Bayesian Networks: Belief Propagation in Singly Connected Networks

Huizhen Yu

janey.yu@cs.helsinki.fi
Dept. Computer Science, Univ. of Helsinki

Probabilistic Models, Spring, 2010

Outline

Belief Propagation

In Chains

In Trees

In Singly Connected Networks

Form of Evidence and Notation

We denote evidence (a finding) of $X = \{X_v, v \in V\}$ by $\mathbf{e}$.

- Formally, we think of $\mathbf{e}$ as a function of x taking values in $\{0, 1\}$, representing a statement that some elements of x are impossible, i.e.,

$$\{x \,|\, \mathbf{e}(x) = 1\}$$

  is the set of possible values of x based on the evidence $\mathbf{e}$. We also refer to this event as $\mathbf{e}$.

- We consider $\mathbf{e}$ that can be written in the *factor form*

$$\mathbf{e}(x) = \prod_{v \in V} \ell_v(x_v), \quad \text{where} \;\; \ell_v(x_v) \in \{0, 1\}.$$

- For $A \subseteq V$, we use $\mathbf{e}_A$ to denote the partial evidence of X_A:

$$\mathbf{e}_A(x_A) = \prod_{v \in A} \ell_v(x_v).$$

Other short-hand notation we will use: $p(x_A \,\&\, \mathbf{e}) = P(X_A = x_A, \mathbf{e})$,

$$p(x_A \,\&\, \mathbf{e}_A \,|\, x_B) = P(X_A = x_A, \mathbf{e}_A \,|\, X_B = x_B) = P(X_A = x_A \,|\, X_B = x_B) \cdot \mathbf{e}_A(x_A),$$

and $p(x_A \,|\, \mathbf{e})$ denotes the conditional PMF of X_A given the event $\mathbf{e}$.

Motivation

Inference tasks we consider here: calculate $p(x_v \,|\, \mathbf{e}), \forall x_v$ and $P(\mathbf{e})$ for P that is directed Markov w.r.t. a DAG G.

- Note that if we know $P(\mathbf{e})$, then we can calculate the posterior probability of a single x given $\mathbf{e}$ easily:

$$p(x \,|\, \mathbf{e}) = p(x \,\&\, \mathbf{e}) / P(\mathbf{e}) = \Big(\prod_{v \in V} p(x_v \,|\, x_{\mathrm{pa}(v)}) \, \ell_v(x_v) \Big) / P(\mathbf{e}).$$

- Since $P(X = x, \mathbf{e}) = p(x) \, \mathbf{e}(x)$, in principle we can calculate

$$P(X_v = x_v, \mathbf{e}) = \sum_{x_{V \setminus \{v\}}} P(X_{V \setminus \{v\}} = x_{V \setminus \{v\}}, X_v = x_v, \mathbf{e}),$$

$$P(\mathbf{e}) = \sum_{x} P(X = x, \mathbf{e}).$$

But such calculation is not easy in most problems when $|V|$ is large.

The function $p(x_v \,|\, \mathbf{e})$ is referred to as the belief of x_v.

Features of the Algorithms to be Introduced

In the algorithms to be introduced, the DAG G is treated also as the architecture for distributed computation:

- Nodes: associated with autonomous processors
- Edges: communication links between processors

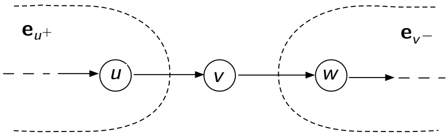
The independence relations represented by the DAG are exploited to separate the total evidence into pieces and streamline the computation.

The algorithms have performance guarantee on DAGs with simple structures – $G^{\sim}$ has no loops. But they have also been used successfully as approximate inference algorithms on loopy graphs.

Outline

Evidence Structure in a Chain

Suppose G is a chain. Consider a vertex v with parent u and child w:



We write $\mathbf{e}$ as three pieces of evidence, $\mathbf{e} = (\mathbf{e}_{u^+}, \mathbf{e}_v, \mathbf{e}_{v^-})$, where

- $\mathbf{e}_{u^+}$: partial evidence of $X_{\text{an}(v)}$
- $\mathbf{e}_v$: partial evidence of X_v
- $\mathbf{e}_{v^-}$: partial evidence of $X_{\text{de}(v)}$

We want to compute $p(x_v \,\&\, \mathbf{e}) = P(X_v = x_v, \mathbf{e})$ for all x_v. Since

$$P(X_{\text{an}(v)}, X_v, X_{\text{de}(v)}) = P(X_{\text{an}(v)}) \cdot P(X_v \,|\, X_u) \cdot P(X_{\text{de}(v)} \,|\, X_v),$$

we have

$$p\big((x_u, x_v) \,\&\, \mathbf{e}\big) = p(x_u \,\&\, \mathbf{e}_{u^+}) \cdot p(x_v \,\&\, \mathbf{e}_v \,|\, x_u) \cdot p(\mathbf{e}_{v^-} \,|\, x_v).$$

If v can get the first and third terms from u and w respectively, then v can calculate its marginal $p(x_v \,\&\, \mathbf{e})$ by summing over x_u.

Message Passing in a Chain

If node v receives

- from parent u the probabilities of x_u and partial evidence $\mathbf{e}_{u^+}$ on u's side:
$$\pi_{u,v}(x_u) = p(x_u \,\&\, \mathbf{e}_{u^+}), \quad \forall x_u;$$
- from child w the likelihoods of x_v based on the partial evidence $\mathbf{e}_{v^-}$ on w's side:
$$\lambda_{w,v}(x_v) = p(\mathbf{e}_{v^-} \,|\, x_v), \quad \forall x_v,$$

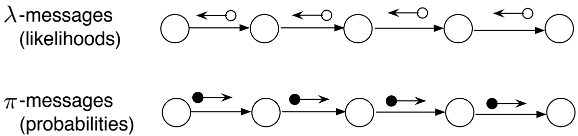
then node v can calculate

$$p(x_v \,\&\, \mathbf{e}) = \sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v) \cdot \lambda_{w,v}(x_v), \quad \forall x_v.$$

What u and w need from v in order to calculate their marginal probabilities?

- Parent u needs for all x_u, the likelihood of x_u based on $\mathbf{e}_{u^-} = (\mathbf{e}_v, \mathbf{e}_{v^-})$:
$$\lambda_{v,u}(x_u) = p(\mathbf{e}_{u^-} \,|\, x_u) = \sum_{x_v} p(x_v \,\&\, \mathbf{e}_v \,|\, x_u) \cdot p(\mathbf{e}_{v^-} \,|\, x_v)$$
$$= \sum_{x_v} p(x_v \,|\, x_u)\, \ell_v(x_v) \cdot \lambda_{w,v}(x_v).$$
- Child w needs for all x_v, the probability of x_v and $\mathbf{e}_{v^+} = (\mathbf{e}_{u^+}, \mathbf{e}_v)$:
$$\pi_{v,w}(x_v) = p(x_v \,\&\, \mathbf{e}_{v^+}) = \sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v).$$

Algorithm Summary



Each node v

- when receiving the message $\lambda_{w,v}$ from its child, sends to its parent u

$$\lambda_{v,u}(x_u) = \sum_{x_v} p(x_v \,|\, x_u)\, \ell_v(x_v) \cdot \lambda_{w,v}(x_v), \quad \forall x_u;$$

- when receiving the message $\pi_{u,v}$ from its parent, sends to its child w

$$\pi_{v,w}(x_v) = \sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v), \quad \forall x_v;$$

- when receiving both messages, calculates

$$p(x_v \,\&\, \mathbf{e}) = \sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v) \cdot \lambda_{w,v}(x_v), \quad \forall x_v,$$

$$P(\mathbf{e}) = \sum_{x_v} p(x_v \,\&\, \mathbf{e}), \qquad p(x_v \,|\, \mathbf{e}) = p(x_v \,\&\, \mathbf{e}) / P(\mathbf{e}).$$

Outline

Evidence Structure in a Rooted Tree

Suppose G is a rooted tree. Then $G^m = G^{\sim}$.

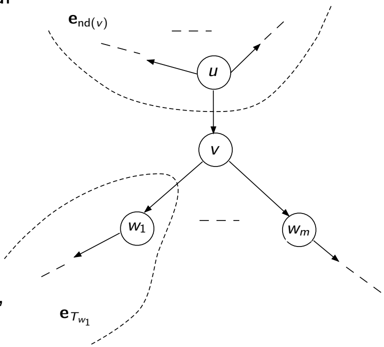
Consider a vertex v with parent u and children $w_1, \ldots, w_m$:

We write the total evidence $\mathbf{e}$ as several pieces of evidence,

$$\mathbf{e} = (\mathbf{e}_{\mathrm{nd}(v)}, \mathbf{e}_v, \mathbf{e}_{T_{w_1}}, \ldots, \mathbf{e}_{T_{w_m}}),$$

where

- $\mathbf{e}_{\mathrm{nd}(v)}$: partial evidence of $X_{\mathrm{nd}(v)}$
- $\mathbf{e}_v$: partial evidence of X_v
- $\mathbf{e}_{T_w}, w \in \mathrm{ch}(v)$: partial evidence of the variables associated with the subtree T_w rooted at w, i.e., $X_{\{w\} \cup \mathrm{de}(w)}$



Since $P(X_{\mathrm{nd}(v)}, X_v, X_{\mathrm{de}(v)}) = P(X_{\mathrm{nd}(v)}) \cdot P(X_v \mid X_u) \cdot \prod_{w \in \mathrm{ch}(v)} P(X_{T_w} \mid X_v)$,

$$p\big((x_u, x_v) \,\&\, \mathbf{e}\big) = p(x_u \,\&\, \mathbf{e}_{\mathrm{nd}(v)}) \cdot p(x_v \,\&\, \mathbf{e}_v \mid x_u) \cdot \prod_{w \in \mathrm{ch}(v)} p(\mathbf{e}_{T_w} \mid x_v).$$

Message Passing in a Rooted Tree

From

$$p\big((x_u, x_v)\,\&\,\mathbf{e}\big) = p(x_u\,\&\,\mathbf{e}_{\mathrm{nd}(v)}) \cdot p(x_v\,\&\,\mathbf{e}_v \,|\, x_u) \cdot \prod_{w \in \mathrm{ch}(v)} p(\mathbf{e}_{T_w} \,|\, x_v).$$

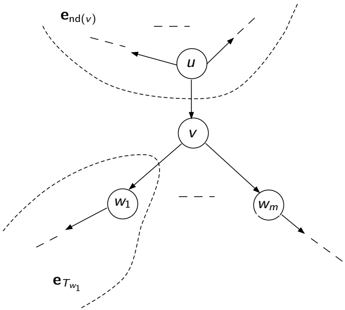
we see that if v receives

- from parent u the probabilities of x_u and evidence $\mathbf{e}_{\mathrm{nd}(v)}$ for all x_u:

$$\pi_{u,v}(x_u) = p(x_u\,\&\,\mathbf{e}_{\mathrm{nd}(v)}), \quad \forall x_u;$$

- from every child w the likelihoods of all x_v based on the evidence $\mathbf{e}_{T_w}$:

$$\lambda_{w,v}(x_v) = p(\mathbf{e}_{T_w} \,|\, x_v), \quad \forall x_v,$$



then node v can calculate

$$p(x_v\,\&\,\mathbf{e}) = \sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v) \cdot \prod_{w \in \mathrm{ch}(v)} \lambda_{w,v}(x_v).$$

Message Passing in a Rooted Tree

What do nodes u and w need from v in order to calculate their marginals?

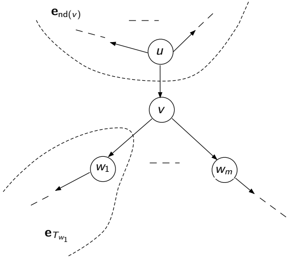
- Parent u needs the likelihoods of x_u based on $\mathbf{e}_{T_v}$ for all x_u:

$$\lambda_{v,u}(x_u) = \sum_{x_v} p(x_v \,\&\, \mathbf{e}_v \,|\, x_u) \cdot \prod_{w \in \mathrm{ch}(v)} p(\mathbf{e}_{T_w} \,|\, x_v)$$
$$= \sum_{x_v} p(x_v \,|\, x_u)\, \ell_v(x_v) \cdot \prod_{w \in \mathrm{ch}(v)} \lambda_{w,v}(x_v).$$

- Child w needs for all x_v, the probability of x_v and

$$\mathbf{e}_{\mathrm{nd}(w)} = \left(\mathbf{e}_{\mathrm{nd}(v)}, \left(\mathbf{e}_{T_{w'}}\right)_{w' \in \mathrm{ch}(v)\setminus\{w\}}\right):$$

$$\pi_{v,w}(x_v) = p(x_v \,\&\, \mathbf{e}_{\mathrm{nd}(w)})$$
$$= \left(\sum_{x_u} p(x_u \,\&\, \mathbf{e}_{\mathrm{nd}(v)}) \cdot p(x_v \,\&\, \mathbf{e}_v \,|\, x_u)\right) \cdot \prod_{w' \in \mathrm{ch}(v)\setminus\{w\}} p(\mathbf{e}_{T_{w'}} \,|\, x_v)$$
$$= \left(\sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v)\right) \cdot \prod_{w' \in \mathrm{ch}(v)\setminus\{w\}} \lambda_{w',v}(x_v).$$



Algorithm Summary

Each node v

- sends to its parent u

$$\lambda_{v,u}(x_u) = \sum_{x_v} p(x_v \,|\, x_u)\, \ell_v(x_v) \cdot \prod_{w \in \mathrm{ch}(v)} \lambda_{w,v}(x_v), \quad \forall x_u;$$

- sends to its child w

$$\pi_{v,w}(x_v) = \Big(\sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v) \Big) \cdot \prod_{w' \in \mathrm{ch}(v) \setminus \{w\}} \lambda_{w',v}(x_v), \quad \forall x_v;$$

- when receiving all messages, calculates

$$p(x_v \,\&\, \mathbf{e}) = \Big(\sum_{x_u} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_u)\, \ell_v(x_v) \Big) \cdot \prod_{w \in \mathrm{ch}(v)} \lambda_{w,v}(x_v) \quad \forall x_v,$$

$$P(\mathbf{e}) = \sum_{x_v} p(x_v \,\&\, \mathbf{e}), \qquad p(x_v \,|\, \mathbf{e}) = p(x_v \,\&\, \mathbf{e}) / P(\mathbf{e}).$$

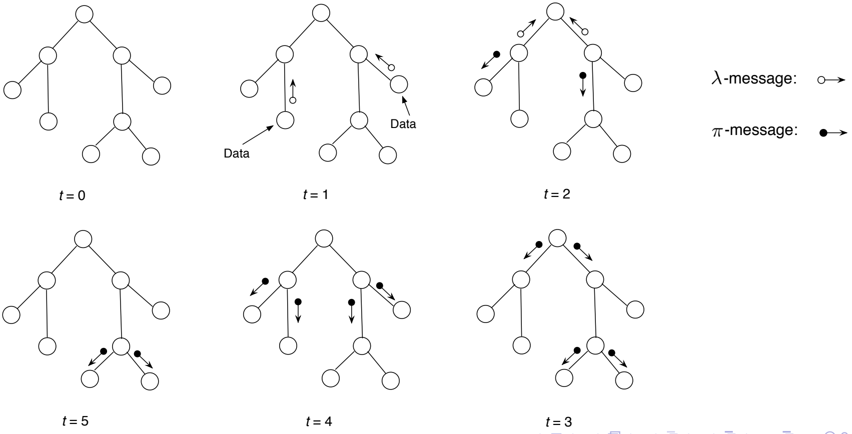
Message passing schemes:

(i) Each node can send a message to a linked node if it has received messages from all the other linked nodes.

(ii) Each node can send updated messages to linked nodes whenever it gets a new message from some node.

Illustration of Parallel Updating

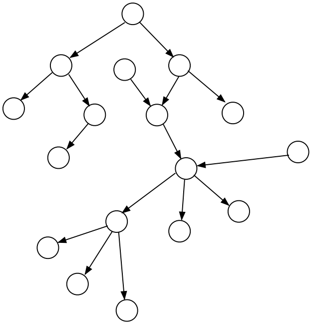
From J. Peal's book, 1988:
At time 0, each node of the tree has calculated its own marginal.
At time 1, two new pieces of evidence arrive and trigger new messages.
After time 5, all nodes have updated their marginals incorporating the new evidence.



Outline

Definition of a Singly Connected Network

Definition: a DAG G is *singly connected*, if its undirected version $G^{\sim}$ is a tree. Such a G is also called a *polytree*.



In a polytree G:

- Each node can have multiple parents and children.
- But there is only one trail between each pair of nodes.

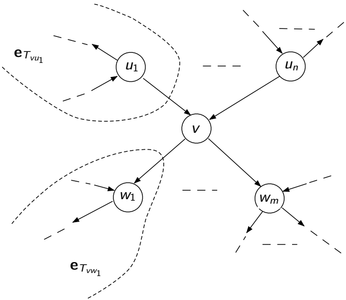
Consider a vertex v with parents $u_1, \ldots u_n$ and children $w_1, \ldots, w_m$. When v is viewed as the center, the branch of the polytree containing one of its parents or children is a sub-polytree. Denote

- $T_{vu_i}, i = 1 \ldots, n$: the sub-polytree containing the node u_i, resulting from removing the edge (u_i, v);
- $T_{vw_i}, i = 1, \ldots, m$: the sub-polytree containing the node w_i, resulting from removing the edge (v, w_i).

Evidence Structure in a Singly Connected Network

For a sub-polytree T, denote

- X_T: the variables associated with nodes in T
- $\mathbf{e}_T$: the partial evidence of X_T



We have

$$P(X_{T_{vu_1}}, \ldots, X_{T_{vu_n}}) = P(X_{T_{vu_1}}) \cdots P(X_{T_{vu_n}}),$$

and

$$P(X_{T_{vw_1}}, \ldots, X_{T_{vw_m}} \mid X_v) = P(X_{T_{vw_1}} \mid X_v) \cdots P(X_{T_{vw_m}} \mid X_v).$$

(Why? We may argue this using (DG) or d-separation – the latter is also simple in this case because there is only one trail between each pair of nodes.)

Therefore,

$$p\Big(\big(x_{\mathrm{pa}(v)}, x_v\big) \,\&\, \mathbf{e}\Big) = \Big(\prod_{u \in \mathrm{pa}(v)} p\big(x_u \,\&\, \mathbf{e}_{T_{vu}}\big)\Big) \cdot p\big(x_v \,\&\, \mathbf{e}_v \mid x_{\mathrm{pa}(v)}\big) \cdot \prod_{w \in \mathrm{ch}(v)} p(\mathbf{e}_{T_{vw}} \mid x_v).$$

Message Passing in a Singly Connected Network

From

$$p\Big(\big(x_{\mathrm{pa}(v)}, x_v\big) \,\&\, \mathbf{e}\Big) = \Big(\prod_{u\in \mathrm{pa}(v)} p(x_u \,\&\, \mathbf{e}_{T_{vu}})\Big) \cdot p\big(x_v \,\&\, \mathbf{e}_v \,|\, x_{\mathrm{pa}(v)}\big) \cdot \prod_{w \in \mathrm{ch}(v)} p(\mathbf{e}_{T_{vw}} \,|\, x_v).$$

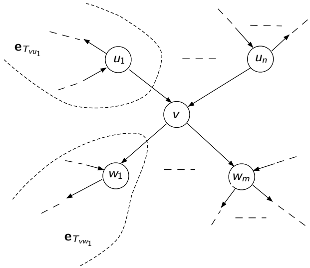
we see that v can calculate its marginal if it receives messages

- $\pi_{u,v}$ from all parents, where

$$\pi_{u,v}(x_u) = p(x_u \,\&\, \mathbf{e}_{T_{vu}}), \quad \forall x_u;$$

- and $\lambda_{w,v}$ from all children, where

$$\lambda_{w,v}(x_v) = p(\mathbf{e}_{T_{vw}} \,|\, x_v), \quad \forall x_v.$$



Then, $p(x_v \,\&\, \mathbf{e})$ is given by

$$p(x_v \,\&\, \mathbf{e}) = \Big(\sum_{x_{\mathrm{pa}(v)}} \prod_{u\in\mathrm{pa}(v)} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_{\mathrm{pa}(v)})\, \ell_v(x_v)\Big) \cdot \prod_{w\in\mathrm{ch}(v)} \lambda_{w,v}(x_v)$$

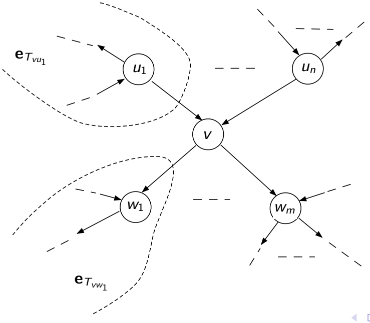
Message Passing in a Singly Connected Network

What do parents need from v in order to calculate their marginals?

- A parent u needs the likelihoods of all x_u based on the partial evidence $\mathbf{e}_{T_{uv}}$ from the sub-polytree on v's side with respect to u:

$$\lambda_{v,u}(x_u) = \sum_{x_v} \sum_{x_{\mathrm{pa}(v)\setminus\{u\}}} p\big(x_v \,\&\, \mathbf{e}_v \,|\, x_{\mathrm{pa}(v)}\big) \cdot \prod_{u' \in \mathrm{pa}(v)\setminus\{u\}} p(x_{u'} \,\&\, \mathbf{e}_{T_{vu'}}) \cdot \prod_{w \in \mathrm{ch}(v)} p(\mathbf{e}_{T_{vw}} \,|\, x_v)$$

$$= \sum_{x_v} \Big(\sum_{x_{\mathrm{pa}(v)\setminus\{u\}}} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big)\, \ell_v(x_v) \cdot \prod_{u' \in \mathrm{pa}(v)\setminus\{u\}} \pi_{u',v}(x_{u'}) \Big) \cdot \prod_{w \in \mathrm{ch}(v)} \lambda_{w,v}(x_v).$$

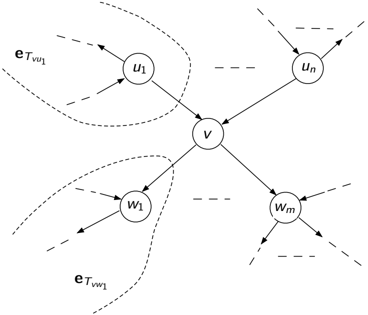


Message Passing in a Singly Connected Network

What do children need from v in order to calculate their marginals?

- A child w needs for all x_v, the probability of x_v and the partial evidence $\mathbf{e}_{T_{wv}}$ from the sub-polytree on v's side with respect to w:

$$p(x_v \,\&\, \mathbf{e}_{T_{wv}}) = \sum_{x_{\mathrm{pa}(v)}} p\big(x_v \,\&\, \mathbf{e}_v \,|\, x_{\mathrm{pa}(v)}\big) \cdot \prod_{u \in \mathrm{pa}(v)} p\big(x_u \,\&\, \mathbf{e}_{T_{vu}}\big) \prod_{w' \in \mathrm{ch}(v) \setminus \{w\}} p(\mathbf{e}_{T_{vw'}} \,|\, x_v)$$
$$= \Big(\sum_{x_{\mathrm{pa}(v)}} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big)\, \ell_v(x_v) \cdot \prod_{u \in \mathrm{pa}(v)} \pi_{u,v}(x_u) \Big) \cdot \prod_{w' \in \mathrm{ch}(v) \setminus \{w\}} \lambda_{w',v}(x_v).$$



Algorithm Summary

Each node v

- sends to each u of its parents

$$\lambda_{v,u}(x_u) = \sum_{x_v} \sum_{x_{\mathrm{pa}(v)\setminus\{u\}}} p(x_v \,|\, x_{\mathrm{pa}(v)})\, \ell_v(x_v) \cdot \prod_{u' \in \mathrm{pa}(v)\setminus\{u\}} \pi_{u',v}(x_{u'}) \cdot \prod_{w \in \mathrm{ch}(v)} \lambda_{w,v}(x_v), \ \forall x_u;$$

- sends to each w of its children

$$\pi_{v,w}(x_v) = \prod_{w' \in \mathrm{ch}(v)\setminus\{w\}} \lambda_{w',v}(x_v) \cdot \sum_{x_{\mathrm{pa}(v)}} p(x_v \,|\, x_{\mathrm{pa}(v)})\, \ell_v(x_v) \cdot \prod_{u \in \mathrm{pa}(v)} \pi_{u,v}(x_u), \quad \forall x_v;$$

- when receiving all messages from parents and children, calculates

$$p(x_v \,\&\, \mathbf{e}) = \Big(\prod_{w \in \mathrm{ch}(v)} \lambda_{w,v}(x_v) \Big) \cdot \sum_{x_{\mathrm{pa}(v)}} \prod_{u \in \mathrm{pa}(v)} \pi_{u,v}(x_u) \cdot p(x_v \,|\, x_{\mathrm{pa}(v)})\, \ell_v(x_v), \quad \forall x_v,$$

$$P(\mathbf{e}) = \sum_{x_v} p(x_v \,\&\, \mathbf{e}), \qquad p(x_v \,|\, \mathbf{e}) = p(x_v \,\&\, \mathbf{e}) / P(\mathbf{e}).$$

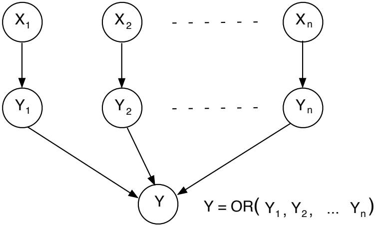
Message passing schemes:

(i) Each node can send a message to a linked node if it has received messages from all the other linked nodes.

(ii) Each node can send updated messages to linked nodes whenever it gets a new message from some node.

Example of Noisy-Or Gate

$$x_i, y_i, y \in \{0,1\}.$$
$$P(X_i = 1) = p_i,$$
$$P(Y_i = 1 \,|\, X_i = 0) = 0,$$
$$P(Y_i = 1 \,|\, X_i = 1) = 1 - q_i.$$
$$p(y \,|\, x_1, \ldots, x_n)$$
$$= \begin{cases} \prod_{i:x_i=1} q_i, & \text{if } \; y = 0; \\ 1 - \prod_{i:x_i=1} q_i, & \text{if } \; y = 1. \end{cases}$$



Express the message $\pi_{X_i,Y_i}(x_i)$ in the vector form $\left[\pi_{X_i,Y_i}(1)\,,\, \pi_{X_i,Y_i}(0)\right]$:

$$\pi_{X_i,Y_i} = \left[\, p_i \,,\, 1 - p_i \,\right].$$

Similarly, express $\pi_{Y_i,Y}(y_i)$ as $\left[\pi_{Y_i,Y}(1)\,,\, \pi_{Y_i,Y}(0)\right]$:

$$\pi_{Y_i,Y}(y_i) = \sum_{x_i \in \{0,1\}} \pi_{X_i,Y_i}(x_i) p(y_i \,|\, x_i), \qquad \text{so} \quad \pi_{Y_i,Y} = \left[\, p_i(1-q_i) \,,\, p_i q_i + (1-p_i) \,\right].$$

Example of Noisy-Or Gate

Suppose $\mathbf{e} : \{Y = 1\}$ is received. Then, Y sends a message $\lambda_{Y,Y_i} = \left[\, \lambda_{Y,Y_i}(1) \,,\, \lambda_{Y,Y_i}(0) \,\right]$ to each Y_i, where

$$\lambda_{Y,Y_i}(y_i) = \sum_{k \neq i} \sum_{y_k \in \{0,1\}} p(1 \,|\, y_1, \ldots, y_n) \prod_{j \neq i} \pi_{Y_j,Y}(y_j).$$

(What are these values?)

Subsequently, each Y_i sends to X_i the message $\lambda_{Y_i,X_i}(x_i)$:

$$\lambda_{Y_i,X_i}(1) = (1 - q_i) \cdot \lambda_{Y,Y_i}(1) + q_i \cdot \lambda_{Y,Y_i}(0), \quad \lambda_{Y_i,X_i}(0) = \lambda_{Y,Y_i}(0).$$

Each X_i can calculate its marginal and posterior probability of $X_i = 1$ as

$$P(X_i = 1, \mathbf{e}) = P(X_i = 1) \cdot \lambda_{Y_i,X_i}(1),$$
$$P(X_i = 0, \mathbf{e}) = P(X_i = 0) \cdot \lambda_{Y_i,X_i}(0),$$
$$P(X_i = 1 \,|\, \mathbf{e}) = \frac{p_i \cdot \lambda_{Y_i,X_i}(1)}{p_i \cdot \lambda_{Y_i,X_i}(1) + (1 - p_i) \cdot \lambda_{Y_i,X_i}(0)}.$$

Generalizations and Further Reading

- Find most probable configurations: $\max_x p(x \,\&\, \mathbf{e})$
- Conditioning:

  $G^\sim$ is not a tree. We condition on certain variables to create several singly connected networks and then fuse together the calculated results.
- Loopy belief propagation:

  $G^\sim$ is not a tree, but we apply the message passing algorithm any way. Algorithm variants and convergence analysis are active research topics.

Further reading:

1. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, 1988. Chap. 4.