Pairwise HMM

Taikai Takeda* and ...

Waseda University, Japan
`{297.1951}@gmail.com`

Abstract. .

1 Pairwise Alignment

Pairwise Alignment is to aligne two sequences of, for example, DNA, by inserting gaps inbetween the elements in the sequences. The goal of this task is to maximize a score of the alignment so that we can choose the best alignment of all the possible ones.

...

2 HMM

HMM (Hidden Markov Model) has been widely used for from gene alignment to speech recognitions. Let us introduce simple HMM before presenting Pairwise HMM.

2.1 Formulation

Let $\mathcal{D} = \boldsymbol{X} = (X_1, ..., X_T)$ and $\boldsymbol{Z} = (Z_1, ..., Z_T)$ be, respectively, observed and hidden random variables. Let $\mathcal{A}$ be a set of simbols and a set of hidden states be $\mathcal{S}$. Input data is a set of sequences, $\boldsymbol{x} = (\boldsymbol{x}^1, ..., \boldsymbol{x}^N)$ where n-th sequence $\boldsymbol{x}^n \in \mathcal{A}^{T_n}$ is T_n is the length of the sequence. Similarly, hidden states are denoted as $\boldsymbol{z} = (\boldsymbol{z}^1, ..., \boldsymbol{z}^N)$ where n-th sequence $\boldsymbol{z}^n \in \mathcal{S}^{T_n}$. Note that we sometimes omit superscript n when concentrating on a single sequence for the sake of notational simplicity. The corresponding joint disribution has the form

$$p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) = p(Z_1\boldsymbol{\alpha}) \prod_{t=2}^{T} p(Z_t|Z_{t-1}, \boldsymbol{\beta}) \prod_{t=1}^{T} p(X_t|Z_t, \boldsymbol{\phi}) \tag{1}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}\}$. $p(Z_1|\boldsymbol{\alpha})$, $p(Z_t|Z_{t-1}, \boldsymbol{\beta})$ and $p(X_t|Z_t, \boldsymbol{\phi})$ are an initial probability, a transition probability, an emission probability, respectively. They are described, respectively, as $p(Z_1 = k, \boldsymbol{\alpha}) = \alpha_k$, $p(Z_t = k|Z_{t-1} = j, \boldsymbol{\beta}) = \beta_{jk}$ and $p(X_t|Z_t = k, \boldsymbol{\phi}) = p(X_t|\phi_k)$. $\boldsymbol{\alpha} = \{\alpha\}_k$ is $K-$dimensional vector and $\boldsymbol{\beta} = \{\beta\}_{jk}$ is $K \times K$ matrix where K is the number of hidden states.

---

*

2.2 Forward-Backward Algorithm

We here discuss how to compute the smoothed marginal $p(z_t = j|\boldsymbol{x})$ and tghe smoothed two-sliced marginal $p(z_{t-1}, z_t|\boldsymbol{x})$.[1]

Taking a look at the graphical model in Fig(), we can see conditioning on z_t eable to decompose joint distribution into two parts: the past and the future.

$$p(z_t = k|\boldsymbol{x}) \propto p(z_t = k, \boldsymbol{x}_{t+1:T}|\boldsymbol{x}_{1:t}) \propto p(z_t = k|\boldsymbol{x}_{1:t})p(\boldsymbol{x}_{t+1:T}|z_t = k) \quad (2)$$

Let us define forward variables $f_{t,k} = p(z_t = k|\boldsymbol{x}_{1:t})$, the bilief of the state given all the previous sequence. Also, define backward variables $b_{t,k} = p(\boldsymbol{x}_{t+1:T}|z_t = k)$, the conditional likelihood of future evidence give the hidden staetes z_t. Forward variables are efficiently computed using dynamic programming. The base case and the recursive relationship is given as follows:

$$
\begin{aligned}
f_{t,k} &= p(z_t = k|\boldsymbol{x}_{1:t}) \\
&= \frac{p(z_t = k, x_t|\boldsymbol{x}_{1:t-1})}{p(x_t|\boldsymbol{x}_{1:t-1})} \\
&= \frac{p(x_t|z_t = k, \cancel{\boldsymbol{x}_{1:t-1}})p(z_t = k|\boldsymbol{x}_{1:t-1})}{p(x_t|\boldsymbol{x}_{1:t-1})} \\
&= \frac{p(x_t|z_t = k)\sum_{j=1}^{K} p(z_t = k|z_{t-1} = j)p(z_{t-1} = j|\boldsymbol{x}_{1:t-1})}{p(x_t|\boldsymbol{x}_{1:t-1})} \\
&= p(x_t|z_t = k)\sum_{j=1}^{K} \beta_{j,k} f_{t-1,k} \quad (3)
\end{aligned}
$$

$$f_{1,k} = p(z_1 = k) = \alpha_k \quad (4)$$

[2] Similarly, backward variables are computed using following equations:

$$
\begin{aligned}
b_{t-1,j} &= p(\boldsymbol{x}_{t:T}|z_{t-1} = j) \\
&= \sum_{j=1}^{K} p(\boldsymbol{x}_{t:T}, z_t = j|z_{t-1} = j) \\
&= \sum_{j=1}^{K} p(z_t = k|z_{t-1} = j)p(\boldsymbol{x}_{t:T}|z_t = k, \cancel{z_{t-1} = j}) \\
&= \sum_{j=1}^{K} p(z_t = k|z_{t-1} = j)p(x_t|z_t = k)p(\boldsymbol{x}_{t+1:T}|z_t = k) \\
&= \sum_{j=1}^{K} \beta_{j,k}\psi_{t,k}b_{t,k} \quad (5)
\end{aligned}
$$

---

[1] Note that in online setting, we can only compute $p(z_t = j|\boldsymbol{x}_{1:t})$, so called filtered marginal, but we concentrate on the offline setting here.

[2] MEMO: should we define emission notation e.g. $\psi_{t,k}$

Now, we can compute smoothed posterior using forward and backward variables. Let us denote smoothed posterior $\gamma_{t,k} = p(z_t = k | \boldsymbol{x}_{1:T})$ and smoothed two-sliced marginal $\xi_{t,j,k} = p(z_{t-1}, z_t | \boldsymbol{x})$

$$\gamma_{t,k} \propto f_{t,k} b_{t,k} \tag{6}$$

Also, smoothed two-sliced marginal is computed as follows:

$$
\begin{aligned}
\xi_{t,j,k} &= p(z_{t-1}, z_t | \boldsymbol{x}_{1:T}) \\
&\propto p(z_t, z_{t-1}, \boldsymbol{x}_{t:T} | \boldsymbol{x}_{1:t-1}) \\
&= p(z_{t-1} | \boldsymbol{x}_{1:t-1}) p(z_t, \boldsymbol{x}_{t:T} | z_{t-1}, \cancel{\boldsymbol{x}_{1:t-1}}) \\
&= p(z_{t-1} | \boldsymbol{x}_{1:t-1}) p(z_t | z_{t-1}) p(\boldsymbol{x}_{t:T} | z_t, \cancel{z_{t-1}}) \\
&= p(z_{t-1} | \boldsymbol{x}_{1:t-1}) p(z_t | z_{t-1}) p(x_t | z_t) p(\boldsymbol{x}_{t+1:T} | z_t, \cancel{x_t}) \\
&= f_{t-1,j} p(z_t | z_{t-1}) p(x_t | z_t) b_{t,k}
\end{aligned}
\tag{7}
$$

2.3 Parameter Optimizations via EM

The paramters can be learned from the daaset using EM (Expectation Maximization), which is also called Baum-Welch specifically for HMM. Likelihood function $l(\boldsymbol{\theta}) = p(\boldsymbol{X}|\boldsymbol{\theta}) = \sum_Z p(\boldsymbol{X}, \boldsymbol{Z})$ is hard to optimize because it includes partition function over all the possible states of the hidden states.

The complete data log likelyhood $l_c(\boldsymbol{\theta})$ is written down as

$$l_c(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left[\ln p(z_1^n | \boldsymbol{\alpha}) + \sum_{t=2}^{T} \ln p(z_t^n | z_{t-1}^n, \boldsymbol{\beta}) + \sum_{t=1}^{T} \ln p(x_t^n | z_t^n, \boldsymbol{\phi}) \right] \tag{8}$$

The auxilary function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$, the expected complete log likelihood, is given by

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}) &= E_{p(\boldsymbol{Z}|\boldsymbol{x}, \boldsymbol{\theta}^{old})}[l_c(\boldsymbol{\theta})] \\
&= \sum_{n=1}^{N} \left[\ln p(z_1^n | \boldsymbol{\alpha}) + \sum_{t=2}^{T} \ln p(z_t^n | z_{t-1}^n, \boldsymbol{\beta}) + \sum_{t=1}^{T} \ln p(x_t^n | z_t^n, \boldsymbol{\phi}) \right]
\end{aligned}
\tag{9}
$$

...

2.4 Viterbi decoding

Choose the optimal sequence of hiden states. ...

3 Pairwise HMM

Pairwise HMM (PHMM) is probablistic generative model used for pairwise sequence alignment. Given transition and emission probability distributions, it

can compute likelyhood of 'similarity' as well as the most probable alignment. Furthermore, it is possible to optimize parameters by iterative procedure (EM algorithm). This one of discrete HMMs, but different from them as the length of hidden states changes dinamically according to the alighment. We will introduce PHMM using analogy to simple HMM.

3.1 Formulation

Let input Let $\mathcal{A}$ be a set of simbols. For DNA alignment, $\mathcal{A} = \{\text{"}A\text{"}, \text{"}T\text{"}, \text{"}G\text{"}, \text{"}C\text{"}\}$. Input data is two sequences, $\boldsymbol{x} = (x^1, ...x^{T_x}) \in \mathcal{A}^{T_x}$ and $\boldsymbol{y} = (y^1, ..., y^{T_y}) \in \mathcal{A}^{T_y}$ where T_x and T_y are the length of sequences of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. A set of hidden states is $\mathcal{S}$. In the simpliest case shown in Fig (), $\mathcal{S} = \{\text{"}M\text{"}, \text{"}X\text{"}, \text{"}Y\text{"}\}$. Hidden random variablesre denoted as $\boldsymbol{z} = (z^1, ..., z^{T_z})$

3.2 EM