自然语言处理

概率: 从随机试验中的事件到实数域的映射函数，用以表示事件发生的可能性.

估计概率值：相同的情况下重复试验N次，当N足够大时，用相对频率近似概率值

条件概率：如果A,B是样本空间的两个事件，在已知事件B发生的情况下，事件A的概率。

条件概率计算：（贝叶期法则）

随机变量：一个随机试验可能有多种不同的结果，到底会出现哪一种，存在一定的概率。简单地说，就是试验结果的函数 (也称为随机变量的概率分布)

在自然语言处理中，一般以句子为处理单位，为了简化问题的复杂性，通常假设一个句子的出现独立于它前面的其他语句，句子的概率近似地被认为符合二项式分布。

联合概率分布和条件概率分布

贝叶斯决策理论

信息论

信息熵：信息量的量化，即描述一个随机变量的不确定性的数量。物理意义：表示信源X每发一个符号（不论发什么符号）所提供的平均信息量。熵越大，不确定性越大（即信息量越大），那么正确估计其值的可能性越小。

在已知部分知识的前提下，关于未知分布最合理的推断应该是符合已知知识最不确定或最大随机的推断.

最大熵的应用：选择模型

自然语言处理中，通常的做法是，根据已知的样本设计特征函数，假设存在k个特征函数

，它们都在建模过程中对输出有影响，那么，所建立的模型应满足所有这些特征的约束，即所建立的模型p应该属于这k个特征函数约束下所产生的所有模型的集合C.使熵H(p)值最大的模型用来推断某种语言现象存在的可能性，或者作为进行某种处理操作的可靠性依据。

联合熵：描述一对随机变量平均所需要的信息量。

长度为n信息，每一个字符的熵:

互信息：反映的是在知道Y的值以后，X的不确定性的减少量。可以理解为Y的值透露了多少关于X的信息量。互信息体现了两变量之间的依赖程度：如果I(X,Y)>>0,表示X与Y是高度相关的话，如果I(X,Y)=0，表明X和Y是相互独立的.

相对熵:也称KL距离，是衡量相同事件空间里两个概率分布相对差距的测度.

交叉熵：用来衡量估计模型与真实概率分布之间差异情况的。等于真实概率分布的熵+估计模型概率分布和真实概率分布之间的相对熵。在设计模型时，目的是使交叉熵最小，从而使模型最接近真实的概率分布. 设计语言模型时，一般用困惑度来代替交叉熵衡量语言模型的好坏。困惑度是交叉熵的指数函数.

噪声信道模型

信息熵可以定量地估计信息源每发送一个符号所提供的平均信息量。

一般情况下，在信号传输的过程中都要进行双重性处理：一方面要对编码进行压缩，尽量消除所有的冗余； 另一方面又要通过增加一定的可控冗余以保障输入信号经地噪声信道传输以后可以很好地恢复原状。

在自然语言处理中，不需要进行编码，一种自然语言的句子可以视为已编码的符号序列，但需要进行解码，使观察到的输出序列更接近于输入。

模拟信道模型，在自然语言处理中，很多问题都可以归结为在给定输出O（可能含有误传信息）的情况下，如何从所有可能的输入I中求解最有可能的那个，即求出使p(I|O)最大的I所有为输入I。用贝叶斯公式求解：

举例：如何想把一个法语句子f翻译成英语e,那么相应的翻译信道模型就是假定法语句子f作为信道模型的输出，它原本就是一个英语句子e，但通过噪声通道传输时被改变成了法语句子f.那么现在需要做的就是如何根据概率p(e)和p(f|e)的计算求出最接近原始英文句子e的解