HMM

Sep 2022

ruanyl@buaa.edu.cn

Assumptions

$$o_t \perp \{o_1, ..., o_{t-1}, o_{t+1}, ..., o_\tau, s_1, ..., s_{t-1}, s_{t+1}, ..., s_\tau\} | s_t.$$

$$s_{t+1} \perp \{o_1, ..., o_t\} | s_t.$$

Forward algorithm

$$
\begin{aligned}
&P(o_1, ..., o_t, s_t = x_j) \\
&= P(o_t | o_1, ..., o_{t-1}, s_t = x_j) P(o_1, ..., o_{t-1}, s_\tau = x_j) \\
&= P(o_t | s_t = x_j) \sum_i P(o_1, ..., o_{t-1}, s_{t-1} = x_i, s_t = x_j) \\
&= P(o_t | s_t = x_j) \sum_i P(s_t = x_j | o_1, ..., o_{t-1}, s_{t-1} = x_i) P(o_1, ..., o_{t-1}, s_{t-1} = x_i) \\
&= P(o_t | s_t = x_j) \sum_i P(s_t = x_j | s_{t-1} = x_i) P(o_1, ..., o_{t-1}, s_{t-1} = x_i).
\end{aligned}
$$

$$p_t(i) = P(o_1, ..., o_t, s_t = x_i), \;\; \alpha_{ij} = P(s_t = x_j | s_{t-1} = x_i), \;\; \beta_j(o_t) = P(o_t | s_t = x_j),$$

$$p_t(j) = \left(\sum_i p_{t-1}(i) \alpha_{ij} \right) \beta_j(o_t)$$

$$P(o_1, ..., o_t) = \sum_j p_t(j).$$

Backward algorithm

$$
\begin{aligned}
&P(o_{t+1}, ..., o_\tau | s_t = x_i) \\
&= \sum_j P(o_{t+1}, ..., o_\tau, s_{t+1} = x_j | s_t = x_i) \\
&= \sum_j P(o_{t+2}, ..., o_\tau | o_{t+1}, s_{t+1} = x_j, s_t = x_i) P(o_{t+1}, s_{t+1} = x_j | s_t = x_i) \\
&= \sum_j P(o_{t+2}, ..., o_\tau | s_{t+1} = x_j) P(o_{t+1} | s_{t+1} = x_j) P(s_{t+1} = x_j | s_t = x_i).
\end{aligned}
$$

$$q_t(i) = P(o_{t+1}, ..., o_\tau | s_t = x_i),\ \ q_\tau(i) = 1,$$

$$q_t(i) = \sum_j \alpha_{ij} \beta_j(o_{t+1}) q_{t+1}(j).$$

$$P(o_1, ..., o_\tau) = \sum_i q_0(i) P(S_0 = x_i).$$

Baum-Welch algorithm

$$\{o_{i1}, \dots, o_{i\tau}, s_{i1}, \dots, s_{i\tau}\}_{i=1}^{N}$$

$$\theta = (A, B, \pi)$$

Let m be the number of hidden states

n be the number of observable states

$$s_{it} \in \{e_1, \dots, e_m\}, \text{ here } e_k = (0, \dots, \underset{the\ k\text{-th}}{1}, \dots, 0)^T \in \mathbb{R}^m.$$

$$o_{it} \in \{e_1, \dots, e_n\}, \text{ here } e_l = (0, \dots, \underset{the\ l\text{-th}}{1}, \dots, 0)^T \in \mathbb{R}^n.$$

Baum-Welch algorithm

$$\{o_{i1}, \ldots, o_{i\tau}, s_{i1}, \ldots, s_{i\tau}\}_{i=1}^{N}$$

$$\theta = (A, B, \pi)$$

Let m be the number of hidden states

n be the number of observable states

$$s_{it} \in \{e_1, \ldots, e_m\}, \text{ here } e_k = (0, \ldots, \underset{the\ k\text{-th}}{1}, \ldots, 0)^T \in \mathbb{R}^m.$$

$$o_{it} \in \{e_1, \ldots, e_n\}, \text{ here } e_l = (0, \ldots, \underset{the\ l\text{-th}}{1}, \ldots, 0)^T \in \mathbb{R}^n.$$

Baum-Welch algorithm

$$\log A \triangleq (\log a_{ij})_{ij}, \ \frac{1}{A} \triangleq \left(\frac{1}{a_{ij}}\right)_{ij}, \ A^B \triangleq \prod_{i,j} a_{ij}^{b_{ij}}, \ A \circledast B \triangleq \sum_{i,j} a_{ij} b_{ij}.$$

e_{ij} is a matrix with 1 at ij and 0 otherwise.

'Complete data'

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{N} p(o_{i1}, ..., o_{i\tau}, s_{i1}, ..., s_{i\tau} | \theta) = \prod_{i=1}^{N} \pi^{s_{i1}} B^{s_{i1} \cdot o_{i1}^T} A^{s_{i1} \cdot s_{i2}^T} B^{s_{i2} \cdot o_{i2}^T} \cdots A^{s_{i,\tau-1} \cdot s_{i\tau}^T} B^{s_{i\tau} \cdot o_{i\tau}^T} \\ &= \prod_{i=1}^{N} \pi^{s_{i1}} B^{s_{i1} \cdot o_{i1}^T} \prod_{t=2}^{\tau} A^{s_{i,t-1} \cdot s_{it}^T} B^{s_{it} \cdot o_{it}^T}. \end{aligned}$$

Baum-Welch algorithm

$$\log L(\theta) = \sum_{i=1}^{N} \left(s_{i1} \circledast \log \pi + s_{i1} \cdot o_{i1}^{T} \circledast \log B + \sum_{t=2}^{\tau} \left(s_{i,t-1} \cdot s_{it}^{T} \circledast \log A + s_{it} \cdot o_{it}^{T} \circledast \log B \right) \right)$$

$$\sum_{i=1}^{N} \left(s_{i1} \circledast \log \pi + s_{i1} \cdot o_{i1}^{T} \circledast \log B + \sum_{t=2}^{\tau} \left(s_{i,t-1} \cdot s_{it}^{T} \circledast \log A + s_{it} \cdot o_{it}^{T} \circledast \log B \right) \right) +$$
$$\lambda \left(\sum_{k=1}^{m} \pi_k - 1 \right) + \left(\sum_{j=1}^{m} \xi_j \left(\sum_{l=1}^{n} b_{jl} - 1 \right) \right) + \left(\sum_{j=1}^{m} \eta_j \left(\sum_{k=1}^{m} a_{jk} - 1 \right) \right)$$

$$\frac{\partial L}{\partial \pi_k} = \sum_{i=1}^{N} \left(\delta(s_{i1} - e_k) s_{i1} \circledast \frac{1}{\pi} + \lambda \right) = 0,$$

$$\sum_{i=1}^{N} \delta(s_{i1} - e_k) + N \lambda \pi_k = 0,$$

Baum-Welch algorithm

$$\sum_{i=1}^{N}\left(s_{i1} \circledast \log \pi + s_{i1} \cdot o_{i1}^{T} \circledast \log B + \sum_{t=2}^{\tau}\left(s_{i,t-1} \cdot s_{it}^{T} \circledast \log A + s_{it} \cdot o_{it}^{T} \circledast \log B\right)\right) +$$
$$\lambda\left(\sum_{k=1}^{m} \pi_k - 1\right) + \left(\sum_{j=1}^{m} \xi_j \left(\sum_{l=1}^{n} b_{jl} - 1\right)\right) + \left(\sum_{j=1}^{m} \eta_j \left(\sum_{k=1}^{m} a_{jk} - 1\right)\right)$$

$$\frac{\partial L}{\partial \pi_k} = \sum_{i=1}^{N}\left(\delta(s_{i1} - e_k) s_{i1} \circledast \frac{1}{\pi} + \lambda\right) = 0,$$

$$\sum_{i=1}^{N} \delta(s_{i1} - e_k) + N\lambda\pi_k = 0,$$

$$\pi_k = \frac{\sum_{i=1}^{N} \delta(s_{i1} - e_k)}{N}$$

Baum-Welch algorithm

$$\sum_{i=1}^{N}\left(s_{i1} \circledast \log \pi + s_{i1} \cdot o_{i1}^{T} \circledast \log B + \sum_{t=2}^{\tau}\left(s_{i,t-1} \cdot s_{it}^{T} \circledast \log A + s_{it} \cdot o_{it}^{T} \circledast \log B\right)\right) +$$

$$\lambda\left(\sum_{k=1}^{m} \pi_k - 1\right) + \left(\sum_{j=1}^{m} \xi_j \left(\sum_{l=1}^{n} b_{jl} - 1\right)\right) + \left(\sum_{j=1}^{m} \eta_j \left(\sum_{k=1}^{m} a_{jk} - 1\right)\right)$$

$$\frac{\partial L}{\partial b_{jl}} = \sum_{i=1}^{N}\left(\sum_{t=1}^{\tau} \delta\left(s_{it} \cdot o_{it}^{T} - e_{jl}\right) s_{it} \cdot o_{it}^{T} \circledast \frac{1}{B} + \xi_j\right) = 0,$$

$$\sum_{i=1}^{N}\sum_{t=1}^{\tau} \delta\left(s_{it} \cdot o_{it}^{T} - e_{jl}\right) + N\xi_j b_{jl} = 0.$$

$$\xi_j = -\tau$$

$$b_{jl} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{\tau} \delta\left(s_{it} \cdot o_{it}^{T} - e_{jl}\right)}{N\tau}$$

Baum-Welch algorithm

$$\sum_{i=1}^{N}\left(s_{i1}\circledast\log\pi + s_{i1}\cdot o_{i1}^{T}\circledast\log B + \sum_{t=2}^{\tau}\left(s_{i,t-1}\cdot s_{it}^{T}\circledast\log A + s_{it}\cdot o_{it}^{T}\circledast\log B\right)\right)+$$
$$\lambda\left(\sum_{k=1}^{m}\pi_k - 1\right) + \left(\sum_{j=1}^{m}\xi_j\left(\sum_{l=1}^{n}b_{jl} - 1\right)\right) + \left(\sum_{j=1}^{m}\eta_j\left(\sum_{k=1}^{m}a_{jk} - 1\right)\right)$$

$$\frac{\partial L}{\partial a_{jk}} = \sum_{i=1}^{N}\left(\sum_{t=2}^{\tau}\delta\left(s_{i,t-1}\cdot s_{it}^{T} - e_{jk}\right)s_{i,t-1}\cdot s_{it}^{T}\circledast\frac{1}{A} + \eta_j\right) = 0.$$

$$a_{jk} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{\tau}\delta\left(s_{i,t-1}\cdot s_{it}^{T} - e_{jk}\right)}{N(\tau-1)}$$

Baum-Welch algorithm

a single data sample $(o_1, ..., o_\tau, s_1, ..., s_\tau)$

$$x = (o_1, ..., o_\tau), z = (s_1, ..., s_\tau).$$

$$\begin{aligned}\pi_{k,EM} &= E_{Z|x,\theta'}(\delta(S_1 - e_k)) = \sum_z \delta(s_1 - e_k) p(z|x,\theta') \\ &= \sum_{s_1} \delta(s_1 - e_k) p(s_1|x,\theta') = p(s_1 = e_k|x,\theta').\end{aligned}$$

$$b_{jl,EM} = \frac{\sum_{t=1}^{\tau} E_{Z|x,\theta'}\big(\delta\big(S_t \cdot o_t^T - e_{jl}\big)\big)}{\sum_{l=1}^{n} \sum_{t=1}^{\tau} E_{Z|x,\theta'}\big(\delta\big(S_t \cdot o_t^T - e_{jl}\big)\big)} = \frac{\sum_{t=1}^{\tau} p(s_t = e_j|x,\theta')\delta(o_t - e_l)}{\sum_{t=1}^{\tau} p(s_t = e_j|x,\theta')}.$$

$$a_{jk,EM} = \frac{\sum_{t=2}^{\tau} E_{Z|x,\theta'}\big(\delta\big(S_{t-1} \cdot S_t^T - e_{jk}\big)\big)}{\sum_{k=1}^{m} \sum_{t=2}^{\tau} E_{Z|x,\theta'}\big(\delta\big(S_{t-1} \cdot S_t^T - e_{jk}\big)\big)} = \frac{\sum_{t=2}^{\tau} p(s_{t-1} = e_j, s_t = e_k|x,\theta')}{\sum_{t=2}^{\tau} p(s_{t-1} = e_j|x,\theta')}.$$

Baum-Welch algorithm

$$p(s_1 = e_k | x, \theta'),\ p(s_{t-1} = e_j, s_t = e_k | x, \theta')$$

Backward and forward algorithm

Baum-Welch algorithm

$$Q(\theta|\theta') = \sum_z p(x, z|\theta') \log p(x, z|\theta)$$

$$\begin{aligned}
Q(\theta|\theta') = &\sum_z (s_1 \circledast \log \pi) p(x, z|\theta') + \sum_z \left(\sum_{t=1}^{\tau} s_t \cdot o_t^T \circledast \log B \right) p(x, z|\theta') + \\
&\sum_z \left(\sum_{t=2}^{\tau} s_{t-1} \cdot s_t^T \circledast \log A \right) p(x, z|\theta') \\
&= \sum_{s_1} (s_1 \circledast \log \pi) p(x, s_1|\theta') + \sum_{t=1}^{\tau} \sum_{s_t} (s_t \cdot o_t^T \circledast \log B) p(x, s_t|\theta') + \\
&\sum_{t=2}^{\tau} \sum_{s_{t-1}, s_t} (s_{t-1} \cdot s_t^T \circledast \log A) p(x, s_{t-1}, s_t|\theta').
\end{aligned}$$

Baum-Welch algorithm

$$
\begin{aligned}
Q(\theta|\theta') = &\sum_z (s_1 \circledast \log \pi) p(x, z|\theta') + \sum_z \left(\sum_{t=1}^{\tau} s_t \cdot o_t^T \circledast \log B \right) p(x, z|\theta') + \\
&\sum_z \left(\sum_{t=2}^{\tau} s_{t-1} \cdot s_t^T \circledast \log A \right) p(x, z|\theta') \\
&= \sum_{s_1} (s_1 \circledast \log \pi) p(x, s_1|\theta') + \sum_{t=1}^{\tau} \sum_{s_t} (s_t \cdot o_t^T \circledast \log B) p(x, s_t|\theta') + \\
&\sum_{t=2}^{\tau} \sum_{s_{t-1}, s_t} (s_{t-1} \cdot s_t^T \circledast \log A) p(x, s_{t-1}, s_t|\theta').
\end{aligned}
$$