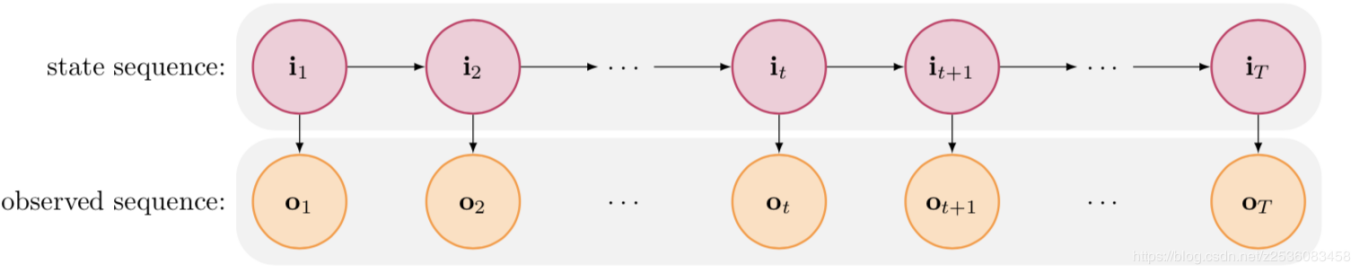
# 14.隐马尔可夫模型

## 14.1.隐马尔可夫模型概念

**隐马尔可夫模型（Hidden Markov Model，简称HMM）**是一种著名的有向图模型，主要用于时序数据建模，在语音识别、自然语言处理等领域有广泛应用。

下面以李航的《统计学习方法》来进行隐马尔可夫模型的讲解

**定义：**隐马尔可夫模型是关于时序的概率模型，描述由一个隐藏的马尔可夫链随机生成不可观测的状态随机序列，再由各个状态生成一个观测从而产生观测随机序列的过程。隐藏的马尔可夫链随机生成的状态的序列，称为状态序列（state sequence）；每个状态生成一个观测，而由此产生的观测的随机序列，称为观测序列（observation sequence）。序列的每一个位置又可以看作是一个时刻。



**隐马尔可夫模型中的数学符号定义**

是所有种可能的状态的集合：

是所有种可能的观测的集合：

是长度为的状态序列，是对应的观测序列：，

是状态转移概率矩阵：其中，，

是观测概率矩阵：其中，，

是初始状态概率向量：其中，，

隐马尔可夫模型由初始状态概率向量、状态转移概率矩阵和观测概率矩阵决定。和决定状态序列，决定观测序列。因此，隐马尔可夫模型可以用三元符号表示，即：



**两个基本假设：**

1.齐次马尔可夫性假设，即假设隐藏的马尔可夫链在任意时刻的状态只依赖于其前一时刻的状态，与其他时刻的状态及观测无关，也与时刻无关：



2.观测独立性假设，即假设任意时刻的观测只依赖于该时刻的马尔可夫链的状态，与其他观测及状态无关：



**三个基本问题：**

1.概率计算问题：给定模型和观测序列，计算在模型下观测序列出现的概率；

2.学习问题：已知观测序列，估计模型参数，使得在该模型下观测序列概率最大，即用极大似然估计的方法估计参数；

3.预测问题：也称为解码问题，已知模型和观测序列，求对给定观测序列条件概率最大的状态序列，即给定观测序列，求最有可能的对应的状态序列。

## 14.2.概率计算问题

这里使用三种算法进行求解，分别为：直接计算法、前向算法、后向算法

**（1）直接计算法**

对于求最直接的方法就是按照概率公式直接计算，即：





其中，表示给定模型参数时，产生状态序列的概率：



表示给定模型参数且状态序列为时，产生观测序列的概率：



所以



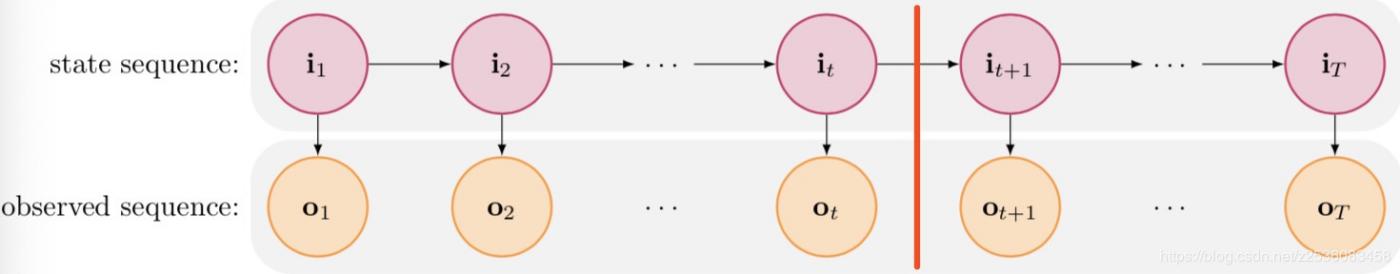


其中，共有种可能，计算的时间复杂度为，所以上式整体的时间复杂度为，显然这种算法是不可行的。

（2）前向算法

前向算法采用动态规划的思想，首先定义前向概率：给定隐马尔可夫模型，定义到时刻t部分观测序列为且状态为的概率为前向概率，记作：





根据前向概率的定义可推得：



于是求解的问题被转化为了求解前向概率的问题。由前向概率的定义可知：







以此类推，可得如下递推公式



因此可以递推求得：



将上式所求结果代回



即可求得

此时来计算一下时间复杂度：

对于任意，由于有种取值，所以需要计算次，在计算每次时，需要对进行次求和（假设已经算出），所以计算的时间复杂度为,而总共有次计算，所以时间复杂度为，从指数级别的复杂度降成了多项式级别的复杂度。

例（盒子和球模型）假设有3个盒子，每个盒子里面都装有红白两种颜色的球，盒子里的红白球数如下表所示：

|  |  |  |  |
| --- | --- | --- | --- |
| 盒子 | 1 | 2 | 3 |
| 红球数 | 5 | 4 | 7 |
| 白球数 | 5 | 6 | 3 |

按照下面的方法抽球，产生一个球的颜色的观测序列：首先以0.2、0.4、0.4的概率从1、2、3号盒子中选取一个盒子，从这个盒子里随机抽出1个球，记录其颜色后放回，接着按以下概率选取下一个盒子

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| 1 | 0.5 | 0.2 | 0.3 |
| 2 | 0.3 | 0.5 | 0.2 |
| 3 | 0.2 | 0.3 | 0.5 |

确定转移的盒子后，再从盒子里随机抽出1个球，记录其颜色后放回，如此下去，重复3次，最终得到的观测序列为={红，白，红}，适用前向算法计算。

解：该例子是一个典型的隐马尔可夫模型，由题意可得隐马尔可夫模型的参数（三要素）为：

首先，计算初值







然后，进行递推计算













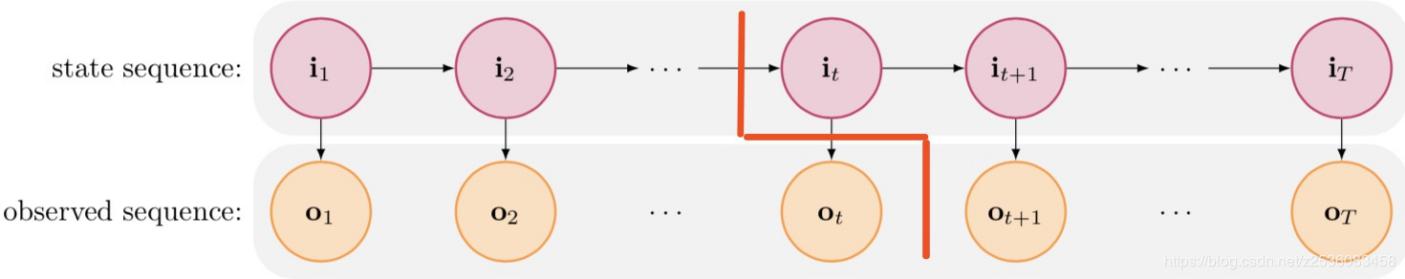
最后，可得



（3）后向算法

同前向算法类似，也是采用动态规划的思想，首先定义后向概率：给定隐马尔可夫模型，定义在时刻t状态为的条件下，从t+1到T的部分观测序列为的概率为后向概率，记作：











以此类推，可得如下递推公式



根据递推公式可求得，又



从而也可求得

后向算法和前向算法类似，时间复杂度都是。

综上可以看出前向算法和后向算法都是先计算局部概率，然后递推到全局，每一时刻的概率计算都会用上前一时刻计算出的结果，整体的时间复杂度大约为，明显小于直接计算法的。

为了后续计算的方便，这里首先利用前向和后向概率公式可以得到关于单个状态和两个状态概率的一些计算公式：

1.给定模型参数和观测，在时刻t处于状态的概率，记为：

可以通过前向概率和后向概率进行计算，推导如下：



又由前向概率和后向概率的定义可知：



所以



2.给定模型参数和观测，在时刻t处于状态且在时刻t+1处于状态的概率，记为：



可以通过前向概率和后向概率进行计算，推导如下：



又







所以



## 14.3.学习问题

这里使用两种算法进行求解，分别为：监督学习算法、Baum-Welch算法（无监督学习）

（1）监督学习方法：

假设已给出训练数据包含个长度相同的观测序列和对应的状态序列，那么可以利用极大似然估计法来估计隐马尔可夫模型的参数，具体方法如下：

转移概率的估计：，其中，为样本中时刻t处于状态而到时刻t+1移到状态的频数。

观测概率的估计：，其中，为样本中状态为，其对应观测为的频数

初始状态概率的估计为个样本中初始状态为的频率.

由于监督学习需要使用标注的训练数据，而人工标注训练数据往往代价很高，有时就会利用无监督学习的方法。

（2）Baum-Welch算法

如果只有观测序列数据，而没有状态序列数据，那么隐马尔可夫模型就是一个含有隐变量的概率模型：



如果要对它进行参数估计，则可以采用EM算法来实现。

确定完全数据的对数似然函数：

此时观测数据为，未观测数据为，则完全数据为，完全数据的对数似然函数为：



其中，所以可以进一步推得：





E步：求函数



其中，是隐马尔可夫模型参数的当前估计值，是要极大化的隐马尔可夫模型参数。为了便于后续计算，函数还可以作如下恒等变形：





由于接下来仅极大化，可以看做常数项进行略去，所以函数可以进一步化简为：







M步：极大化函数，由于要极大化的参数在上式中单独地出现在3个项中，所以只需对各项分别极大化。

求：函数中的第1项可以写成：



展开求和符号中的得



这里，中的指第一个求和符号中的，不受第二个求和符号约束，所以可以提出来





由于需要满足约束，利用拉格朗日乘子法，写出拉格朗日函数



对拉格朗日函数关于求偏导并令结果为0：



对于求偏导，所以求和符号可以去掉，得





利用，对上式两边关于求和可得：









将其代回可得：



解得



其中，表示给定模型参数和观测，在时刻t处于状态的概率。

求：函数中的第2项可以写成：



由于在中只和、有关，所以单独将这两项拆开得



这里，不受求和约束，所以可以提出来





由于需要满足约束，同样利用拉格朗日乘子法，写出拉格朗日函数



对拉格朗日函数关于求偏导并令结果为0



对于求偏导，所以和、相关的求和符号可以去掉，得





利用，对上式两边关于求和可得：







将其代回 可得



解得



分子分母同时除以



其中  表示给定模型参数和观测，在时刻t处于状态且在时刻t+1处于状态的概率。

求：函数中的第3项可以写成：



同样，由于在中和有关，所以单独将这项拆开得







由于需要满足约束，同样利用拉格朗日乘子法，写出拉格朗日函数



对拉格朗日函数关于求偏导并令结果为0



对于求偏导，所以和、相关的求和符号可以去掉，但是这里由于中受前面求和符号的影响，的取值为，表示状态为时，观测为的概率，所以只有当时，其导数才有值，否则导数为0，但是由于并不知道中哪个为的项，所以可以引入指示函数，其值为，因此上式可变为





利用，对上式两边关于求和可得：



在中，肯定可以找到一个，使得指示函数的值为1，而其他个指示函数的值都为0，因此，上式可以变为





将其代回可得



解得



分子分母同时除以



这里，将指示函数放在了求和符号中，表示在求和时只遍历那些的项。

## 14.4.预测问题

这里使用两种算法进行求解，分别为：近似算法、维特比算法

（1）近似算法

近似算法思想：在每个时刻t选择在该时刻最有可能出现的状态，从而得到一个状态序列，将它作为预测的结果。具体算法如下：

给定隐马尔可夫模型入和观测序列，在时刻t处于状态的概率是



在每一时刻t最有可能的状态是

，

从而得到状态序列

近似算法的优点是计算简单，其缺点是不能保证预测的状态序列整体是最有可能的状态序列，因为预测的状态序列可能有实际不发生的部分。事实上，上述方法得到的状态序列中有可能存在转移概率为0的相邻状态，即对某些时。尽管如此，近似算法仍然是有用的。

（2）维特比算法

维特比算法实际是用动态规划解隐马尔可夫模型预测问题，即用动态规划求概率最大路径，这时一条路径对应着一个状态序列。具体算法如下：

定义在时刻t状态为的所有单个路径中概率最大值为

，

由上述定义可知：







依次类推可得如下递推公式：



只记录了取得概率最大的值，但是这个概率是从哪个节点过来的并不知道，因此需要记录是从哪个节点转移到当前节点的

定义在时刻t状态为的所有单个路径中概率最大的路径的第t-1个结点为



因此，取，则，，...，

例：基于前向算法的例题，已知观测序列为：={红，白，红}，试求最优状态序列，即最优路径

解：同样，首先获取模型参数为：

首先，进行初始化。在t=1时，对每一个状态，，求状态为观测为红的概率，记此概率为，则

， 

代入实际数据可得

， ， 

记，

在t=2时，对每个状态，，求在t=1时状态为观测为红并在t=2时状态为观测为白的路径的最大概率，记此最大概率为，则



同时，对每个状态，，记录概率最大路径的前一个状态：

，

计算：



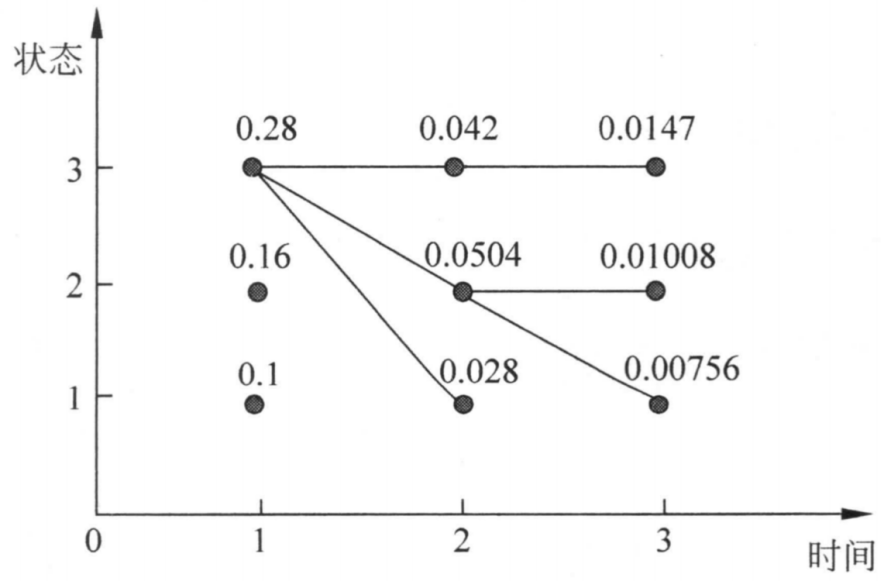








同样，在t=3时：

综上，以表示最优路径的概率，则



最优路径的终点是：



由最优路径的终点，逆向找到：

在t=2时， 

在t=1时， 

于是求得最优路径，即最优状态序列。