COS 513: FOUNDATIONS OF PROBABILISTIC MODELING LECTURE 6

WONHO KIM, CHRIS PARK

1. The benefits of Junction Tree

Junction Tree gives us

- A factoral representation of the joint probability distribution
- Clique potentials that are marginals

2. Clique trees

Given a clique tree with cliques $\mathcal{C}$ and separators $\mathcal{S}$, the joint probability distribution is defined as follows:

$$p(x) = \frac{\prod_C \Psi_C(x_C)}{\prod_S \Phi_S(x_S)} \tag{1}$$

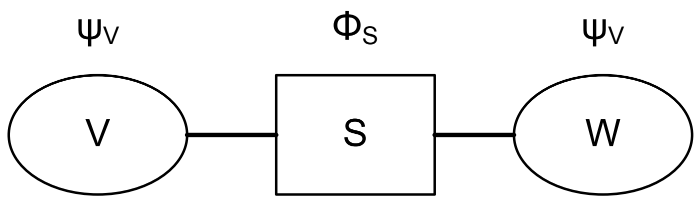
where $\Psi_C(x_C)$ is the potential for a clique C, and $\Phi_S(x_S)$ is the potential for a separator S. After Junction Tree algorithm, the clique potential $\Psi_C(x_C)$ becomes the marginal probability for clique C. Thus, we are able to achieve a representation that is a product of marginals, and yet is also a representation of the joint probability.

After we initialize maximal cliques, we first initialize the separator potentials to 1. The good thing is that we maintain global consistency by setting separator potentials to one. The basic idea of Junction Tree algorithm is that we adjust clique potentials $\Psi_C(x_C)$ to obtain marginals (local marginals), and adjust separator potentials $\Phi_S(x_S)$ to maintain the joint probability distribution $p(x)$.

3. Local consistency

3.1. **Definition.** In clique trees, cliques can overlap, so the same node can appear in multiple neighboring cliques. If the clique potentials are marginals, they *have to* agree on common nodes. In other words, common nodes should have the same marginals.

FIGURE 1. Clique tree



3.2. **Potential update ($V \rightarrow W$).** The local consistency is achieved by exchange of information between neighboring cliques. Suppose that we have two cliques V and W and suppose that V and W have non-empty intersection S (see Figure 1). The cliques V and W have potentials Ψ_V and Ψ_W, and S has a potential Φ_S that we initialize to one. The joint probability distribution $p(x)$ for this clique tree is as follows:

$$p(x) = \frac{\Psi_V \cdot \Psi_W}{\Phi_S} \tag{2}$$

We first update W based on V (Information is passed $V \rightarrow W$).

$$\Phi_S^* = \sum_{V \setminus S} \Psi_V \tag{3}$$

$$\Psi_W^* = \frac{\Phi_S^*}{\Phi_S} \Psi_W \tag{4}$$

Φ_S^* and Ψ_W^* are updated potentials for Φ_S and Ψ_W

Claim: The joint probability distribution is invariant while updating potentials ($V \rightarrow W$).

$$\begin{aligned} p(x) &= \frac{\Psi_W^* \cdot \Psi_V^*}{\Phi_S^*} \\ &= \frac{\Phi_S^* \cdot \Psi_W \cdot \Psi_V}{\Phi_S \cdot \Phi_S^*} \\ &= \frac{\Psi_V \cdot \Psi_W}{\Phi_S} \end{aligned} \tag{5}$$

The last line in (5) is equal to (2), which means that the joint probability distribution has not changed by the update.

3.3. **Potential update ($W \to V$).** Now we update V based on W (Information is passed $W \to V$).

$$\Phi_S^{**} = \sum_{W \setminus S} \Psi_W^* \tag{6}$$

$$\Psi_V^{**} = \frac{\Phi_S^{**}}{\Phi_S^*} \Psi_V^* \tag{7}$$

$$\Psi_W^{**} = \Psi_W^* \tag{8}$$

Like in (5), we can show that the joint probability distribution does not change after this update.

3.4. **Proof of local consistency.** To prove that the clique tree has local consistency, we have to show that nodes in the separator S have the same marginals.

$$\begin{aligned} \sum_{V \setminus S} \Psi_V^{**} &= \frac{\Phi_S^{**}}{\Phi_S^*} \cdot \sum_{V \setminus S} \Psi_V^* \\ &= \frac{\Phi_S^{**}}{\Phi_S^*} \cdot \Phi_S^* \\ &= \Phi_S^{**} \end{aligned} \tag{9}$$

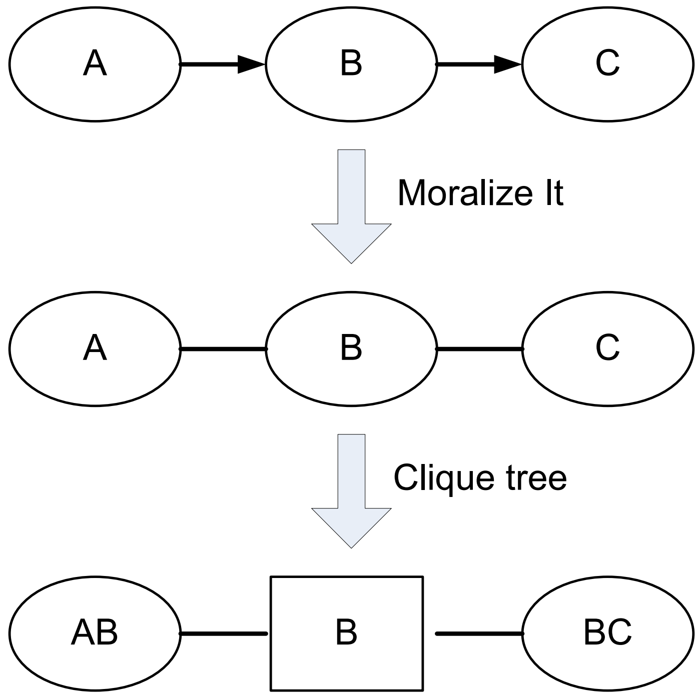
$$\begin{aligned} \sum_{W \setminus S} \Psi_W^{**} &= \sum_{W \setminus S} \Psi_W^* \\ &= \Phi_S^{**} \end{aligned} \tag{10}$$

From (9) and (10), we can see that the following equation is satisfied.

$$\sum_{V \setminus S} \Psi_V^{**} == \sum_{W \setminus S} \Psi_W^{**} \tag{11}$$

We have shown that the potentials Ψ_V^{**} and Ψ_W^{**} are consistent with respect to their intersection S; that is, they have the same marginals.

FIGURE 2. Constructing a clique tree from a directional graph



3.5. **Directional graphs.** By moralizing, we can construct clique trees from directional graphs (See Figure 2).

The potentials Ψ_{AB}, Ψ_{BC}, and Φ_B are defined as followings:

$$\begin{aligned} \Psi_{AB} &= p(A) \cdot p(B|A) \\ \Psi_{BC} &= p(C|B) \\ \Phi_B &= 1 \end{aligned} \tag{12}$$

Like in Section 3, to keep local consistency, we update the potentials by exchange of information between neighboring cliques. First, we update Ψ_{BC} based on Ψ_{AB}.

$$\begin{aligned} \Phi_B^* &= \sum_a p(a, B) \\ &= p(B) \end{aligned} \tag{13}$$

$$\Psi^*_{BC} = \frac{p(B)}{1} \cdot p(C|B) = p(B, C) \tag{14}$$

$$\Psi^*_{AB} = p(A, B) \tag{15}$$

Now, we see that the clique potentials have become marginal probabilities. The backward phase in this case ($BC \to AB$) is vacuous.

3.6. **Introducing evidence.** Now consider the case in which evidence is observed. Suppose that all nodes are binary and we are given the evidence ($A = 1$) in Figure 2. We have

$$\Phi^*_B = p(A = 1, B) \tag{16}$$

Performing the update ($AB \to BC$) yields

$$\begin{aligned} \Psi^*_{BC} &= p(A = 1, B) \cdot p(C|B) \\ &= p(A = 1, B, C) \end{aligned} \tag{17}$$

Once again the backward pass ($BC \to AB$) is vacuous in this case. Thus our potentials are as follows

$$\begin{aligned} \Psi^{**}_{AB} &= p(A = 1, B) \\ \Phi^{**}_{B} &= p(A = 1, B) \\ \Psi^{**}_{BC} &= p(A = 1, B, C) \end{aligned} \tag{18}$$

We see that we have obtained marginals as before, and evidence is present in all terms. The potentials are unnormalized marginals. Normalizing gives us conditionals $p(B|A = 1)$, $p(B|A = 1)$, and $p(B, C|A = 1)$.

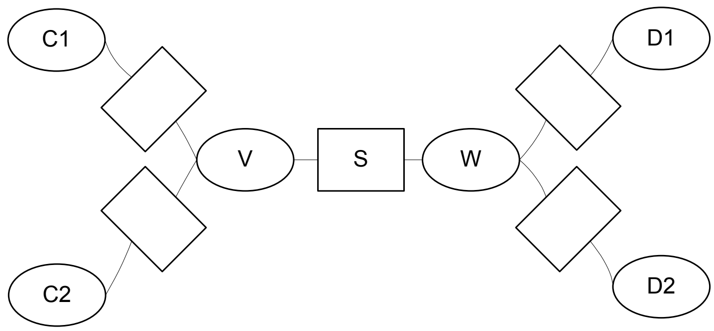
4. MULTIPLE OVERLAPPING CLIQUES.

4.1. **Message passing protocol.** So far we have been looking at clique trees with only one separator set. In practice, we would want to expand our algorithm to more complicated clique trees. Then, we must address the following three questions.

(1) How do we construct appropriate clique tree?
(2) How to perform multiple updates without ruining local consistency?
(3) Prove that the resulting Φ and Ψ are marginals as in the case with one separator set.

Lets first focus on question (2), consider the following clique tree, in figure 3.

FIGURE 3. Clique tree with explicit representation of the separators



In figure 3, as discussed in the previous section, *passing a message* from v to p consists the following steps.

v $\rightarrow$ w $\equiv$

(1) Update Φ_S^* from Ψ_v
(2) Update Ψ_w from Φ_S^* and Φ_S

However, we must decide when a given clique is allowed to pass a message to one of its neighbors. This problem is solved by the message passing protocol.

Message passing protocol. *A clique can send a message to a neighbor clique only when it has received messages from its other neighbors.*

We claim that the message passing protocol maintains local consistency. To examine the correctness of the message passing protocol, consider when w has received messages from all its neighbors D1 and D2, and is sending a message passing from w to v.

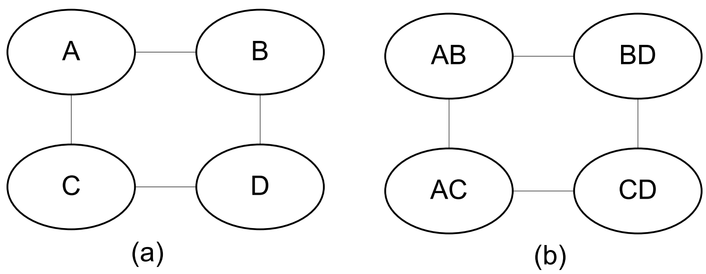
There can be two cases to consider,

- Case 1, v has received messages from its other neighbors and has sent a message to w. Local consistency is ok, since no other messages will be sent.
- Case 2, v has not yet sent a message to w. When message passing w to v creates Φ_S^*, later v will pass a message to w thus creating Φ_S^{**}. This message utilizes the stored marginal Φ_S^* so that consistency is maintained.

4.2. **Junction tree.** So far we have seen that given a clique tree and using the message passing protocol we can achieve local consistency between clique potentials. However, although we have shown that the message passing protocol enforces local consistency, there is no guarantee of global consistency. *Global consistency we define as, for any two nodes sharing the same variables, the marginal should be the same.*

For example, Consider the graphical model in figure 4.(a) and a particular choice of clique tree is shown in figure 4.(b).

FIGURE 4. (a) An undirected graphical model and (b) a corresponding clique tree.



We cannot guarantee global consistency because C occurs in 2 non-neighboring cliques. More formally, no guarantee that $\sum_D \Phi_{CD} = \sum_A \Phi_{AC}$. In order to guarantee global consistency, we must limit ourselves to a subset of clique trees called Junction trees.

Junction tree properties. *For every pair of cliques v and w, all cliques on the unique path between v and w contain $v \cup w$.*

We claim that on Junction trees, local consistency implies global consistency. If a node A appears in two cliques in a Junction tree, then cliques along the path are pair wise consistent with respect to A due to local consistency, then they must also be jointly consistent with respect to A.

One step further, in a Junction tree not only do we want global consistency, but also marginals on clique potentials. More formally,

Theorem. *Let p(x) be represented by clique potentials and separator potentials on a Junction tree. When the Junction tree algorithm terminates, the potentials are local marginal probabilities.*

$\Psi_C(x_C) = p(x_C)$

$\Phi_S(x_S) = p(x_S)$

We will prove this theorem next lecture.