Table of Contents

- 8 隐马尔科夫模型
  - 8.1 隐马尔科夫模型定义
  - 8.2 概率计算算法
  - 8.3 学习算法
  - 8.4 预测算法

8 隐马尔科夫模型

8.1 隐马尔科夫模型定义

状态集合

$$Q = \{q_1, q_2, \ldots, q_N\} \quad |Q| = N$$

观测集合

$$V = \{v_1, v_2, \ldots, v_M\} \quad |V| = M$$

状态序列

$$I = \{i_1, i_2, \ldots, i_t, \ldots, i_T\} \quad i_t \in Q \quad (t = 1, 2, \ldots, T)$$

观测序列

$$O = \{o_1, o_2, \ldots, o_t, \ldots, o_T\} \quad o_t \in V \quad (t = 1, 2, \ldots, T)$$

状态转移矩阵

$$A = \left[a_{ij}\right]_{N \times N}$$

在t时刻处于状态q_i的条件下，在$t+1$时刻转移到状态q_j的概率

$$a_{ij} = P\left(i_{t+1} = q_j | i_t = q_i\right) \quad (i = 1, 2, \ldots, N) \quad (j = 1, 2, \ldots, M)$$

观测概率矩阵

$$B = \left[b_j(k)\right]_{N \times M}$$

在t时刻处于状态q_i的条件下，生成观测v_k的概率

$$b_j(k) = P\left(o_t = v_k | i_t = q_j\right) \quad (k = 1, 2, \ldots, M) \quad (j = 1, 2, \ldots, N)$$

初始概率向量

$$\pi = (\pi_i)$$

在时刻$t = 1$处于状态q_i的概率

$$\pi_i = P(i_1 = q_i) \quad (i = 1, 2, \ldots, N)$$

隐马尔科夫模型

$$\lambda = (A, B. \pi)$$

隐马尔科夫模型基本假设：

1. 齐次马尔科夫性假设：在任意时刻t的状态只依赖于时刻$t-1$的状态。
$$P\left(i_t|i_{t-1},o_{t-1},\ldots,i_1,o_1\right)=P\left(i_t|i_{t-1}\right)\quad(t=1,2,\ldots,T)$$
2. 观测独立性假设：任意时刻t的观测只依赖于时刻t的状态。
$$P\left(o_t|i_T,o_T,i_{T-1},o_{T-1},\ldots,i_{t+1},o_{t+1},i_t,i_{t-1},o_{t-1},\ldots,i_1,o_1\right)=P\left(o_t|i_t\right)\quad(t=1,2,\ldots,T)$$

观测序列生成算法:
输入：隐马尔科夫模型$\lambda=(A,B.\pi)$,观测序列长度T;
输出：观测序列$O=\{o_1,o_2,\ldots,o_t,\ldots,o_T\}$;

1. 由初始概率向量π产生状态i_1;
2. $t=1$;
3. 由状态i_t的观测概率分布$b_j(k)$生成o_t;
4. 由状态i_t的状态转移概率分布$a_{i_ti_{t+1}}$生成状态$i_{t+1}\quad(i_{t+1}=1,2,\ldots,N)$;
5. $t=t+1$；如果$t<T$，转至3.；否则，结束。

隐马尔科夫模型的3个基本问题：

1. 概率计算：已知$\lambda=(A,B.\pi)$和$O=\{o_1,o_2,\ldots,o_t,\ldots,o_T\}$，计算$P(O|\lambda)$
2. 学习：已知$O=\{o_1,o_2,\ldots,o_t,\ldots,o_T\}$，计算 $\lambda^*=\arg\max P(O|\lambda)$
3. 预测（编码）：已知$\lambda=(A,B.\pi)$和$O=\{o_1,o_2,\ldots,o_t,\ldots,o_T\}$，计算 $I^*=\arg\max P(I|O\lambda)$

8.2 概率计算算法

前向概率
$$\alpha_t(i)=P\left(o_1,o_2,\ldots,o_t,i_t=q_i|\lambda\right)$$

给定模型λ，时刻t部分观测序列为$o_1,o_2,\ldots,o_t$且状态为q_i的概率。

前向概率递推计算
$$\alpha_t(i)=P\left(o_1,o_2,\ldots,o_t,i_t=q_i|\lambda\right)=P\left(\boxed{i_t=q_i,o_1^t}\right)$$
λ后面全部省略，不代表不存在，简写。

qi有N个可能取值
$$=\sum_{j=1}^{N}P\left(i_{t-1}=q_j,\boxed{i_t=q_i,o_1^{t-1},o_t}\right)$$
概率加法规则P(X)=在Y的所有条件求和∑P(X,Y)

概率乘法规则:P(X,Y)=P(Y|X)P(X)
$$=\sum_{j=1}^{N}\underbrace{P\left(i_t=q_i,o_t|i_{t-1}=q_j,o_1^{t-1}\right)}_{Y|X}\cdot\boxed{P\left(i_{t-1}=q_j,o_1^{t-1}\right)}$$
X

最后面的若干0怎么没了？
齐次马尔科夫性假设：当前状态只与前一状态相关
$$=\sum_{j=1}^{N}\boxed{P\left(i_t=q_i,o_t|i_{t-1}=q_j\right)}\cdot\boxed{\alpha_{t-1}(j)}$$
$$=\sum_{j=1}^{N}\underbrace{\boxed{P\left(o_t|i_t=q_i,i_{t-1}=q_j\right)}}_{Y|X}\cdot\underbrace{P\left(i_t=q_i|i_{t-1}=q_j\right)}_{X}\cdot\alpha_{t-1}(j)$$
概率乘法规则

观测独立性假设
当前观测只与当前状态有关
$$=\sum_{j=1}^{N}\boxed{b_i\left(o_t\right)}\cdot a_{ji}\cdot\alpha_{t-1}(j)$$

概率计算

$$
\begin{aligned}
P(O|\lambda) &= P\left(o_1^T|\lambda\right) \\
&\overset{\text{概率加法}}{=} \sum_{i=1}^{N} P\left(o_1^T, i_T = q_i\right) \\
&\overset{\text{定义：}\alpha t(i)=P(o1,...ot,it=qi|\lambda)}{=} \sum_{i=1}^{N} \alpha_T(i)
\end{aligned}
$$

观测序列概率计算的前向算法：
输入：隐马尔科夫模型λ,观测序列O;
输出：观测序列概率$P(O|\lambda)$;

1. 初值

$$\alpha_1(i) = \pi_i b_i(o_1) \quad (t = 1, 2, \ldots, N)$$

2. 递推 对$t = 1, 2, \ldots, T-1$

$$\alpha_{t+1}(i) = \sum_{j=1}^{N} b_i(o_{t+1}) \cdot a_{ji} \cdot \alpha_t(j) \quad (t = 1, 2, \ldots, N)$$

3. 终止

$$P(O|\lambda) = \sum_{j=1}^{N} \alpha_T(i)$$

后向概率

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \ldots, o_T | i_t = q_i \lambda)$$

给定模型λ，时刻t状态为q_i的条件下，从时刻$t+1$到时刻T的部分观测序列为$o_{t+1}, o_{t+2}, \ldots, o_T$的概率。

后向概率递推计算

$$
\begin{aligned}
\beta_t(i) &= P(o_{t+1}, o_{t+2}, \ldots, o_T | i_t = q_i, \lambda) = P\left(o_{t+1}^T | i_t = q_i\right) \\
&= \frac{P\left(o_{t+1}^T, i_t = q_i\right)}{P(i_t = q_i)} \\
&= \frac{\sum_{j=1}^{N} P\left(o_{t+1}^T, i_t = q_i, i_{t+1} = q_j\right)}{P(i_t = q_i)} \\
&= \sum_{j=1}^{N} \frac{P\left(o_{t+1}^T | i_t = q_i, i_{t+1} = q_j\right) \cdot P\left(i_t = q_i, i_{t+1} = q_j\right)}{P(i_t = q_i)} \\
&= \sum_{j=1}^{N} P\left(o_{t+1}^T | i_{t+1} = q_j\right) \cdot \frac{P\left(i_{t+1} = q_j | i_t = q_i\right) \cdot P(i_t = q_i)}{P(i_t = q_i)} \\
&= \sum_{j=1}^{N} P\left(o_{t+2}^N, o_{t+1} | i_{t+1} = q_j\right) \cdot a_{ij} \\
&= \sum_{j=1}^{N} P\left(o_{t+2}^T | i_{t+1} = q_j\right) \cdot P\left(o_{t+1} | i_{t+1} = q_j\right) \cdot a_{ij} \\
&= \sum_{j=1}^{N} \beta_{t+1}(j) \cdot b_j(o_{t+1}) \cdot a_{ij}
\end{aligned}
$$

概率计算

$$
\begin{aligned}
P(O|\lambda) &= P\left(o_1^T|\lambda\right) \\
&= \sum_{i=1}^{N} P\left(o_1^T, i_1 = q_i\right) \\
&= \sum_{i=1}^{N} P(i_1 = q_i) \cdot P(o_1|i_1 = q_i) \cdot P\left(o_2^T|i_1 = q_i\right) \\
&= \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(i)
\end{aligned}
$$

观测序列概率计算的后向算法：
输入：隐马尔科夫模型λ,观测序列O;
输出：观测序列概率$P(O|\lambda)$；

1. 初值
$$\beta_T(i) = 1 \quad (t = 1, 2, \ldots, N)$$
2. 递推 对$t = T-1, T-2, \ldots, 1$
$$\beta_t(i) = \sum_{j=1}^{N} \beta_{t+1}(j) \cdot b_j(o_{t+1}) \cdot a_{ij} \quad (t = 1, 2, \ldots, N)$$
3. 终止
$$P(O|\lambda) = \sum_{j=1}^{N} \pi_i b_i(o_1) \beta_1(i)$$

$P(O|\lambda)$的前向概率、后向概率的表示

$$
\begin{aligned}
&P(O|\lambda) = P\left(o_1^T\right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} P\left(o_1^t, o_{t+1}^T, i_t = q_i, i_{t+1} = q_j\right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} P\left(o_1^t, i_t = q_i, i_{t+1} = q_j\right) P\left(o_{t+1}^T|i_{t+1} = q_j\right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} P\left(o_1^t, i_t = q_i\right) P\left(i_{t+1} = q_j|i_t = q_i\right) P\left(o_{t+1}^T|i_{t+1} = q_j\right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} P\left(o_1^t, i_t = q_i\right) P\left(i_{t+1} = q_j|i_t = q_i\right) P\left(o_{t+1}|i_{t+1} = q_j\right) P\left(o_{t+2}^T|i_{t+1} = q_j\right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \qquad t = 1, 2, \cdots, T-1
\end{aligned}
$$

给定模型λ和观测O，在时刻t处于状态q_i的概率

$$
\begin{aligned}
\gamma_t(i) &= P(i_t = q_i | O, \lambda) \\
&= \frac{P(i_t = q_i, O|\lambda)}{P(O|\lambda)} \\
&= \frac{P(i_t = q_i, O|\lambda)}{\sum_{j=1}^N (i_t = q_i, O|\lambda)} \\
&= \frac{P\left(o_1^t, i_t = q_i\right) P\left(o_{t+1}^T | i_t = q_i\right)}{\sum_{j=1}^N P\left(o_1^t, i_t = q_i\right) P\left(o_{t+1}^T | i_t = q_i\right)} \\
&= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(i)\beta_t(i)}
\end{aligned}
$$

给定模型λ和观测O，在时刻t处于状态q_i且在时刻$t+1$处于状态q_j的概率

$$
\begin{aligned}
\xi_t(i,j) &= P\left(i_t = q_i, i_{t+1} = q_j | O, \lambda\right) \\
&= \frac{P\left(i_t = q_i, i_{t+1} = q_j, O|\lambda\right)}{P(O|\lambda)} \\
&= \frac{P\left(i_t = q_i, i_{t+1} = q_j, O|\lambda\right)}{\sum_{i=1}^N \sum_{j=1}^N P\left(i_t = q_i, i_{t+1} = q_j, O|\lambda\right)} \\
&= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}
\end{aligned}
$$

在观测O下状态i出现的期望

$$
\sum_{t=1}^{T} \gamma_t(i) = \sum_{t=1}^{T} P(i_t = q_i | O, \lambda)
$$

在观测O下由状态i转移的期望

$$
\sum_{t=1}^{T-1} \gamma_t(i) = \sum_{t=1}^{T-1} P(i_t = q_i | O, \lambda)
$$

在观测O下由状态i转移到状态j的期望

$$
\sum_{t=1}^{T-1} \xi_t(i,j) = \sum_{t=1}^{T-1} P\left(i_t = q_i, i_{t+1} = q_j | O, \lambda\right)
$$

8.3 学习算法

将观测序列作为观测数据O,将状态序列作为隐数据I，则应马尔科夫模型是含有隐变量的概率模型

$$
P(O|\lambda) = \sum_I P(O|I,\lambda) P(I|\lambda) \quad \text{加法准则 + 条件概率}
$$

完全数据

$$
(O, I) = (o_1, o_2, \cdots, o_T, i_1, i_2, \cdots, o_T)
$$

完全数据的对数似然函数

$$
\log P(O, I|\lambda)
$$

$Q\left(\lambda, \overline{\lambda}\right)$函数

$$
\begin{aligned}
Q\left(\lambda, \overline{\lambda}\right) &= E_I\left[\log P(O, I|\lambda)\,|O, \overline{\lambda}\right] \\
&= \sum_I \log P(O, I|\lambda)\, P\left(I|O, \overline{\lambda}\right) \\
&= \sum_I \log \frac{P(O, I|\lambda)\, \boxed{P\left(O, I|\overline{\lambda}\right)}}{P\left(O|\overline{\lambda}\right)}
\end{aligned}
$$

分子分母同乘P(O|λ_)再联合概率

其中，$\overline{\lambda}$是隐马尔科夫模型参数的当前估计值，λ是隐马尔科夫模型参数。

由于对最大化$Q\left(\lambda, \overline{\lambda}\right)$函数，$P\left(O|\overline{\lambda}\right)$为常数因子，
以及

$$
P(O, I|\lambda) = \pi_{i_1} b_{i_1}(o_1)\, a_{i_1 i_2} b_{i_2}(o_2) \cdots a_{i_{T-1} i_T} b_T(o_T)
$$

所以求$Q\left(\lambda, \overline{\lambda}\right)$函数对$\lambda$的最大

$$
\lambda = \arg\max Q\left(\lambda, \overline{\lambda}\right) \Leftrightarrow \arg\max \sum_I \log P(O, I|\lambda)\, P\left(O, I|\overline{\lambda}\right)
$$

$$
= \boxed{\sum_I \log \pi_{i_1} P\left(O, I|\overline{\lambda}\right)} + \boxed{\sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}}\right) P\left(O, I|\overline{\lambda}\right)} + \boxed{\sum_I \left(\sum_{t=1}^{T} \log b_{i_t}(o_t)\right) P\left(O, I|\overline{\lambda}\right)}
$$

对三项分别进行极大化：

1.

$$
\begin{aligned}
&\max \sum_{I} \log \pi_{i_1} P\left(O, I|\overline{\lambda}\right) = \sum_{i=1}^{N} \log \pi_{i_1} P\left(O, i_1 = i|\overline{\lambda}\right) \\
&s.t. \sum_{i=1}^{N} \pi_i = 1
\end{aligned}
$$

构造拉格朗日函数，对其求偏导，令结果为0

$$
\frac{\partial}{\partial \pi_i}\left[\sum_{i=1}^{N} \log \pi_{i_1} P\left(O, i_1 = i|\overline{\lambda}\right) + \gamma\left(\sum_{i=1}^{N} \pi_i - 1\right)\right] = 0
$$

得

$$
\begin{aligned}
&P\left(O, i_1 = i|\overline{\lambda}\right) + \gamma \pi_i = 0 \\
&\sum_{i=1}^{N}\left[P\left(O, i_1 = i|\overline{\lambda}\right) + \gamma \pi_i\right] = 0 \\
&\sum_{i=1}^{N} P\left(O, i_1 = i|\overline{\lambda}\right) + \gamma \sum_{i=1}^{N} \pi_i = 0 \\
&P\left(O|\overline{\lambda}\right) + \gamma = 0 \\
&\gamma = -P\left(O|\overline{\lambda}\right)
\end{aligned}
$$

代入$P\left(O, i_1 = i|\overline{\lambda}\right) + \gamma \pi_i = 0$，得

$$
\begin{aligned}
\pi_i &= \frac{P\left(O, i_1 = i|\overline{\lambda}\right)}{P\left(O|\overline{\lambda}\right)} \\
&= \gamma_1(i)
\end{aligned}
$$

2.

$$
\begin{aligned}
&\max \sum_{I}\left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}}\right) P\left(O, I|\overline{\lambda}\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \log a_{ij} P\left(O, i_t = i, i_{t+1} = j|\overline{\lambda}\right) \\
&s.t. \sum_{j=1}^{N} a_{ij} = 1
\end{aligned}
$$

得

$$
\begin{aligned}
a_{ij} &= \frac{\sum_{t=1}^{T-1} P\left(O, i_t = i, i_{t+1} = j|\overline{\lambda}\right)}{\sum_{t=1}^{T-1} P\left(O, i_t = i|\overline{\lambda}\right)} \\
&= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}
\end{aligned}
$$

3.

$$
\begin{aligned}
&\max \sum_{I}\left(\sum_{t=1}^{N} \log b_{i_t}(o_t)\right) P\left(O, I|\overline{\lambda}\right) = \sum_{j=1}^{N} \sum_{t=1}^{T} \log b_j(o_t) P\left(O, i_t = j|\overline{\lambda}\right) \\
&s.t. \sum_{k=1}^{M} b_j(k) = 1
\end{aligned}
$$

Baum-Welch算法：
输入：观测数据$O=(o_1, o_2, \cdots, o_T)$
输出：隐马尔科夫模型参数

1. 初始化
   对$n=0$，选取$a_{ij}^{(0)}, b_j(k)^{(0)}, \pi_i^{(0)}$，得到模型$\lambda^{(0)}=\left(a_{ij}^{(0)}, b_j(k)^{(0)}, \pi_i^{(0)}\right)$
2. 递推
   对$n=1,2,\cdots,$

$$\begin{aligned} a_{ij}^{(n+1)} &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\ b_j(k)^{(n+1)} &= \frac{\sum_{t=1, o_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \\ \pi_i^{(n+1)} &= \frac{P\left(O, i_1=i \mid \overline{\lambda}\right)}{P\left(O \mid \overline{\lambda}\right)} \end{aligned}$$

   其中，右端各值按观测数据$O=(o_1, o_2, \cdots, o_T)$和模型$\lambda^{(n)}=\left(A^{(n)}, B^{(n)}, \pi^{(n)}\right)$计算。
3. 终止
   得到模型$\lambda^{(n+1)}=\left(A^{(n+1)}, B^{(n+1)}, \pi^{(n+1)}\right)$

8.4 预测算法

在时刻t状态为i的所有单个路径$(i_1, i_2, \cdots, i_t)$中概率最大值

$$\delta_t(i)=\max_{i_1, i_2, \cdots, i_{t-1}} P(i_t=i, i_{t-1}, \cdots, i_1, o_t, \cdots, o_1 \mid \lambda) \qquad i=1,2,\cdots,N$$

得递推公式

$$\begin{aligned} &\delta_{t+1}(i)=\max_{i_1, i_2, \cdots, i_t} P(i_{t+1}=i, i_t, \cdots, i_1, o_{t+1}, \cdots, o_1 \mid \lambda) \\ &=\max_{1 \leq j \leq N}\left[\max_{i_1, i_2, \cdots, i_{t-1}} P(i_{t+1}=i, i_t=j, i_{t-1}, \cdots, i_1, o_{t+1}, o_t, \cdots, o_1 \mid \lambda)\right] \\ &=\max_{1 \leq j \leq N}\left[\max_{i_1, i_2, \cdots, i_{t-1}} P(i_{t+1}=i, i_t=j, i_{t-1}, \cdots, i_1, o_t, o_{t-1}, \cdots, o_1 \mid \lambda) P(o_{t+1} \mid i_{t+1}=i, \lambda)\right] \\ &=\max_{1 \leq j \leq N}\left[\max_{i_1, i_2, \cdots, i_{t-1}} P(i_t=j, i_{t-1}, \cdots, i_1, o_t, o_{t-1}, \cdots, o_1 \mid \lambda) P(i_{t+1}=i \mid i_t=j, \lambda) P(o_{t+1} \mid i_{t+1}=i, \lambda)\right] \\ &=\max_{1 \leq j \leq N}\left[\delta_t(j) a_{ji}\right] b_i(o_{t+1}) \qquad i=1,2,\cdots,N \end{aligned}$$

在时刻t状态为i的所有单个路径$(i_1, i_2, \cdots, i_t)$中概率最大值的路径的第$t-1$个结点

$$\psi_t(i)=\arg\max_{1 \leq j \leq N}\left[\delta_{t-1}(j) a_{ji}\right] \qquad i=1,2,\cdots,N$$

维特比算法：
输入：模型$\lambda = (A, B, \pi)$和观测数据$O = (o_1, o_2, \cdots, o_T)$
输出：最优路径$I^* = \left(i_1^*, i_2^*, \cdots, i_T^*\right)$

1. 初始化

$$\begin{aligned} & \delta_1(i) = \pi_i b_i(o_1) \qquad i = 1, 2, \cdots, N \\ & \psi_1(i) = 0 \end{aligned}$$

2. 递推
对$t = 2, 3, \cdots, T$

$$\begin{aligned} & \delta_t(i) = \max_{1 \leq j \leq N} \left[\delta_{t-1}(j) a_{ji}\right] b_i(o_t) \qquad i = 1, 2, \cdots, N \\ & \psi_t(i) = \arg\max_{1 \leq j \leq N} \left[\delta_{t-1}(j) a_{ji}\right] \qquad i = 1, 2, \cdots, N \end{aligned}$$

3. 终止

$$\begin{aligned} P^* &= \max_{1 \leq j \leq N} \delta_T(i) \\ i_T^* &= \arg\max_{1 \leq j \leq N} \left[\delta_T(i)\right] \end{aligned}$$

4. 最优路径回溯
对$t = T-1, T-2, \cdots, 1$

$$i_t^* = \psi_{t+1}\left(i_{t+1}^*\right)$$

求得最优路径$I^* = \left(i_1^*, i_2^*, \cdots, i_T^*\right)$