Probabilistic Partial Evaluation: Exploiting rule structure in probabilistic inference

David Poole

University of British Columbia

Overview

- Belief Networks
- Variable Elimination Algorithm
- Parent Contexts & Structured Representations
- Structure-preserving inference
- Conclusion

Belief (Bayesian) Networks

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1} \ldots x_1)$$

$$= \prod_{i=1}^{n} P(x_i | \pi_{x_i})$$

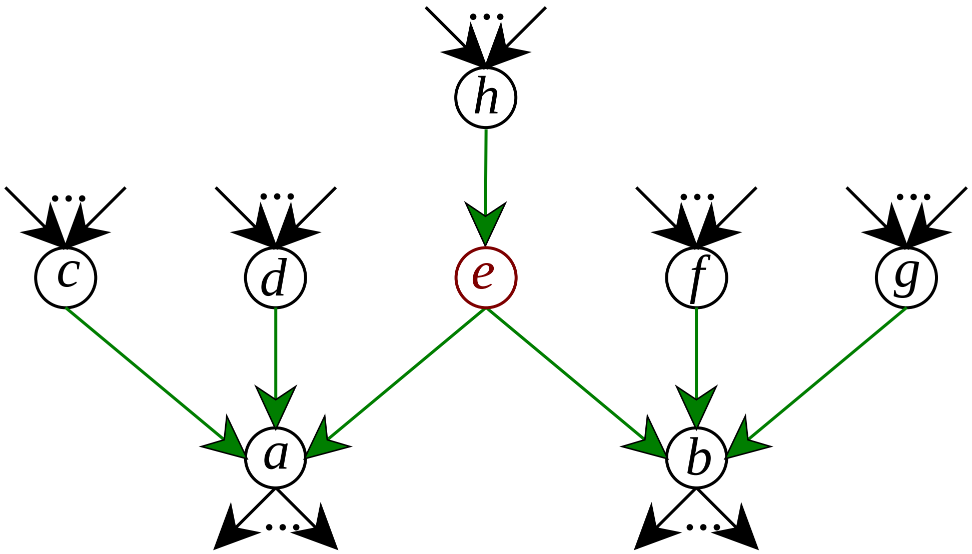
π_{x_i} are parents of x_i: set of variables such that the predecessors are independent of x_i given its parents.

Variable Elimination Algorithm

Given: Bayesian Network,
Query variable,
Observations,
Elimination ordering on remaining variables

1. set observed variables
2. sum out variables according to elimination ordering
3. renormalize

Summing Out a Variable

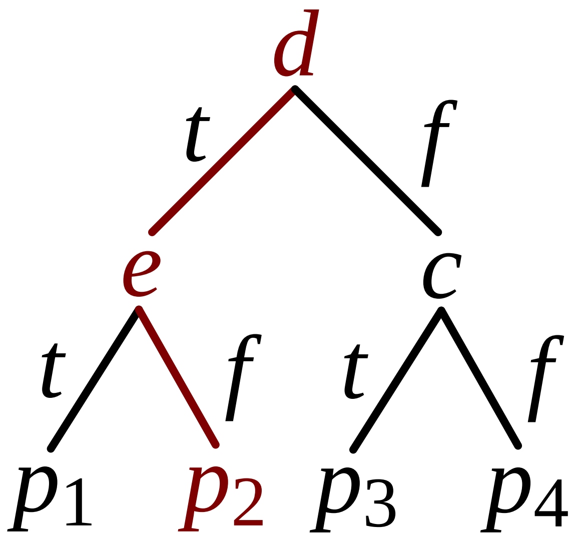


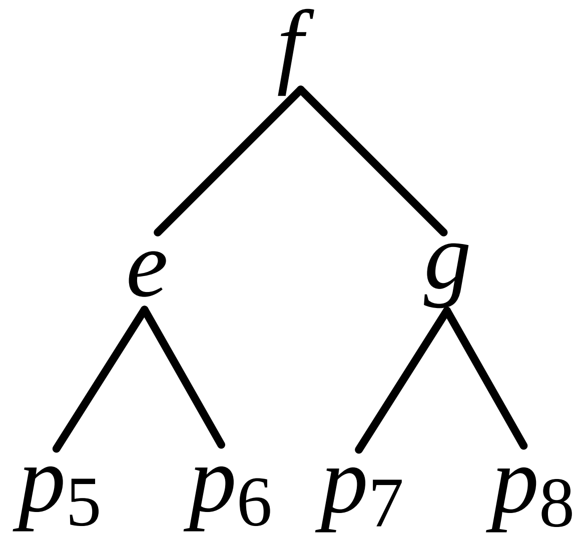
Sum out *e*:

$$\left.\begin{array}{l} P(a|c,d,e) \\ P(b|e,f,g) \\ P(e|h) \end{array}\right\} P(a,b|c,d,f,g,h)$$

Structured Probability Tables

$P(a|c,d,e)$ $P(b|e,f,g)$





$$p_2 = P(a = t|d = t \wedge e = f)$$

Eliminating e, preserving structure

- We only need to consider a & b together when $d = true \wedge f = true$.
  In this context c & g are irrelevant.

- In all other contexts we can consider a & b separately.

- When $d = false \wedge f = false$, e is irrelevant. In this context the probabilities shouldn't be affected by eliminating e.

Contextual Independence

Given a set of variables C, a **context on C** is an assignment of one value to each variable in C.

Suppose X, Y and C are disjoint sets of variables. X and Y are **contextually independent given context $c \in val(C)$** if

$$P(X|Y{=}y_1 \wedge C{=}c) = P(X|Y{=}y_2 \wedge C{=}c)$$

for all $y_1, y_2 \in val(Y)$ such that $P(y_1 \wedge c) > 0$ and $P(y_2 \wedge c) > 0$.

Parent Contexts

A **parent context** for variable x_i is a context c for a subset of the predecessors for x_i such that **x_i is contextually independent of the other predecessors given c.**

For variable x_i & assignment $x_{i-1}{=}v_{i-1}, \ldots, x_1{=}v_1$ of values to its preceding variables, there is a parent context $\pi_{x_i}^{v_{i-1}\ldots v_1}$.

$$
\begin{aligned}
& P(x_1{=}v_1, \ldots, x_n{=}v_n) \\
= \; & \prod_{i=1}^{n} P(x_i{=}v_n | x_{i-1}{=}v_{i-1}, \ldots, x_1{=}v_1) \\
= \; & \prod_{i=1}^{n} P(x_i{=}v_i | \pi_{x_i}^{v_{i-1}\ldots v_1})
\end{aligned}
$$

Idea behind probabilistic partial evaluation

- Maintain "rules" that are statements of probabilities in contexts.
- When eliminating a variable, you can ignore all *rules* that don't involve that variable.
- This wins when a variable is only in few parent contexts.
- Eliminating a variable looks like resolution!

Rule-based representation of our example

$$a \leftarrow d \wedge e : p_1$$

$$a \leftarrow d \wedge \overline{e} : p_2$$

$$a \leftarrow \overline{d} \wedge c : p_3$$

$$a \leftarrow \overline{b} \wedge \overline{c} : p_4$$

$$b \leftarrow f \wedge e : p_5$$

$$b \leftarrow f \wedge \overline{e} : p_6$$

$$b \leftarrow \overline{f} \wedge g : p_7$$

$$b \leftarrow \overline{f} \wedge \overline{g} : p_8$$

$$e \leftarrow h : p_9$$

$$e \leftarrow \overline{h} : p_{10}$$

Eliminating e

$$a \leftarrow d \wedge e : p_1 \qquad b \leftarrow f \wedge e : p_5$$

$$a \leftarrow d \wedge \overline{e} : p_2 \qquad b \leftarrow f \wedge \overline{e} : p_6$$

$$a \leftarrow \overline{d} \wedge c : p_3 \qquad b \leftarrow \overline{f} \wedge g : p_7$$

$$a \leftarrow \overline{b} \wedge \overline{c} : p_4 \qquad b \leftarrow \overline{f} \wedge \overline{g} : p_8$$

$$e \leftarrow h : p_9$$

$$e \leftarrow \overline{h} : p_{10}$$

unaffected by eliminating e

Variable partial evaluation

If we are eliminating e, and have rules:

$$x \leftarrow y \wedge e : p_1$$

$$x \leftarrow y \wedge \overline{e} : p_2$$

$$e \leftarrow z : p_3$$

- no other rules compatible with y contain e in the body
- y & z are compatible contexts,

we create the rule:

$$x \leftarrow y \wedge z : p_1 p_3 + p_2(1 - p_3)$$

Splitting Rules

A rule

$$a \leftarrow b : p_1$$

can be split on variable d, forming rules:

$$a \leftarrow b \wedge d : p_1$$

$$a \leftarrow b \wedge \overline{d} : p_1$$

Why Split?

If there are different contexts for a given e and for a given $\overline{e}$, you need to split the contexts to make them directly comparable:

$$a \leftarrow b \wedge e : p_1 \qquad \left\langle \begin{array}{l} a \leftarrow b \wedge c \wedge e : p_1 \\ a \leftarrow b \wedge \overline{c} \wedge e : p_1 \end{array} \right.$$

$$a \leftarrow b \wedge c \wedge \overline{e} : p_2$$

$$a \leftarrow b \wedge \overline{c} \wedge \overline{e} : p_3$$

Combining Heads

Rules

$$a \leftarrow c : p_1$$

$$b \leftarrow c : p_2$$

where a and b refer to different variables, can be combined producing:

$$a \wedge b \leftarrow c : p_1 p_2$$

Thus in the context with a, b, and c all true, the latter rule can be used instead of the first two.

Splitting Compatible Bodies

~~$a \leftarrow d \wedge e : p_1$~~

$a \leftarrow d \wedge f \wedge e : p_1$

$a \leftarrow d \wedge \overline{f} \wedge e : p_1$

~~$a \leftarrow d \wedge \overline{e} : p_2$~~

$a \leftarrow d \wedge f \wedge \overline{e} : p_2$

$a \leftarrow d \wedge \overline{f} \wedge \overline{e} : p_2$

~~$b \leftarrow f \wedge e : p_5$~~

$b \leftarrow d \wedge f \wedge e : p_5$

$b \leftarrow \overline{d} \wedge f \wedge e : p_5$

~~$b \leftarrow f \wedge \overline{e} : p_6$~~

$b \leftarrow d \wedge f \wedge \overline{e} : p_6$

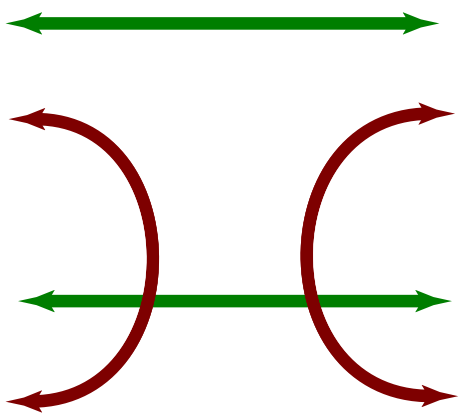
$b \leftarrow \overline{d} \wedge f \wedge \overline{e} : p_6$

$e \leftarrow h : p_9$

$e \leftarrow \overline{h} : p_{10}$

Combining Rules

$$a \leftarrow d \wedge f \wedge e : p_1 \qquad b \leftarrow d \wedge f \wedge e : p_5$$
$$a \leftarrow d \wedge \overline{f} \wedge e : p_1 \qquad b \leftarrow \overline{d} \wedge f \wedge e : p_5$$



$$a \leftarrow d \wedge f \wedge \overline{e} : p_2 \qquad b \leftarrow d \wedge f \wedge \overline{e} : p_6$$
$$a \leftarrow d \wedge \overline{f} \wedge \overline{e} : p_2 \qquad b \leftarrow \overline{d} \wedge f \wedge \overline{e} : p_6$$

$$e \leftarrow h : p_9$$
$$e \leftarrow \overline{h} : p_{10}$$

Result of eliminating *e*

The resultant rules encode the probabilities of $\{a, b\}$ in the contexts:

$d \wedge f \wedge h$,

$d \wedge f \wedge \overline{h}$

For all other contexts we consider a and b separately.

The resulting number of rules is 24.

Tree structured probability for $P(a, b|c, d, f, g, h, i)$ has 72 leaves. (Same as number of rules if a and b are combined in all contexts).

VE has a table of size 256.

Evidence

We can set the values of all evidence variables before summing out the remaining non-query variables.

Suppose $e_1{=}o_1 \wedge \ldots \wedge e_s{=}o_s$ is observed:

- Remove any rule that contains $e_i{=}o'_i$, where $o_i \neq o'_i$ in the body.
- Remove any term $e_i{=}o_i$ in the body of a rule.
- Replace any $e_i{=}o'_i$, where $o_i \neq o'_i$, in the head of a rule *false*.
- Replace any $e_i{=}o_i$ in the head of a rule by *true*.

In rule heads, use $true \wedge a \equiv a$, and $false \wedge a \equiv false$.

Conclusions

- New notion of parent context $\Longrightarrow$ rule-based representation for Bayesian networks.
- New algorithm for probabilistic inference that preserves rule-structure.
- Exploits more structure than tree-based representations of conditional probability.
- Allows for finer-grained approximation than in a Bayesian network.