Graphical Models

Clique trees & Belief Propagation

Siamak Ravanbakhsh

Winter 2018
Learning objectives

- message passing on clique trees
- its relation to variable elimination
- two different forms of belief propagation
Recap: variable elimination (VE)

- marginalize over a subset - e.g., $P(J)$
- expensive to calculate (why?) $\sum_{C,D,I,G,S,L,H} P(C, D, I, G, S, L, J, H)$
- use the factorized form of $P$

$$\sum_{C,D,I,G,S,L} P(D|C) P(G|D,I) P(S|I) P(L|G) P(J|L,S) P(H|G,J)$$
Recap: variable elimination (VE)

- marginalize over a subset - e.g., \( P(H, J) \)
- expensive to calculate (why?) \( \sum_{C,D,I,G,S,L} P(C, D, I, G, S, L) \)
- use the factorized form of \( P \)

\[
\sum_{C,D,I,G,S,L} P(D|C)P(G|D,I)P(S|I)P(L|G)P(J|L,S)P(H|G,J)
\]

\[\phi_2(H, G, J)\]

think of this as a factor/potential
same treatment of
- Bayes-nets
- Markov nets

for inference

*note that they do not encode the same CIs*
Recap: variable elimination (VE)

- marginalize over a subset - e.g., $P(H, J)$
- expensive to calculate (why?) $\sum_{C, D, I, G, S, L} P(C, D, I, G, S, L)$
- use the factorized form of $P$

$$\sum_{C, D, I, G, S, L} \phi_1(D, C) \phi_2(G, D, I) \phi_3(S, I) \phi_4(L, G) \phi_5(J, L, S) \phi_6(H, G, J)$$

$$= \ldots \sum_I \phi_3(S|I) \sum_D \phi_2(G, D, I) \sum_C \phi_1(D, C)$$

- repeat this

$$= \ldots \sum_I \phi_3(S, I) \sum_D \phi_2(G, D, I) \psi_1(D)$$

$$\psi_1(D) \quad \psi_1(D, C')$$

$$\psi_2(G, I) \quad \psi_2(G, I, D)$$
Recap: variable elimination (VE)

- marginalize over a subset - e.g., $P(J)$
- expensive to calculate (why?) $\sum_{C,D,I,G,S,L,H} P(C, D, I, G, S, L, J, H)$
- eliminate variables in some order
Recap: variable elimination (VE)

- eliminate variables in some order
- creates a chordal graph
- maximal cliques are the factors created during VE ($\psi_t$)

Order: C, D, I, H, G, S, L
Clique-tree

- summarize the VE computation using a clique-tree

- clusters are maximal cliques (factors that are marginalized)

\[ C_i = \text{Scope}[\psi_i] \]

\[ P(J) = \ldots \sum_I P(S|I) \sum_D P(G|D, I) \sum_C P(D|C) \]

\[ \psi_1'(D) \]

\[ \psi_1(D, C) \]
Clique-tree

- summarize the VE computation using a clique-tree

- clusters are maximal cliques (factors that are marginalized)
- sepsets are the result of marginalization over cliques

\[ C_i = \text{Scope}[\psi_i] \]
\[ S_{i,j} = \text{Scope}[\psi'_i] \]
\[ S_{i,j} = C_i \cap C_j \]
Clique-tree: **properties**

- a tree $\mathcal{T}$ from clusters $C_i$ and sepsets $S_{i,j} = C_i \cap C_j$

- **family-preserving property:** $\alpha(\phi) = j$
  - each factor $\phi$ is associated with a cluster $C_j$ s.t. $\text{Scope}[\phi] \subseteq C_j$
Clique-tree: properties

• a tree $\mathcal{T}$ from clusters $C_i$ and sepsets $S_{i,j} = C_i \cap C_j$

• family-preserving property: $\alpha(\phi) = j$
  ▪ each factor $\phi$ is associated with a cluster $C_j$ s.t. $\text{Scope}[\phi] \subseteq C_j$

• running intersection property:
  ▪ if $X \in C_i, C_j$ then $X \in C_k$ for $C_k$ in the path $C_i \rightarrow \ldots \rightarrow C_j$
VE as message passing

- think of VE as sending messages

\[
\begin{align*}
\delta_{1 \rightarrow 2}(D) & : \sum_C \psi_1(C_1) \\
\delta_{2 \rightarrow 3}(G,I) & : \sum_D \psi_2(C_2) \times \delta_{1 \rightarrow 2} \\
\delta_{3 \rightarrow 4}(G,S) & : \sum_I \psi_3(C_3) \times \delta_{2 \rightarrow 3} \\
\delta_{4 \rightarrow 5}(G,J) & : \sum_J \psi_4(C_4)
\end{align*}
\]
**VE as message passing**

- think of VE as sending messages

\[ \psi_i(C_i) \triangleq \prod_{\phi: \phi(\phi) = i} \phi \]

**calculate the product of factors in each clique**

**send messages from the leaves towards a root:**

\[ \delta_{i \rightarrow j}(S_{i,j}) = \sum_{C_i \subseteq S_{i,j}} \psi_i(C_i) \prod_{k \in Nb_{i \rightarrow j}} \delta_{k \rightarrow i}(S_{i,k}) \]
message passing

- think of VE as sending messages

\[ \delta_{i \rightarrow j}(S_{i,j}) = \sum_{C \in S_{i,j}} \psi_i(C_i) \prod_{k \in N b_{i \rightarrow j}} \delta_{k \rightarrow i}(S_{i,k}) \]

- send messages from the leaves towards a root:

\[ \phi = \sum_{\psi \in (i \rightarrow j)} ~ \prod_{\phi \in F \setminus (i \rightarrow j)} \phi \]

- the message is the marginal from one side of the tree
message passing

- think of VE as sending messages

\[
\delta(S_i) \triangleq \psi(C_i) \\
\delta(S_i)_{i \rightarrow j} = \sum_{C_i \sim S_i} \psi_i(C_i) \prod_{k \in Nb_i \sim j} \delta_{k \rightarrow i}(S_{i,k}) = \sum_{\mathcal{V}_{<i \rightarrow j}} \prod_{\phi \in \mathcal{F}_{<i \rightarrow j}} \phi
\]

- send messages from the leaves towards a root:

\[
\beta_r(C_r) \triangleq \psi_r(C_r) \prod_{k \in Nb_r} \delta_{k \rightarrow r}(S_{r,k})
\]

- the belief at the root clique is proportional to the marginal

\[
\beta_r(C_r) \propto \sum_{X \sim C_i} P(X)
\]
message passing: **downward pass**

- what if we continue sending messages?

  *(from the root to leaves)*

- clique $i$ sends a message to clique $j$ when received messages from all the **other** neighbors $k$
message passing: **downward pass**

- what if we continue sending messages?
  (from the root to leaves)

- sum-product **belief propagation** (BP)

  \[
  \delta_{i \rightarrow j}(S_{i,j}) = \sum_{C_i - S_{i,j}} \psi_i(C_i) \prod_{k \in N_{b_j} - j} \delta_{k \rightarrow i}(S_{i,k})
  \]

  \[
  \mu_{i,j}(S_{i,j}) \triangleq \delta_{i \rightarrow j}(S_{i,j}) \delta_{j \rightarrow i}(S_{i,j})
  \]

  \[
  \beta_i(C_i) \triangleq \psi_i(C_i) \prod_{k \in N_{b_i}} \delta_{k \rightarrow i}(S_{i,k})
  \]

  async. message update

  marginals

  for any clique (not only root)
Clique-tree & queries

What **type of queries** can we answer?

- **marginals** over subset of cliques \( P(A) \ A \subseteq C_i \)
- updating the beliefs after **new evidence** \( P(A \mid E^{(t)} = e^{(t)}) \ A \subseteq C_i, E \subseteq C_j \)
  - multiply the *(previously calibrated)* beliefs \( \beta(C_i) \mathbb{I}(E^{(t)} = e^{(t)}) \)
  - propagate to recalibrate
Clique-tree & queries

What **type of queries** can we answer?

- marginals over subset of cliques \( P(A) \quad A \subseteq C_i \)
- updating the beliefs after *new* evidence \( P(A \mid E^{(t)} = e^{(t)}) \quad A \subseteq C_i, E \subseteq C_j \)
  - multiply the *(previously calibrated)* beliefs \( \beta(C_i) \mathbb{I}(E^{(t)} = e^{(t)}) \)
  - propagate to recalibrate
- marginals outside cliques: \( P(A, B) \quad A \subseteq C_i, B \subseteq C_j \)
  - define a super-clique that has both A,B
  - a more efficient alternative?
- partition function \( Z \)
Chordal graph and clique-tree

any chordal graph gives a clique-tree

how to get a chordal graph? triangulation

- use the chordal graph from VE
  - min-neighbor, min-fill ...
- or find the optimal chordal graph
  - smallest tree-width NP-hard
  - also smallest max-clique
Chordal graph and clique-tree

Chordal graph = Markov \( \cap \) Bayesian networks

from MRF to Bayes-net:

- triangulate
- build a clique-tree
- **within cliques:** fully connected directed edges
- **between cliques:** from a root to leaves
Building a clique-tree: example

input

triangulated

clique-tree

from: wainwright & jordan
clique-tree quiz

what clique-tree to use here?
what are the sepsets?
cost of exact inference?
Summary

- **VE as message passing** in a clique-tree
- **clique-tree**: running intersection & family preserving
- **belief propagation** updates:
  - message update
  - belief update
- **types** of queries
- how to **build** a clique-tree for exact inference
bonus slides
Clique-tree: calibration

represent $P$ using marginals:
\[
\frac{\prod_i \beta_i}{\prod_{i,j \in E} \mu_{i,j}} = \frac{\prod_i \psi_i \prod_{k \rightarrow i} \delta_{k \rightarrow i}}{\prod_{i,j \in E} \delta_{i \rightarrow j} \delta_{j \rightarrow i}} = \prod_i \psi_i = \tilde{P}
\]

how about arbitrary assignments? \(\beta_i, \mu_{i,j} \ \forall i, j \in \mathcal{E}\)

an assignment is calibrated iff

\(BP\) produces calibrated beliefs

\[
\mu_{i,j}(S_{i,j}) = \sum_{C_i - S_{i,j}} \beta_i(C_i) = \sum_{C_j - S_{i,j}} \beta_j(C_j)
\]

for calibrated beliefs these "arbitrary assignments" have to be marginals

\[
\tilde{P}(X) \propto \frac{\prod \beta_i(C_i)}{\prod_{i,j \in E} \mu_{i,j}(S_{i,j})} \iff \beta_i(C_i) \propto P(C_i)
\]
BP: an alternative update

message update

\[ \delta_{i \rightarrow j}(S_{i,j}) = \sum_{C_i} \psi_i(C_i) \prod_{k \in N_{bi} - j} \delta_{k \rightarrow i}(S_{i,k}) \]

calculate the beliefs in the end

\[ \beta_i(C_i) = \psi_i(C_i) \prod_{k \in N_{bi}} \delta_{k \rightarrow i}(S_{i,k}) \]

belief update

since \[ \delta_{i \rightarrow j}(S_{i,j}) = \frac{\sum_{C_i} \beta_i(C_i)}{\delta_{j \rightarrow i}(S_{i,j})} \] we can update the beliefs instead of messages
BP: an alternative update

belief update

initialize $\beta_i \leftarrow \psi_i = \prod_{\phi : x(\phi) = i} \phi$, $\mu_{i,j} \leftarrow 1$

until convergence:

pick some $(i,j) \in E$

$\hat{\mu}_{i,j} \leftarrow \sum_{C_i} S_{i,j} \beta_i$ // $\hat{\mu}_{i,j} = \delta_{i \rightarrow j}^{\text{new}} \delta_{j \rightarrow i}^{\text{old}}$

$\beta_j \leftarrow \beta_j \frac{\hat{\mu}_{i,j}}{\mu_{i,j}}$ // $\frac{\hat{\mu}_{i,j}}{\mu_{i,j}} = \frac{\delta_{i \rightarrow j}^{\text{new}} \delta_{j \rightarrow i}^{\text{old}}}{\delta_{i \rightarrow j}^{\text{old}} \delta_{j \rightarrow i}^{\text{old}}}$

$\mu_{i,j} \leftarrow \hat{\mu}_{i,j}$

at convergence, beliefs are calibrated and so they are $\propto$ marginals

$\sum_{C_i} S_{i,j} \beta_i(C_i) = \sum_{C_j} S_{i,j} \beta_j(C_j)$