Sample of HSMM 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 13, 2018

1 Graphical model

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

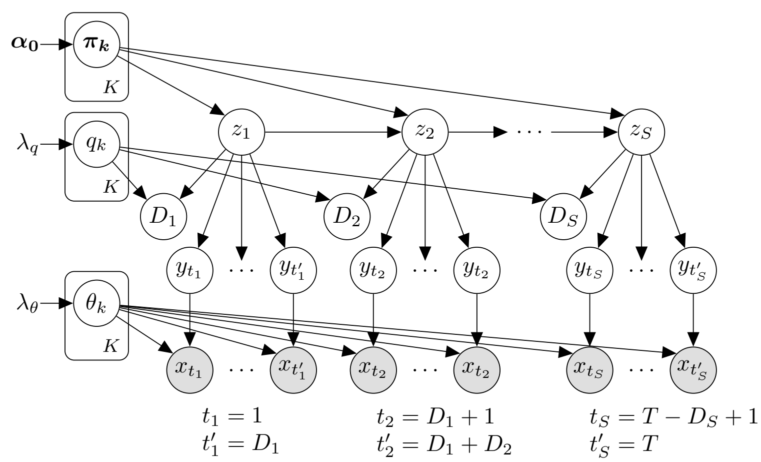

Figure 1: Graphical model of HSMM.

The probabilistic generative process of HMM is shown below.

$$
\begin{aligned}
\boldsymbol{\pi_0} &\sim \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha_0}) && (1)\\
\boldsymbol{\pi_k} &\sim \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha_0}) && (2)\\
\theta_k &\sim p(\theta|\lambda_\theta) && (3)\\
q_k &\sim p(q|\lambda_q) && (4)\\
z_1 &\sim \mathrm{Cat}(z|\boldsymbol{\pi_0}) && (5)\\
z_s &\sim \mathrm{Cat}(z|\boldsymbol{\pi_{z_{s-1}}}) && (6)\\
D_s &\sim p(D|q_{z_s}) && (7)\\
t_s &= 1+\sum_{s'=1}^{s-1} D_{s'} && (8)\\
t'_s &= \sum_{s'=1}^{s} D_{s'} && (9)\\
&= t_s + D_s - 1 && (10)\\
y_{t_s:t'_s} &= z_s && (11)\\
x_t &\sim p(x|\theta_{y_t}) && (12)
\end{aligned}
$$

$$
\begin{aligned}
k &= 1,2,\dots,K && (13)\\
s &= 1,2,\dots,S && (14)\\
t &= 1,\dots,T && (15)
\end{aligned}
$$

Here, Dir represents the Dirichlet distribution, Cat represents the categorical distribution. And, you can use the any emission distribution $p(x|\theta_t)$ and the prior distribution $p(\theta|\lambda_\theta)$. Similary, you can use the any duration distribution of super-state $p(D|q_k)$ and the prior distribution $p(q|\lambda_q)$. 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.

$$
\begin{aligned}
p(z_{1:S}, D_{1:S}|x_{1:T}) &= \prod_{s=1}^{S} \{p(z_s, D_s|z_{1:s-1}, D_{1:s-1}, x_{1:T})\} &(16)\\
&= \prod_{s=1}^{S} \{p(z_s|z_{1:s-1}, D_{1:s-1}, x_{1:T})p(D_s|z_{1:s}, D_{1:s-1}, x_{1:T})\} &(17)
\end{aligned}
$$

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

$$
\begin{aligned}
p(z_s = i|z_{1:s-1}, D_{1:s-1}, x_{1:T}) &\overset{z_s}{\propto} p(x_{1:T}|z_{1:s-1}, z_s = i, D_{1:s-1})p(z_s = i|z_{1:s-1}, D_{1:s-1}) &(18)\\
&\overset{z_s}{\propto} p(x_{t_s:T}|z_s = i)p(z_s = i|z_{s-1}) &(19)\\
&= \beta^*_{t_s}(i)p(z_s = i|z_{s-1}) &(20)
\end{aligned}
$$

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

$$
\begin{aligned}
p(D_s|z_{1:s-1}, z_s = i, D_{1:s-1}, x_{1:T}) &= p(D_s|z_s = i, x_{t_s:T}) &(21)\\
&= \frac{p(D_s, x_{t_s:T}|z_s = i)}{p(x_{t_s:T}|z_s = i)} &(22)\\
&= \frac{p(x_{t_s:t_s+d-1}|z_s = i, D_s = d)p(x_{t_s+d:T}|z_s = i)p(D_s = d|z_s = i)}{p(x_{t_s:T}|z_s = i)} &(23)\\
&= \frac{p(x_{t_s:t'_s}|z_s = i, D_s = d)p(x_{t_{s+1}:T}|z_s = i)p(D_s = d|z_s = i)}{p(x_{t_s:T}|z_s = i)} &(24)\\
&= \frac{p(x_{t_s:t'_s}|z_s = i, D_s = d)\beta_{t_{s+1}}(i)p(D_s = d|z_s = i)}{\beta^*_{t_s}(i)} &(25)
\end{aligned}
$$

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.

$$
\begin{aligned}
\beta_{t_s}(i) &\overset{\text{def}}{=} p(x_{t_s:T}|z_{s-1} = i) &(26)\\
\beta^*_{t_s}(i) &\overset{\text{def}}{=} p(x_{t_s:T}|z_s = i) &(27)\\
\beta_T(i) &\overset{\text{def}}{=} 1 &(28)
\end{aligned}
$$

In addition, those are calculated as follows.

$$
\begin{aligned}
\beta_{t_s}(i) &\overset{\text{def}}{=} p(x_{t_s:T}|z_{s-1} = i) &(29)\\
&= \sum_j p(x_{t_s:T}, z_s = j|z_{s-1} = i) &(30)\\
&= \sum_j p(x_{t_s:T}|z_s = j, z_{s-1} = i)p(z_s = j|z_{s-1} = i) &(31)\\
&= \sum_j p(x_{t_s:T}|z_s = j)p(z_s = j|z_{s-1} = i) &(32)\\
&= \sum_j \beta^*_{t_s}(j)p(z_s = j|z_{s-1} = i) &(33)
\end{aligned}
$$

$$
\begin{aligned}
\beta^*_{t_s}(i) &\stackrel{\text{def}}{=} p(x_{t_s:T}|z_s = i) && (34)\\
&= \sum_{d=1}^{T-t_s} p(x_{t_s:T}, D_s = d|z_s = i) && (35)\\
&= \sum_{d=1}^{T-t_s} p(x_{t_s:T}|z_s = i, D_s = d)p(D_s = d|z_s = i) && (36)\\
&= \sum_{d=1}^{T-t_s} p(x_{t_s:t_s+d-1}|z_s = i, D_s = d)p(x_{t_s+d:T}|z_s = i, D_s = d)p(D_s = d|z_s = i) && (37)\\
&= \sum_{d=1}^{T-t_s} p(x_{t_s:t'_s}|z_s = i, D_s = d)p(x_{t_{s+1}:T}|z_s = i)p(D_s = d|z_s = i) && (38)\\
&= \sum_{d=1}^{T-t_s} p(x_{t_s:t'_s}|z_s = i, D_s = d)\beta_{t_{s+1}}(i)p(D_s = d|z_s = i) && (39)\\
t'_s &= t_s + d - 1 && (40)\\
t_{s+1} &= t_s + d && (41)
\end{aligned}
$$

Summarize, the formulas about backward message are as follows.

$$
\begin{aligned}
\beta_{t_s}(i) &\stackrel{\text{def}}{=} p(x_{t_s:T}|z_{s-1} = i) && (42)\\
\beta^*_{t_s}(i) &\stackrel{\text{def}}{=} p(x_{t_s:T}|z_s = i) && (43)\\
\beta_T(i) &\stackrel{\text{def}}{=} 1 && (44)
\end{aligned}
$$

$$
\begin{aligned}
\beta_{t_s}(i) &= \sum_j \beta^*_{t_s}(j)p(z_s = j|z_{s-1} = i) && (45)\\
\beta^*_{t_s}(i) &= \sum_{d=1}^{T-t_s} p(x_{t_s:t'_s}|z_s = i, D_s = d)\beta_{t_{s+1}}(i)p(D_s = d|z_s = i) && (46)\\
t'_s &= t_s + d - 1 && (47)\\
t_{s+1} &= t_s + d && (48)
\end{aligned}
$$