隐马尔可夫模型（**Hidden Markov Model**，**HMM**）是可以用于标注问题的统计学模型，描述由隐藏的马尔可夫链随机生成观测序列的过程，属于生成模型。在语音识别、自然语言处理、生物信息、模式识别等领域有着广泛的应用。

1 隐马尔可夫模型的基本概念

1.1 隐马尔可夫模型的定义

隐马尔可夫模型是关于时序的概率模型，描述了一个隐藏的马尔可夫链生成不可观测的状态随机序列，再由各个状态生成对应观测的过程。

定义相关符号

- 设**Q**是所有可能的状态集合，$Q = q_1, q_2, \dots, q_N$，**N**是可能的状态数
- 设**V**是所有可能的观测集合，$V = v_1, v_2, \dots, v_M$，**M**是可能的观测数
- **I**是长度为**T**的状态序列，$I = i_1, i_2, \dots, i_T$
- **O**是长度为**T**的观测序列，$O = o_1, o_2, \dots, o_T$

隐马尔可夫模型主要由三部分组成

- 初始状态概率分布π**(**R^N**)**
- 状态转移概率矩阵**A(**R^N **x** R^N**)**
- 观测生成概率矩阵**B(**R^N **x** R^M**)**

隐马尔可夫模型作了两个基本假设

- 齐次马尔可夫性假设（状态）：假设任意**t**时刻的状态只依赖于前一时刻**t-1**的状态
- 观测独立性假设（序列）：假设任意**t**时刻的观测只依赖于**t**时刻的状态

隐马尔可夫模型可以用于标注问题，这时状态对应着标记。标注问题是给定观测序列去预测对应的状态序列（标记序列），可以假设标注问题的数据是由隐马尔可夫模型生成的，这样我们可以利用隐马尔可夫模型的学习与预测算法进行序列标注。

1.2 隐马尔可夫的生成过程

输入：**A**，**B**，π，观测序列长度**T**

输出：观测序列

Step1：令**t = 1**

Step2：按照初始状态概率分布π随机生成状态i_t

Step3：根据观测概率矩阵**B**生成**t**时刻的观测o_t

Step4：根据状态概率转移矩阵**A**生成**t + 1**时刻的状态i_{t+1}

Step5：令**t = t + 1**，如果**t < T**，转到**Step3**，否则终止

1.3 隐马尔可夫的3个基本研究问题

- 概率计算问题（模型定义），即给定模型的参数$\lambda = (A, B, \pi)$，计算$P(O|\lambda)$的概率
- 学习问题（模型参数求解，训练），给定观测**O**，估计模型参数$\lambda = (A, B, \pi)$，通过极大似然估计求解模型参数，即最大化$P(O|\lambda)$进行求解
- 预测问题（预测），序列标注问题上也称解码问题，给定模型参数$\lambda = (A, B, \pi)$和观测序列**O**，求最大$P(I|O, \lambda)$对应的状态序列

2 概率计算方法

2.1 直接计算法

为了计算$P(O|\lambda)$，枚举所有长度可能为**T**的状态序列，一共有N^T这么多条状态序列，其中任意一条状态序列**I**的长度为**T**，对所有的状态序列进行求和（相当于对联合概率$P(O, I|\lambda)$进行积分，求边缘概率），即$\sum_I P(O|I, \lambda)P(I|\lambda)$时间复杂度为$O(N^T * T)$，不可行，需要引入前向或后向算法

2.2 前向算法

定义前向概率

给定模型参数λ，定义从开头到**t**时刻的观测序列为$o_1, o_2, \ldots, o_t$且**t**时刻状态为q_i的概率为前向概率，记作$\alpha_t(i) = P(o_1, o_2, \ldots, o_t, i_t = q_i|\lambda)$

接下来可以递推地求得前向概率$\alpha_t(i)$以及观测序列概率$P(O|\lambda)$

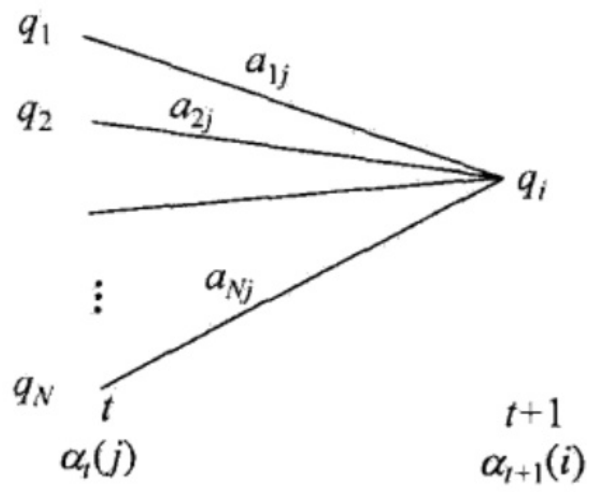
输入：隐马尔可夫模型参数λ，观测序列O

输出：观测序列概率$P(O|\lambda)$

Step1：初始化，$\alpha_1(i) = \pi_i b_i(o_1)$（**t = 1**时刻，状态为$q_i$且由$q_i$状态生成$o_1$观测的概率）

Step2：递推，$\alpha_{t+1}(i) = [\sum_{j=1}^N \alpha_t(j)\alpha_{ji}]b_i(o_{t+1})$，请见下图

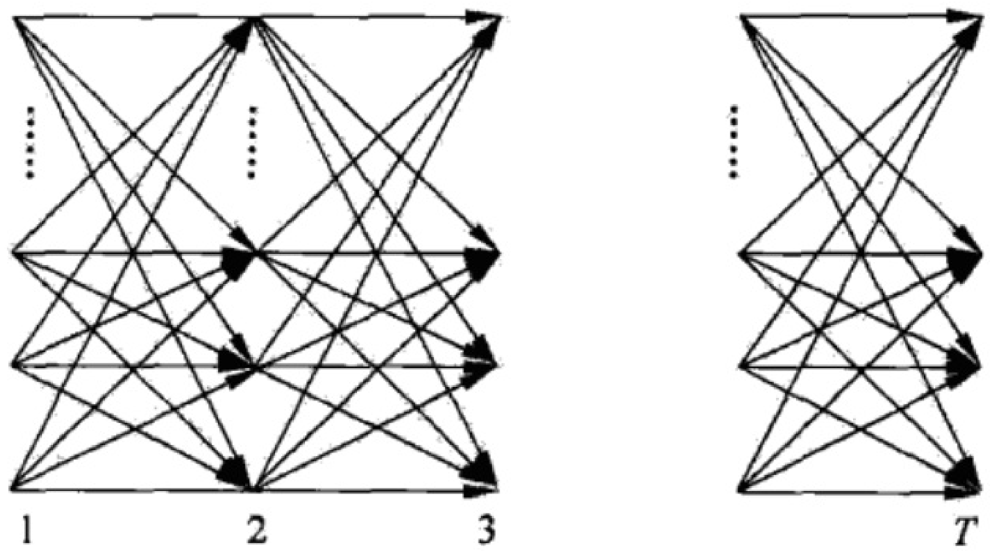
Step3：$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$（**T**时刻的状态有**N**种可能，对这**N**种可能进行求和，$\alpha_T(i) = P(o_1, o_2, \ldots, o_T, i_T = q_i|\lambda)$，求和之后就是$P(o_1, o_2, \ldots, o_T|\lambda)$）



下面分析下它的复杂度

前向算法高效的关键是其局部计算前向概率，然后利用路径结构将前向概率“推算”全局，得到 $P(O|\lambda)$

具体地，在时刻$t = 1, 2, \ldots, T-1$，计算$\alpha_{t+1}(i)$的**N**个值（$i = 1, 2, \ldots, N$），而每个 $\alpha_{t+1}(i)$的计算利用到前一时刻的**N**个$\alpha_t(j)$，减少计算量的原因在于每一次计算直接引入前一个时刻的计算结果，避免重复计算（如下图，如果要计算**T**时刻q_i的概率的话，左边有N^{T-1}这么多条路径可以到达q_i，是指数级增长的），前向算法复杂度为$O(N^2 * T)$（**t**时刻的q_i需要利用**t-1**时刻的**N**个状态，**t**时刻一共需要计算**N**个q_i，一共有**T**个时刻）



2.3 后向算法

与前向相反，不再具体阐述

3 学习算法

这里主要介绍有标记的序列，即有监督学习的方法

假设所有训练样本的长度都相同，训练样本：$\{(O_1, I_1), (O_2, I_2), \ldots, (O_S, I_S)\}$

通过极大似然函数进行估计，也就是统计样本中以下三个参数

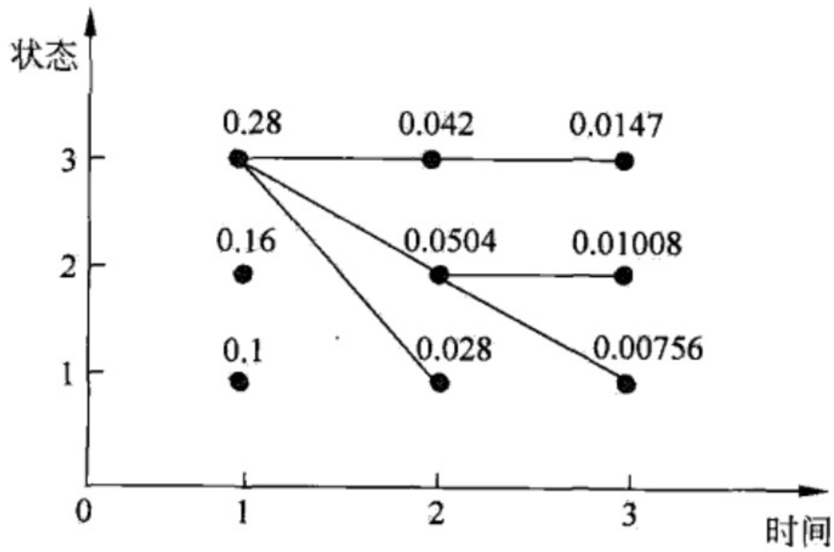
- 状态转移概率a_{ij}，统计状态i转移到状态j的样本数量，归一化
- 观测生成概率$b_j(k)$，统计状态j且观测为k的数量，归一化
- 初始状态概率π_i，统计第一个状态是i的数量，归一化

4 预测算法

解决隐马尔可夫模型的预测问题常用维特比算法(Viterbi)，即用动态规划求概率最大的路径，这条概率最大的路径对应着一个状态序列。维特比算法的大致思想就是，通过前向概率计算t时刻，每个状态i的最大概率，并且记录使得状态i取最大时的前驱状态j，从t=1开始计算，在T时刻，选取一个概率最大的状态i^*，通过前驱一直回溯，从而得到最优状态序列。下面引入两个关键变量δ和ψ

- 定义在t时刻以状态i为终点的最大概率为$\delta_t(i) = \max_j[\delta_{t-1}(j)a_{ji}]b_i(o_t)$
- 定义使得t时刻以状态i为终点的最大概率的前一个状态$\psi_t(i) = \mathrm{argmax}_j[\delta_{t-1}(j)a_{ji}]$

如下图所示，$\delta_3(2) = 0.01008$的意思是，在第3个时刻以状态2为终点的最大概率（在第3个时刻，到达状态2这个结点的路径有很多，我们需要的是最大的那条，图中只画出了最大的，即0.28->0.0504->0.01008），$\psi_3(2) = 2$的意思是，这个最大概率发生时的前一个状态是2



参考资料

1.《统计学习方法》第10章 隐马尔可夫模型