CSCE 5214 - Supplement for the lecture Unsupervised Learning From Examples

Slide 41 of the lecture "Unsupervised Learning From Examples" provides equations of the forward computation of the joined probability of HMM state and associated event at time t from the one at time t-1. The following sequence of equations detailed the algorithm.

Background

Conditional probability equation

$$P(x,y) = P(x|y)p(y) \tag{1}$$

Chain rule equation

$$P(x_1, x_2, x_3, ..x_n) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1)... \tag{2}$$

That is,

$$P(x_1, x_2, x_3, ..x_n) = \prod_i^n P(x_i|x_1..x_{i-1}) \tag{3}$$

Forward believe update

We aim to compute the probability of a hidden state x at t by induction. That is compute $P(x_t, e_{1:t})$.

Let x_t and x_{t-1} be two successive states and let e_t and e_{t-1} be the associated observed events respectively. The probability of the 4 events is given by $P(e_t, x_t, e_{1:t-1}x_{t-1})$. By applying the chain rule on the joint probability we obtain Equation 4.

$$P(e_t, x_t, e_{1:t-1}, x_{t-1}) = P(x_{t-1})P(e_{1:t-1}/x_{t-1})P(x_t|e_{1:t-1}, x_{t-1})P(e_t|x_t, e_{1:t-1}, x_{t-1}) \tag{4}$$

In HMM, we assume that x_t depends on x_{t-1} but is independent of states $x_\infty, .., x_{t-2}, ..x_1$. We also assume that e_t depends on x_t but is independent of events $e_\infty, ..e_1$. Specifically

1. x_t is independent of $e_{1:t-1}$ (e.g., the weather of today is independent of whether the guard brought or not their umbrella in the previous days)

2. e_t is independent of $e_{1:t-1}$ (e.g., the chance for the guard to bring their umbrella today is independent of bringing the umbrella in the previous days)

3. e_t is independent of x_{t-1} (e.g., the chance for the guard to bring their umbrella today is independent of the weather of yesterday)

Using these assumptions, we can transform Eq. 4 to:

$$P(e_t, x_t, e_{1:t-1}x_{t-1}) = P(x_{t-1})P(e_{1:t-1}/x_{t-1})P(x_t|x_{t-1})P(e_t|x_t) \tag{5}$$

Events sequence $e_{1:t}$ is $(e_t, e_{1:t-1})$. Thus,

$$P(x_t, e_{1:t}) = P(x_t, e_t, e_{1:t-1}) \tag{6}$$

Now, let's marginalize[1] over x_{t-1}

$$P(x_t, e_{1:t}) = \sum P(x_t, x_{t-1}, e_t, e_{1:t-1}) \tag{7}$$

Applying Eq. 5 gives:

$$P(x_t, e_{1:t}) = \sum_{x_{t-1}} P(x_{t-1})P(e_{1:t-1}|x_{t-1})P(x_t|x_{t-1})P(e_t|x_t) \tag{8}$$

Use conditional probability of Eq. 1

$$P(e_{1:t-1}, x_{t-1}) = P(e_{1:t-1}|x_{t-1})P(x_{t-1}) \tag{9}$$

Plug Eq. 9 into Eq 8 to obtain

$$P(x_t, e_{1:t}) = \sum_{x_{t-1}} P(x_{t-1}, e_{1:t-1})P(x_t/x_{t-1})P(e_t/x_t) \tag{10}$$

Factor out $P(e_t/x_t)$

$$P(x_t, e_{1:t}) = P(e_t/x_t) \sum_{x_{t-1}} P(x_{t-1}, e_{1:t-1})P(x_t/x_{t-1}) \tag{11}$$

$\square$

Now:

$$P(x_t|e_{1:t}) = P(x_t, e_{1:t})P(e_{1:t}) \tag{12}$$

With Normalization

$$P(x_t|e_{1:t}) \propto P(x_t, e_{1:t}) \tag{13}$$

[1] https://en.wikipedia.org/wiki/Marginal$_{d}istribution$