Inference in Bayesian Networks
Variable Elimination Algorithm

Blake VanBerlo

Lecture 9

Readings: RN 13.4. PM 8.4.

Outline

Learning Goals

By the end of the lecture, you should be able to

- Explain how we can perform probabilistic inference more efficiently using the variable elimination algorithm.
- Define factors. Manipulate factors using operations restrict, sum out, multiply and normalize.
- Describe/trace/implement the variable elimination algorithm for calculating a prior or a posterior probability given a Bayesian network.
- Explain how the elimination ordering affects the complexity of the variable elimination algorithm.
- Identify the variables that are irrelevant to a query.

Learning Goals

Why Use the Variable Elimination Algorithm

The Variable Elimination Algorithm

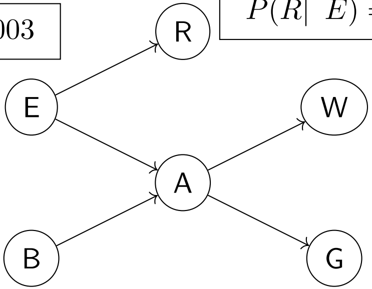
Factors Affecting the Complexity of VEA

Revisiting the Learning goals

A Bayesian Network for the Holmes Scenario

$$P(E) = 0.0003$$

$$P(R|\neg E) = 0.0002$$
$$P(R|\ E) = 0.9$$



$$P(W|\neg A) = 0.4$$
$$P(W|\ A) = 0.8$$

$$P(G|\neg A) = 0.04$$
$$P(G|\ A) = 0.4$$

$$P(B) = 0.0001$$

$$P(A|\neg B \land \neg E) = 0.01$$
$$P(A|\neg B \land \ E) = 0.2$$
$$P(A|\ B \land \neg E) = 0.95$$
$$P(A|\ B \land \ E) = 0.96$$

Answering a Question

What is the probability that a burglary is happening given that Dr. Watson and Mrs. Gibbon both call?

$$P(B|w \wedge g)$$

- Query variables: B
- Evidence variables: W and G
- Hidden variables: E, A, and R.

Notice the new notation: capital letters for random variables and lowercase letters for values.

Answering the query using the joint distribution

Evaluate $P(B|w \wedge g)$ in terms of known distributions from the Bayesian network.

Answering the query using the joint distribution

Evaluate $P(B|w \wedge g)$ in terms of known distributions from the Bayesian network.

$\rightarrow$ Following the approach from Lecture 6:

$$
\begin{aligned}
P(B|w \wedge g) &= \frac{P(B \wedge w \wedge g)}{P(w \wedge g)} \\
&= \frac{P(B \wedge w \wedge g)}{P(b \wedge w \wedge g) + P(\neg b \wedge w \wedge g)} \\
&\propto P(B \wedge w \wedge g) \\
&\propto \sum_e \sum_a \sum_r P(B \wedge e \wedge a \wedge w \wedge g \wedge r) \\
&\propto \sum_e \sum_a \sum_r P(B)P(e)P(a|B \wedge e)P(w|a)P(g|a)P(r|e)
\end{aligned}
$$

Q #1: Number of operations using the joint distribution

How many addition and multiplication operations do we need to calculate the expression below?

$$\sum_{e}\sum_{a}\sum_{r} P(B)P(e)P(a|B \wedge e)P(w|a)P(g|a)P(r|e)$$

(A) ≤ 10

(B) 11-20

(C) 21-40

(D) 41-60

(E) ≥ 61

Q #1: Number of operations using the joint distribution

How many addition and multiplication operations do we need to calculate the expression below?

$$\sum_e \sum_a \sum_r P(B)P(e)P(a|B \wedge e)P(w|a)P(g|a)P(r|e)$$

(A) ≤ 10

(B) 11-20

(C) 21-40

(D) 41-60

(E) ≥ 61

$\rightarrow$ Correct answer is (D). 47 operations.

Answering the query using variable elimination algorithm

$$\sum_{e}\sum_{a}\sum_{r} P(B)P(e)P(a|B\wedge e)P(w|a)P(g|a)P(r|e)$$

$\rightarrow$ We should move the summations as much to the right of the expression as possible to reduce the overall # of operations.

$$\begin{aligned}
&= \sum_{e}\sum_{a}\sum_{r} P(B)P(e)P(a|B\wedge e)P(w|a)P(g|a)P(r|e) \\
&= P(B)\sum_{e} P(e)\sum_{a} P(a|B\wedge e)P(w|a)P(g|a)\sum_{r} P(r|e) \\
&= P(B)\sum_{e} P(e)\sum_{a} P(a|B\wedge e)P(w|a)P(g|a)
\end{aligned}$$

Q #2: Number of operations via the variable elimination algorithm

How many addition and multiplication operations do we need to calculate the expression below?

$$P(B) \sum_{e} P(e) \sum_{a} P(a|B \wedge e) P(w|a) P(g|a)$$

(A) ≤ 10

(B) 11-20

(C) 21-40

(D) 41-60

(E) ≥ 61

Q #2: Number of operations via the variable elimination algorithm

How many addition and multiplication operations do we need to calculate the expression below?

$$P(B) \sum_e P(e) \sum_a P(a|B \wedge e) P(w|a) P(g|a)$$

(A) ≤ 10

(B) 11-20

(C) 21-40

(D) 41-60

(E) ≥ 61

$\rightarrow$ Correct answer is (B) 11-20. 14 operations in total.

Inference by enumeration

We can compute any conditional probability by summing terms from the joint distribution.

- If the value of a variable is observed, calculate the product of its conditional probability multiplied by the rest.
- Otherwise, it must be a hidden variable. Sum it out by calculating the sum of the products of its conditional probability multiplied by the rest.

Inference by Enumeration

Algorithm 1 Enumerate(vars, bn, evidence)

if vars is empty **then return** 1.0
$Y \leftarrow$ first in vars
if Y has value y in evidence **then**
    return $P(y|parents(Y)) \times \text{Enumerate}(REST(\text{vars}), \text{bn}, \text{evidence})$
else
    return $\sum_y P(y|parents(Y)) \times \text{Enumerate}(REST(\text{vars}), \text{bn}, \text{evidence} \cup \{y\})$

Inference by Enumeration

A problem with the Enumeration algorithm:

$P(W|A)P(G|A)$ and $P(W|\neg A)P(G|\neg A)$ are computed twice, once for each value of E.

How can we avoid these wasted computations?

Introducing the Variable Elimination Algorithm

- Performing probabilistic inference is challenging.

  → Computing the posterior distribution of one or more query variables given some evidence is #NP. Estimate the posterior probability in a Bayesian network within an absolute error is already NP-hard. No general efficient implementation.

- Exact and approximate inferences.

  → Compute the probabilities exactly.
  Naive approach: enumerate all the worlds consistent with the evidence. Do better below.

- The variable elimination algorithm uses dynamic programming and exploits the conditional independence.

  → Do the calculations once and save the results for later.

  Factors and operations.

Factors

- A function from some random variables to a number.
- $f(X_1, \ldots, X_j)$: a factor f on variables $X_1, \ldots, X_j$.
- A factor can represent a joint or a conditional distribution. For example, $f(X_1, X_2)$ can represent $P(X_1 \wedge X_2)$, $P(X_1|X_2)$ or $P(X_1 \wedge X_3 = v_3|X_2)$.
- Define a factor for every conditional probability distribution in the Bayes net.

$\rightarrow$ Every conditional probability distribution in the Bayes net is a factor.

$f(B), f(E), f(A, B, E), f(R, E), f(W, A), f(G, A)$

$P(B), P(E), P(A|B \wedge E), P(R|E), P(W|A), P(G|A)$

$$P(B \wedge w \wedge g)$$
$$= P(B) \sum_a P(w|a)P(g|a) \sum_e P(e)P(a|B \wedge e)$$

Restrict a factor

Restrict a factor.

- Assign each evidence variable to its observed value.
- Restricting $f(X_1, X_2, \ldots, X_j)$ to $X_1 = v_1$, produces a new factor $f(X_1 = v_1, X_2, \ldots, X_j)$ on $X_2, \ldots, X_j$.
- $f(X_1 = v_1, X_2 = v_2, \ldots, X_j = v_j)$ is a number.

→ Restrict $f(W, A)$ to $W = t$. Restrict $f(G, A)$ to $G = t$.

$$
\begin{aligned}
&P(B \wedge w \wedge g) \\
&= P(B) \sum_a P(w|a) P(g|a) \sum_e P(e) P(a|B \wedge e)
\end{aligned}
$$

Restrict a factor

$f_1(X,Y,Z)$:

X	Y	Z	val
t	t	t	0.1
t	t	f	0.9
t	f	t	0.2
t	f	f	0.8
f	t	t	0.4
f	t	f	0.6
f	f	t	0.3
f	f	f	0.7

$\rightarrow$

$f_2(Y,Z)$:

Y	Z	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$f_3(Y)$:

Y	val
t	0.9
f	0.8

$f_4() = 0.8$

- What is $f_2(Y,Z) = f_1(x,Y,Z)$?
- What is $f_3(Y) = f_2(Y,\neg z)$?
- What is $f_4() = f_3(\neg y)$?

Sum out a variable

Sum out a variable.

Summing out X_1 with domain $\{v_1, \ldots, v_k\}$ from factor $f(X_1, \ldots, X_j)$, produces a factor on $X_2, \ldots, X_j$ defined by:

$$(\sum_{X_1} f)(X_2, \ldots, X_j) = f(X_1 = v_1, \ldots, X_j) + \cdots + f(X_1 = v_k, \ldots, X_j)$$

$\rightarrow$ Sum out a and e.

$$\begin{aligned} & P(B \wedge w \wedge g) \\ & = P(B) \sum_a P(w|a) P(g|a) \sum_e P(e) P(a|B \wedge e) \end{aligned}$$

Sum out a variable

$f_1(X, Y, Z)$:

X	Y	Z	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

$\rightarrow$ $f_2(X, Z)$:

X	Z	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

What is $f_2(X, Z) = \sum_Y f_1(X, Y, Z)$?

Multiplying factors

Multiply two factors together.

The **product** of factors $f_1(X, Y)$ and $f_2(Y, Z)$, where Y are the variables in common, is the factor $(f_1 \times f_2)(X, Y, Z)$ defined by:

$$(f_1 \times f_2)(X, Y, Z) = f_1(X, Y) * f_2(Y, Z).$$

$\rightarrow$

$$P(B \wedge w \wedge g)$$
$$= P(B) \sum_a P(w|a) P(g|a) \sum_e P(e) P(a|B \wedge e)$$

Multiplying factors

f_1:

X	Y	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

f_2:

Y	Z	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$\to f_1 \times f_2$:

X	Y	Z	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

What is $f_1(X,Y) \times f_2(Y,Z)$?

Normalize a factor

- Convert it to a probability distribution.
- Divide each value by the sum of all the values.

f_1:

Y	val
t	0.2
f	0.6

Variable elimination algorithm

To compute $P(X_q | X_{o_1} = v_1 \land \ldots \land X_{o_j} = v_j)$:

- **Construct a factor** for each conditional probability distribution.
- **Restrict** the observed variables to their observed values.
- Eliminate each hidden variable X_{h_j}.
  - **Multiply** all the factors that contain X_{h_j} to get new factor g_j.
  - **Sum out** the variable X_{h_j} from the factor g_j.
- **Multiply** the remaining factors.
- **Normalize** the resulting factor.

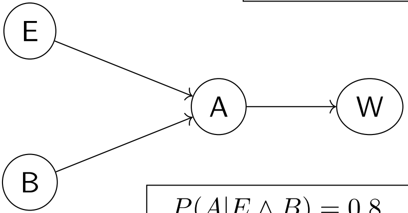
Example of VEA

Given a portion of the Holmes network below,
calculate $P(B|\neg A)$ using the variable elimination algorithm.

Eliminate the hidden variables in reverse alphabetical order.

$$P(E) = 0.1$$

$$P(W|A) = 0.8$$
$$P(W|\neg A) = 0.4$$



$$P(B) = 0.3$$

$$P(A|E \land B) = 0.8$$
$$P(A|E \land \neg B) = 0.2$$
$$P(A|\neg E \land B) = 0.7$$
$$P(A|\neg E \land \neg B) = 0.1$$

Learning Goals

Why Use the Variable Elimination Algorithm

The Variable Elimination Algorithm

Factors Affecting the Complexity of VEA

Revisiting the Learning goals

Effect of The Elimination Ordering

In general, VEA is exponential in space and time.

The complexity of VEA depends on:

- The size of the CPT in the Bayesian network.
- The size of the largest factor during algorithm execution.

Effect of the elimination ordering on algorithm complexity:

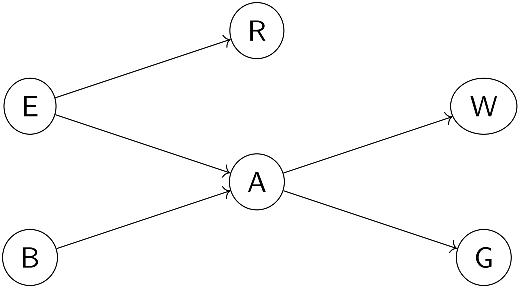
- Every order yields a valid algorithm.
- Different orderings yields different intermediate factors.

Examples of Good and Bad Orderings

Suppose that we want to calculate $P(G)$.
What factors do we produce if we

- Eliminate R first?
- Eliminate A first?

Which ordering leads to worse complexity for VEA?



Choose an Elimination Ordering

How should we choose an order of eliminating the hidden variables?

- Intractable to determine the optimal order.

- A good greedy heuristic: Eliminate the variable that minimizes the size of the next factor.

- For a *polytree:* work outside in.

  $\rightarrow$ Singly-connected network.

  At most one un-directed path between any two nodes.

Complexity of VEA in Polytrees

- A **polytree** is a single-connected network in which there is at most one undirected path between any two nodes.
- The time and space complexity of the exact inference in polytrees is linear in the size of the network (# entries in the conditional probability tables)

Irrelevant Variables

- Certain variables have no impact on the query.
- Every variable that is not an ancestor of a query variable or evidence variable is irrelevant to the query.

Revisiting the Learning Goals

By the end of the lecture, you should be able to

- Explain how we can perform probabilistic inference more efficiently using the variable elimination algorithm.
- Define factors. Manipulate factors using operations restrict, sum out, multiply and normalize.
- Describe/trace/implement the variable elimination algorithm for calculating a prior or a posterior probability given a Bayesian network.
- Explain how the elimination ordering affects the complexity of the variable elimination algorithm.
- Identify the variables that are irrelevant to a query.