

Sample of HMM parameters estimation using Gibbs sampling method

Ryo Ozaki
Ritsumeikan University
Graduate School of Information Science and Engineering
Emergent Systems Laboratory
ryo.ozaki@em.ci.ritsumei.ac.jp

December 12, 2018

1 Graphical model

This section shows the graphical model of hidden semi-Markov model (HSMM) which were used this paper.

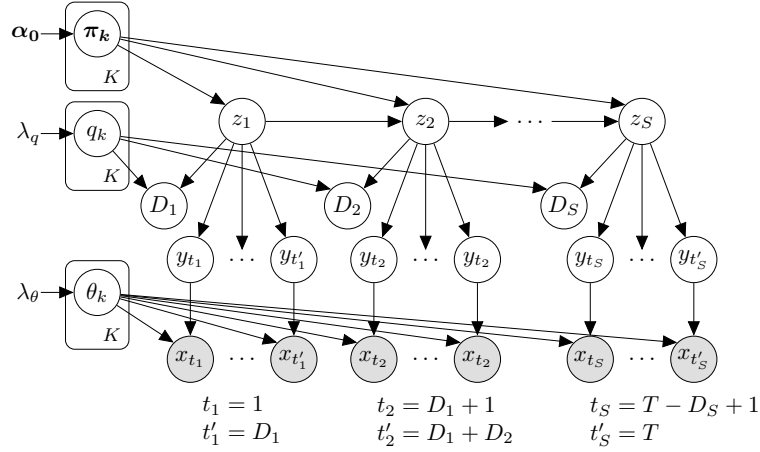


Figure 1: Graphical model of HSMM.

The probabilistic generative process of HMM is shown below.

$$\boldsymbol{\pi}_0 \sim \text{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}_0) \quad (1)$$

$$\boldsymbol{\pi}_k \sim \text{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}_0) \quad (2)$$

$$\theta_k \sim p(\theta|\lambda_\theta) \quad (3)$$

$$q_k \sim p(q|\lambda_q) \quad (4)$$

$$z_1 \sim \text{Cat}(z|\boldsymbol{\pi}_0) \quad (5)$$

$$z_s \sim \text{Cat}(z|\boldsymbol{\pi}_{\mathbf{z}_{s-1}}) \quad (6)$$

$$D_s \sim p(D|q_{z_s}) \quad (7)$$

$$t_s = 1 + \sum_{s'=1}^{s-1} D_{s'} \quad (8)$$

$$t'_s = \sum_{s'=1}^s D_{s'} \quad (9)$$

$$= t_s + D_s - 1 \quad (10)$$

$$y_{t_s:t'_s} = z_s \quad (11)$$

$$x_t \sim p(x|\theta_{y_t}) \quad (12)$$

$$k = 1, 2, \dots, K \quad (13)$$

$$s = 1, 2, \dots, S \quad (14)$$

$$t = 1, \dots, T \quad (15)$$

Here, Dir represents the Dirichlet distribution, Cat represents the categorical distribution. And, you can use the any emission distribution $p(x|\theta_t)$ and prior distribution $p(\theta|\lambda_\theta)$. For example, set $p(x|\theta_t)$ to multivariate normal distribution and set $p(\theta|\lambda_\theta)$ to normal-inverse-Wishart distribution.

2 Posterior distribution

This section shows the posterior distributions.

2.1 Posterior distribution of z_s and D_s

When sample the super-state sequence $z_{1:S}$ and the duration of super-states $D_{1:S}$, basically we use the blocked Gibbs sampling. In this section, we shows the sampling algorithm using blocked Gibbs sampling.

In the blocked Gibbs sampling of HSMM, the super-state sequence $z_{1:S}$ and the duration of super-states $D_{1:S}$ are sampled by the conditional posterior distribution $p(z_{1:S}, D_{1:S}|x_{1:T}, \boldsymbol{\pi}_{0:K}, \theta_{1:K}, q_{1:K})$. The distribution is little bit redundancy, so, we don't write the $\boldsymbol{\pi}_{0:K}$, $\theta_{1:K}$ and $q_{1:K}$ often in condition part. Here, $p(z_{1:S}, D_{1:S}|x_{1:T})$ can be described as follows.

$$p(z_{1:S}, D_{1:S} | x_{1:T}) = \prod_{s=1}^S \{p(z_s, D_s | z_{1:s-1}, D_{1:s-1}, x_{1:T})\} \quad (16)$$

$$= \prod_{s=1}^S \{p(z_s | z_{1:s-1}, D_{1:s-1}, x_{1:T}) p(D_s | z_{1:s}, D_{1:s-1}, x_{1:T})\} \quad (17)$$

And then, the term regarding sampling the z_s is as follows.

$$p(z_s = i | z_{1:s-1}, D_{1:s-1}, x_{1:T}) \stackrel{z_s}{\propto} p(x_{1:T} | z_{1:s-1}, z_s = i, D_{1:s-1}) p(z_s = i | z_{1:s-1}, D_{1:s-1}) \quad (18)$$

$$\stackrel{z_s}{\propto} p(x_{t_s:T} | z_s = i) p(z_s = i | z_{s-1}) \quad (19)$$

$$= \beta_{t_s}^*(i) p(z_s = i | z_{s-1}) \quad (20)$$

And, the term regarding sampling the D_s is as follows.

$$p(D_s | z_{1:s-1}, z_s = i, D_{1:s-1}, x_{1:T}) = p(D_s | z_s = i, x_{t_s:T}) \quad (21)$$

$$= \frac{p(D_s, x_{t_s:T} | z_s = i)}{p(x_{t_s:T} | z_s = i)} \quad (22)$$

$$= \frac{p(x_{t_s:t_s+d-1} | z_s = i, D_s = d) p(x_{t_s+d:T} | z_s = i) p(D_s = d | z_s = i)}{p(x_{t_s:T} | z_s = i)} \quad (23)$$

$$= \frac{p(x_{t_s:t'_s} | z_s = i, D_s = d) p(x_{t_{s+1}:T} | z_s = i) p(D_s = d | z_s = i)}{p(x_{t_s:T} | z_s = i)} \quad (24)$$

$$= \frac{p(x_{t_s:t'_s} | z_s = i, D_s = d) \beta_{t_{s+1}}^*(i) p(D_s = d | z_s = i)}{\beta_{t_s}^*(i)} \quad (25)$$

Here, the $\beta_{t_s}(i)$ and $\beta_{t_s}^*$ are backward message in HSMM. The definition of backward message is as follows. Simply said, the backward message $\beta_{t_s}(i)$ is the probability about the $x_{t_s:T}$ was generated after if i -th super-state was beginning at t_s .

$$\beta_{t_s}(i) \stackrel{\text{def}}{=} p(x_{t_s:T} | z_{s-1} = i) \quad (26)$$

$$\beta_{t_s}^*(i) \stackrel{\text{def}}{=} p(x_{t_s:T} | z_s = i) \quad (27)$$

$$\beta_T(i) \stackrel{\text{def}}{=} 1 \quad (28)$$

In addition, those are calculated as follows.

$$\beta_{t_s}(i) \stackrel{\text{def}}{=} p(x_{t_s:T} | z_{s-1} = i) \quad (29)$$

$$= \sum_j p(x_{t_s:T}, z_s = j | z_{s-1} = i) \quad (30)$$

$$= \sum_j p(x_{t_s:T} | z_s = j, z_{s-1} = i) p(z_s = j | z_{s-1} = i) \quad (31)$$

$$= \sum_j p(x_{t_s:T} | z_s = j) p(z_s = j | z_{s-1} = i) \quad (32)$$

$$= \sum_j \beta_{t_s}^*(j) p(z_s = j | z_{s-1} = i) \quad (33)$$

$$\beta_{t_s}^*(i) \stackrel{\text{def}}{=} p(x_{t_s:T}|z_s = i) \quad (34)$$

$$= \sum_{d=1}^{T-t_s} p(x_{t_s:T}, D_s = d|z_s = i) \quad (35)$$

$$= \sum_{d=1}^{T-t_s} p(x_{t_s:T}|z_s = i, D_s = d)p(D_s = d|z_s = i) \quad (36)$$

$$= \sum_{d=1}^{T-t_s} p(x_{t_s:t_s+d-1}|z_s = i, D_s = d)p(x_{t_s+d:T}|z_s = i, D_s = d)p(D_s = d|z_s = i) \quad (37)$$

$$= \sum_{d=1}^{T-t_s} p(x_{t_s:t'_s}|z_s = i, D_s = d)p(x_{t_{s+1}:T}|z_s = i)p(D_s = d|z_s = i) \quad (38)$$

$$= \sum_{d=1}^{T-t_s} p(x_{t_s:t'_s}|z_s = i, D_s = d)\beta_{t_{s+1}}(i)p(D_s = d|z_s = i) \quad (39)$$

$$t'_s = t_s + d - 1 \quad (40)$$

$$t_{s+1} = t_s + d \quad (41)$$

Summarize, the formulas about backward message are as follows.

$$\beta_{t_s}(i) \stackrel{\text{def}}{=} p(x_{t_s:T}|z_{s-1} = i) \quad (42)$$

$$\beta_{t_s}^*(i) \stackrel{\text{def}}{=} p(x_{t_s:T}|z_s = i) \quad (43)$$

$$\beta_T(i) \stackrel{\text{def}}{=} 1 \quad (44)$$

$$\beta_{t_s}(i) = \sum_j \beta_{t_s}^*(j)p(z_s = j|z_{s-1} = i) \quad (45)$$

$$\beta_{t_s}^*(i) = \sum_{d=1}^{T-t_s} p(x_{t_s:t'_s}|z_s = i, D_s = d)\beta_{t_{s+1}}(i)p(D_s = d|z_s = i) \quad (46)$$

$$t'_s = t_s + d - 1 \quad (47)$$

$$t_{s+1} = t_s + d \quad (48)$$