Lecture 21:

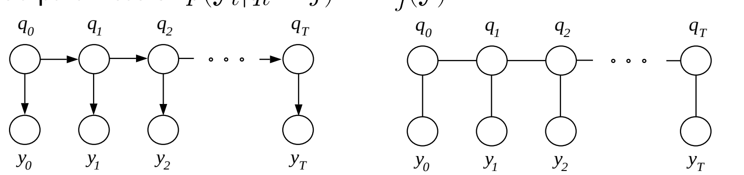
Junction Tree Derivation of HMM Inference

Sam Roweis

March 19, 2004

HMM Graphical Model

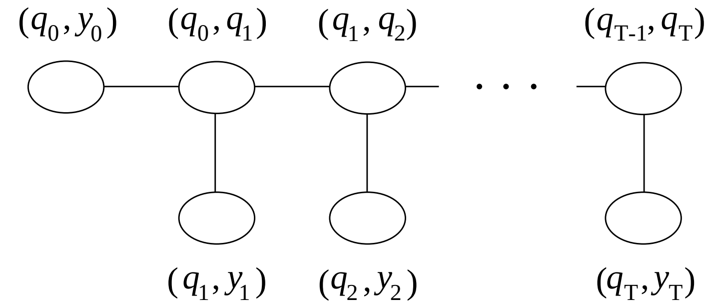
- Hidden states q_t, observations $\mathbf{y}_t$.
- Transition parameters: $p(q_{t+1} = j | q_t = i) = S_{ij}$
- Output parameters: $p(\mathbf{y}_t | q_t = j) = A_j(\mathbf{y})$



- Moralization easy: each node has a single parent.
- Triangulation easy: moralized graph has no cycles.

HMM Junction Tree

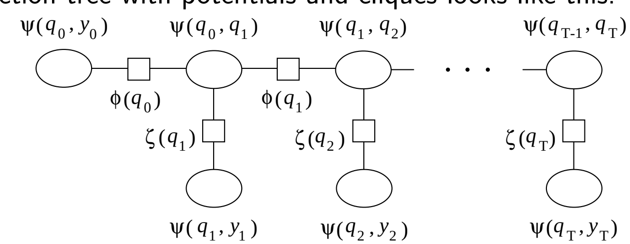
- Cliques of moralized-triangulated: (q_t, q_{t+1}) and $(q_t, \mathbf{y}_t)$.
- Many maximal spanning trees, so many junction trees.
  For standard algorithms, select this one:



- Other spanning trees lead to other algorithms.

Cliques and Potentials

- The junction tree with potentials and cliques looks like this:

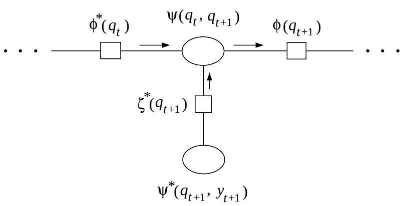


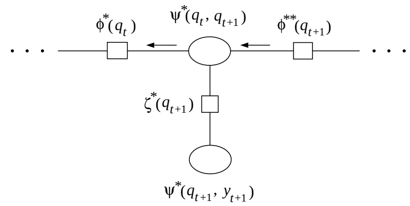
- Initialization:

$$
\begin{aligned}
\psi(q_0, \mathbf{y}_0) &= p(q_0) p(\mathbf{y}_0 | q_0) = \pi_{q_0} A_{q_0}(\mathbf{y}_0) \\
\psi(q_t, q_{t+1}) &= p(q_{t+1} || q_t) = S_{q_t, q_{t+1}} \\
\psi(q_t, \mathbf{y}_t) &= p(\mathbf{y}_t | q_t) = A_{q_t}(\mathbf{y}_t) \\
\phi(\cdot) &= 1 \\
\xi(\cdot) &= 1
\end{aligned}
$$

Message Passing (no evidence)

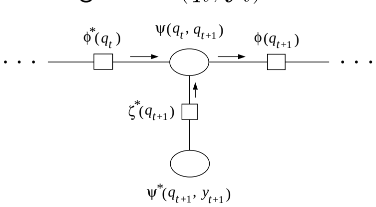
- Select (q_{T-1}, q_T) as the root.
- CollectEvidence(root) generates: observation messages upwards from $(q_t, \mathbf{y}_t)$ to (q_{t-1}, q_t); and backbone messages from (q_{t-1}, q_t) to (q_t, q_{t+1}).
- DistributeEvidence(root) generates: correction messages downwards from (q_{t-1}, q_t) to $(q_t, \mathbf{y}_t)$; and backwards from (q_t, q_{t+1}) to (q_{t-1}, q_t).

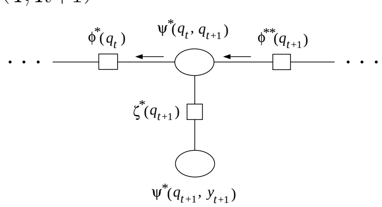




Message Passing (no evidence)

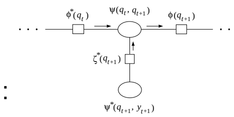
- Upwards messages: $\sum_{\mathbf{y}_t} \psi(q_t, \mathbf{y}_t) = \sum_{\mathbf{y}_t} p(\mathbf{y}_t | q_t) = 1$ so the separator potential $\xi^*(q_t) = 1$ is unchanged by marginalization.
- Upwards messages have no effect when no evidence is observed.
- Backbone messages: $\phi^*(q_0) = \sum_{\mathbf{y}_0} \psi(q_0, \mathbf{y}_0) = P(q_0)$
  $\psi^*(q_0, q_1) = \psi(q_0, q_1)\phi^*(q_0) = P(q_0, q_1)$ etc...
- All backbone potentials get converted to marginals in Collect phase. Backwards Distribute phase has no effect on ϕ.
- Distribute converts $\xi(q_t)$ into marginal $P(q_t)$ and $\psi(q_t, \mathbf{y}_t)$ into marginals $P(q_t, \mathbf{y}_t)$. No effect on $\psi(q, q_{t+1})$.





Message Passing with evidence – collect

- First set the ψ potentials to introduce evidence: $\psi(q_t, \mathbf{y}_t) = A_{q_t}\delta(\mathbf{y}_t - \bar{\mathbf{y}}_t)$.
- Now run Collect:
  - Marginalizing gives $\sum_{\mathbf{y}_t} \psi(q_t, \mathbf{y}_t) = A_{q_t}(\bar{\mathbf{y}}_t)$. Thus, separator $\xi^*(q_t) = p(\bar{\mathbf{y}}_t | q_t)$ for fixed $\bar{\mathbf{y}}_t$.
  - Consider update factors passed to (q_t, q_{t+1}):
    $\psi^*(q_t, q_{t+1}) = \psi(q_t, q_{t+1})\phi^*(q_t)\xi^*(q_{t+1})$.
    $\psi^*(q_t, q_{t+1}) = S_{q_t, q_{t+1}}\phi^*(q_t)P(\mathbf{y}_{t+1}|q_{t+1})$.
  - Initialize with $\phi^*(q_0) = p(\bar{\mathbf{y}}_0|q_0)p(q_0)$.
  - Now we can continue along the chain:
    $\phi^*(q_{t+1}) = \sum_{q_t} \psi^*(q_t, q_{t+1}) = \sum_{q_t} S_{q_t, q_{t+1}}\phi^*(q_t)P(\mathbf{y}_{t+1}|q_{t+1})$



- Notice: $\phi^*(q_t) = \alpha_t = p(\mathbf{y}_0^t, q_t)$
  We have recovered the α recursion automatically.
- After collect, how do we comput $L = p(\mathbf{Y})$?

Check of ϕ^*

- Check that $\phi^*(q_t) = P(\mathbf{y}_0^t, q_t)$.
- Initially, $\phi^*(q_0) = p(\bar{\mathbf{y}}_0|q_0)p(q_0)$.
- By induction:

$$\begin{aligned}\phi^*(q_{t+1}) &= \sum_{q_t} S_{q_t, q_{t+1}}\phi^*(q_t)P(\mathbf{y}_{t+1}|q_{t+1}) \\ &= \sum_{q_t} P(q_{t+1}|q_t)P(\mathbf{y}_0^t, q_t)P(\mathbf{y}_{t+1}|q_{t+1}) \\ &= \sum_{q_t} P(\mathbf{y}_0^t, \mathbf{y}_{t+1}, q_t, q_{t+1}) \\ &= P(\mathbf{y}_0^{t+1}, q_{t+1})\end{aligned}$$

- After collect, $\psi^*(q_{t-1}, q_t) = p(\mathbf{y}_0^t, q_{t-1}, q_t)$.

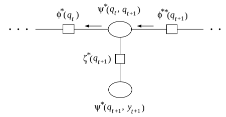
Message Passing with evidence – distribute

- The Distribute call generates backwards updates:

$$\psi^{**}(q_t, q_{t+1}) = \psi^{*}(q_t, q_{t+1}) \frac{\phi^{**}(q_{t+1})}{\phi^{*}(q_{t+1})}$$

$$\phi^{**}(q_t) = \sum_{q_{t+1}} \frac{\psi^{*}(q_t, q_{t+1})}{\phi^{*}(q_{t+1})} \phi^{**}(q_{t+1})$$

$$\phi^{**}(q_t) = \sum_{q_{t+1}} \frac{\psi^{*}(q_t, q_{t+1})}{\sum_{q_t} \psi^{*}(q_t, q_{t+1})} \phi^{**}(q_{t+1})$$



- Now, $\phi^{**}(q_t) = L\gamma_t = p(q_t, \mathbf{y}_0^T)$. No beta!
- After distribute, $\psi^{**}(q_{t-1}, q_t) = p(\mathbf{y}_0^T, q_{t-1}, q_t)$.

Message Passing – no evidence

- Consider the case when no observations have been made.
- Marginalizing gives $\sum_{\mathbf{y}_t} \psi(q_t, \mathbf{y}_t) = 1$ so separator $\xi^*(q_t)$ does not change. Thus, update factor passed to (q_{t-1}, q_t) is unity and $\psi(q_{t-1}, q_t)$ is also unchanged.
  Leaf messages do nothing when no evidence.
- Subsequent distribute pass does not change backbone, but will convert $\xi(q_t)$ into marginals $p(q_t)$ and potentials $\psi(q_t, \mathbf{y}_t)$ into marginals $p(q_t, \mathbf{y}_t)$.
- Why would you ever want to do this?
  - tells you about generative behaviour
  - can help numerical scaling of algorithms

Recursions

- The basic Collect-Distribute messages allow us to generate a variety of recursions.
- We chose $\phi^*(q_t)$ and $\phi^{**}(q_t)$ which gave the `alpha-gamma` recursions for HMM inference.
- Using root (q_0, q_1) gives `beta` recursions instead of `alpha`.
- A recursion on the update factors $\phi^{**}(q_t)/\phi^{*}(q_t)$ gives the `alpha-beta` algorithm.
- Recursions on $\psi^*(q_{t-1}, q_t)$ and $\psi^{**}(q_{t-1}, q_t)$ directly gives a new algorithm known as `rho-xi`.