



Statistical Computing

Chap. 2.2: Hidden Markov Model

LIU, Ran

March 20, 2025

LIU, Ran - Department of Statistics @BNU

Summary

介绍

学习问题

评估问题

解码问题



LIU, Ran - Department of Statistics @BNU

隐马尔可夫模型 (HMM)

隐马尔可夫模型 (Hidden Markov Model, HMM) 是一种用于建模序列数据的概率模型，它适用于描述具有潜在未观察状态的动态过程。

HMM 由两个基本组成部分构成：

- **隐藏状态 (Hidden States)**: 不可直接观察到的状态，通常用 $X = \{x_1, x_2, \dots, x_T\}$ 表示。
- **观察序列 (Observation Sequence)**: 可观测的数据序列，表示为 $Y = \{y_1, y_2, \dots, y_T\}$ 。

LIU, Ran - Department of Statistics @BNU

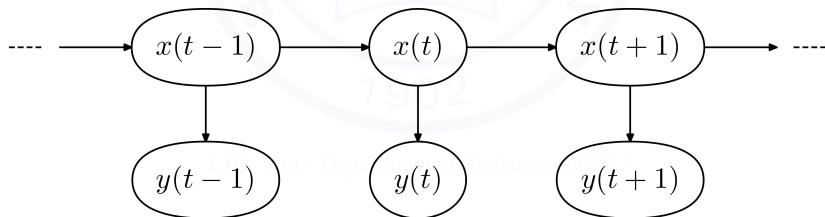
HMM 的基本假设是：

- 1 **马尔可夫性 (Markov Property)**：隐藏状态的转移仅依赖于前一个时间步骤的状态，即：

$$p(x_t | x_{t-1}, x_{t-2}, \dots) = p(x_t | x_{t-1}).$$

- 2 **观测独立性**：当前的观测值仅依赖于当前的隐藏状态：

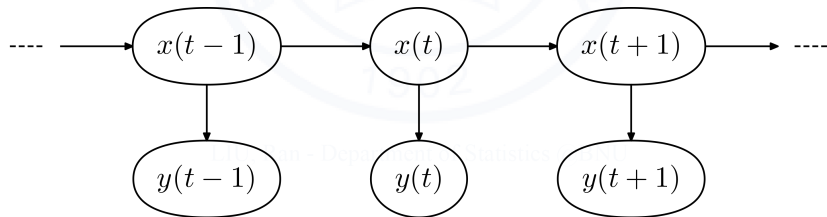
$$p(y_t | x_t, x_{t-1}, \dots) = p(y_t | x_t).$$



数据生成过程

在 HMM 中，观测数据的生成过程如下：

- 1 根据初始状态概率分布 Π 选择初始状态 x_1 。
- 2 在状态 x_1 下，根据观察概率矩阵 B 生成观测值 y_1 。
- 3 根据状态转移概率矩阵 A ，由 x_1 生成下一个状态 x_2 。
- 4 在状态 x_2 下，用 B 生成观测值 y_2 。
- 5 依次重复该过程，直到生成完整的观测序列 Y 。



HMM 的基本要素

HMM 由以下几个核心要素完全定义：

- **隐藏状态集合 (Hidden State Set)**：可能的隐藏状态集合，表示为 $S = \{s_1, s_2, \dots, s_N\}$ 。
- **观测值集合 (Observation Set)**：可能的观测值集合，表示为 $O = \{o_1, o_2, \dots, o_K\}$ 。

Liu, Ran - Department of Statistics @BNU

HMM 的核心参数

- **初始状态概率分布 (Initial State Distribution)**: 设 $\Pi = \{\pi_i\}$, 其中:

$$\pi_i = p(x_1 = s_i),$$

表示系统初始时处于状态 s_i 的概率。

- **状态转移概率矩阵 (State Transition Probability Matrix)**: 设 $A = \{a_{ij}\}$, 其中:

$$a_{ij} = p(x_t = s_j | x_{t-1} = s_i),$$

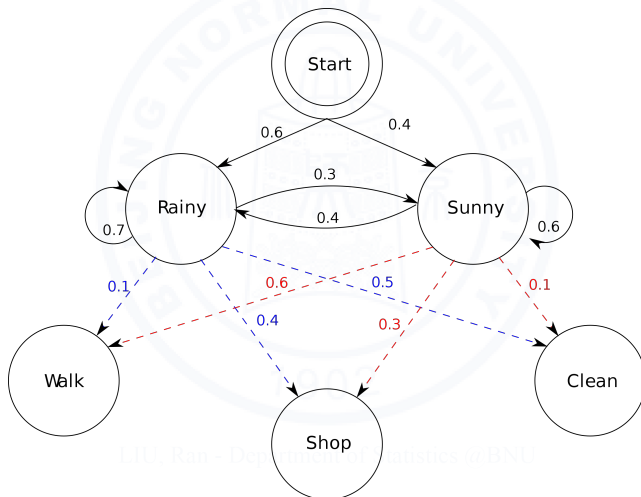
表示从状态 s_i 转移到状态 s_j 的概率。

- **观察概率矩阵 (Observation Probability Matrix, Emission Distribution)**: 设 $B = \{b_{ij}\}$, 其中:

$$b_{ij} = p(y_t = o_j | x_t = s_i),$$

表示在隐藏状态 s_i 下观测到 o_j 的概率。

下图表示在不同天气下不同行为的概率，假设隐藏状态 $X = \{\text{Rainy}, \text{Sunny}\}$ 。
观测值 $Y = \{\text{Walk}, \text{Shop}, \text{Clean}\}$ 。



HMM 的三个基本问题

HMM 需要解决以下三个基本问题：

- 1 **学习问题 (Learning Problem)**: 给定观测序列 $\{y_1, y_2, \dots, y_T\}$, 估计模型参数 $\theta = (A, B, \Pi)$ 。
- 2 **评估问题 (Evaluation Problem)**: 给定模型参数 θ 和观测序列 Y , 计算观测序列的概率:

$$p(Y|\theta) = p(y_1, y_2, \dots, y_T|A, B, \Pi)$$

- 3 **解码问题 (Decoding Problem)**: 给定模型参数 θ 和观测序列 Y , 找到最可能的隐藏状态序列 $X^* = \{x_1^*, x_2^*, \dots, x_T^*\}$, 即:

$$X^* = \arg \max_X p(X|Y, \theta)$$

Summary

介绍

学习问题

评估问题

解码问题



LIU, Ran - Department of Statistics @BNU

解决学习问题

学习问题 (Learning Problem): 给定观测序列 $Y = (y_1, y_2, \dots, y_T)$, 估计 HMM 模型参数 $\theta = (A, B, \pi)$ 。

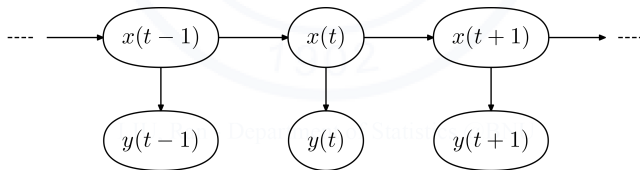
Baum-Welch 算法 (基于期望-最大化 (Expectation-Maximization, EM) 算法): 该算法通过前向和后向迭代计算条件概率, 得到 Q 函数, 再优化更新参数估计值。

LIU, Ran - Department of Statistics @BNU

完全数据的似然函数

令 $\theta = (A, B, \Pi)$, 根据 HMM 的数据生成方式, 可写出完全数据的似然函数: (黑板推导 10-21)

$$\begin{aligned}
 p(X, Y | \theta) &= p_{\theta}(x_1)p_{\theta}(y_1|x_1) \cdot p_{\theta}(x_2|x_1)p_{\theta}(y_2|x_2) \\
 &\quad \times p_{\theta}(x_3|x_2)p_{\theta}(y_3|x_3) \cdot p_{\theta}(x_4|x_3)p_{\theta}(y_4|x_4) \dots \\
 &= p_{\theta}(x_1)p_{\theta}(y_1|x_1) \prod_{t=1}^{T-1} p_{\theta}(x_{t+1}|x_t)p_{\theta}(y_{t+1}|x_{t+1}).
 \end{aligned}$$



参数化对数似然函数

取对数后，得到对数似然函数：

$$\ell(\theta|X, Y) = \log p_{\theta}(x_1) + \sum_{t=1}^{T-1} \log p_{\theta}(x_{t+1}|x_t) + \sum_{t=1}^T \log p_{\theta}(y_t|x_t).$$

将概率用参数表达，对数似然函数可写为：

$$\begin{aligned} \ell(\theta|X, Y) = & \sum_{i=1}^N I(x_1 = s_i) \log \pi_i \\ & + \sum_{t=1}^{T-1} \sum_{i=1}^N \sum_{j=1}^N I(x_t = s_i, x_{t+1} = s_j) \log a_{ij} \\ & + \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^K I(x_t = s_i, y_t = o_j) \log b_{ij}. \end{aligned}$$

E 步：计算 Q 函数

在 EM 算法中，E 步（期望步）计算 Q 函数：

$$Q(\theta \mid \theta^{(l)}) = E[\ell(\theta \mid X, Y) \mid Y, \theta^{(l)}]$$

$$= \sum_{i=1}^N p(x_1 = s_i \mid Y, \theta^{(l)}) \log \pi_i \quad (1)$$

$$+ \sum_{t=1}^{T-1} \sum_{i=1}^N \sum_{j=1}^N p(x_t = s_i, x_{t+1} = s_j \mid Y, \theta^{(l)}) \log a_{ij} \quad (2)$$

$$+ \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^K I(y_t = o_j) p(x_t = s_i \mid Y, \theta^{(l)}) \log b_{ij}. \quad (3)$$

接下来，我们分别计算这些条件概率。

我们先计算边际后验概率 $p(x_t | Y, \theta^{(l)})$ 。我们将 Y 分成给定 x_t 条件下独立的两部分：

$$y_1, \dots, y_t \mid x_t \mid y_{t+1}, \dots, y_T$$

由贝叶斯公式，我们有：

$$\begin{aligned} p(x_t | Y, \theta^{(l)}) &\propto p(x_t, Y | \theta^{(l)}) \\ &\propto p(y_1, y_2, \dots, y_t, x_t | \theta^{(l)}) \\ &\quad \times p(y_{t+1}, y_{t+2}, \dots, y_T | y_1, y_2, \dots, y_t, x_t, \theta^{(l)}) \\ &\propto p(y_1, y_2, \dots, y_t, x_t | \theta^{(l)}) \quad (a) \\ &\quad \times p(y_{t+1}, y_{t+2}, \dots, y_T | x_t, \theta^{(l)}). \quad (b) \end{aligned}$$

其中：

- (a) 部分：表示前向概率，即从初始状态到 x_t 生成观测的概率。
- (b) 部分：表示后向概率，即从 x_t 继续生成剩余观测的概率。

前向算法 (Forward Algorithm)

我们先来看 (a) 式，定义前向概率：

$$\alpha_i^{(l)}(t) := p(y_1, y_2, \dots, y_t, x_t = s_i \mid \theta^{(l)}).$$

当 t 等于 1 时，我们有

$$p(y_1, x_1 = s_i \mid \theta^{(l)}) = p(y_1 \mid x_1 = s_i, \theta^{(l)})p(x_1 = s_i \mid \theta^{(l)})$$

因此初始条件为 $\alpha_i^{(l)}(1) = b_i^{(l)}(y_1)\pi_i^{(l)}$ ，观测概率也可写为 $b_i^{(l)}(y_t) = \sum_{k=1}^K I(y_t = o_k)b_{ik}^{(l)}$ ，接下来推导 $\alpha_i^{(l)}(t)$ 的递归公式。考虑第 t 个：

$$\begin{aligned}\alpha_i^{(l)}(t) &= p(y_1, y_2, \dots, y_t, x_t = s_i \mid \theta^{(l)}) \\ &= p(y_t \mid y_1, y_2, \dots, y_{t-1}, x_t = s_i, \theta^{(l)}) \cdot p(y_1, y_2, \dots, y_{t-1}, x_t = s_i \mid \theta^{(l)}).\end{aligned}$$

Liu, Ran - Department of Statistics @BNU

其中观测 y_t 仅依赖于当前隐藏状态 x_t :

$$p(y_t|y_1, \dots, y_{t-1}, x_t = s_i, \theta^{(l)}) = p(y_t|x_t = s_i, \theta^{(l)}) = b_i^{(l)}(y_t).$$

为了凑 $\alpha_j^{(l)}(t-1)$, 我们插入 x_{t-1} :

$$\alpha_i^{(l)}(t) = b_i^{(l)}(y_t) \sum_{j=1}^N p(y_1, \dots, y_{t-1}, x_{t-1} = s_j, x_t = s_i | \theta^{(l)})$$

(提出 x_t , 并利用 HMM 性质)

$$\begin{aligned} &= b_i^{(l)}(y_t) \sum_{j=1}^N p(y_1, \dots, y_{t-1}, x_{t-1} = s_j | \theta^{(l)}) p(x_t = s_i | x_{t-1} = s_j, \theta^{(l)}) \\ &= b_i^{(l)}(y_t) \sum_{j=1}^N \alpha_j^{(l)}(t-1) a_{ji}^{(l)}. \end{aligned}$$

Liu, Ran - Department of Statistics @BNU

前向计算流程：

$$\{\alpha_i^{(l)}(1)\}_i \rightarrow \{\alpha_i^{(l)}(2)\}_i \rightarrow \cdots \rightarrow \{\alpha_i^{(l)}(T)\}_i, \quad i = 1, 2, \dots, N.$$

它计算的是：

$$\begin{aligned} \{p(y_1, x_1 = s_i \mid \theta^{(l)})\}_i &\rightarrow \{p(y_1, y_2, x_2 = s_i \mid \theta^{(l)})\}_i \rightarrow \cdots \\ &\rightarrow \{p(y_1, y_2, \dots, y_T, x_T = s_i \mid \theta^{(l)})\}_i. \end{aligned}$$

LIU, Ran - Department of Statistics @BNU

后向算法 (Backward Algorithm)

再来看 (b), 定义后向概率:

$$\beta_i^{(l)}(t) = p(y_{t+1}, y_{t+2}, \dots, y_T \mid x_t = s_i, \theta^{(l)}).$$

我们定义初始为 $\beta_i^{(l)}(T) = 1$ 。当 t 等于 $T-1$ 时, 我们有:

$$\begin{aligned} \beta_i^{(l)}(T-1) &= p(y_T \mid x_{T-1} = s_i, \theta^{(l)}) \\ &= \sum_j p(y_T \mid x_T = s_j, \theta^{(l)}) p(x_T = s_j \mid x_{T-1} = s_i, \theta^{(l)}) \\ &= \sum_{j=1}^N b_j^{(l)}(y_T) \beta_j^{(l)}(T) a_{ij}^{(l)} \end{aligned}$$

接下来我们推导迭代公式。

考虑第 t 个:

$$\beta_i^{(l)}(t) = p(y_{t+1}, y_{t+2}, \dots, y_T | x_t = s_i, \theta^{(l)})$$

(为了凑 $\beta_j^{(l)}(t+1)$, 我们插入 x_{t+1})

$$= \sum_{j=1}^N p(y_{t+1}, y_{t+2}, \dots, y_T, x_{t+1} = s_j | x_t = s_i, \theta^{(l)})$$

(提出 x_{t+1}, y_{t+1})

$$= \sum_{j=1}^N p(y_{t+2}, \dots, y_T | x_{t+1} = s_j, \theta^{(l)}) p(y_{t+1}, x_{t+1} = s_j | x_t = s_i, \theta^{(l)}).$$

根据 HMM 的性质, 我们有:

$$\begin{aligned} \beta_i^{(l)}(t) &= \sum_{j=1}^N \beta_j^{(l)}(t+1) p(y_{t+1} | x_{t+1} = s_j, \theta^{(l)}) p(x_{t+1} = s_j | x_t = s_i, \theta^{(l)}) \\ &= \sum_{j=1}^N b_j^{(l)}(y_{t+1}) \beta_j^{(l)}(t+1) a_{ij}^{(l)}. \end{aligned}$$

后向计算流程 ($\{\beta_i^{(l)}(T)\}_i$ 因为不能用原先的定义, 所以定义固定为 1):

$$\{\beta_i^{(l)}(T-1)\}_i \rightarrow \cdots \rightarrow \{\beta_i^{(l)}(1)\}_i, \quad i = 1, 2, \dots, N.$$

它计算的是:

$$\{p(y_T \mid x_{T-1} = s_i, \theta^{(l)})\}_i \rightarrow \cdots \rightarrow \{p(y_2, y_3, \dots, y_T \mid x_1 = s_i, \theta^{(l)})\}_i.$$

LIU, Ran - Department of Statistics @BNU

利用前向-后向概率，我们可以计算：

- 单个状态的后验概率（即 $\gamma_i^{(l)}(t)$ ）：

$$p(x_t = s_i | Y, \theta^{(l)}) \propto \alpha_i^{(l)}(t) \beta_i^{(l)}(t).$$

定义：

$$\gamma_i^{(l)}(t) := p(x_t = s_i | Y, \theta^{(l)}) = \frac{\alpha_i^{(l)}(t) \beta_i^{(l)}(t)}{\sum_{i=1}^N \alpha_i^{(l)}(t) \beta_i^{(l)}(t)}.$$

- 两个相邻状态的联合后验概率（即 $\xi_{ij}^{(l)}(t)$ ）：

$$\begin{aligned} \xi_{ij}^{(l)}(t) &:= p(x_t = s_i, x_{t+1} = s_j | Y, \theta^{(l)}) \\ &\propto p(x_t = s_i, x_{t+1} = s_j, Y | \theta^{(l)}). \end{aligned}$$

还是将 Y 划分成条件独立的两部分，上式等于：

$$\begin{aligned}
 & p(x_t = s_i, x_{t+1} = s_j, Y \mid \theta^{(l)}) \\
 &= p(y_1, y_2, \dots, y_t, x_t = s_i \mid \theta) \cdot p(y_{t+1}, y_{t+2}, \dots, y_T, x_{t+1} = s_j \mid x_t = s_i, \theta) \\
 & \text{(将第二个部分的 } x_{t+1}, y_{t+1} \text{ 提出，再根据 HMM 性质)} \\
 &= p(y_1, y_2, \dots, y_t, x_t = s_i \mid \theta) \cdot p(y_{t+2}, \dots, y_T \mid x_t = s_i, \theta) \\
 & \quad \cdot p(y_{t+1} \mid x_{t+1} = s_j) p(x_{t+1} = s_j \mid x_t = s_i) \\
 &= \alpha_i^{(l)}(t) \beta_j^{(l)}(t+1) b_j^{(l)}(y_{t+1}) a_{ij}^{(l)}.
 \end{aligned}$$

因此，计算 $\xi_{ij}^{(l)}(t)$ ：

$$\xi_{ij}^{(l)}(t) = \frac{\alpha_i^{(l)}(t) \beta_j^{(l)}(t+1) b_j^{(l)}(y_{t+1}) a_{ij}^{(l)}}{\sum_i \sum_j \alpha_i^{(l)}(t) \beta_j^{(l)}(t+1) b_j^{(l)}(y_{t+1}) a_{ij}^{(l)}}.$$

LIU, Ran - Department of Statistics @BNU

回到最初的三个条件分布，我们有

$$\begin{aligned}
 (1) &= \sum_{i=1}^N p(x_1 = s_i \mid Y, \theta^{(l)}) \log \pi_i \\
 &= \sum_{i=1}^N \gamma_i^{(l)}(1) \log \pi_i \\
 (2) &= \sum_{t=1}^{T-1} \sum_{i=1}^N \sum_{j=1}^N p(x_t = s_i, x_{t+1} = s_j \mid Y, \theta^{(l)}) \log a_{ij} \\
 &= \sum_{t=1}^{T-1} \sum_{i=1}^N \sum_{j=1}^N \xi_{ij}^{(l)}(t) \log a_{ij} \\
 (3) &= \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^K I(y_t = o_j) p(x_t = s_i \mid Y, \theta^{(l)}) \log b_{ij} \\
 &= \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^K I(y_t = o_j) \gamma_i^{(l)}(t) \log b_{ij}
 \end{aligned}$$

最大化 $Q(\theta \mid \theta^{(l)})$ ，使用拉格朗日乘子法，求导后得到参数的估计值：

- 初始状态概率：

$$\frac{\partial Q}{\partial \pi_i} = \frac{\gamma_i^{(l)}(1)}{\pi_i} \implies \pi_i^{(l+1)} = \frac{\gamma_i^{(l)}(1)}{\sum_{i=1}^N \gamma_i^{(l)}(1)}.$$

- 状态转移概率：

$$\frac{\partial Q}{\partial a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_{ij}^{(l)}(t)}{a_{ij}} \implies a_{ij}^{(l+1)} = \frac{\sum_{t=1}^{T-1} \xi_{ij}^{(l)}(t)}{\sum_{j=1}^N \sum_{t=1}^{T-1} \xi_{ij}^{(l)}(t)}.$$

- 观测概率：

$$\begin{aligned} \frac{\partial Q}{\partial b_{ik}} &= \frac{\sum_{t=1}^T \gamma_i^{(l)}(t) I(y_t = o_k)}{b_{ik}} \\ \implies b_{ik}^{(l+1)} &= \frac{\sum_{t=1}^T \gamma_i^{(l)}(t) I(y_t = o_k)}{\sum_{k=1}^K \sum_{t=1}^T \gamma_i^{(l)}(t) I(y_t = o_k)}. \end{aligned}$$

若观测概率的建模比较复杂，比如为高斯混合模型，则将 $\sum_{j=1}^K I(y_t = o_j) b_{ij}$ 还原成 $b_i(y_t)$ ，其中观测输出的参数设为 ψ_i ，观测概率密度函数为 $f_{\psi_i}(y_t)$ 。我们写出 Q 函数关于观测概率的部分：

$$\begin{aligned} (3) &= \sum_{t=1}^T \sum_{i=1}^N p(x_t = s_i \mid Y, \theta^{(l)}) \log b_i(y_t) \\ &= \sum_{t=1}^T \sum_{i=1}^N \gamma_i^{(l)}(t) \log f_{\psi_i}(y_t) \end{aligned}$$

则关于观测概率的参数 ψ_i 的梯度就为：

$$\frac{\partial Q}{\partial \psi_i} = \frac{\sum_{t=1}^T \gamma_i^{(l)}(t) \nabla f_{\psi_i}(y_t)}{f_{\psi_i}(y_t)}$$

令其等于 0 可得相应的解析更新公式。

如果使用神经网络建模观测概率，可所有状态都共享同一网络参数，但输入变为状态和观测值，即 $f_\psi(s_i, y_t)$ ，Q 函数相应的部分变为

$$\begin{aligned} (3) &= \sum_{t=1}^T \sum_{i=1}^N p(x_t = s_i \mid Y, \theta^{(l)}) \log b_i(y_t) \\ &= \sum_{t=1}^T \sum_{i=1}^N \gamma_i^{(l)}(t) \log f_\psi(s_i, y_t) \end{aligned}$$

相应的梯度为：

$$\frac{\partial Q}{\partial \psi} = \sum_{t=1}^T \sum_{i=1}^N \gamma_i^{(l)}(t) \frac{\nabla f_\psi(s_i, y_t)}{f_\psi(s_i, y_t)}$$

可利用梯度下降更新网络参数。

Baum-Welch 算法流程

- Step 1: 初始化设定初始参数 $\theta_0 = (\Pi, A, B)$ 。
- Step 2: E 步骤 (Expectation Step) 使用前向-后向算法计算 $\alpha_i^{(l)}(t)$ 和 $\beta_i^{(l)}(t)$, 以及条件期望 $\gamma_i^{(l)}(t)$ 和 $\xi_{ij}^{(l)}(t)$ 。
- Step 3: M 步骤 (Maximization Step) 计算新的参数 $\theta^{(l+1)} = (\Pi^{(l+1)}, A^{(l+1)}, B^{(l+1)})$, 即更新 $\pi_i^{(l+1)}$, $a_{ij}^{(l+1)}$ 和 $b_{ik}^{(l+1)}$ 。
- Step 4: 重复迭代直到参数 $\theta^{(l)}$ 收敛。

Liu, Ran - Department of Statistics @BNU

Summary

介绍

学习问题

评估问题

解码问题



LIU, Ran - Department of Statistics @BNU

解决评估问题

在 HMM (Hidden Markov Model) 中, 评估问题 (Evaluation Problem) 指的是:

- 给定模型参数 $\theta = (\Pi, A, B)$ 和观测序列 $Y = (y_1, y_2, \dots, y_T)$, 计算该观测序列出现的概率:

$$p(Y|\theta) = p(y_1, y_2, \dots, y_T | A, B, \Pi).$$

- 枚举所有可能的隐藏状态序列 X 计算该概率的复杂度为 $\mathcal{O}(N^T)$, 使用前向-后向算法 (Forward-Backward Algorithm) 可以降低复杂度至 $\mathcal{O}(N^2T)$ 。(N 是状态种类数。)

Liu, Ran - Department of Statistics @BNU

回顾前向概率定义：

$$\alpha_i(t) = p(y_1, y_2, \dots, y_t, x_t = s_i | \theta),$$

递推公式：

- 初始化 ($t = 1$): $\alpha_i(1) = \pi_i b_i(y_1), \quad i = 1, 2, \dots, N.$
- 递推 ($t = 2$ 到 T):

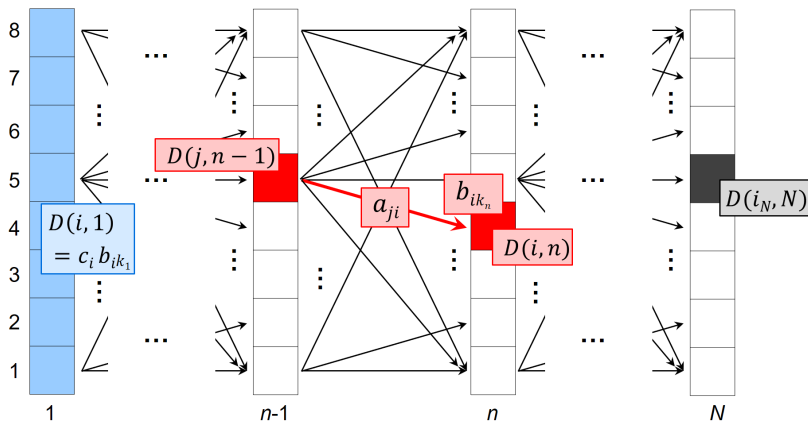
$$\alpha_j(t) = \sum_{i=1}^N \alpha_i(t-1) a_{ij} b_j(y_t), \quad j = 1, 2, \dots, N.$$

- 终止 (计算观测序列概率):

$$p(Y|\theta) = \sum_{i=1}^N p(y_1, y_2, \dots, y_T, x_T = s_i | \theta) = \sum_{i=1}^N \alpha_i(T).$$

复杂度为 $\mathcal{O}(N^2T)$

Figure 5.27 from [Müller, FMP, Springer 2015]



回顾后向概率定义：

$$\beta_i(t) = p(y_{t+1}, y_{t+2}, \dots, y_T | x_t = s_i, \theta),$$

递推公式：

- 初始化 ($t = T$): $\beta_i(T) = 1, \quad i = 1, 2, \dots, N.$
- 递推 ($t = T - 1$ 到 1):

$$\beta_i(t) = \sum_{j=1}^N a_{ij} b_j(y_{t+1}) \beta_j(t+1), \quad i = 1, 2, \dots, N.$$

- 终止 (计算观测序列概率):

$$\begin{aligned} p(Y|\theta) &= \sum_{i=1}^N p(y_1, y_2, \dots, y_T, x_1 = s_i | \theta) \\ &= \sum_{i=1}^N p(x_1 = s_i | \theta) p(y_1 | x_1 = s_i, \theta) p(y_2, \dots, y_T | x_1 = s_i, \theta) \\ &= \sum_{i=1}^N \pi_i b_i(y_1) \beta_i(1). \end{aligned}$$

前向-后向算法的等价性

如果要验证：

$$\begin{aligned} p(Y|\theta) &= \sum_{i=1}^N \alpha_i(T) \quad (\text{前向算法}) \\ &= \sum_{i=1}^N \pi_i b_i(y_1) \beta_i(1) \quad (\text{后向算法}). \end{aligned}$$

两种方法均可用于计算 $p(Y|\theta)$ ，在不同情况下可选择最优方法。

LIU, Ran - Department of Statistics @BNU

Summary

介绍

学习问题

评估问题

解码问题



LIU, Ran - Department of Statistics @BNU

解决解码问题

解码问题 (Decoding Problem): 给定模型参数 $\theta = (\Pi, A, B)$ 和观测序列 $Y = (y_1, y_2, \dots, y_T)$, 找到最可能的隐藏状态序列:

$$X^* = \{x_1^*, x_2^*, \dots, x_T^*\},$$

即:

$$X^* = \arg \max_X p(X|Y, \theta)$$

枚举所有可能的隐藏状态序列 X 计算其后验概率, 复杂度为 $\mathcal{O}(N^T)$, 不可行。维特比算法 (Viterbi Algorithm) 提供了一种动态规划方法, 将复杂度降低至 $\mathcal{O}(N^2T)$ 。

Liu, Ran - Department of Statistics @BNU

维特比算法

我们希望找到最可能的隐藏状态序列：

$$X^* = \arg \max_X p(X|Y, \theta).$$

根据条件概率定义公式，可写为：

$$X^* = \arg \max_X \frac{p(X, Y|\theta)}{p(Y|\theta)} = \arg \max_X p(X, Y|\theta).$$

由于分母 $p(Y|\theta)$ 与 X 无关，因此等价于最大化联合概率：

$$p(X, Y|\theta) = p(x_1, y_1, x_2, y_2, \dots, x_T, y_T|\theta).$$

Liu, Ran - Department of Statistics @BNU

为简便起见，我们定义符号：

$$x_{1:t} = \{x_1, x_2, \dots, x_t\}, \quad y_{1:t} = \{y_1, y_2, \dots, y_t\}.$$

我们定义：

$$\begin{aligned} \delta_j(t) &= \max_{x_1, x_2, \dots, x_{t-1}} p(x_1, x_2, \dots, x_t = s_j, y_1, y_2, \dots, y_t | \theta), \\ &= \max_{x_{1:(t-1)}} p(x_{1:(t-1)}, x_t = s_j, y_{1:t} | \theta) \end{aligned}$$

即在时刻 t 处于状态 s_j ，并且到达该状态的最优路径的概率。

注意到

$$\alpha_j(t) = p(x_t = s_j, y_{1:t} | \theta) = \sum_{x_{1:(t-1)}} p(x_{1:(t-1)}, x_t = s_j, y_{1:t} | \theta)$$

所以递推公式类似，将求和换成求最大值。

考虑 $\delta_j(t)$ 的递推关系：(黑板推导)

$$\delta_j(t) = \max_{x_{1:(t-1)}} p(x_{1:(t-1)}, x_t = s_j, y_{1:t} | \theta)$$

(利用条件概率公式，提出 $x_{1:(t-1)}, y_{1:(t-1)}$ 边际分布)

$$= \max_{x_{1:(t-1)}} p(x_t = s_j, y_t | x_{1:(t-1)}, y_{1:(t-1)}, \theta) p(x_{1:(t-1)}, y_{1:(t-1)} | \theta)$$

(由于条件独立，第一项的条件可去除 $y_{1:(t-1)}$)

$$= \max_{x_{1:(t-1)}} p(x_t = s_j, y_t | x_{t-1}, \theta) p(x_{1:(t-1)}, y_{1:(t-1)} | \theta)$$

(拆分第一项)

$$= \max_{x_{1:(t-1)}} p(x_t = s_j | x_{t-1}, \theta) p(y_t | x_t = s_j, \theta) p(x_{1:(t-1)}, y_{1:(t-1)} | \theta)$$

($p(y_t | x_t = s_j, \theta)$ 提到 max 外面，将 $x_{1:(t-1)}$ 拆分成 $x_{1:(t-2)}$ 和 x_{t-1})

$$= b_j(y_t) \max_{i=1,2,\dots,N} \max_{x_{1:(t-2)}} p(x_t = s_j | x_{t-1} = s_i, \theta) p(x_{1:(t-2)}, x_{t-1} = s_i, y_{1:(t-1)} | \theta)$$

$$= b_j(y_t) \max_{i=1,2,\dots,N} [\delta_i(t-1) a_{ij}].$$

这就是维特比递推公式 (跟前向算法唯一不同的是求和变成了 max):

$$\delta_j(t) = \max_{i=1,2,\dots,N} [\delta_i(t-1) a_{ij}] b_j(y_t).$$

回溯求最优路径

令 $\psi_j(t)$ 表示在时刻 t 处于状态 s_j 时, 从前一时刻 $t - 1$ 传递到当前状态 s_j 的最优前驱状态的索引, 即:

$$\psi_j(t) = \arg \max_{i=1,2,\dots,N} [\delta_i(t-1)a_{ij}].$$

$\psi_j(t)$ 记录了在最优路径中, 状态 s_j 在时刻 t 是从哪个状态 s_i 转移而来的, 用于回溯最优隐藏状态序列。

- 最优终点状态:

$$x_T^* = \arg \max_{i=1,2,\dots,N} \delta_i(T).$$

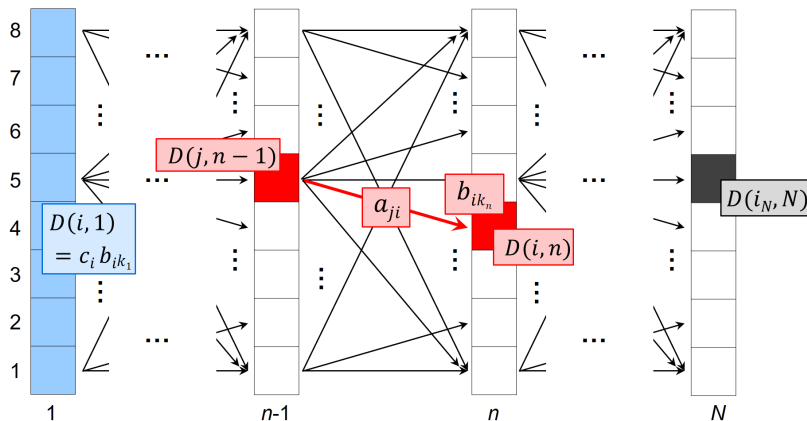
- 递归回溯:

$$x_t^* = \psi_{x_{t+1}^*}(t+1), \quad t = T-1, T-2, \dots, 1.$$

最终得到最优隐藏状态序列:

$$X^* = (x_1^*, x_2^*, \dots, x_T^*).$$

Figure 5.27 from [Müller, FMP, Springer 2015]



LIU, Ran - Department of Statistics @BNU

- Step 1: 初始化计算 $\delta_i(1)$, 记录初始状态概率:

$$\delta_i(1) = \pi_i b_i(y_1), \quad i = 1, 2, \dots, N.$$

- Step 2: 递推计算计算所有 $\delta_j(t)$:

$$\delta_j(t) = \max_{i=1,2,\dots,N} [\delta_i(t-1) a_{ij}] b_j(y_t).$$

同时记录路径索引:

$$\psi_j(t) = \arg \max_{i=1,2,\dots,N} [\delta_i(t-1) a_{ij}].$$

- Step 3: 终止计算 P^* 并确定最优路径的终点状态:

$$x_T^* = \arg \max_{i=1,2,\dots,N} \delta_i(T).$$

- Step 4: 回溯通过 $\psi_j(t)$ 递归回溯, 求出最优状态序列 X^* :

$$x_t^* = \psi_{x_{t+1}^*}(t+1), \quad t = T-1, T-2, \dots, 1.$$

HMM 在语音识别中的应用

隐马尔可夫模型 (HMM) 是一种常用于语音识别的概率模型，适用于处理时间序列数据，如语音信号的特征序列。

- 隐藏状态 (Hidden States): 对应音素或音素子状态，如 $/h/$, $/e/$, $/l/$, $/o/$ 。
- 观测值 (Observations): 对应语音的特征向量，如 Mel 频率倒谱系数 (MFCC)。
- 初始状态概率: 定义语音开始时处于某个音素的概率。
- 状态转移概率: 定义音素之间的转换概率，如从 $/h/$ 到 $/e/$ 。
- 观测概率 (发射概率): 定义在某个音素状态下，特征向量出现的概率。

Liu, Ran - Department of Statistics @BNU

语音识别过程：

- 输入语音信号，提取 MFCC 特征序列。
- 训练 HMM 参数：
 - 通过 Baum-Welch 算法估计的转移概率和发射概率。
- 使用 HMM 计算最优状态序列：
 - 通过 Viterbi 算法计算最可能的音素序列。
 - 生成对应的文本输出 (Speech-to-Text)。

HMM 在语音识别中用于建模音素序列，通过 Viterbi 解码找到最可能的文本输出，是自动语音识别 (ASR) 系统的重要组成部分。

Liujun Ran - Department of Statistics @BNU