Proofs of Entropic Forward-Backward algorithms for HMM2 and HMM-CN

Elie Azraf

This document is a supplementary material of the paper *Introduction to the Hidden Neural Markov Chain framework*, to appear for publication at the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021). The paper link will be set at publication time. We keep the same notations as in the paper.

We introduce a new notation, for each $t \in \{1, ..., T\}, \lambda_i \in \Lambda_X, y_t \in \Omega_Y$: $b_i(y_t) = p(y_t | x_t = \lambda_i)$.

Proof of the EFB algorithm for HMM2

The EFB algorithm for HMM2 aims to compute, for each $p(x_t = \lambda_i | y_{1:T})$.
We can show, with the law of HMM2, for each $t > 1$:

$$p(x_t = i | y_{1:T}) = \frac{\sum_j \alpha_t^{2'}(j,i)\beta_t^{2'}(j,i)}{\sum_k \sum_j \alpha_t^{2'}(j,k)\beta_t^{2'}(j,k)}$$

with:

$$\alpha_t^{2'}(j,i) = p(x_{t-1} = \lambda_j, x_t = \lambda_i, y_{1:t})$$
$$\beta_t^{2'}(j,i) = p(y_{t+1:T} | x_{t-1} = \lambda_j, x_t = \lambda_i)$$

$\alpha^{2'}$ can be computed with the following recursion:

- For $t = 2$:

$$\alpha_2^{2'}(j,i) = \pi(j) b_j(y_1) a_j(i) b_i(y_2)$$

- For $2 \leq t < T$:

$$\alpha_{t+1}^{2'}(j,i) = \sum_k \alpha_t^{2'}(k,j) a_{k,j}^2(i) b_i(y_{t+1})$$

And $\beta^{2'}$ with the following one:

- For $t = T$:

$$\beta_T^{2'}(j,i) = 1$$

- For $2 \leq t < T$:

$$\beta_t^{2'}(j,i) = \sum_k \beta_{t+1}^{2'}(i,k) a_{j,i}^2(k) b_k(y_{t+1})$$

We can show, for each $2 \leq t \leq T$:

$$\alpha_t^2(j,i) = \frac{\alpha_t^{2'}(j,i)}{p(y_1)p(y_2)...p(y_t)} \tag{1}$$

$$\beta_t^2(j,i) = \frac{\beta_t^{2'}(j,i)}{p(y_{t+1})p(y_{t+2})...p(y_T)} \tag{2}$$

For $t = 2$,

$$\begin{aligned} \alpha_2^{2'}(j,i) &= p(y_1, x_1 = \lambda_j, y_2, x_2 = \lambda_i) \\ &= p(y_1) L_{y_1}(j) p(y_2) a_j(i) \frac{L_{y_2}(i)}{\pi(i)} \end{aligned}$$

Therefore, (1) is true for $t = 2$. We suppose (1) for t, and we prove it for $t+1$:

$$\alpha_{t+1}^2(j,i) = \sum_k \frac{\alpha_t^{2'}(k,j)}{p(y_1)p(y_2)...p(y_t)} a_{k,j}^2(i) \frac{b_i(y_{t+1})}{p(y_{t+1})} = \frac{\alpha_{t+1}^{2'}(j,i)}{p(y_1)p(y_2)...p(y_t)p(y_{t+1})}$$

About (2), it is true for $t = T$. We suppose (2) true for $t+1 < T$, and we prove it for t:

$$\beta_t^2(j,i) = \sum_k \frac{\beta_{t+1}^{2'}(i,k)}{p(y_{t+2})...p(y_T)} a_{j,i}^2(k) \frac{b_k(y_{t+1})}{p(y_{t+1})} = \frac{\beta_t^{2'}(j,i)}{p(y_{t+1})p(y_{t+2})...p(y_T)}$$

(1) and (2) and proved for each t.
Therefore,

$$p(x_t = \lambda_i | y_{1:T}) = \frac{\alpha_t^{2'}(i)\beta_t^{2'}(i)}{\sum_j \alpha_t^{2'}(j)\beta_t^{2'}(j)} = \frac{\alpha_t^2(i)\beta_t^2(i)}{\sum_j \alpha_t^2(j)\beta_t^2(j)}$$

Which ends the proof of EFB algorithm for HMM2.

Proof of the EFB algorithm for HMM-CN

With the HMM-CN, $(x_{1:t-1}, y_{1:t-1})$ and $(x_{t+1:T}, y_{t+1:T})$ are independent conditionally on (x_t, y_t), and thus we have:

$$p(x_t = \lambda_i | y_{1:T}) = \frac{\alpha_t^{CN'}(i)\beta_t^{CN'}(i)}{\sum_j \alpha_t^{CN'}(j)\beta_t^{CN'}(j)}$$

with:

$$\alpha_t^{CN'}(i) = p(x_t = \lambda_i, y_{1:t})$$
$$\beta_t^{CN'}(i) = p(y_{t+1:T}|x_t = \lambda_i, y_t)$$

$\alpha^{CN'}$ can be computed with the following recursion:

- For $t = 1$:

$$\alpha_1^{CN'}(i) = \pi(i)p(y_t|x_1 = \lambda_i)$$

- For $1 \leq t < T$:

$$\alpha_{t+1}^{CN'}(i) = \sum_j \alpha_t^{CN'}(j)I_{j,y_t}(i)p(y_{t+1}|x_t = \lambda_j, x_{t+1} = \lambda_i)$$

And $\beta^{CN'}$ with the following one:

- For $t = T$

$$\beta_T^{CN'}(i) = 1$$

- For $1 \leq t < T$:

$$\beta_t^{CN'}(i) = \sum_j I_{i,y_t}(j)p(y_{t+1}|x_t = \lambda_i, x_{t+1} = \lambda_j)\beta_{t+1}^{CN'}(j)$$

We can show:

$$\alpha_t^{CN}(i) = \frac{\alpha_t^{CN'}(i)}{p(y_1)p(y_2)...p(y_t)} \tag{3}$$
$$\beta_t^{CN}(i) = \frac{\beta_t^{CN'}(i)}{p(y_{t+1})p(y_{t+1})...p(y_T)} \tag{4}$$

(3) is true for $t = 1$. We suppose (3) true for t, and we prove it for $t + 1$:

$$\begin{aligned}\alpha_{t+1}^{CN}(i) = & \frac{1}{p(y_1)...p(y_t)} \sum_j \alpha_t^{CN'}(j)I_{j,y_t}(i)\times \\ & \frac{p(x_{t+1} = i|y_{t+1})p(x_t = \lambda_j|x_{t+1} = \lambda_i, y_{t+1})}{p(x_t = \lambda_j, x_{t+1} = \lambda_i)} \\ & = \frac{1}{p(y_1)...p(y_t)p(y_{t+1})} \sum_j \alpha_t^{CN'}(j)I_{j,y_t}(i)\frac{p(x_t = \lambda_j, x_{t+1} = \lambda_i, y_{t+1})}{p(x_t = \lambda_j, x_{t+1} = \lambda_i)} \\ & = \frac{\alpha_{t+1}^{CN'}(i)}{p(y_1)...p(y_t)p(y_{t+1})}\end{aligned}$$

Therefore, (3) is proved for all t.
The proof of (4) follows the same reasoning. (4) is true for $t = T$. We suppose (4) true for $t+1$, and we prove it a t:

$$\begin{aligned}
\beta_t^{CN}(i) = & \frac{1}{p(y_{t+2})...p(y_T)} \sum_j \beta_{t+1}^{CN'}(j) I_{i,y_t}(j) \times \\
& \frac{p(x_{t+1} = \lambda_j | y_{t+1}) p(x_t = \lambda_i | x_{t+1} = \lambda_j, y_{t+1})}{p(x_t = \lambda_i, x_{t+1} = \lambda_j)} \\
& = \frac{1}{p(y_{t+1})p(y_{t+2})...p(y_T)} \sum_j \beta_{t+1}^{CN'}(j) I_{i,y_t}(j) \frac{p(x_t = \lambda_i, x_{t+1} = \lambda_j, y_{t+1})}{p(x_t = \lambda_i, x_{t+1} = \lambda_j)} \\
& = \frac{\beta_t^{CN'}(i)}{p(y_{t+1})p(y_{t+2})...p(y_T)}
\end{aligned}$$

, which prove (4) for all t.
Therefore,

$$p(x_t = \lambda_i | y_{1:T}) = \frac{\alpha_t^{CN}(i) \beta_t^{CN}(i)}{\sum_j \alpha_t^{CN}(j) \beta_t^{CN}(j)}$$

Which ends the proof of EFB algorithm for HMM-CN.