

CS221 Fall 2018 Homework [car]

SUNet ID: prabhjot

Name: Prabhjot Singh Rai

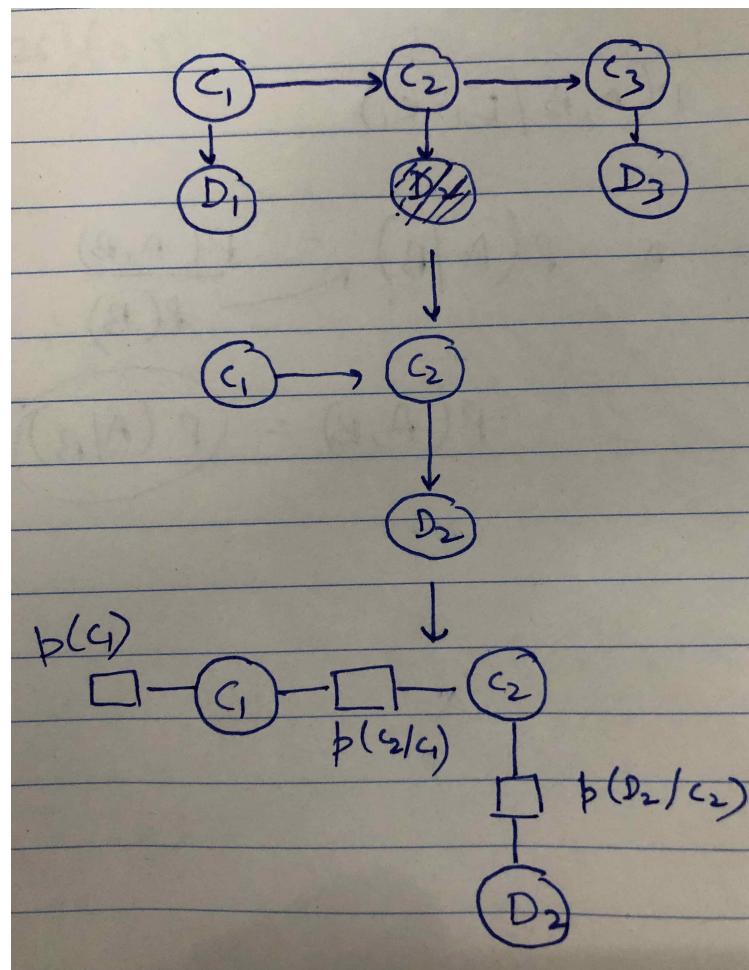
By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1

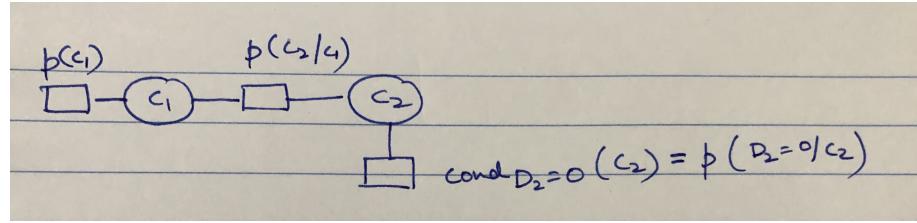
- (a) Step 1: Remove variables that are not ancestors

Step 2: Converting to factor graph

Step 1 and step 2 are shown in diagram below:



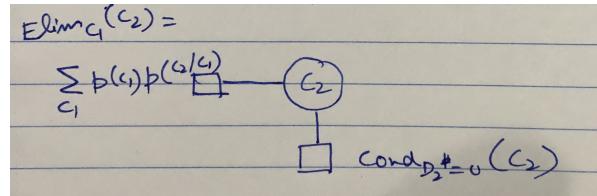
Step3: Conditioning on $D_2 = 0$



Condition variable D_2 on value $D_2 = 0$, replacing it with a factor $\text{cond}_{D_2=0}(C_2)$, we get

$$\begin{array}{ll} \text{cond}_{D_2=0}(C_2) & C_2 \\ 1 - \eta & 0 \\ \eta & 1 \end{array}$$

Step4: Eliminate C_1



$$\begin{aligned} \text{elim}_{C_1}(C_2) &= \sum_{C_1} p(C_1)p(C_2|C_1) \\ &= 0.5 \sum_{C_1} p(C_2|C_1) \end{aligned}$$

This is given from the below table:

$$\begin{array}{ll} \text{elim}_{C_1}(C_2) & C_2 \\ 0.5(1 - \epsilon + \epsilon) = 0.5 & 0 \\ 0.5(\epsilon + 1 - \epsilon) = 0.5 & 1 \end{array}$$

Therefore, now that we know $\text{elim}_{C_1}(C_2)$ and $\text{cond}_{D_2=0}(C_2)$,

$$p(C_2/D_2 = 0) = \text{elim}_{C_1}(C_2) * \text{cond}_{D_2=0}(C_2)$$

$$\begin{array}{ll} p(C_2/D_2 = 0) & C_2 \\ 0.5(1 - \eta) & 0 \\ 0.5\eta & 1 \end{array}$$

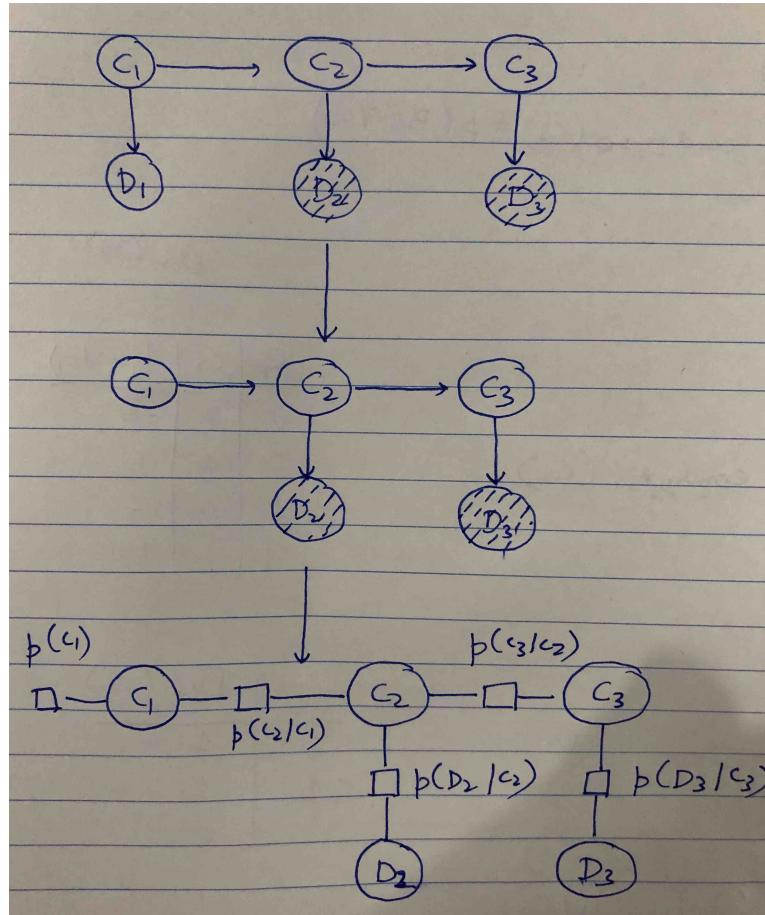
Hence, the given query,

$$\begin{aligned} p(C_2 = 1 / D_2 = 0) &= \frac{0.5\eta}{0.5\eta + 0.5(1 - \eta)} \\ &= \eta \end{aligned}$$

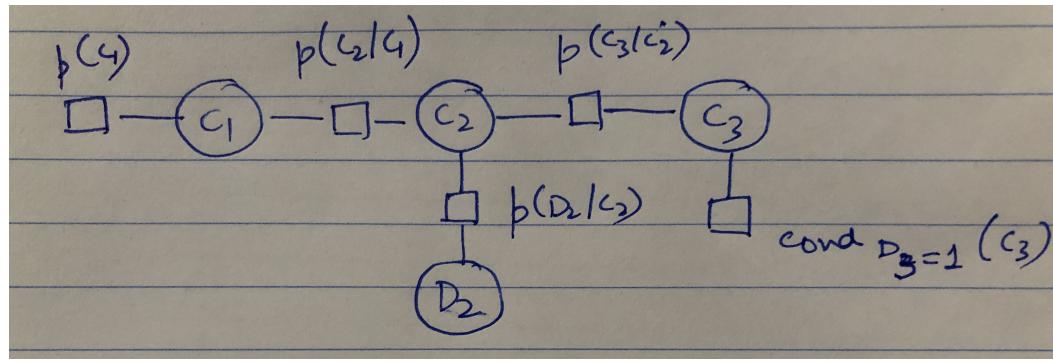
(b) **Step1:** Remove variables that are not ancestors

Step2: Converting to factor graph

Step 1 and step 2 are illustrated below:



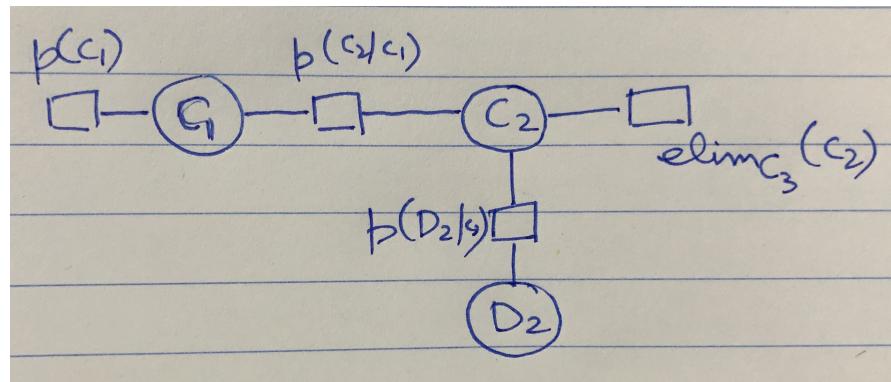
Step3: Conditioning on $D_3 = 1$



Conditioning on variable D_3 , and replacing it with a factor $\text{cond}_{D_3=1}(C_3)$, we get

$$\begin{array}{ll} \text{cond}_{D_3=1}(C_3) & C_3 \\ \eta & 0 \\ 1 - \eta & 1 \end{array}$$

Step4: Eliminating C_3



Defining function $\text{elim}_{C_3}(C_2)$ in order to eliminate node C_3 as

$$\text{elim}_{C_3}(C_2) = \sum_{C_3} \text{cond}_{D_3=1}(C_3)p(C_3|C_2)$$

The probability distribution $p(C_3|C_2)$ is given by:

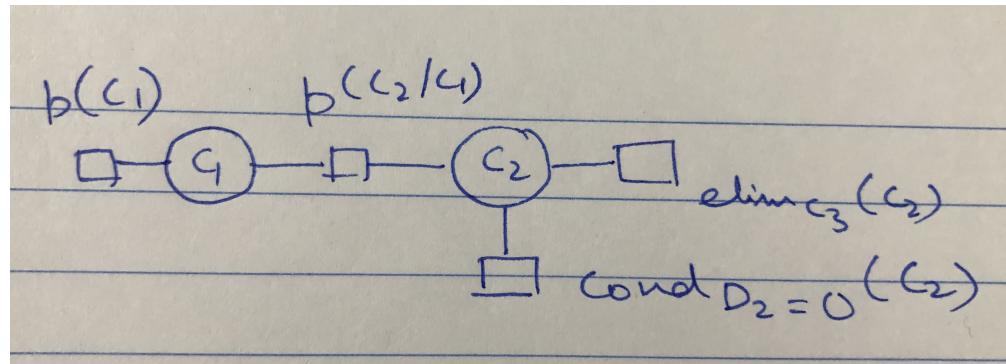
C2	C3	$p(C_3 C_2)$
0	0	$1 - \epsilon$
0	1	ϵ
1	0	ϵ
1	1	$1 - \epsilon$

The probability distribution $\text{cond}_{D_3=1}(C_3)$ is defined in Step 3.

Combining both and substituting in equation 1, and doing summation over values of C_3 , we will have probability distribution of $\text{elim}_{C_3}(C_2)$ is given by:

$$\begin{array}{ll} C_2 & \text{elim}_{C_3}(C_2) \\ 0 & (1 - \epsilon)\eta + \epsilon(1 - \eta) \\ 1 & \epsilon\eta + (1 - \eta)(1 - \epsilon) \end{array}$$

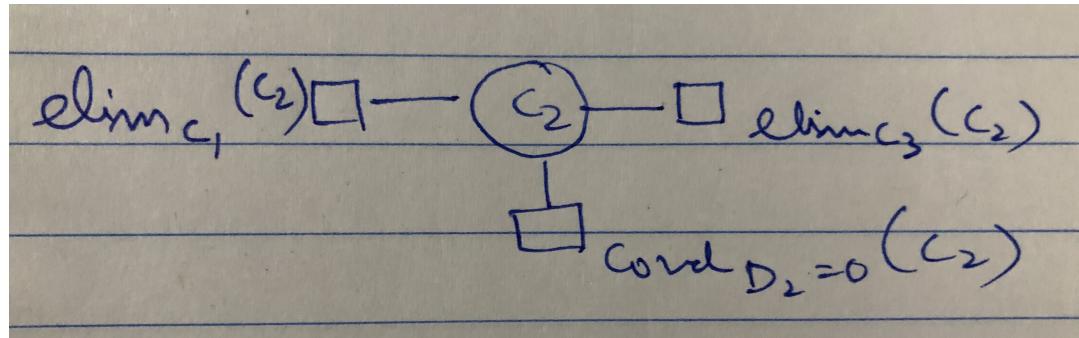
Step5: Conditioning on $D_2 = 0$



Condition variable D_2 on value $D_2 = 0$, replacing it with a factor $\text{cond}_{D_2=0}(C_2)$, we get

$$\begin{array}{ll} \text{cond}_{D_2=0}(C_2) & C_2 \\ 1 - \eta & 0 \\ \eta & 1 \end{array}$$

Step6: Eliminate C_1



$$\begin{aligned} \text{elim}_{C_1}(C_2) &= \sum_{C_1} p(C_1)p(C_2/C_1) \\ &= 0.5 \sum_{C_1} p(C_2/C_1) \end{aligned}$$

This is given from the below table:

$\text{elim}_{C_1}(C_2)$	C_2
$0.5(1 - \epsilon + \epsilon) = 0.5$	0
$0.5(\epsilon + 1 - \epsilon) = 0.5$	1

Step7: Combining all factors of C_2

Therefore, now that we know $\text{elim}_{C_1}(C_2)$, $\text{cond}_{D_2=0}(C_2)$ and $\text{elim}_{C_3}(C_1)$,

$$p(C_2/D_2 = 0, D_3 = 1) = \text{elim}_{C_1}(C_2) * \text{cond}_{D_2=0}(C_2) * \text{elim}_{C_3}(C_2)$$

$p(C_2/D_2 = 0, D_3 = 1)$	C_2
$0.5((1 - \epsilon)\eta + \eta(1 - \epsilon))(1 - \eta)$	0
$0.5(\epsilon\eta + (1 - \eta)(1 - \epsilon))\eta$	1

Therefore,

$$\begin{aligned} P(C_2 = 1/D_2 = 0, D_3 = 1) &= \frac{0.5(\epsilon\eta + (1 - \eta)(1 - \epsilon))\eta}{0.5(\epsilon\eta + (1 - \eta)(1 - \epsilon))\eta + 0.5((1 - \epsilon)\eta + \epsilon(1 - \eta))(1 - \eta)} \\ &= \frac{(\epsilon\eta + (1 - \eta)(1 - \epsilon))\eta}{(\epsilon\eta + (1 - \eta)(1 - \epsilon))\eta + ((1 - \epsilon)\eta + \epsilon(1 - \eta))(1 - \eta)} \end{aligned}$$

(c) i.

$$\begin{aligned} P(C_2 = 1/D_2 = 0) &= 0.2 \\ P(C_2 = 1/D_2 = 0, D_3 = 1) &= 0.4157 \end{aligned}$$

- ii. Adding second sensor reading increased the probability from 0.2 to 0.4157. Since D_3 is equal to 1, it means we observed the location to be 1 at location 3. This would increase the probability of $C_3 = 1$ since the emission probability $p(d_t/c_t)$ favours similar values with higher probability. $C_3 = 1$ increases the probability of $C_2 = 1$, since the transition probability $p(c_t/c_{t-1})$ favours same location with higher probability.
- iii. Both the probabilities would be same when the sensor reading at D_3 doesn't matter. This won't matter when the transition probabilities $p(c_t/c_{t-1})$ are equal meaning no matter what is the value of c_3 out of all the possible values, we will get constant transition probability. This would happen when $\epsilon = 1 - \epsilon$, therefore when $\epsilon = 0.5$.

Problem 5

- (a) Simplifying the equation $P(C_{11}, C_{12}|E_1)$ using the Bayes Theorem, we have

$$P(C_{11}, C_{12}|E_1) = P(C_{11}/E_1|C_{12}/E_1) * P(C_{12}/E_1)$$

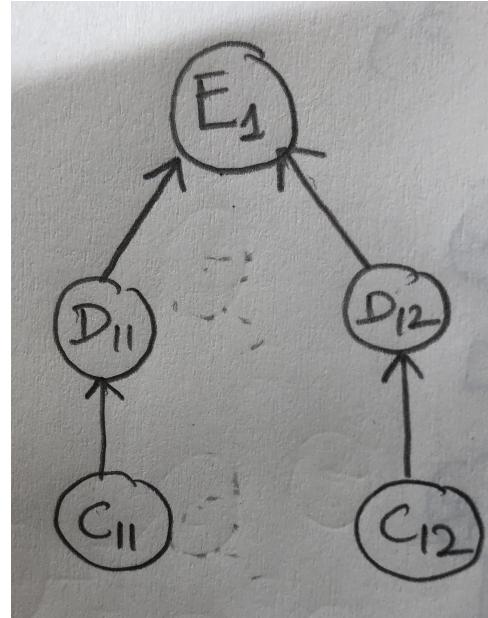
Since the location of both the cars are independent of each other (C_{11} doesn't depend in any way on C_{12} and vice versa), therefore, C_{11}/E_1 and C_{12}/E_1 are independent events.

$$P(C_{11}, C_{12}|E_1) = P(C_{11}/E_1) * P(C_{12}/E_1)$$

For the given question, we have to compute $P(C_{11}, C_{12}/E_1 = e_1)$. This means we are given a permutation e_1 out of possible domain values of E_1 , which means observed distance of car 1 is e_{11} and observed distance of car 2 is e_{12} .

$$\begin{aligned} P(C_{11}, C_{12}|E_1 = e_1) &= P(C_{11}/E_1 = e_1) * P(C_{12}/E_1 = e_1) \\ &\propto p(c_{11})p_N(e_{11}, \|a_1 - c_{11}\|, \sigma^2) * p(c_{12})p_N(e_{12}, \|a_1 - c_{12}\|, \sigma^2) \end{aligned}$$

Here's what the Bayesian Network would look like:



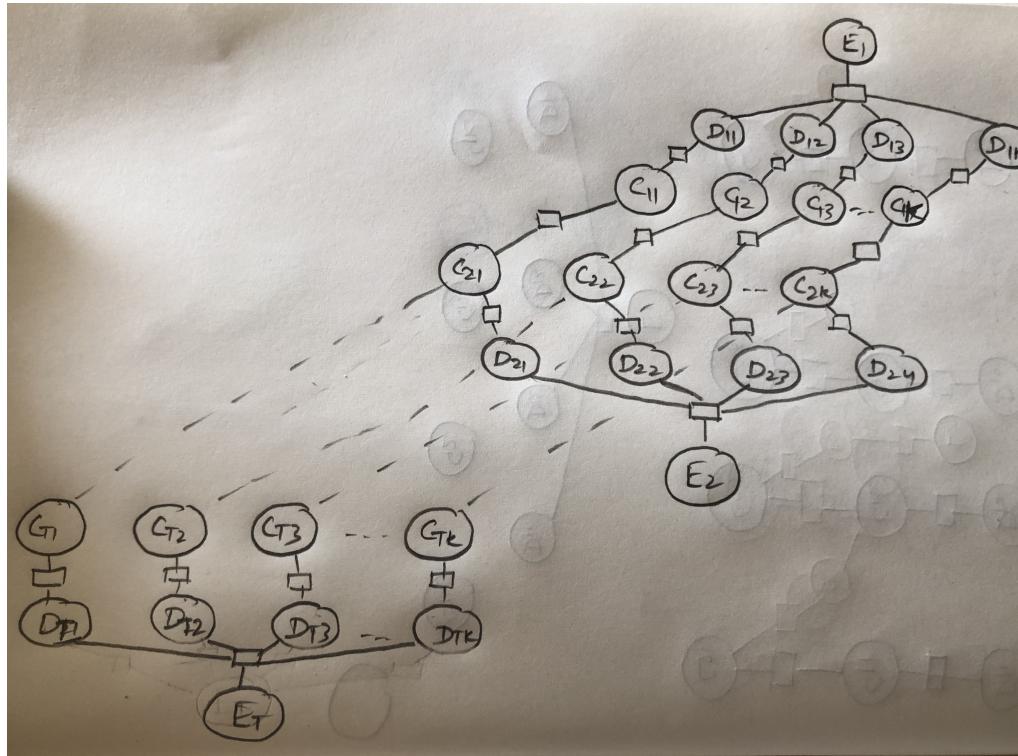
- (b) As per the solution in 5a, the joint probability would be defined as

$$\begin{aligned} P(C_{11} = c_{11}, C_{12} = c_{12} \dots C_{1K} = c_{1k}|E_1 = e_1) \\ \propto p(c_{11})p_N(e_{11}, \|a_1 - c_{11}\|, \sigma^2) * p(c_{12})p_N(e_{12}, \|a_1 - c_{12}\|, \sigma^2) \dots p(c_{1K})p_N(e_{1K}, \|a_1 - c_{1K}\|, \sigma^2) \\ \propto p(c_{1i})^K \prod_{j=1}^K p_N(e_{1j}, \|a_1 - c_{1j}\|, \sigma^2) \end{aligned}$$

Since $e((e_{11}, e_{12} \dots e_{1k}))$ is one set of readings from the list of all readings in E_1 , therefore, in order to find the maximum value of the above equation, the minimum number of assignments to find maximum probability would be the total number of possible permutations of different distances observed, namely $(D_{11}, D_{12} \dots D_{1k})$. For each permutation, we can compute the value of the above product and find out the permutation for which the product is the maximum. Therefore, permutations of $(D_{11}, D_{12} \dots D_{1k})$ is: D_{11} can be assigned in k ways, D_{12} can be assigned in $k - 1$ ways and so on. Therefore,

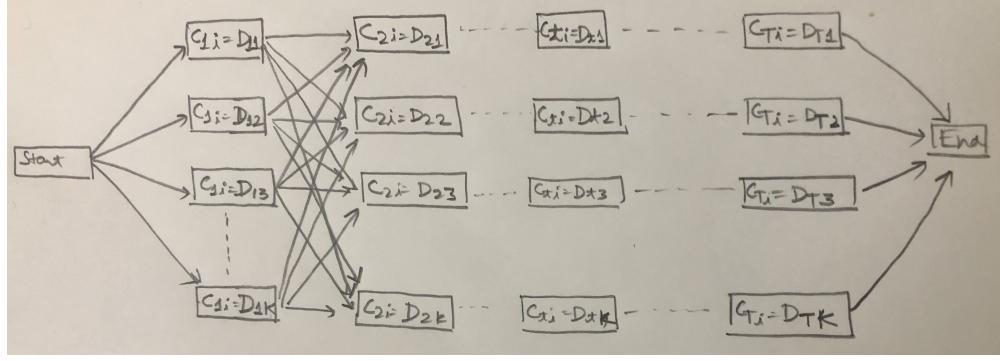
$$\begin{aligned}\text{Minimum assignments} &= k * (k - 1) * (k - 2) \dots (2)(1) \\ &= k!\end{aligned}$$

- (c) Converting to factor graph, it would look like the following for the posterior distribution over all K car locations at all T time steps conditioned on all sensor readings:



Even when we start eliminating variables along the $D_{11}, C_{11}, \dots, C_{T1}, D_{T1}, \dots, D_{1K}, C_{1K}, \dots, C_{TK}, D_{TK}$, we will still have factors with maximum arity as K which are connected with every E_t where t varies from 1 to T . Therefore, the tree width of this factor graph is K .

- (d) In order to compute $p(c_{ti}/e_1, e_2, \dots, e_t)$, this would be a smoothening query in HMM. Domain of C_{ti} would be the number of readings we get since those are exact values with equal probability of being assigned to C_{ti} , therefore has the domain $\{D_{t1}, D_{t2}, D_{t3}, \dots, D_{tK}\}$. Following would be the lattice representation of this scenario:



We are interested in calculating the total number of paths which will pass through a particular value of C_{ti} , let's assume D_{t1} . This can be solved through Forward-Backward algorithm.

Forward

$$F_i(h_i) = \sum_{c_{(t-1)i}} F_{i-1}(c_{(t-1)i}) w(c_{(t-1)i}, c_{ti})$$

Backward

$$B_i(h_i) = \sum_{c_{(t+1)i}} B_{i+1}(c_{(t+1)i}) w(c_{ti}, c_{(t+1)i})$$

Finally we define $S_i(c_{ti}) = F_i(h_i)B_i(h_i)$, since it would be the sum of weights of paths from start to end. Finally we would normalise over the sum of the weights of all the possible paths from start to end.

The algorithm would be as follows:

Compute $F_1, F_2, F_3, \dots, F_t$

Compute $B_t, B_{t-1} \dots, B_1$

Compute S_i for each i and normalise.