



信息论与编码 (第四章 信息论基础) 学习笔记

4.1 引言

信息论是由美国科学家香农在1948年通过其论文《通信的数学理论》奠定的理论基础。

消息与信息

- **消息** : 由符号、文字、数字、语音或图像等组成的序列，是信息的物理载体。
- **信息** : 是消息所包含的**内涵**。信息的本质是用来**消除不确定性**的东西。
 - **形式化**: 通信的基本任务是在接收端从形式上恢复出发送端发出的消息，不关心对消息内容的理解。
 - **不确定性**: 在收到消息前，消息的内容对接收者来说是随机的、不确定的。通信的作用就是通过传递消息，消除或部分消除这种不确定性。
 - **信息量**在数量上就等于获得消息前后“不确定性”的消除量。

4.2 信息的度量

信息量的概念

1. **某消息的信息量 = 获得该消息后不确定性的消除量。**
2. 不确定性与事件发生的**概率**有关。事件发生的概率越小，其不确定性就越大，一旦发生，所提供的信息量也越大。
3. 信息量的度量应满足**可加性**。即两个独立事件同时发生的信息量，等于它们各自信息量之和。

4.2.1 离散信源信息的度量 (自信息)

- **自信息定义**: 对于一个离散消息，其信息量（自信息）定义为：

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

其中 $P()$ 为消息发生的概率。

- **单位**: 信息量的单位与对数的底有关。
 - 底为2: **比特 (bit)**  (默认单位)
 - 底为e: 奈特 (nit)
 - 底为10: 哈特 (hart)

-  **示例**:

已知某信源的概率场为:

() $3/8, 1/4, 1/4, 1/8$

各符号统计独立。

1. 计算各符号信息量:

$$(0) = \log_2(3/8) \approx 1.415$$

$$(1) = \log_2(1/4) = 2$$

$$(2) = \log_2(1/4) = 2$$

$$(3) = \log_2(1/8) = 3$$

2. 计算符号序列 S="113200" 的信息量:

由于各符号独立，信息量可加。

$$(S) = (1) + (1) + (3) + (2) + (0) + (0)$$

$$(S) = 2 + 2 + 3 + 2 + 1.415 + 1.415 = 11.83$$

4.2.2 离散信源的平均信息量 (信源的熵)

- **熵的定义:** 熵是信源输出的**平均信息量**，表示信源在统计意义上每个符号所含信息量的平均值。

- **熵的公式:**

$$H() = - \sum_{i=1}^N P_i \log_2 P_i$$

- **单位:** 比特/符号。

-  **核心思想:** 熵描述了一个信源的**平均不确定性**。熵越大的信源，其输出同样个数的符号，所含的信息量就越大。

-  **示例:**

求上一示例中信源的熵。

$$H() = (0) \log_2(0) + (1) \log_2(1) + (2) \log_2(2) + (3) \log_2(3)$$

$$H() = \frac{3}{8} \log_2 \frac{3}{8} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{8} \log_2 \frac{1}{8}$$

$$H() = \frac{3}{8}(1.415) + \frac{1}{4}(2) + \frac{1}{4}(2) + \frac{1}{8}(3) \approx 1.906 \text{ 比特/符号}$$

4.2.3 离散信源的最大熵定理

- **定理:** 当离散信源X中所有符号**等概率**分布时，其熵 $H()$ 取最大值。
- **最大熵公式:**
$$H_{\max} = \log_2 N$$
其中 N 是信源中不同符号的个数。
- **物理意义:** 当信源取等概分布时，其不确定性最大。
- **应用:** 如果能让每个符号/码元平均携带最大的信息量，可以对信源进行变换，使其变换后的符号集具有等概或近似等概的特性。

4.2.4 联合熵与条件熵

- **联合熵 $H(X, Y)$:**

描述一个联合信源 (X, Y) 的平均不确定性。

$$H(X, Y) = - \sum_{i,j} P_{ij} \log_2 P_{ij}$$

- **重要性质:** 当 X 和 Y **统计独立**时, $H(X, Y) = H(X) + H(Y)$ 。
- **条件熵 $H(Y|X)$:**
在已知随机变量 X 的条件下, 随机变量 Y 仍具有的不确定性 (平均信息量)。
 $H(Y|X) = -\sum_{i=1}^N \sum_{j=1}^M P_{ij} \log P_{ij}$
 - **重要性质:** $H(Y|X) \leq H(Y)$ 。这表明, **知识总是有益的**, 知道一个变量 X 的信息, 总会使得另一个相关变量 Y 的不确定性减小或不变。

4.3 离散信道及容量

4.3.1 信道模型

- **离散信道模型:** 其特性可以用一个**转移概率矩阵**来描述, 记为 P 。
 $P = [P_{ij}]$
表示当信道输入为 i 时, 输出为 j 的概率。
- **无记忆信道:** 指信道的输出仅与**当前**的输入有关, 与之前的输入无关。

4.3.2 互信息量

- **互信息定义:** 收到消息 Y 后, 关于 X 的不确定性的**消除量**。
 $I(X; Y) = H(X) - H(X|Y)$
其中 $H(X)$ 是先验不确定性, $H(X|Y)$ 是在收到 Y 后的后验不确定性。
- **平均互信息量:** 信道每传输一个符号, 在接收端获得的**平均信息量**。
 $I(X; Y) = \sum_{i=1}^N \sum_{j=1}^M P_{ij} \log \frac{P_{ij}}{P_i P_j}$
 - **重要性质:** 平均互信息量具有**非负性**, 即 $I(X; Y) \geq 0$ 。

4.3.3 熵函数与平均互信息量间的关系

这些关系非常重要, 可以用韦恩图来辅助理解:

- $H(X, Y) = H(X) + H(Y|X)$ (知道 Y 后, X 不确定性的减小量)
- $H(X, Y) = H(Y) + H(X|Y)$ (知道 X 后, Y 不确定性的减小量)
- $H(X, Y) = H(X) + H(Y)$
- $H(X, Y) = H(X) + H(Y) = H(X|Y) + H(Y|X)$
- $H(X, Y) = H(X) + H(Y)$ (仅当 X, Y 独立时取等号)

4.3.4 离散信道的容量

- **信道容量 C :** 定义为离散信道的**最大信息传输速率**。

$$C = \max_{P_X} I(X; Y) \times R_s$$

其中 R_s 是符号传输速率 (单位: 符号/秒)。信道容量的单位是**比特/秒**。

- **匹配信源**: 能使信道达到其容量的那个**信源概率分布** () 称之为与信道相匹配的信源。
- **⚠ 注意点**:
 - 信道容量 是信道的**固有属性**，由信道的转移概率决定，与信源无关。
 - 而实际的平均信息传输速率 () 则同时与信道和信源有关。

4.3.5 离散无记忆对称信道

- **定义**: 转移概率矩阵中**各行各列**均具有相同的元素集的信道。
- **特性**:
 1. 条件熵 () 是一个**常数**，仅由信道特性决定，与信源统计特性无关。
 2. 若输入信源符号**等概分布**，则输出符号也等概分布。
- **信道容量**:

对于对称信道，当输入为**等概分布**时，信息传输速率达到最大值，即信道容量。

$$C = (\log N ()) \times R_s$$

其中 N 是输出符号的个数， () 是信道的条件熵 (一个常数)。

🌟 本章学习总结

本章作为信息论的基础，核心是围绕“**不确定性**”的度量展开的。

1. **从单个事件出发** ➡ 我们学习了用**自信息** () 来量化一个事件发生所带来的信息量，它与概率成反比。
2. **扩展到整个信源** ➡ 我们引入了**熵** () 的概念，它衡量了一个信源输出符号的平均信息量，也就是信源的平均不确定性。我们还知道了，**等概率分布**的信源熵最大。
3. **考虑信源间的关系** ➡ 通过**联合熵** () 和**条件熵** ()，我们能够量化多个信源之间以及它们相互提供信息后不确定性的变化。
4. **研究信息在信道中的传递** ➡ **互信息** () 成为了关键，它衡量了信道输入和输出之间的关联程度，即通过信道成功传递的平均信息量。
5. **衡量信道的终极能力** ➡ 最后，我们定义了**信道容量**，它是信道能够达到的理论信息传输速率上限。对于给定的信道，通过寻找一个“**匹配信源**”，就可以实现这个最大速率。

总之，本章为我们提供了一套完整的数学工具，用于定量分析信息、信源和信道，是后续学习编码理论的基石。