Ling 5801: Lecture Notes 15
From Probability Models to Sequence Models

Contents

15.1 Repeated trials

We can extend probability models to include unbounded number of trials.

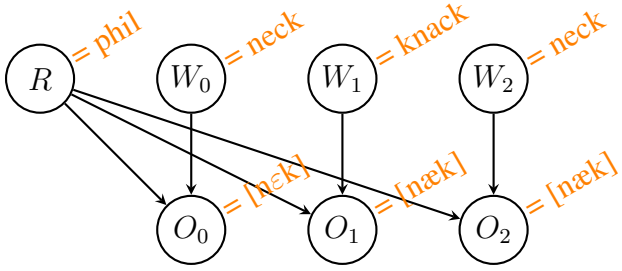
This is well defined if models are re-used. This is called ***stationarity***.

For example:

$$\langle R \times W_0 \times \cdots \times W_T \times O_0 \times \cdots \times O_T,\ 2^{R \times W_0 \times \cdots \times W_T \times O_0 \times \cdots \times O_T},\ \mathsf{P} \rangle$$
$$\text{where } R = \{\text{ohio}, \text{phil}\}, W_t = \{\text{neck}, \text{knack}\}, O_t = \{[\text{nɛk}], [\text{næk}]\},$$

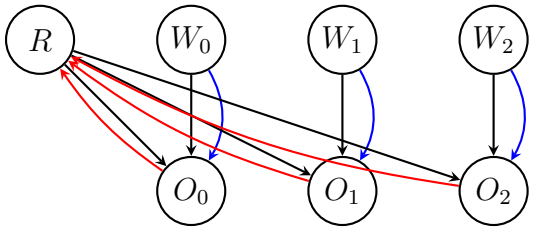
$$\mathsf{P}_{\theta_R}(R) \stackrel{\text{def}}{=} \mathsf{P}(R)$$
$$\mathsf{P}_{\theta_W}(W_t \mid R, W_0, \ldots, W_{t-1}) \stackrel{\text{def}}{=} \mathsf{P}(W_t)$$
$$\mathsf{P}_{\theta_O}(O_t \mid R, W_0, \ldots, W_T, O_0, \ldots, O_{t-1}) \stackrel{\text{def}}{=} \mathsf{P}(O_t \mid R, W_t)$$

Graphically (with example values diagonally):



Inference:

$$\mathsf{P}_{\theta_{Wag}}(r) = \sum_{w_0,w_1,w_2,o_0,o_1,o_2} \mathsf{P}_{\theta_{Wag}}(r, w_0, w_1, w_2, o_0, o_1, o_2)$$
$$\stackrel{\text{def}}{=} \mathsf{P}(r) \cdot \left(\sum_{r\ o_0,w_0} \left(\mathsf{P}(w_0) \right)_{w_0} \cdot \mathsf{P}(o_0 \mid r, w_0) \right) \cdot \left(\sum_{r\ o_1,w_1} \left(\mathsf{P}(w_1) \right)_{w_1} \cdot \mathsf{P}(o_1 \mid r, w_1) \right) \cdot \left(\sum_{r\ o_2,w_2} \left(\mathsf{P}(w_2) \right)_{w_2} \cdot \mathsf{P}(o_2 \mid r, w_2) \right)$$



Complexity:

$\mathcal{O}(n \cdot |\mathcal{D}_R| \cdot |\mathcal{D}_W| \cdot |\mathcal{D}_O|)$

15.2 Interdependence

Models with unbounded number of variables can have variables be interdependent.

This kind of model is still stationary.

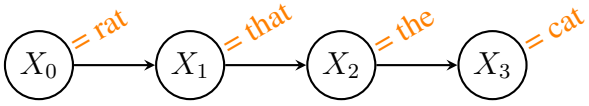
For example:

1. ***Markov chain***: a simple sequence of sucessively dependent variables:

$$\langle X_0 \times \cdots \times X_T, 2^{X_0 \times \cdots \times X_T}, \mathsf{P} \rangle$$
$$\mathsf{P}_{\theta_{X_0}}(X_0) \stackrel{\text{def}}{=} \mathsf{P}(X_0)$$
$$\mathsf{P}_{\theta_X}(X_t \mid X_0, \ldots, X_{t-1}) \stackrel{\text{def}}{=} \mathsf{P}(X_t \mid X_{t-1})$$

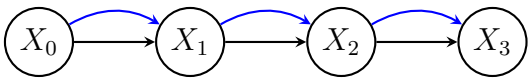
   Note π_X for first variable, θ_X subsequently.

   Graphically:



   Inference:

$$\mathsf{P}_{\theta_{MC}}(x_3) = \sum_{x_0,x_1,x_2} \mathsf{P}_{\theta_{MC}}(x_0, x_1, x_2, x_3)$$
$$\stackrel{\text{def}}{=} \sum_{x_2} \left(\sum_{x_2\ x_1} \left(\sum_{x_1\ x_0} \left(\mathsf{P}(x_0) \right)_{x_0} \cdot \mathsf{P}(x_1 \mid x_0) \right) \cdot \mathsf{P}(x_2 \mid x_1) \right) \cdot \mathsf{P}(x_3 \mid x_2)$$



Complexity:

$\mathcal{O}(n)$ (if X observed)

2. ***Hidden Markov model***:

seq. of successively dependent hidden variables ea. w. dependent observation:

$$\langle Y_0 \times \cdots \times Y_T \times X_0 \times \cdots \times X_T, 2^{Y_0 \times \cdots \times Y_T \times X_0 \times \cdots \times X_T}, \mathsf{P} \rangle$$

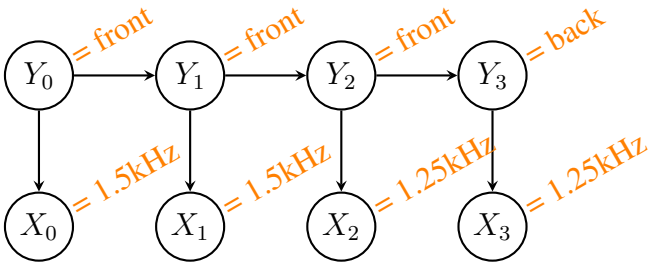
$$\mathsf{P}_{\theta_{Y_0}}(Y_0) \stackrel{\text{def}}{=} \mathsf{P}(Y_0)$$

$$\mathsf{P}_{\theta_Y}(Y_t \mid Y_0, \ldots, Y_{t-1}) \stackrel{\text{def}}{=} \mathsf{P}(Y_t \mid Y_{t-1})$$

$$\mathsf{P}_{\theta_X}(X_t \mid Y_0, \ldots, Y_T, X_0, \ldots, X_{t-1}) \stackrel{\text{def}}{=} \mathsf{P}(X_t \mid Y_t)$$

Note π_Y for first variable, θ_Y subsequently.

Graphically:

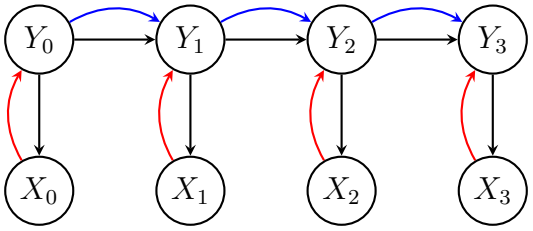


Inference:

$$\mathsf{P}_{\theta_{HMM}}(y_3) = \sum_{x_0,x_1,x_2,x_3,y_0,y_1,y_2} \mathsf{P}_{\theta_{HMM}}(x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3)$$

$$\stackrel{\text{def}}{=} \sum_{y_2} \left(\sum_{y_2} \sum_{y_1} \left(\sum_{y_1} \sum_{y_0} \left(\mathsf{P}(y_0) \cdot \left(\sum_{y_0} \sum_{x_0} \mathsf{P}(x_0 \mid y_0) \right) \right) \cdot \mathsf{P}(y_1 \mid y_0) \cdot \left(\sum_{y_1} \sum_{x_1} \mathsf{P}(x_1 \mid y_1) \right) \right) \cdot \right.$$

$$\left. \mathsf{P}(y_2 \mid y_1) \cdot \left(\sum_{y_2} \sum_{x_2} \mathsf{P}(x_2 \mid y_2) \right) \right) \cdot \mathsf{P}(y_3 \mid y_2) \cdot \left(\sum_{y_3} \sum_{x_3} \mathsf{P}(x_3 \mid y_3) \right)$$



Algorithm:

# initialize table of possible states at each time step using start states
for y_0 in Y:
  $V[0, y_0] = \mathsf{P}_{\pi_Y}(y_0) \cdot \mathsf{P}_{\theta_X}(x_0 \mid y_0)$

# for each possible state y_{t-1} in V at time t, for each y_{t-1}, x_{t-1}, y_t in M, add y_t
for each t in $1..T$:
 for each y_{t-1} in Y:
  for each y_t in Y:
   $V[t, y_t] = V[t, y_t] + (V[t-1, y_{t-1}] \cdot \mathsf{P}_{\theta_Y}(y_t \mid y_{t-1}) \cdot \mathsf{P}_{\theta_X}(x_t \mid y_t))$

Complexity:

$\mathcal{O}(n \cdot |\mathcal{D}_Y| \cdot |\mathcal{D}_Y|)$ (if X observed)

15.3 Example filtering (estimation of last hidden variable)

For example, in an HMM for estimating front/back tongue position with parameters:

$\pi_Y =$

front	back
.5	.5

$\theta_Y =$

Y_{t-1}	front	back
front	.8	.2
back	.2	.8

(encodes inertia in tongue position)

$\theta_X =$

Y_t	.5kHZ	.75kHZ	1kHz	1.25kHz	1.5kHz	1.75kHz	2kHz
front	0	0	.1	.2	.4	.2	.1
back	.1	.2	.4	.2	.1	0	0

we obtain forward messages calculating $\mathsf{P}(Y_t, X_{1..t})$, marginalizing out all $Y_{1..t-1}$:

t	X_t	$\mathsf{P}(Y_t = \text{front}, X_{1..t})$	$\mathsf{P}(Y_t = \text{back}, X_{1..t})$
0	1.5kHz	$.5 \cdot .4 = .2$	$.5 \cdot .1 = .05$
1	1.5kHz	$.2 \cdot .8 \cdot .4 + .05 \cdot .2 \cdot .4 = .068$	$.2 \cdot .2 \cdot .1 + .05 \cdot .8 \cdot .1 = .008$
2	1.25kHz	$.068 \cdot .8 \cdot .2 + .008 \cdot .2 \cdot .2 = .0112$	$.068 \cdot .2 \cdot .2 + .008 \cdot .8 \cdot .2 = .004$

15.4 Most likely sequence estimation

Subst semiring: $\langle \mathbb{R}_0^\infty, +, \cdot, 0, 1 \rangle$ to $\langle \mathbb{R}_0^\infty \times (\mathcal{D}_Y \times \mathcal{D}_X)^*, \text{max-argmax}, \text{prod-pair}, \langle 0, \langle\rangle\rangle, \langle 1, \langle\rangle\rangle \rangle$

where:

$$p, x \text{ max-argmax } q, y = \begin{cases} \text{if } p > q & : p, x \\ \text{otherwise} & : q, y \end{cases}$$

p, x prod-pair $q, y = p \cdot q, \langle x, y \rangle$

Algorithm:

# initialize table of possible states at each time step using start states
for y_0 in Y:
  $V[0, y_0] = \mathsf{P}_{\pi_Y}(y_0), y_0$ prod-pair $\mathsf{P}_{\theta_X}(x_0 \mid y_0), x_0$

# for each possible state y_{t-1} in V at time t, for each y_{t-1}, x_{t-1}, y_t in M, add y_t
for each t in $1..T$:
  for each y_{t-1} in Y:
    for each y_t in Y:
      $V[t, y_t] = V[t, y_t]$ max-argmax $(V[t-1, y_{t-1}]$ prod-pair $\mathsf{P}_{\theta_Y}(y_t \mid y_{t-1}), y_t$ prod-pair $\mathsf{P}_{\theta_X}(x_t \mid y_t), x_t)$

Complexity:

$\mathcal{O}(n \cdot |\mathcal{D}_Y| \cdot |\mathcal{D}_Y|)$ (if X observed)

Example: obtain Viterbi MLS messages w. maximal $\mathsf{P}(y_{1..t}, x_{1..t})$ ending at each y_t:

t	X_t	Y_t = front	Y_t = back
0	1.5	$.5 \cdot .4 = .2$, ⟨front,1.5⟩	$.5 \cdot .1 = .05$, ⟨back,1.5⟩
1	1.5	$\max(.2 \cdot .8 \cdot .4, .05 \cdot .2 \cdot .4) = .064$, ⟨⟨front,1.5⟩, ⟨front,1.5⟩⟩	$\max(.2 \cdot .2 \cdot .1, .05 \cdot .8 \cdot .1) = .004$, ⟨⟨back,1.5⟩, ⟨back,1.5⟩⟩
2	1.25	$\max(.064 \cdot .8 \cdot .2, .004 \cdot .2 \cdot .2) = .01024$, ⟨⟨⟨front,1.5⟩, ⟨front,1.5⟩⟩, ⟨front,1.25⟩⟩	$\max(.064 \cdot .2 \cdot .2, .004 \cdot .8 \cdot .2) = .00256$, ⟨⟨⟨front,1.5⟩, ⟨front,1.5⟩⟩, ⟨back,1.25⟩⟩

15.5 Weighted finite-state automaton

1. Semiring substitution from $\langle \{T, F\}, \vee, \wedge, F, T \rangle$ to $\langle \mathbb{R}_0^\infty, +, \cdot, 0, 1 \rangle$ allows weighted FSA:

   # initialize table of possible states at each time step using start states
   for each q_t in Q:
     $V[0, q_t] = S[q_t]$

   # for each possible state t_{t-1} in V at time t, for each q_{t-1}, x_{t-1}, q_t in M, add q_t
   for each t in $1..T$:
     for each q_{t-1} in Q:
       for each q_t in Q:
         $V[t, q_t] = V[t, q_t] + (V[t-1, q_{t-1}] \cdot M[q_{t-1}, x_{t-1}, q_t])$

2. Now replacing S with probabilities π_Q and factoring M into θ_X and θ_Q gives:

   # initialize table of possible states at each time step using start states
   for q_t in Q:
     $V[0, q_t] = \mathsf{P}_{\pi_Q}(q_t)$

   # for each possible state q_{t-1} in V at time t, for each q_{t-1}, x_{t-1}, q_t in M, add q_t
   for each t in $1..T$:

for each q_{t-1} in Q:
  for each q_t in Q:
    $V[t, q_t] = V[t, q_t] + (V[t-1, q_{t-1}] \cdot \mathsf{P}_{\theta_X}(x_{t-1} \mid q_{t-1}) \cdot \mathsf{P}_{\theta_Q}(q_t \mid q_{t-1}, x_{t-1}))$

where: $\mathsf{P}_{\theta_X}(x_{t-1} \mid q_{t-1}) = \sum_{q_t} M[q_{t-1}, x_{t-1}, q_t]$

$$\mathsf{P}_{\theta_Q}(q_t \mid q_{t-1}, x_{t-1}) = \frac{M[q_{t-1}, x_{t-1}, q_t]}{\sum_{q_t} M[q_{t-1}, x_{t-1}, q_t]}$$

3. This recognizer can be expressed as a probability model:

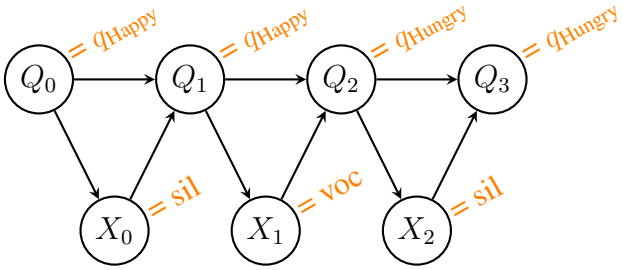
$$\langle Q_0 \times X_0 \times Q_1 \times \cdots \times X_{T-1} \times Q_T, 2^{Q_0 \times X_0 \times Q_1 \times \cdots \times X_{T-1} \times Q_T}, \mathsf{P} \rangle$$

$$\mathsf{P}_{\theta_{Q_0}}(Q_0) \stackrel{\text{def}}{=} \mathsf{P}(Q_0)$$

$$\mathsf{P}_{\theta_X}(X_t \mid Q_0, X_0, Q_1, \ldots, X_{t-1}, Q_t) \stackrel{\text{def}}{=} \mathsf{P}(X_t \mid Q_t)$$

$$\mathsf{P}_{\theta_Q}(Q_t \mid Q_0, X_0, Q_1, \ldots, X_{t-1}, Q_{t-1}, X_{t-1}) \stackrel{\text{def}}{=} \mathsf{P}(Q_t \mid Q_{t-1}, X_{t-1})$$

Graphically:

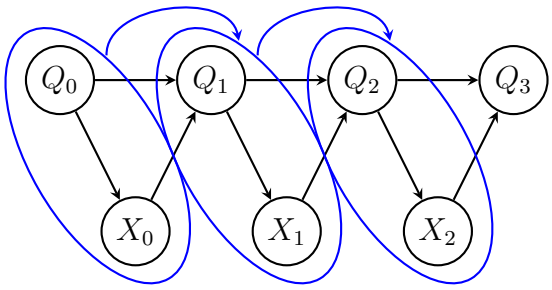


Inference:

$$\mathsf{P}_{\theta_{WFSA}}(q_3) = \sum_{x_0,x_1,x_2,q_0,q_1,q_2} \mathsf{P}_{\theta_{WFSA}}(x_0, x_1, x_2, q_0, q_1, q_2, q_3)$$

$$\stackrel{\text{def}}{=} \sum_{q_2} \left(\sum_{q_2,x_2} \left(\sum_{q_1,x_1} \left(\sum_{q_0,x_0} \left(\mathsf{P}(q_0) \cdot \mathsf{P}(x_0 \mid q_0) \right) \cdot \mathsf{P}(q_1 \mid q_0, x_0) \cdot \mathsf{P}(x_1 \mid q_1) \right) \cdot \right. \right.$$

$$\left. \left. \mathsf{P}(q_2 \mid q_1, x_1) \cdot \mathsf{P}(x_2 \mid q_2) \right) \cdot \mathsf{P}(q_3 \mid q_2, x_2) \right)$$

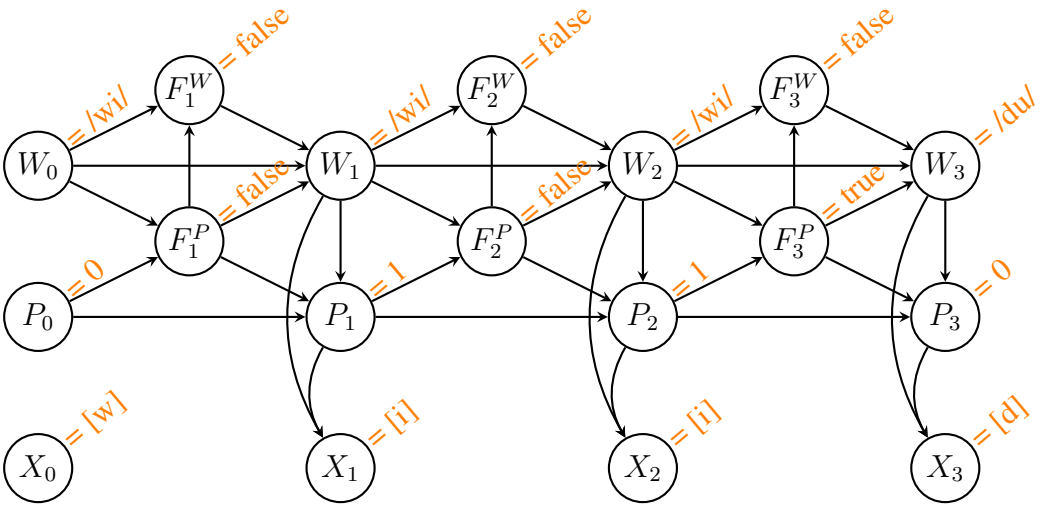


Complexity:

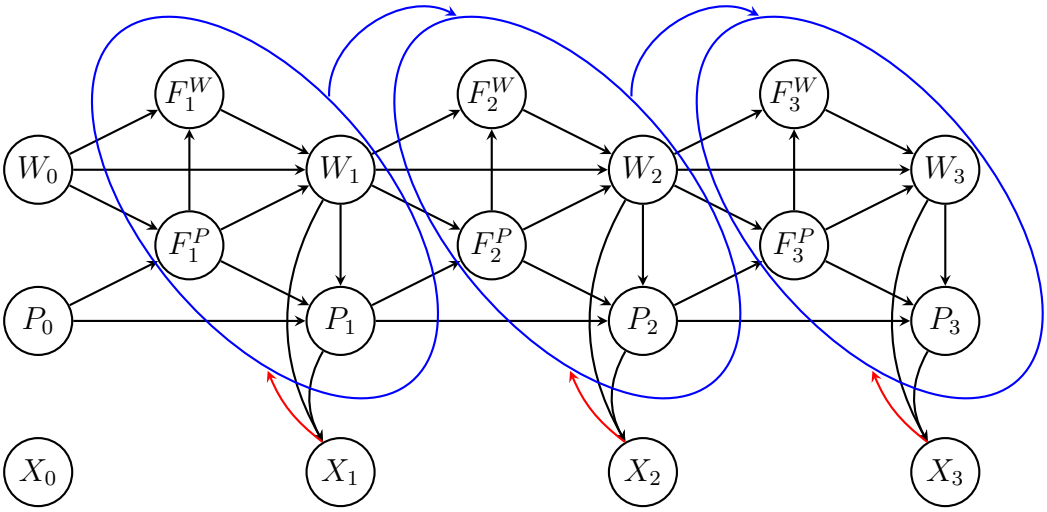
$\mathcal{O}(n \cdot |\mathcal{D}_Q| \cdot |\mathcal{D}_Q|)$ (if X observed)

15.6 Factored sequence model for speech recognition

Synchronize high-/low-level sequences with true/false 'final state' variables:



Inference:



Complexity:

$\mathcal{O}(n \cdot |\mathcal{D}_W| \cdot |\mathcal{D}_P| \cdot |\mathcal{D}_F| \cdot |\mathcal{D}_F| \cdot |\mathcal{D}_W| \cdot |\mathcal{D}_P|)$ (if X observed)

15.7 Weighted PDA (bounded as hierarchic HMM)

1. Semiring substitution from $\langle \{T, F\}, \vee, \wedge, F, T \rangle$ to $\langle \mathbb{R}_0^\infty, +, \cdot, 0, 1 \rangle$ allows weighted PDA:

   for each time step t:
   for each previous state q_{t-1}^d and store $q_{t-1}^{1..d-1}$ (where $d-1$ is the depth of the store):

i) for each final state f_t^d and current state q_t^d: (expand +) trans

$$V[t, q_t^d q_{t-1}^{1..d-1}] = V[t, q_t^d q_{t-1}^{1..d-1}] + (V[t-1, q_{t-1}^d q_{t-1}^{1..d-1}] \cdot M[q_{t-1}^d, \epsilon, x_{t-1}, f_t^d, \epsilon] \cdot M[f_t^d, q_{t-1}^{d-1}, \epsilon, q_t^d, q_{t-1}^{d-1}])$$

ii) for each previous state at deeper level q_{t-1}^{d+1}, final state f_t^{d+1} and current state q_t^{d+1}:

$$V[t, q_t^{d+1} q_t^d q_{t-1}^{1..d-1}] = V[t, q_t^{d+1} q_t^d q_{t-1}^{1..d-1}] + (V[t-1, q_{t-1}^d q_{t-1}^{1..d-1}] \cdot M[q_{t-1}^d, \epsilon, \epsilon, q_{t-1}^{d+1}, q_{t-1}^d] \cdot M[q_{t-1}^{d+1}, \epsilon, x_{t-1}, f_t^{d+1}, \epsilon] \cdot M[f_t^{d+1}, q_{t-1}^d, \epsilon, q_t^{d+1}, q_{t-1}^d])$$

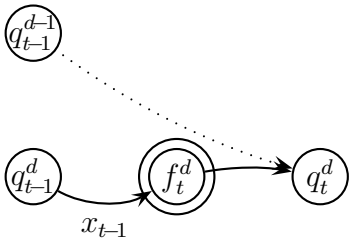
iii) for each final state f_t^d and current state q_t^{d-1}: (expand +) trans + reduce

$$V[t, q_t^{d-1} q_{t-1}^{1..d-2}] = V[t, q_t^{d-1} q_{t-1}^{1..d-2}] + (V[t-1, q_{t-1}^d q_{t-1}^{d-1} q_{t-1}^{1..d-2}] \cdot M[q_{t-1}^d, \epsilon, x_{t-1}, f_t^d, \epsilon] \cdot M[f_t^d, q_{t-1}^{d-1}, \epsilon, q_t^{d-1}, \epsilon])$$

iv) for each previous state at deeper level q_{t-1}^{d+1} and final state f_t^{d+1} and current state q_t^d:

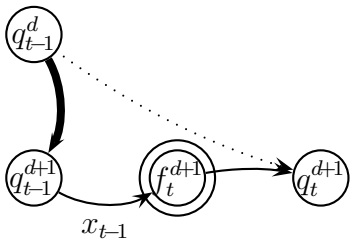
$$V[t, q_t^d q_t^{d-1} q_{t-1}^{1..d-2}] = V[t, q_t^d q_t^{d-1} q_{t-1}^{1..d-2}] + (V[t-1, q_{t-1}^d q_{t-1}^{d-1} q_{t-1}^{1..d-2}] \cdot M[q_{t-1}^d, \epsilon, \epsilon, q_{t-1}^{d+1}, q_{t-1}^d] \cdot M[q_{t-1}^{d+1}, \epsilon, x_{t-1}, f_t^{d+1}, \epsilon] \cdot M[f_t^{d+1}, q_{t-1}^d, \epsilon, q_t^d, \epsilon])$$

i)



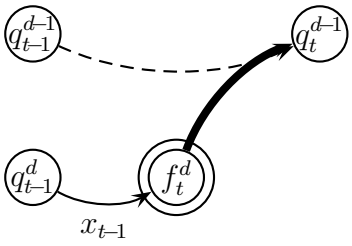
(trans only)

ii)

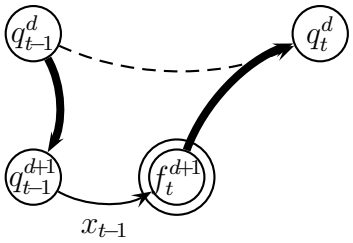


(expand + trans)

iii)



(trans + reduce)

iv) 

(expand + trans + reduce)

2. Now replacing S with probabilities π_Q and factoring M into θ_X, θ_F, and θ_Q gives:

for each time step t:
for each previous state q^d_{t-1} and store $q^{1..d-1}_{t-1}$ (where $d-1$ is the depth of the store):
i) for each final state f^d_t and current state q^d_t: (expand +) trans

$$V[t, q^d_t q^{1..d-1}_{t-1}] = V[t, q^d_t q^{1..d-1}_{t-1}] + (V[t-1, q^d_{t-1} q^{1..d-1}_{t-1}] \cdot \mathsf{P}_{\theta_X}(x_{t-1} \mid q^d_{t-1}) \cdot \mathsf{P}_{\theta_F}(f^d_t \mid -, \_, q^d_{t-1}, x_{t-1}) \cdot \mathsf{P}_{\theta_Q}(q^d_t \mid f^d_t, -, q^{d-1}_{t-1}, \_))$$

ii) for each previous state at deeper level q^{d+1}_{t-1}, final state f^{d+1}_t and current state q^{d+1}_t:

$$V[t, q^{d+1}_t q^d_t q^{1..d-1}_{t-1}] = V[t, q^{d+1}_t q^d_t q^{1..d-1}_{t-1}] + (V[t-1, q^d_{t-1} q^{1..d-1}_{t-1}] \cdot \mathsf{P}_{\theta_X}(x_{t-1} \mid q^d_{t-1}) \cdot \mathsf{P}_{\theta_F}(f^{d+1}_t \mid -, q^d_{t-1}, -, x_{t-1}) \cdot \mathsf{P}_{\theta_Q}(q^{d+1}_t \mid f^{d+1}_t, -, q^d_{t-1}, \_))$$

iii) for each final state f^d_t and current state q^{d-1}_t: (expand +) trans + reduce

$$V[t, q^{d-1}_t q^{1..d-2}_{t-1}] = V[t, q^{d-1}_t q^{1..d-2}_{t-1}] + (V[t-1, q^d_{t-1} q^{d-1}_{t-1} q^{1..d-2}_{t-1}] \cdot \mathsf{P}_{\theta_X}(x_{t-1} \mid q^d_{t-1}) \cdot \mathsf{P}_{\theta_F}(f^d_t \mid -, \_, q^d_{t-1}, x_{t-1}) \cdot \mathsf{P}_{\theta_Q}(q^{d-1}_t \mid -, f^d_t, \_, q^{d-1}_{t-1}))$$

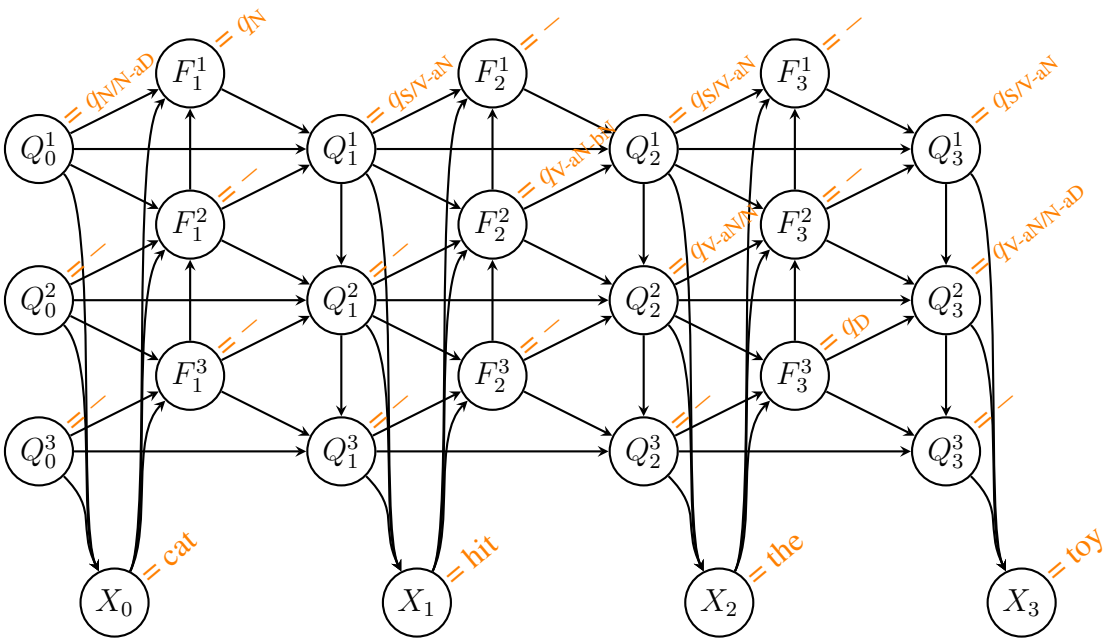
iv) for each previous state at deeper level q^{d+1}_{t-1} and final state f^{d+1}_t and current state q^d_t:

$$V[t, q^d_t q^{d-1}_t q^{1..d-2}_{t-1}] = V[t, q^d_t q^{d-1}_t q^{1..d-2}_{t-1}] + (V[t-1, q^d_{t-1} q^{d-1}_{t-1} q^{1..d-2}_{t-1}] \cdot \mathsf{P}_{\theta_X}(x_{t-1} \mid q^d_{t-1}) \cdot \mathsf{P}_{\theta_F}(f^{d+1}_t \mid -, q^d_{t-1}, -, x_{t-1}) \cdot \mathsf{P}_{\theta_Q}(q^d_t \mid -, f^{d+1}_t, \_, q^d_{t-1}))$$

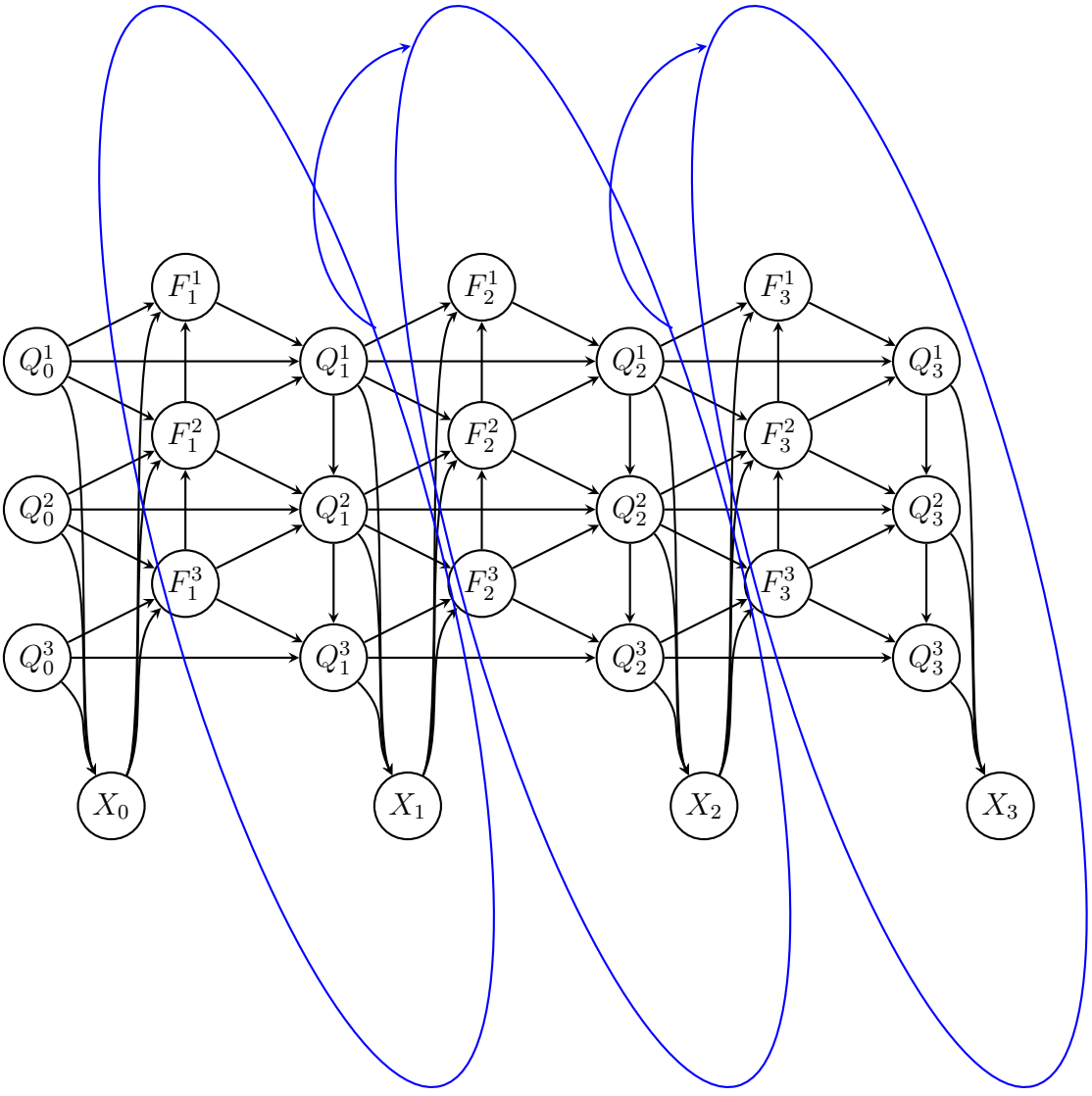
where:

$$\begin{aligned}
\mathsf{P}_{\theta_X}(x_{t-1} \mid q^d_{t-1}) &= \textstyle\sum_{f^d_t} M[q^d_{t-1}, \epsilon, x_{t-1}, f^d_t, \epsilon] \\
&\quad + \textstyle\sum_{f^{d+1}_t, q^{d+1}_{t-1}} M[q^d_{t-1}, \epsilon, \epsilon, q^{d+1}_{t-1}, q^d_{t-1}] \cdot M[q^{d+1}_{t-1}, \epsilon, x_{t-1}, f^{d+1}_t, \epsilon] \\
\mathsf{P}_{\theta_F}(f^d_t \mid -, _, q^d_{t-1}, x_{t-1}) &= M[q^d_{t-1}, \epsilon, x_{t-1}, f^d_t, \epsilon] \\
\mathsf{P}_{\theta_F}(f^{d+1}_t \mid -, q^d_{t-1}, -, x_{t-1}) &= \textstyle\sum_{q^{d+1}_{t-1}} M[q^d_{t-1}, \epsilon, \epsilon, q^{d+1}_{t-1}, q^d_{t-1}] \cdot M[q^{d+1}_{t-1}, \epsilon, x_{t-1}, f^{d+1}_t, \epsilon] \\
\mathsf{P}_{\theta_Q}(q^d_t \mid -, f^{d+1}_t, _, q^d_{t-1}) &= M[f^{d+1}_t, q^d_{t-1}, \epsilon, q^d_{t-1}, \epsilon] \\
\mathsf{P}_{\theta_Q}(q^d_t \mid f^d_t, -, q^{d-1}_t, _) &= M[f^d_t, q^{d-1}_{t-1}, \epsilon, q^d_t, q^{d-1}_t] \text{ where } q^{d-1}_{t-1} = q^{d-1}_t
\end{aligned}$$

3. This recognizer can be expressed as a probability model:


Q_0^1
Q_0^2
Q_0^3
F_1^1
F_1^2
F_1^3
Q_1^1
Q_1^2
Q_1^3
F_2^1
F_2^2
F_2^3
Q_2^1
Q_2^2
Q_2^3
F_3^1
F_3^2
F_3^3
Q_3^1
Q_3^2
Q_3^3
X_0
X_1
X_2
X_3
= $q_{N/N\text{-}aD}$
= q_N
= $q_{S/V\text{-}aN}$
= $q_{V\text{-}aN\text{-}by}$
= $q_{V\text{-}aN/N}$
= q_D
= $q_{V\text{-}aN/N\text{-}aD}$
= cat
= hit
= the
= toy
= _

Inference:



Complexity:

$\mathcal{O}(n \cdot |\mathcal{D}_Q|^D \cdot |\mathcal{D}_F|^D \cdot |\mathcal{D}_Q|^D)$ (if X observed)

Some observations:

- generates incremental probabilities (good for predicting reading times)
- connects one rule per word (good for efficient/human-like semantic composition)