义上的信源平均符号熵。
- H_1、H_2、H_3、……、H_N都是近似的平均符号熵。它们是从无穷长的消息序列中截取的一段，计算序列熵时必然忽略了该序列与序列外的符号的关联，称之为“截短近似”。
- 往往把产生长度为N的序列的信源称为N次扩展信源。

[例3]某有记忆信源发出符号时相邻符号的条件概率为$p(0|0)=0.8$；$p(0|1)=0.1$；求它的最大熵H_0、平均符号熵H_1、H_2、H_3和极限熵H_∞。

解：（1）最大熵H_0

假设信源等概且无记忆，才能达到最大熵：

$H_0=\log m=\log 2=1$ bit/符号。

（2）平均符号熵H_1

先计算信源发送单个符号$X=(0,1)$的概率$p(X)$.

由：$p(0|0)=0.8$ 知：$p(1|0)=0.2$；

由：$p(0|1)=0.1$ 知：$p(1|1)=0.9$；

利用：

$$p(x_2)=\sum_{x_1}p(x_1x_2)=\sum_{x_1}p(x_1)p(x_2\mid x_1)$$

当x_2取0时：$p(0)=p(0)p(0\mid 0)+p(1)p(0\mid 1)=0.8p(0)+0.1p(1)$

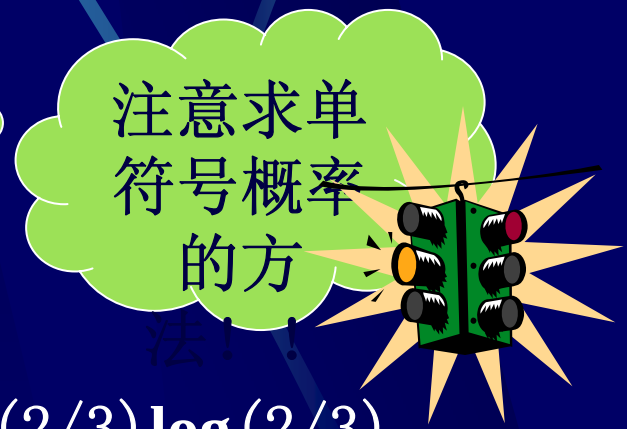
即为：　　$0.2p(0)=0.1p(1)$

连同归一化条件：$p(0)+p(1)=1$

解得：　　$p(0)=1/3$；　$p(1)=2/3$；

于是：$H_1=H(X)=-(1/3)\log(1/3)-(2/3)\log(2/3)$

$=0.9183$ bit/符号。



（3）平均符号熵H_2

信源发出长度为2的序列

X_1X_2=(00, 01, 10, 11)，

其概率可由$p(x_1x_2)=p(x_1)p(x_2|x_1)$计算：

$p(X_1X_2)$==(0.2667, 0.0667, 0.0667, 0.6)

$H(X_1X_2)$=-0.2667log0.2667-.0667log0.0667

-0.0667log0.0667-0.6log0.6

=1.4716（bit/符号）。

H_2 = $H(X_1X_2)/2$ = 0.7359 bit/符号。

（4）平均符号熵H_3

信源发出长度为3的序列

$$X_1X_2X_3=(000, 001, 010, 011, 100, 101, 110, 111),$$

其概率可由：$p(x_1x_2x_3)=p(x_1)\,p(x_2|x_1)\,p(x_3|x_1x_2)$

$=p(x_1)\,p(x_2|x_1)\,p(x_3|x_2)$计算。

注意这里！

计算结果是：$p(X_1X_2X_3)=$

$=(0.2133, 0.0533, 0.0067, 0.06, 0.0533, 0.0133, 0.06, 0.54)$

所以：$H(X_1X_2X_3)=2.0249$ bit/符号

$H_3= H(X_1X_2X_3)/3= 0.6750$ bit/符号。

显然，符合$H_0 \geqslant H_1 \geqslant H_2 \geqslant H_3$的规律。

（5）极限熵H_∞：

二阶条件熵：

$$H(X_2 \mid X_1) = -\sum_{x_1}\sum_{x_2} p(x_1x_2)\log p(x_2 \mid x_1)$$

$$= -0.2667\log 0.8 - 0.0667\log 0.2$$

$$-0.0667\log 0.1 - 0.6\log 0.9 = 0.5533 \quad \text{bit/符号}$$

$N \geqslant 2$以后，条件熵不再随着序列长度而变化，所以$N \to \infty$时：$H_\infty = 0.5533$ bit/符号

[温旧引新]

第3讲2007.9.11.

- 单符号信息熵： $H(X)=-\sum_{i=1}^{m} p_i \log p_i$
- 序列的信息熵：

$$H(X_1X_2\cdots X_N)=-\sum_{x_1=a_1}^{a_m}\sum_{x_2=a_1}^{a_m}\cdots\sum_{x_N=a_1}^{a_m} p(x_1x_2\cdots x_N)\log p(x_1x_2\cdots x_N)$$

- 无记忆时： $= H(X_1)+H(X_2)+H(X_3)+\cdots\cdots+H(X_N)$
- 有记忆时： $= H(X_1) + H(X_2|X_1) + H(X_3|X_1X_2)$

$$+ \cdots\cdots + H(X_N|X_1X_2\cdots\cdots X_{N-1})$$

- 定义序列的平均单符号熵：

$$H_N = H(X_1X_2\cdots\cdots X_N) / N$$

- 它们的大小关系为：

$$\begin{array}{ccccccccc} H(X_1) & \geqslant & H(X_2|X_1) & \geqslant & H(X_3|X_1X_2) & \geqslant \cdots\cdots & \geqslant & H(X_N|X_1X_2\cdots\cdots X_{N-1}) \\ \| & & \wedge\!\vee & & \wedge\!\vee & & & \wedge\!\vee \\ H_1 & \geqslant & H_2 & \geqslant & H_3 & \geqslant \cdots\cdots & \geqslant & H_N \end{array}$$

- 无记忆信源取等号，有记忆信源取大于号。
- 实际信源$N \to \infty$，二个极限就一样大了，叫做极限熵：

$$H_\infty = \lim_{N\to\infty} H_N = \lim_{N\to\infty} H(X_N \mid X_1X_2\cdots X_{N-1})$$

- 有限长序列都是无限长实际序列的各阶“截断”近似。
- 等概无记忆信源有最大熵 $H_0 = \log m > H_1$
- 结论是：$H_0 \geqslant H_1 \geqslant H_2 \geqslant H_3 \geqslant \cdots\cdots \geqslant H_N \geqslant H_\infty$

1.2.4 马尔科夫信源的信息熵

1. 马尔科夫信源的数学模型和定义：

- 鉴于“记忆”一般随距离而衰减，当两符号距离较远时关联可以忽略，提出马尔科夫(Markov)近似模型。
- 假设距离小于或等于*N+1*的符号之间存在关联，距离大于*N+1*的符号之间不存在关联，*N+1*为关联长度。
  - 一阶马尔科夫信源，关联长度为2，符号两两相关，
  - 二阶马尔科夫信源，关联长度为3，符号三三相关，
  - ……
  - N-1阶马尔科夫信源关联长度为N，
  - N阶马尔科夫信源关联长度为N+1。

- 马尔科夫信源是一个有限记忆长度的平稳（概率空间不随序列位置而变化的信源）信源。
- 马尔科夫近似优于截短近似！优越之处在于它处理的对象仍然是整个信源序列，只是合理地假设了关联的程度，而没有像截短近似那样硬性地切断关联，去处理较短的序列段。
- 请注意，N次扩展信源与N阶马尔科夫信源有什么区别

- 由于相互关联而被锁链在一起的无穷长序列，如何计算信息熵？
- 状态：把排在指定符号前面与它相关联的N个符号的符号串定义为该符号的状态。
- 状态数：当信源是m个符号的集合时，N阶马尔科夫信源的状态数目共有$L=m^N$种。
- 只要知道L个状态下分别发出m个符号的条件概率，无限长序列中的全部关联就都清楚了。
- 于是求解这个无穷长的互相连锁的链的复杂问题就被简化为求解L个状态决定m个符号的简单问题。

[例4]已知二阶马尔科夫信源的条件概率：

$p(0|00)=p(1|11)=0.8$；$p(0|01)=p(1|10)=0.6$；

求它的状态符号依赖关系和状态转移概率。

解：二阶马氏信源关联长度=3，状态由2符号组成，共有4个状态，分别为：$E_1=00$；$E_2=01$；$E_3=10$；$E_4=11$；

已知的条件概率即是：

$p(0|E_1)=p(1|E_4)=0.8$；$p(0|E_2)=p(1|E_3)=0.6$；

根据归一化条件可求出另外4个状态符号依赖关系为：

$p(1|E_1)=p(0|E_4)=0.2$；$p(1|E_2)=p(0|E_3)=0.4$；

2. 状态转移与稳态概率：

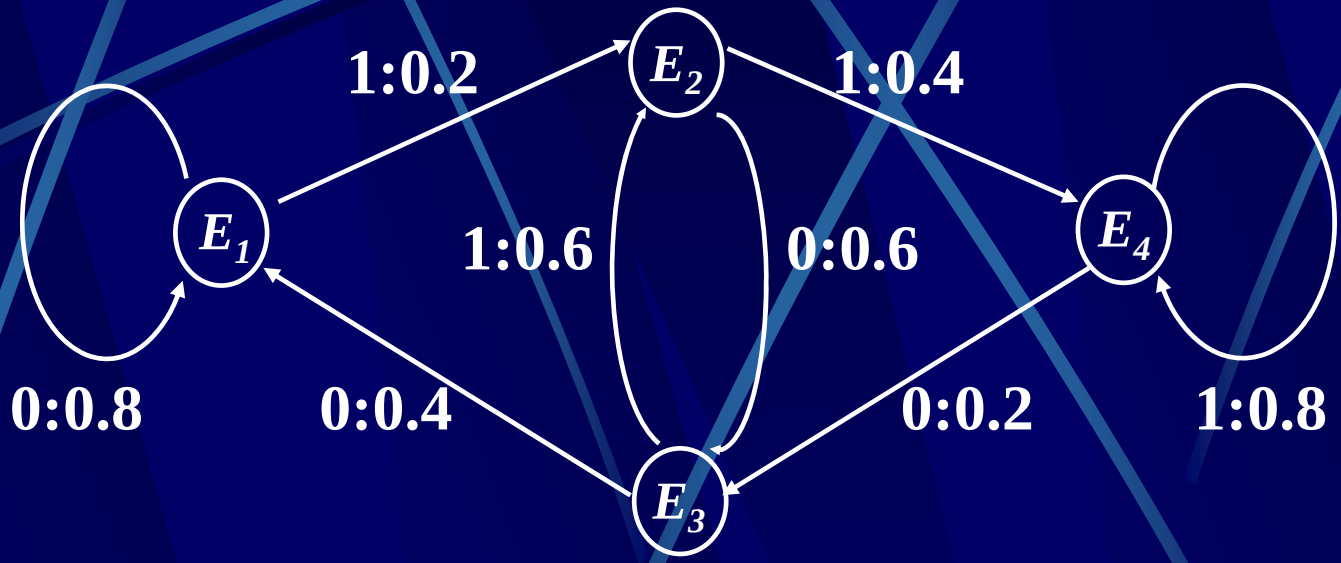
状态转移：随着序列的延伸，当前符号的位置在不断前移，相应的状态也在不断变化，这种变化称为状态转移。新状态由老状态去掉最后那个符号，再接到当前符号上构成。

状态转移概率：信源某一时刻所处状态E_l，由前一时刻$l-1$时信源状态E_{l-1}和新输出的符号x_l唯一确定，即：　$p(x_l / E_{l-1}) = p(E_l / E_{l-1})$

状态转移概率等于相应状态发出相应符号的概率。

原状态	发符号	新状态	状态转移概率
$E_1=00$	发0变为000	$E_1=00$	$P(E_1\|E_1)=p(0\|E_1)=0.8$;
$E_1=00$	发1变为001	$E_2=01$	$P(E_2\|E_1)=p(1\|E_1)=0.2$;
$E_2=01$	发0变为010	$E_3=10$	$P(E_3\|E_2)=p(0\|E_2)=0.6$;
$E_2=01$	发1变为011	$E_4=11$	$P(E_4\|E_2)=p(1\|E_2)=0.4$;
$E_3=10$	发0变为100	$E_1=00$	$P(E_1\|E_3)=p(0\|E_3)=0.4$;
$E_3=10$	发1变为101	$E_2=01$	$P(E_2\|E_3)=p(1\|E_3)=0.6$;
$E_4=11$	发0变为110	$E_3=10$	$P(E_3\|E_4)=p(0\|E_4)=0.2$;
$E_4=11$	发1变为111	$E_4=11$	$P(E_4\|E_4)=p(1\|E_4)=0.8$;

状态转移图：



稳定状态： 随着时间的推移，序列不久就会达到一种动态平衡状态：不断延伸的序列在L种状态之间变来变去，而各种状态出现的概率却不再变化，我们称这种情况为稳定状态。

稳态概率

设：稳态概率分别为$Q(E_i)$　（$i=1,2,\ldots\ldots,L$），

- 由状态转移图可写出它们满足的方程组。
- 因为独立方程只有L-1个，还应加上归一化条件。
- 得到稳态概率方程组：

联立求解

$$Q(E_i)=\sum_{j=1}^{L}p(E_i\mid E_j)Q(E_j)$$

$$\sum_{i=1}^{L}Q(E_i)=1$$

例4的稳态方程组是：

$$\begin{cases} Q(E_1)=0.8Q(E_1)+0.4Q(E_3) \\ Q(E_2)=0.2Q(E_1)+0.6Q(E_3) \\ Q(E_3)=0.6Q(E_2)+0.2Q(E_4) \\ Q(E_4)=0.4Q(E_2)+0.8Q(E_4) \\ Q(E_1)+Q(E_2)+Q(E_3)+Q(E_4)=1 \end{cases}$$

可解得：

$$\begin{cases} Q(E_1)=Q(E_4)=\dfrac{1}{3} \\ Q(E_2)=Q(E_3)=\dfrac{1}{6} \end{cases}$$

3. 稳态符号概率与稳态信息熵:

稳态符号概率： 达到稳态后信源发出某符号的概率不再随时间变化。 由下列方程决定：

$$p(a_k)=\sum_{i=1}^{L}Q(E_i)p(a_k\mid E_i) \qquad (k=1,2,\ldots\ldots,m)$$

例4的稳态符号概率为：

$$\begin{cases} p(0)=\sum_{i=1}^{4}Q(E_i)p(0\mid E_i)=\frac{1}{3}\times 0.8+\frac{1}{6}\times 0.6+\frac{1}{6}\times 0.4+\frac{1}{3}\times 0.2=\frac{1}{2} \\ p(1)=\sum_{i=1}^{4}Q(E_i)p(1\mid E_i)=\frac{1}{3}\times 0.2+\frac{1}{6}\times 0.4+\frac{1}{6}\times 0.6+\frac{1}{3}\times 0.8=\frac{1}{2} \end{cases}$$

对指定的状态E_i，发出各符号的平均不确定程度为：

$$H(X \mid E_i) = -\sum_{k=1}^{m} p(a_k \mid E_i) \log p(a_k \mid E_i)$$

稳态信息熵应为各态不确定度的统计平均：

$$H = \sum_{i=1}^{L} \sum_{k=1}^{m} Q(E_i) H(a_k \mid E_i) = -\sum_{i=1}^{L} Q(E_i) \sum_{k=1}^{m} p(a_k \mid E_i) \log p(a_k \mid E_i)$$

例4的稳态信息熵即为：

$$H = -\frac{1}{3} \times [0.8\log 0.8 + 0.2\log 0.2] \times 2 - \frac{1}{6} \times [0.6\log 0.6 + 0.4\log 0.4] \times 2 =$$

$$= 0.895 \ \text{bit/符号}$$

N阶马尔科夫信源的稳态信息熵：

当前符号$a_k=x_{N+1}$时，状态为：$E_i=(x_1x_2\cdots\cdots x_N)$，

$$Q(E_i)p(a_k|E_i)=p(x_1x_2\cdots\cdots x_N)p(x_{N+1}|x_1x_2\cdots\cdots x_N)$$

$$=p(x_1x_2\cdots\cdots x_{N+1})$$

$$H=-\sum_{i=1}^{L}\sum_{k=1}^{m}Q(E_i)p(a_k|E_i)\log p(a_k|E_i)=$$

$$=-\sum_{x_1=a_1}^{a_m}\sum_{x_2=a_1}^{a_m}\cdots\sum_{x_N=a_1}^{a_m}p(x_1x_2\cdots x_{N+1})\log p(x_{N+1}|x_1x_2\cdots x_N)=$$

$$=H(X_{N+1}|X_1X_2\cdots X_N)$$

因此N阶马尔科夫稳态信息熵等于N+1阶条件熵。

N阶马尔科夫信源的极限熵：

对于N阶马尔科夫信源，关联长度是N+1，N+2以后都不再关联，直至 $N \to \infty$，

因此极限熵：

$$H_\infty = \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1})$$

$$= H(X_{N+1} \mid X_1 X_2 \cdots X_N)$$

结论： N阶马尔科夫信源的极限熵等于N+1阶条件熵，即稳态信息熵。

1.2.5 信源的相对信息率和冗余度

- 等概信源的信息熵可达到最大值$H_0=\log m$，然而实际信源由于非等概和有记忆两个原因，其信息熵远小于H_0
- 以英语为例，26个字母加上空格，共27个符号，$H_0=\log 27=4.76$（比特/符号）。然而各个字母在文章中出现概率是不同的，统计规律见表1.1所示。

英语字母出现概率统计表

a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
6.42	1.27	2.18	3.17	10.31	2.08	1.52	4.67	5.75
j	*k*	*l*	*m*	*n*	*o*	*p*	*q*	*r*
0.08	0.49	0.32	1.98	5.74	6.32	1.52	0.08	4.84
s	*t*	*u*	*v*	*w*	*x*	*y*	*z*	空格
5.14	7.96	2.28	0.83	1.75	0.13	1.64	0.05	18.59

由表不难算出：

$H_1 = H(X) = 4.03$（比特/符号）。

- 在一阶和二阶马尔科夫信源近似下，有人曾求得：H_2= 3.32(比特/符号)，H_3=3.10(比特/符号)。
- 如果再计及词法、句法、语法、修辞等的制约，计入更高阶的关联，极限熵被估计为H_∞=1. 4（比特/符号）。
- H_0代表平均每个英文符号最多所能承载的信息量。H_∞代表英语文章中平均每个信源符号实际所荷载的信息量。拥有27个符号的英文信源，平均每个符号是有能力荷载4.76比特信息的，但实际上它们每符号却只承载着 1.4比特的信息。

这种信息率不饱满的情况在各种信源中普遍存在，因此定义：

·相对信息率： $\mu = H_\infty / H_0$

·信源冗余度（或剩余度）： $\gamma = 1 - \mu$

来反映信源实际所含信息的饱满程度。对于英文， $\mu = 0.29$， $\gamma = 0.71$，冗余是比较大的。

- 汉语也有类似的情况：对10000个汉字统计，最常见的只有140个，出现概率50%；较常用的485个，出现概率35%；一般的1775个，出现概率14.7%；不常见的7600个，出现概率3%。
- 由此算出：H_0=log1000=13.29（bit/汉字）

H_1=10.25（bite/字）

仅算到H_1：　μ=0.77，　γ=0.23，

小结：

- **马尔科夫信源的定义与数学描述：**

关联长度，符号与状态，状态转移，稳态概率，稳态符号概率，稳态信息熵。

- **N阶马尔科夫信源的极限熵：**

$$H_\infty = \lim_{N\to\infty} H(X_N \mid X_1 X_2 \cdots X_{N-1}) = H(X_{N+1} \mid X_1 X_2 \cdots X_N)$$

*N*阶马尔科夫信源的极限熵等于*N*+1阶条件熵，即稳态信息熵。

课后复习题

- 思考题：

哪些因素影响信源消息序列的平均符号熵？为什么序列越长，平均符号熵越小？

- 作业题：

教材第31页习题一第6、9题；

欢迎各位同学拷贝！

- 此课件挂在mail.google.com网站上

用户名：jpkctxyl@gmail.com

密　码：jpkctxyl123456