研究生算法课课堂笔记

上课日期：2016 年 9 月 12 日　　　　　第(2)节课

组长学号及姓名：赵汉宇_1601111272
组员学号及姓名：张雨思_1601214446，蒋捷_1601214437，郭旸泽_ 1601214431

注意：请提交 Word 格式文档

———

第一讲 Information Theory 信息论

1. Information 信息

（1）定义

Information != Knowledge

Information = Surprise

如果事件 i 发生的概率为 Pi，则其信息量大小为 log（1/Pi）。如果事件是广为人知的（Pi 高），那么这个信息则没有什么信息量（例如“太阳明天早上从东方升起”，确定型事件，信息量趋于零）；而如果一个事件是很难发生（Pi 小）的，那么信息量就很大（趋于∞）。信息量大小和 surprise 的程度是正相关的。

（2）举例

① 掷硬币，猜中则给 100 元。若有人提前告诉正确答案，可以付给其 50 元。因为若自己猜，收益期望为 50 元，付超过 50 元就是亏本。

② 如果是掷骰子，则可以付 83 元。可见，掷骰子的结果包含的信息量更大，体现 Information = Surprise 的意义。

③ 若掷硬币，两面的概率分别为 80%和 20%，只付 20 元。因为永远猜 80%的那一面，收益期望为 80 元。

结论：若有 k 种可能性。则 k 的值越大，以及 k 种可能性的概率分布越均匀，则信息量越大。

（3）当信息是相互独立的时候，信息量是可以相加的。但如果不相互独立，两个事件发生的概率之间有相互的影响，故不能直接相加。

2. Entropy 熵

（1）　定义

熵刻画实验结果的平均信息量（期望）可以体现信息的无序程度。其值为概率与信息量的乘积的求和（信息量的加权平均）。

在对信息进行编码时，熵就是平均编码长度。

（2）　性质

① 非负

② 熵与可能性的顺序无关

③ 如果用猜测的概率分布求得的信息量去计算熵值，则不小于真实熵。当且仅当猜测的分布与真实分布相同时取等。

具体的例子：

假设均匀的硬币：0.5*log(1/0.5)*2　小于 0.5*log(1/0,8)+0.5*log(1/0.2)

确实不均匀：0.8log(1/0.8)+0.2log(1/0.2) 小于 0.8*log(1/0.5)+0.2log(1/0.5)

④ 熵值不大于 log k，当且仅当 k 种可能均分时取等。

例：二分查找的理论依据，每次在中间分，能提供的信息量最大。

（3）　举例

①　英语的熵

27 个字符（字母+空格），100,000 个单词（平均长度为 5.5 个字符）

a.　假定连续的字母之间是相互独立的。

假定字符均匀分布，则每个字符的熵是 log27=4.75bits/character。

在真实的字母分布下，每个字符的熵约是 4.03bits/character。因为真实的字母分布不是均等的，有些字母出现频率高，而有些字母出现频率低，降低了熵。

b.　假定连续的单词之间是相互独立的。

假定单词均匀分布，则每个字母的熵是：log(100000)/6.5=2.55bits/character。熵变低是因为形成单词后字母和字母之间有了依赖关系。

在真实的单词分布下，每个字符的熵约是 9.45/6.5=1.45bits/character。原因同 a。

真实的英语中熵还要小得多。说明上下文关系限制了熵。

②　温度和湿度的熵

由于温度（T）和湿度（M）不是独立事件，$H(T, M) \neq H(T) + H(M)$。