The Forward Algorithm

Let x be the event that some specific sequence was generated by a hidden Markov model. The **Forward Algorithm** computes $P(x)$ under the model.

Because many different state paths can give rise to the same sequence x, we must add the probabilities for all possible paths to obtain the full probability of x:

$$P(x) = \sum_{\pi} P(x, \pi)$$

where π is an event in which a specific path was taken through the HMM.

The number of possible paths increases exponentially with the length of the sequence, so brute force evaluation of this probability by enumerating over all paths is not practical. Instead, the Forward algorithm computes this probability efficiently using a dynamic programming strategy.

Description

We build a dynamic programming matrix f such that entry $f_{k,i}$ is defined by the function:

$$f_k(i) = P(x_1, \ldots, x_i,\ \pi_i = k)$$

where x_i is the event that the ith character of the sequence was generated by the HMM.

$f_k(i)$ computes the probability of being in state k having observed the first i characters of the sequence. Thus, the rows of the matrix correspond to the states in the HMM and the columns correspond to each symbol in the sequence x.

We compute the values in the matrix from the top left most element $(0, 0)$ corresponding to the probability of being at the start state before the sequence starts. This probability is simply 1 because we always begin in the start state. We compute each column from left to right and each column is computed from top to bottom.

Each element is calculated using the following recurrence:

$$f_k(i) = e_k(x_i) \sum_{l} f_l(i-1)\ \tau_{l,k}$$

That is, we multiply $e_k(x_i)$, the probability of emitting character x_i from state k, by the summation over the probabilities of being in each state l at $i - 1$ and then transitioning to the current state k. The probability of transitioning from state l to state k is denoted in Equation 2 by $\tau_{l,k}$.

The final probability of the sequence is thus given by,

$$P(x) = \sum_{l} f_l(L)\, \tau_{k,0}$$

where L is the length of the sequence and $\tau_{k,0}$ is the probability of transitioning from state k to the end state.

Example

Assume the following sequence was generated from the HMM in the example:

$$x = \text{TAGA} \tag{1}$$

We wish to compute $P(x)$ using the Forward Algorithm. We start by initializing an empty dynamic programming matrix as seen below:

State, l \ Time Step, t	0	1 (T)	2 (A)	3 (G)	4 (A)
γ_0	**1**	-	-	-	-
γ_1	**0**	-	-	-	-
γ_2	**0**	-	-	-	-
γ_3	**0**	-	-	-	-
γ_4	**0**	-	-	-	-
γ_5	**0**	-	-	-	-

We have fully filled in the probabilities at time step $t = 0$ (before generating the sequence). Entry $f_{\gamma_0}(0) = 1$ indicates that there is a probability of 1.0 that we are in the start state before generating the sequence. Similary, there is a 0.0 probability we are in any of the other states before generating the sequence.

We now compute the values at time step $t = 1$:

$$f_{\gamma_1}(t=1) = e_{\gamma_1}(\mathrm{T}) \cdot \left(f_{\gamma_0}(0)\, \tau_{\gamma_0,\gamma_1} + f_{\gamma_1}(0)\, a\tau_{\gamma_1,\gamma_1} \right)$$

$$= 0.3 \times (0.5 + 0)$$

$$= 0.15$$

$$f_{\gamma_2}(t=1) = e_{\gamma_2}(\mathrm{T}) \cdot \left(f_{\gamma_0}(0)\, \tau_{\gamma_0,\gamma_2} + f_{\pi_2}(0)\, \tau_{\gamma_2,\gamma_2} \right)$$

$$= 0.4 \times (0.5 + 0)$$

$$= 0.2$$

Notice that in the sum $\sum_l f_l(i-1)\, \tau_{l,k}$, we only need to sum over the states l where $\tau_{l,k} \neq 0$. Since the only transitions to state γ_1 are from γ_0 and itself, we only include these states in the summation because the transition probabilities from the other states to γ_1 are 0.

We insert these results into the dynamic programming matrix:

State, l \ Time Step, t	0	1 (T)	2 (A)	3 (G)	4 (A)
γ_0	1	**0**	-	-	-
γ_1	0	**0.15**	-	-	-
γ_2	0	**0.2**	-	-	-
γ_3	0	**0**	-	-	-
γ_4	0	**0**	-	-	-
γ_5	0	**0**	-	-	-

Notice that the probability of being in the rest of states (i.e. γ_0, γ_3, γ_4, and γ_5) at $t = 1$ is zero because there is no path through the HMM for which we can take one transition from the start state and land in these states.

We provide one more computed value at time step $t = 2$:

$$f_{\gamma_1}(t = 2) = e_{\gamma_1}(\mathrm{A}) \cdot \left(f_{\gamma_0}(1)\ \tau_{\gamma_0,\gamma_1} + f_{\gamma_1}(1)\ \tau_{\gamma_1,\gamma_1}\right)$$

$$= 0.4 \times (0 \times 0.5 + 015 \times 0.2)$$

$$= 0.12$$

We insert these two results into the dynamic programming matrix:

Time Step, t / State, l	0	1 (T)	2 (A)	3 (G)	4 (A)
γ_0	1	0	**0**	-	-
γ_1	0	0.15	**0.12**	-	-
γ_2	0	0.2	-	-	-
γ_3	0	0	-	-	-
γ_4	0	0	-	-	-
γ_5	0	0	-	-	-

We continue these computations and fill in the entire matrix. The final result of the algorithm (i.e. the full probability of the sequence under the model) will be the sum of over the products of each entry multiplied by the transition probability from that state to the end state.