Extension of HMM to two states

Aaditya Prakash

February 10, 2015

This document models the HMM extension to two states, and provides relevant formulas, along with proofs of the non-trivial formulas to gurantee the correctness of the said formulas. While there are couple of online references and books where HMM with two, three and arbitrary, 'n' states has been discussed and what it means to extend HMM but none of them provide detailed formulas for application of such a system. This document only gives details on two state HMM, with a possiblity of future update for "n" state HMMs.

1 Notations

I have kept the notations close to as provided in Chapter 6 of "Speech and Language Processing", Second Edition by Martin and Jurafsky.

λ Common term for all HMM parameters. All the probabilties will be conditioned to this term i.e $P(......|\lambda)$
T denotes the total number of time steps
N denotes the total number of states
o_t denotes the observed variable (state) at time step t
q_t denotes the hidden variale (state) at time step t

2 Formulas

2.1 Distribution of $\alpha_t(i, j)$

$\alpha_t(i, j)$ denotes the joint probability distribution of all observed variables until time t and current and the last states.

$$\alpha_t(i, j) = P(o_1, o_2, \ldots, o_t, q_{t-1} = i, q_t = j|\lambda)$$

Where λ is the given HMM parameters.

2.2 Base case of $\alpha_t(i, j)$

Consider the following notation:

$$a_{ij} = P(q_t = j|q_{t-1} = i)$$

$$a_{ijk} = P(q_{t+1} = k|q_t = j, q_{t-1} = i)$$

then base cases are -

$$\alpha_1(i = 0, j) = a_{i=0,j}b_j(o_1)$$

$$\alpha_2(i, j) = \alpha_1(0, j)a_{ij}b_j(o_2)$$

It should be noted that if the second state (q_2) is also allowed to be entry point, then it needs an additional base case of –

$$\alpha_2(i = 0, j) = \sum_{i=1}^{N} \alpha_1(i = 0, j)a_{ij}b_j(o_2)$$

2.3 Inductive step

$$\alpha_t(j, k) = \sum_{i=1}^{N} \alpha_{t-1}(i, j) \times a_{ijk} \times b_k(o_t) \text{ where, } 3 \leq t \leq T$$

2.4 Termination step

$$P(O|\lambda) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_T(i,j) \times a_{iF} \times a_{jF} \times a_{ijF}$$

3 Estimated Expected Transitions

3.1 ξ_t

ξ_t = Probability of being in state k at t+1, j at t and i at t-1

$$\xi_t = P(q_{t+1} = k, q_t = j, q_{t-1} = i|o_1, \ldots o_T) = \frac{\alpha_t(i,j) \times \beta_{t+1}(j,k) \times a_{ijk} b_k(o_{t+1})}{\alpha(q_f)}$$

Derivation provided in § 4.1

3.2 γ_t

γ_t = Probability of being in state i at 't', given all the obervations

$$\gamma_t = P(q_t = i|o_1, \ldots, o_T) = \frac{\sum_j^N \alpha(j,i), \beta(j,i)}{\alpha(q_f)}$$

Derivation provided in § 4.2

where $\alpha(q_f) = P(o_1, \ldots, o_T)$ (joint probability of all observed variables)

*Please note that for γ_t values, the i and j are interchanged in α and β values, because I wanted to keep the input as $q_t = i$, as provided in the problem, whereas convention in textbook is to denote i as antecedent to j.

4 Proofs

4.1 To prove ξ_t

To Prove

$$P(q_{t+1} = k, q_t = j, q_{t-1} = i|o_1, \ldots o_T) = \frac{\alpha_t(i,j) \times \beta_{t+1}(j,k) \times a_{ijk} b_k(o_{t+1})}{\alpha(q_f)}$$

where $\alpha(q_f) = P(o_1, \ldots, o_T)$
which is equivalent to,

$$P(q_{t+1} = k, q_t = j, q_{t-1} = i|o_1, \ldots o_T) \times P(o_1, \ldots, o_T) = \alpha_t(i,j) \times \beta_{t+1}(j,k) \times a_{ijk} b_k(o_{t+1})$$

using, then rule $P(A|B) \times P(B) = P(A,B)$

$$P(q_{t+1} = k, q_t = j, q_{t-1} = i, o_1, \ldots o_T) = \alpha_t(i,j) \times \beta_{t+1}(j,k) \times a_{ijk} \times b_k(o_{t+1})$$

Consider the Left Hand Side,

$$P(q_{t+1} = k, q_t = j, q_{t-1} = i, o_1, \ldots o_T)$$

To shorten the equations, we will use $o_{m:n}$ to denote $o_m, \ldots o_n$
Conditioning on $o_{1:t}$,$q_t = j$ and $q_{t-1} = i$, and using the rule $P(A,B) = P(A|B) \times P(B)$

$$\implies P(o_{(t+1):T}, q_{t+1} = k|o_{1:t}, q_t = j, q_{t-1} = i) \times P(o_{1:t}, q_t = j, q_{t-1} = i)$$

But we know that, $\alpha_t(i,j) = P(o_{1:t}, q_t = j, q_{t-1} = i)$ (Section *§2.1)*

$$\implies P(o_{(t+1):T}, q_{t+1} = k|o_{1:t}, q_t = j, q_{t-1} = i) \times \alpha_t(i,j)$$

Using the **markov assumption** for second order model, $P(o_i|q_{1:T}, o_{1:T}) = P(o_i|q_{i-1}, q_i)$

$$\implies P(o_{(t+1):T}, q_{t+1} = k|q_t = j, q_{t-1} = i) \times \alpha_t(i,j)$$

Conditioning on o_{t+1} and $q_{t+1} = k$, and using the rule $P(A,B|C) = P(A|B) \times P(B|C)$

$$\implies P(o_{(t+2):T}|o_{t+1}, q_{t+1} = k, q_t = j, q_{t-1} = i) \times P(o_{t+1}, q_{t+1} = k|q_t = j, q_{t-1} = i) \times \alpha_t(i,j)$$

Using the **markov assumption** for second order model, $P(o_i|q_{1:T}, o_{1:T}) = P(o_i|q_{i-1}, q_i)$

$$\implies P(o_{(t+2):T}|q_{t+1} = k, q_t = j) \times P(o_{t+1}, q_{t+1} = k|q_t = j, q_{t-1} = i) \times \alpha_t(i, j)$$

But we know that, $\beta_t(i, j) = P(o_{(t+1):T}|q_t = j, q_{t-1} = i)$. Implies, $\beta_{t+1}(j, k) = P(o_{(t+2):T}|q_t = k, q_{t-1} = j)$

$$\implies \beta_{t+1}(j, k) \times P(o_{t+1}, q_{t+1} = k|q_t = j, q_{t-1} = i) \times \alpha_t(i, j)$$

Conditioning on $q_{t+1}$and using the rule $P(A, B|C) = P(A|B) \times P(B|C)$

$$\implies \beta_{t+1}(j, k) \times P(o_{t+1}|q_{t+1} = k, q_{t=j}, q_{t-1} = i) \times P(q_{t+1} = k|q_t = j, q_{t-1} = i) \times \alpha_t(i, j)$$

Using the **markov assumption** for second order model, $P(o_i|q_{1:T}, o_{1:T}) = P(o_i|q_{i-1}, q_i)$

$$\implies \beta_{t+1}(j, k) \times P(o_{t+1}|q_{t+1} = k, q_{t=j}) \times P(q_{t+1} = k|q_t = j, q_{t-1} = i) \times \alpha_t(i, j)$$

But we know from the Emission probability, $b_k(o_{t+1}) = P(o_{t+1}|q_{t+1}, q_t)$

$$\implies \beta_{t+1}(j, k) \times b_k(o_{t+1}) \times P(q_{t+1} = k|q_t = j, q_{t-1} = i) \times \alpha_t(i, j)$$

And we know from the Transition Probability, $a_{ijk} = P(q_{t+1} = k|q_t = j, q_{t-1} = i)$

$$\implies \beta_{t+1}(j, k) \times b_k(o_{t+1}) \times a_{ijk} \times \alpha_t(i, j)$$

On rearranging,

$$\alpha_t(i, j) \times \beta_{t+1}(j, k) \times a_{ijk} \times b_k(o_{t+1})$$

which is equal to the Right Hand Side.

■

4.2 To Prove γ_t

To Prove

$$P(q_t = i|o_1, \ldots, o_T) = \frac{\sum_j^N \alpha(j, i), \beta(j, i)}{\alpha(q_f)}$$

where $\alpha(q_f) = P(o_1, \ldots, o_T)$
which is equivalent to,

$$P(q_t = i|o_1, \ldots o_T) \times P(o_1, \ldots, o_T) = \sum_j^N \alpha(j, i), \beta(j, i)$$

using, then rule $P(A|B) \times P(B) = P(A, B)$

$$P(q_t = i, o_1, \ldots o_T) = \sum_j^N \alpha(j, i), \beta(j, i)$$

Consider the Left Hand Side,

$$P(q_t = i, o_1, \ldots o_T)$$

On marginalizing the $q_{t-1} = j$

$$\implies \sum_{j=1}^N P(q_t = i, q_{t-1} = j, o_{1:T})$$

Using the bayes rule

$$\implies \sum_{j=1}^N P(o_{t+1:T}|o_{1:t}, q_t = i, q_{t-1} = j) \times P(o_{1:t}, q_t = i, q_{t-1} = j)$$

Using the **markov assumption** for second order model, $P(o_i|q_{1:T}, o_{1:T}) = P(o_i|q_{i-1}, q_i)$,

$$\implies \sum_{j=1}^N P(o_{t+1:T}|q_t = i, q_{t-1} = j) \times P(o_{1:t}, q_t = i, q_{t-1} = j)$$

But we know that, $\alpha_t(i,j) = P(o_{1:t}, q_t = j, q_{t-1} = i)$ (Section *§2.1)*

$$\implies \sum_{j=1}^{N} P(o_{t+1:T} | q_t = i, q_{t-1} = j) \times \alpha_t(j,i)$$

And we know that, $\beta_t(i,j) = P(o_{(t+1):T} | q_t = j, q_{t-1} = i)$

$$\implies \sum_{j=1}^{N} \beta(j,i) \times \alpha_t(j,i)$$

which is equal to the Right Hand Side.

$\blacksquare$