Appendix

EM algorithm

Define $\Lambda = \{\pi, b, \lambda\}$, for l step,

1) **E-step**

$$T_{w,k}^{(l)} = P\left(Z_{w,k} = 1 \middle| V_w = v_w, \Lambda^{(l)}\right) = \frac{\pi_k^{(l)} f(V_w; \lambda_k^{(l)}, b_k^{(l)})}{\sum_{s=1}^{K} \pi_s^{(l)} f(V_w; \lambda_s^{(l)}, b_s^{(l)})}$$

$$Q\left(\Lambda, \Lambda^{(l)}\right) = \sum_{w=1}^{W} \sum_{k=1}^{K} T_{k,w}^{(l)} [log\pi_k + logf(V_w; \lambda_s, b_s)]$$

2) **M-step**

$$\hat{\Lambda} = (\hat{\lambda}, \hat{b}) = \text{argmax}(\text{Q}(\Lambda, \Lambda^{(\text{l})}))$$

Firstly, it is easy to see that

$$\pi_k^{(l+1)} = \frac{\sum_{w=1}^{W} T_{k,w}^{(l)}}{\sum_{w=1}^{W} \sum_{k=1}^{K} T_{k,w}^{(l)}} = \frac{1}{W} \sum_{w=1}^{W} T_{k,w}^{(l)}$$

$$z^{(\text{l}+1)}{}_{w,k} = 1 \; when \; k = \overset{argmax}{l} \left(\boldsymbol{E}\left(z_{w,l} \middle| \boldsymbol{O}, \Lambda\right)\right)$$

$$\Lambda^{(l+1)} = \left(\lambda_k^{(l+1)}, b_k^{(l+1)}\right) = \underset{\lambda_k, b_k}{argmax} \sum_{w=1}^{W} z^{(l+1)}{}_{w,k} \{logf(V_w; \lambda_k, b_k)\} \qquad (1)$$

Different from mixture normal model, it is difficult for us to maximum (1) directly. Here, we use Baum-Welch Algorithm to search parameter Λ while maximum (1).

In details, we have

1) **Forward variable for HMM on time:**

$$F_{w,t}^{k}(i) = P\ (v_{w,1}, v_{w,2}, \dots, v_{w,t}, H_{w,t} = i | C_w = k, \Lambda)$$

Therefore we have:

$$F_{w,t+1}^{k}(j)$$

$$= P\left(\nu_{w,t+1}\middle|H_{w,t+1} = j, \quad C_w = k, \quad \lambda\right) \times \sum_{i=0}^{1} F_{w,t}^{k}(i) \times b_{i,j}^{k}$$

$$= \left(\prod_{m=1}^{M} \frac{\left(\lambda_{j,m}^{k}\right)^{\nu_{w,t+1,m}} e^{-\lambda_{j,m}^{k}}}{\nu_{w,t+1,m}!}\right) \sum_{i=0}^{1} F_{w,t}^{k}(i) \times b_{i,j}^{k}$$

And because $H_{w,1} = 1$ for all the w, we have

$$F_{w,1}^{k}(0) = Pr^{k}{}_{1\to\nu_{w,1}}, \text{and } F_{w,1}^{k}(1) = 0$$

Thus, we have

$P\ (\nu_w|C_w = k, \Lambda) = F_{w,3}^{k}(1) + F_{w,3}^{k}(0),$

where $\nu_w = (\nu_{w,1}, \nu_{w,2}, \nu_{w,3})$.

2) Backward variable for HMM on time:

Because $B_{w,t}^{k}(i) = P\left(\nu_{w,t+1}, \nu_{w,t+2}, \ldots, \nu_{w,3}\middle|H_{w,t} = i, C_w = k, \Lambda\right)$, then $\mathrm{B}_{\mathrm{w,t}}^{\mathrm{k}}(\mathrm{i})$ can be got through

$$B_{w,t}^{k}(i) = \sum_{j=0}^{1} B_{w,t+1}^{k}(j) \times b_{i,j}^{k} \times P\left(\nu_{w,t+1}\middle|H_{w,t+1} = j, C_w = k, \lambda\right)$$

$$= \sum_{j=0}^{1} B_{w,t+1}^{k}(j) \times b_{i,j}^{k} \times \left(\prod_{m=1}^{M} \frac{\left(\lambda_{j,m}^{k}\right)^{\nu_{w,t+1,m}} e^{-\lambda_{j,m}^{k}}}{\nu_{w,t+1,m}!}\right)$$

and

$B_{w,3}^{k}(0) = B_{w,3}^{k}(1) = 1$

3) Log-likelihood based on (*π*, *b*, *λ*)

Given total observation $\boldsymbol{O} = (\nu_1, \nu_2, \ldots, \nu_3, \ldots, \nu_W)$

The posteriori expectation of $z_{w,k}$.

$\boldsymbol{E}\left(z_{w,k}\middle|\boldsymbol{O}, \Lambda\right) = \boldsymbol{P}\left(z_{w,k} = 1\middle|\boldsymbol{O}, \Lambda\right) = \boldsymbol{P}\ (C_w = k|\nu_w, \Lambda)$

$$= \frac{P(\nu_w, C_w = k|\Lambda)}{\boldsymbol{P}\,(\nu_w|\Lambda)} = \frac{\left(F_{w,3}^k(1) + F_{w,3}^k(0)\right) \times \pi_k}{\sum_{k=1}^K \left(F_{w,3}^k(1) + F_{w,3}^k(0)\right) \times \pi_k}$$

Define

$$\boldsymbol{T_{w,k}} = \boldsymbol{E}\left(z_{w,k} \middle| \boldsymbol{O}, \Lambda\right) = \boldsymbol{P}\,(C_w = k|\nu_w, \Lambda)$$

Then it is easy to see that

$$Q\left(\Lambda \middle| \Lambda^{(t)}\right) = E_{z|O,\Lambda^{(t)}}(\ln L(\Lambda; O, Z)) = \sum_{w=1}^{W} \sum_{k=1}^{K} T_{w,k} \times \ln P\left(\nu_w, C_w = k \middle| \Lambda^{(t)}\right)$$

$$= \sum_{w=1}^{W} \sum_{k=1}^{K} T_{w,k} \times \ln\left(\left(F_{w,3}^k(1) + F_{w,3}^k(0)\right) \times \pi_k\right)$$

Finally the expectation of log likelihood can be written as

$$\ln\left(P(\Lambda; O)\right) = \ln\left(\prod_{w=1}^{W} P(\nu_w|\Lambda)\right) = \sum_{w=1}^{W} \ln\left(P(\nu_w|\Lambda)\right)$$

$$= \sum_{w=1}^{W} \ln\left(\sum_{k=1}^{K} P(\nu_w, C_w = k|\Lambda)\right)$$

$$= \sum_{w=1}^{W} \ln\left(\sum_{k=1}^{K} \left(\left(F_{w,3}^k(1) + F_{w,3}^k(0)\right) \times \pi_k\right)\right)$$

4) Re-clustering for regions: Estimate *z*, *π*

Given $\boldsymbol{E}\left(z_{w,k} \middle| \boldsymbol{O}, \Lambda\right)$ for each *w*, *k*, the new π_k is

$$\pi'_k = \frac{1}{W} \sum_{w=1}^{W} \boldsymbol{E}\left(z_{w,k} \middle| \boldsymbol{O}, \Lambda\right) = \frac{1}{W} \sum_{w=1}^{W} \frac{\left(F_{w,3}^k(1) + F_{w,3}^k(0)\right) \times \pi_k}{\sum_{k=1}^K \left(F_{w,3}^k(1) + F_{w,3}^k(0)\right) \times \pi_k}$$

$z_{w,k}$ is updated using new π_k by

$z'_{w,k} = 1$, when k = $\underset{l}{\operatorname{argmax}}\, \boldsymbol{E}\left(z_{w,l} \middle| \boldsymbol{O}, \Lambda\right) = \operatorname{argmax}_l\left(P(\nu_w, C_w = l|\Lambda)\right)$.

5) Parameter re-estimation: Baum-Welch Algorithm for estimation of *b*, *λ*

After we calculated forward and backward variables for the HMM on time (1) and (2), it is easy to calculate the joint posteriori probability for each region

$$\xi^k{}_{w,t}(i,j) = p\left\{H_{w,t} = i, H_{w,t+1} = j \,\middle|\, \nu_w, C_w = k, \Lambda\right\}$$
$$= \frac{p\left\{H_{w,t} = i, H_{w,t+1} = j, \nu_w \,\middle|\, C_w = k, \Lambda\right\}}{\sum_{i=0}^{1}\sum_{j=0}^{1} p\left\{H_{w,t} = i, H_{w,t+1} = j, \nu_w \,\middle|\, C_w = k, \Lambda\right\}}, \qquad i,j = 0,1$$

The relative marginal posteriori probability is

$$\gamma^k{}_{w,t}(i) = p\left\{H_{w,t} = i \,\middle|\, \nu_w, C_w = k, \Lambda\right\} = \frac{p\left\{H_{w,t} = i, \nu_w \,\middle|\, C_w = k, \Lambda\right\}}{\sum_{i=0}^{1} p\left\{H_{w,t} = i, \nu_w \,\middle|\, C_w = k, \Lambda\right\}}, \qquad i = 0,1$$

So re-estimate the parameters *b*:

$$b'^k_{i,j} = \frac{\sum_{w=1}^{W}\sum_{t=1}^{2}\left(\xi^k{}_{w,t}(i,j) \times z_{w,k}\right)}{\sum_{w=1}^{W}\sum_{t=1}^{2}\left(\gamma^k{}_{w,t}(i) \times z_{w,k}\right)}$$

With marginal posteriori probability, we can estimate each $H_{w,t}$ by:

$$H_{w,t} = \underset{i}{argmax}\left[\gamma^k{}_{w,t}(i)\right]$$

Then

$$\lambda'^k_{i,m} = \frac{\sum_{w=1}^{W}\sum_{t=1}^{3}\left(\left|i + H_{w,t} - 1\right| \times \nu_{w,t,m} \times z'_{w,k}\right)}{\sum_{w=1}^{W}\sum_{t=1}^{3}\left(\left|i + H_{w,t} - 1\right| \times z'_{w,k}\right)}$$

We propose Baum-Welch to estimate new *b*, λ using the same $z_{w,k}$ until that *b*, λ converge in the condition of current clustering ($z_{w,k}$)

6) Renew the clustering information

When the *b*, λ are converged in (5) step, repeat the algorithm in (3) (4) to renew the clustering information ($z_{w,k}$), and then do those in (5) for optimized *b*, λ for current new clustering. We stop the algorithm when the parameter differences between the last two steps are sufficiently small.

Figure S1. Working flow chart of the model