自然语言处理 第6章阅读笔记

**概率图模型**

在概率图模型中，节点表示变量，节点之间直接相连的边表示相应变量之间的概率关系。

当概率分布P被表示成概率图模型之后，可以用来回答与概率分布P有关的问题，如计算条件概率P(Y|E=e)：在证据e给定的条件下，Y出现的边缘概率。例如，假设S为一个汉语句子，X是句子S切分出来的词序列，那么，汉语句子的分词过程可以看做是推断使得P(X|S)最大的词序列X的分布；在词性标注中，可以看做是在给定序列X的情况下，寻找一组最可能的词性标签分布T，使得P(T|X)最大。【**这些都可以看做是先给出了数据的最终分布情况，求参数使得这种数据分布出现的概率最大，所以这些都可以视为最大后验概率问题。**】

根据图中的边是否有向，可以把概率图分为有向概率图和无向概率图。

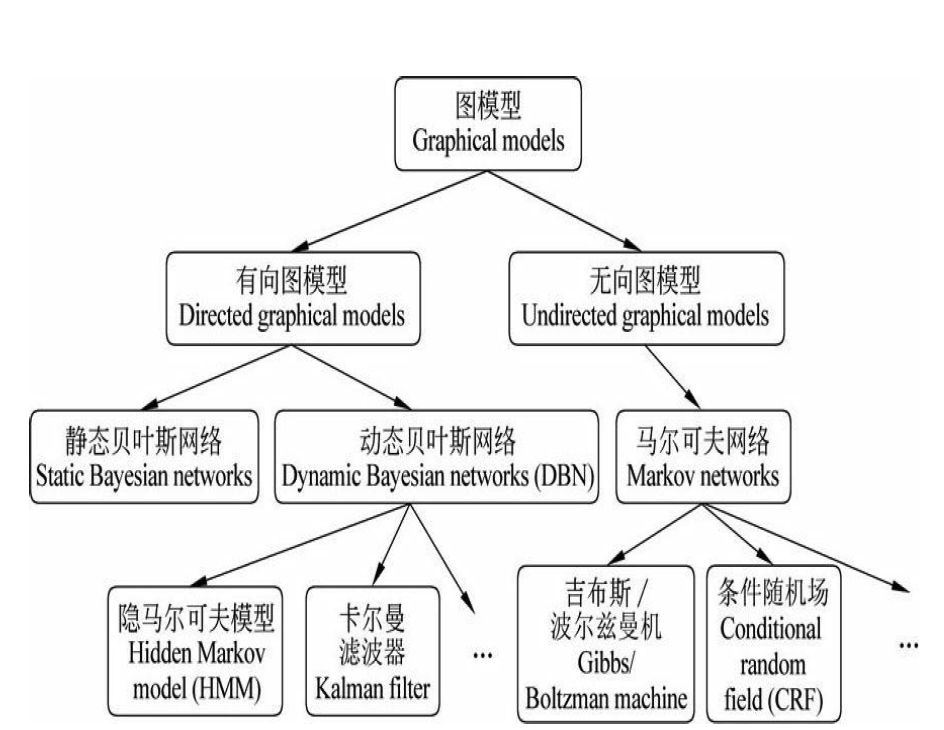


图1：图模型的划分

动态贝叶斯网络（dynamic Bayesian networks , DBN）用于处理随时间变化的动态系统中的推断和预测问题。

隐马尔可夫模型（hidden Markov model, HMM）在语音识别、分词、词性标注中都有应用。

马尔可夫网络（Markov network）又叫做马尔可夫随机场（Markov random field, MRF），马尔可夫网络下的条件随机场（conditional random field, CRF）也常用于序列标注、特征选择、机器翻译等。

玻尔兹曼机（Boltzmann machine）常用于依存句法分析和语义角色标注。

【有个生成式模型和判别式模型的区别，等整章看完再补上】。

**马尔可夫模型**

在介绍隐马尔科夫模型之前，我们先了解马尔可夫模型。

我们常常需要考察一个随机变量序列，这些随机变量之间不是相互独立的，每个随机变量的值依赖于这个序列前面的状态。如果一个系统有N个有限状态,那么随着时间的推移，该系统将从某一个状态转移到另一个状态。为一个随机变量序列，随机变量的取值为状态集S中的某个状态，假定在时间t的状态记为.**那么对该系统的描述通常需要给出当前时刻t的状态与其前面的所有状态的关系：系统在时间t处于状态**的概率取决于其在时间1，2…t-1的状态，该概率为：

****

如果在特定条件下，系统在时间t的状态只与其在时间t-1的状态相关，即：



那么该系统就构成了一个离散的一阶马尔可夫链（Markov chain）。

进一步，如果只考虑式(6-1)独立于时间t的随机过程：



这个随机过程就是马尔可夫模型，其中状态转移概率必须满足以下条件：



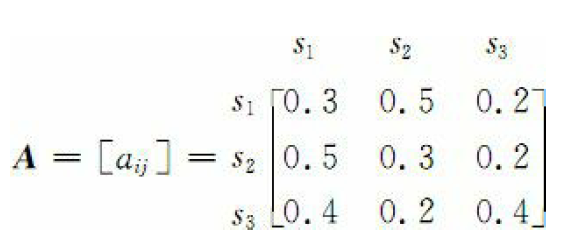
显然，有N个状态的一阶马尔可夫过程共有中状态转移，其个状态转移概率可以表示成一个状态转移矩阵。例如，一段文字中有名词、动词、形容词三类词性出现的情况可以由有三个状态的马尔可夫模型来描述：

状态：名词

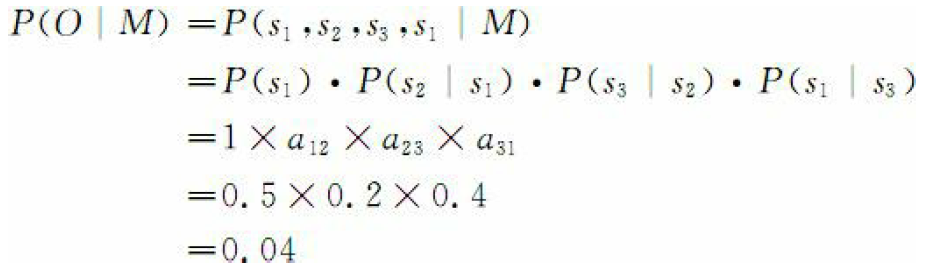
状态：动词

状态：形容词

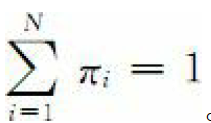
假设状态之间的转移矩阵如下：



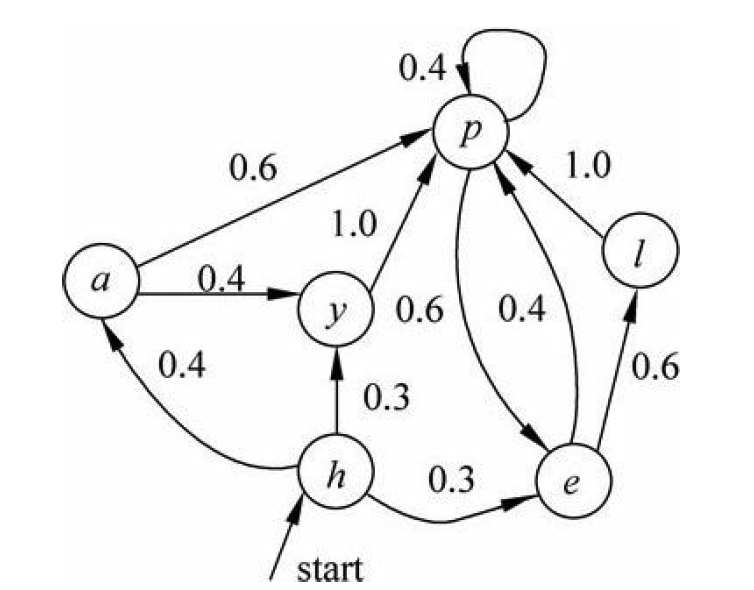
如果在该段文字中某一句子的第一个词为名词，那么根据这一模型M，在该句子中这三类词的出现顺序为O=“名动形名”的概率为：



系统初始化时可以定义一个初始状态的概率向量;并且

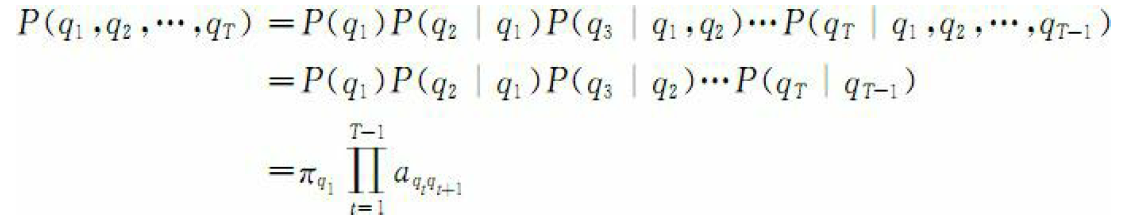


**马尔可夫模型也可以视为随机的有限状态机。如图2所示，圆圈表示状态，状态之间的转移用带箭头的弧表示，弧上的数字为状态转移的概率，初始状态用标记为start的输入箭头表示，假设任何状态都可以作为终止状态，图2中省略了状态转移为0的弧，对于每一个状态来说，发出弧上的概率和为1。从图中可以看出，马尔可夫模型可以看做是一个转移弧上有概率的非确定的有限状态自动机。**

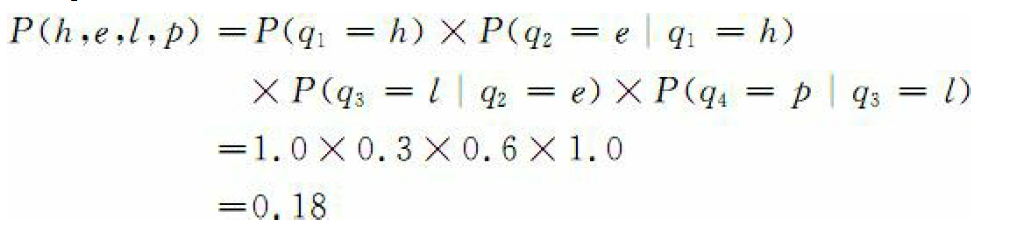
****

**图2：马尔可夫的例子**

一个马尔可夫链的状态序列的概率可以通过计算形成该状态序列的所有状态之间转移弧上的概率乘积而得出，即：



其中。根据图2给出的状态转移概率，我们可以得出：



根据前面第五章的n元语法模型，当n=2时，实际上就是一个马尔可夫模型。但是当n>=3时，就不是一个马尔可夫模型了，因为它不符合马尔可夫模型的基本约束（第N个状态只与前面第N-1个状态相关）。**不过对于n>=3的n元语法模型确定数量的历史来说，可以通过将状态空间描述成多重前面状态的交叉乘积的方式，将其转换为马尔可夫模型。在这种情况下，可以将它称为m阶马尔可夫模型，这里的m是用于预测下一个状态的前面状态的数目，那么n元语法模型就是n-1阶马尔可夫模型。**