CS221 Quiz Solutions

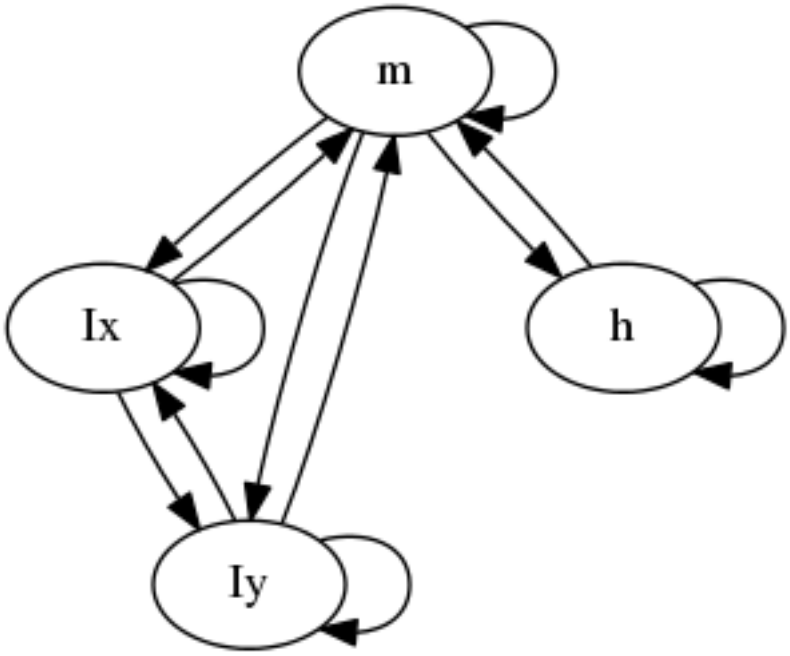
Release v10

Christopher Lee

November 27, 2017

Contents

1. **Detecting hyperconserved regions in a pair of genomes**



(a)

(b) For arriving at $m@X_t, Y_u$ we have to consider the four possible incoming edges

$$\begin{aligned} V(m@X_t, Y_u) = p(X_t, Y_u|m) \max\{&V(m@X_{t-1}, Y_{u-1})p(m|m), \\ &V(h@X_{t-1}, Y_{u-1})p(m|h), \\ &V(I_x@X_{t-1}, Y_u)p(m|I_x), \\ &V(I_y@X_t, Y_{u-1})p(m|I_y)\} \end{aligned}$$

(c) We wish to calculate the posterior probability that letters X_t, Y_u were emitted by state *h* given the entire observation sequences, i.e.

$$p(h@X_t, Y_u|\vec{X}, \vec{Y})$$

This posterior probability assesses the hidden state that we are interested in, taking into account all the observations.

(d) We obtain the posterior probability from Bayes Law in the usual way:

$$p(h@X_t, Y_u|\vec{X}, \vec{Y}) = \frac{p(h@X_t, Y_u, \vec{X}, \vec{Y})}{p(\vec{X}, \vec{Y})}$$

To calculate this efficiently, we split the numerator into a forward part including all the observations up to and including $\vec{X}_{[1,t]}, \vec{Y}_{[1,u]}$ and $h@X_t, Y_u$, and a backward part consisting of all the remaining observations $\vec{X}_{[t+1,n]}, \vec{Y}_{[u+1,m]}$ conditioned on $h@X_t, Y_u$. These two parts each can be calculated via an induction with computation time $O(LS^2)$.

$$p(h@X_t, Y_u, \vec{X}, \vec{Y}) = p(h@X_t, Y_u, \vec{X}_{[1,t]}, \vec{Y}_{[1,u]})p(\vec{X}_{[t+1,n]}, \vec{Y}_{[u+1,m]}|h@X_t, Y_u)$$

The forward probability here only needs to consider two incoming edges:

$$\begin{aligned} p(h@X_t, Y_u, \vec{X}_{[1,t]}, \vec{Y}_{[1,u]}) =&p(h@X_{t-1}, Y_{u-1}, \vec{X}_{[1,t-1]}, \vec{Y}_{[1,u-1]})p(h|h)p(X_t, Y_u|h)+ \\ &p(m@X_{t-1}, Y_{u-1}, \vec{X}_{[1,t-1]}, \vec{Y}_{[1,u-1]})p(h|m)p(X_t, Y_u|h) \end{aligned}$$

The backward probability also only needs to consider two outgoing edges:

$$\begin{aligned} p(\vec{X}_{[t+1,n]}, \vec{Y}_{[u+1,m]}|h@X_t, Y_u) =&p(\vec{X}_{[t+2,n]}, \vec{Y}_{[u+2,m]}|h@X_{t+1}, Y_{u+1})p(h|h)p(X_{t+1}, Y_{u+1}|h)+ \\ &p(\vec{X}_{[t+2,n]}, \vec{Y}_{[u+2,m]}|m@X_{t+1}, Y_{u+1})p(m|h)p(X_{t+1}, Y_{u+1}|m) \end{aligned}$$