Reasoning Under Uncertainty: Belief Network Inference

CPSC 322 – Uncertainty 5

Textbook §10.4

Lecture Overview

Components of a belief network

Definition (belief network)

A belief network consists of:

- a directed acyclic graph with nodes labeled with random variables
- a domain for each random variable
- a set of conditional probability tables for each variable given its parents (including prior probabilities for nodes with no parents).

Relating BNs to the joint

If you have a belief network, you can recover the joint.

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Claim: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|pX_i)$.
  1. For each i, $P(X_i|pX_i) = P(X_i|X_1, \ldots, X_{i-1})$
     - The parents pX_i of X_i are those predecessors of X_i that render X_i independent of the other predecessors.
     - That is, $pX_i \subseteq X_1, \ldots, X_{i-1}$ and
  2. Theorem of probability theory (chain rule): $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1})$

Lecture Overview

Example: Fire Diagnosis

Suppose you want to diagnose whether there is a fire in a building

- you receive a noisy report about whether everyone is leaving the building.
- if everyone *is* leaving, this may have been caused by a fire alarm.
- if there is a fire alarm, it may have been caused by a fire or by tampering
- if there is a fire, there may be smoke

Example: Fire Diagnosis

First you choose the variables. In this case, all are boolean:

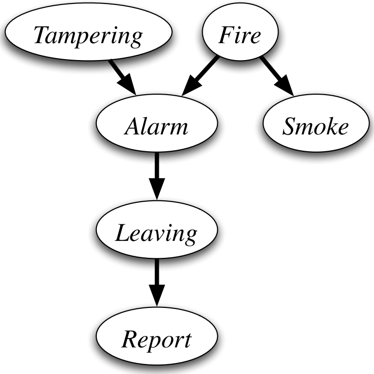
- Tampering is true when the alarm has been tampered with
- Fire is true when there is a fire
- Alarm is true when there is an alarm
- Smoke is true when there is smoke
- Leaving is true if there are lots of people leaving the building
- Report is true if the sensor reports that people are leaving the building

Example: Fire Diagnosis

- Next, you order the variables: $Fire$; $Tampering$; $Alarm$; $Smoke$; $Leaving$; $Report$.
- Now evaluate which variables are conditionally independent given their parents:
  - $Fire$ is independent of $Tampering$ (learning that one is true would not change your beliefs about the probability of the other)
  - $Alarm$ depends on both $Fire$ and $Tampering$: it could be caused by either or both.
  - $Smoke$ is caused by $Fire$, and so is independent of $Tampering$ and $Alarm$ given whether there is a $Fire$
  - $Leaving$ is caused by $Alarm$, and thus is independent of the other variables given $Alarm$.
  - $Report$ is caused by $Leaving$, and thus is independent of the other variables given $Leaving$.

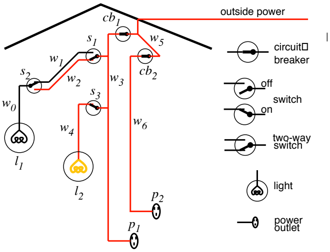
Example: Fire Diagnosis

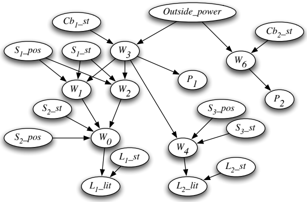
This corresponds to the following belief network:



Of course, we're not done until we also come up with conditional probability tables for each node in the graph.

Example: Circuit Diagnosis





The belief network also specifies:

- The domain of the variables:
  $W_0, \ldots, W_6 \in \{live, dead\}$
  S_1_pos, S_2_pos, and S_3_pos have domain $\{up, down\}$
  S_1_st has $\{ok, upside\_down, short, intermittent, broken\}$.
- Conditional probabilities, including:
  $P(W_1 = live | s_1\_pos = up \land S_1\_st = ok \land W_3 = live)$
  $P(W_1 = live | s_1\_pos = up \land S_1\_st = ok \land W_3 = dead)$
  $P(S_1\_pos = up)$
  $P(S_1\_st = upside\_down)$

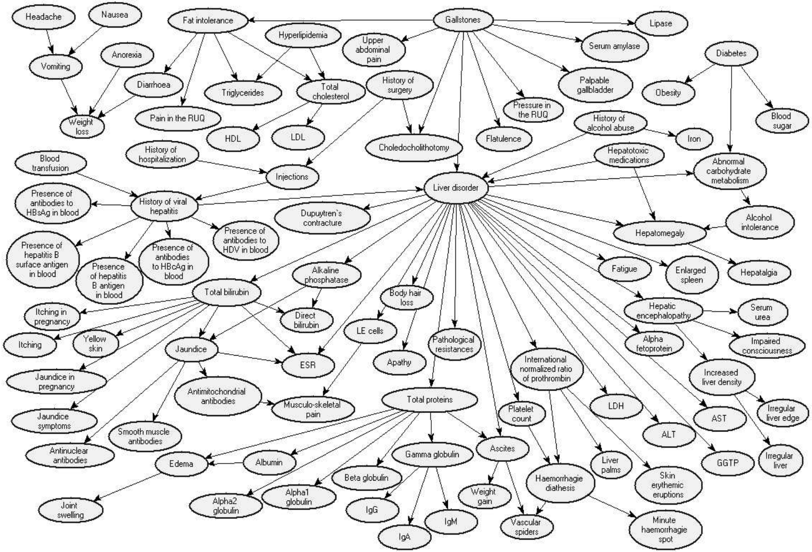
Example: Circuit Diagnosis

The power network can be used in a number of ways:

- Conditioning on the status of the switches and circuit breakers, whether there is outside power and the position of the switches, you can simulate the lighting.
- Given values for the switches, the outside power, and whether the lights are lit, you can determine the posterior probability that each switch or circuit breaker is ok or not.
- Given some switch positions and some outputs and some intermediate values, you can determine the probability of any other variable in the network.

Example: Liver Diagnosis

Source: Onisko et al., 1999

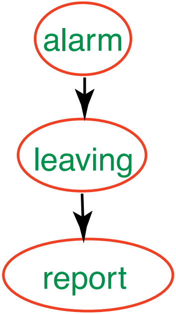


Belief network summary

- A belief network is a directed acyclic graph (DAG) where nodes are random variables.
  - A belief network is automatically acyclic by construction.
- The parents of a node n are those variables on which n directly depends.
- A belief network is a graphical representation of dependence and independence:
  - A variable is conditionally independent of its non-descendants given its parents.

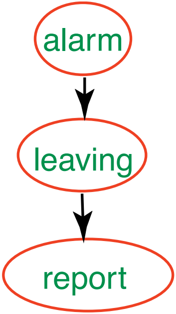
Lecture Overview

Chain



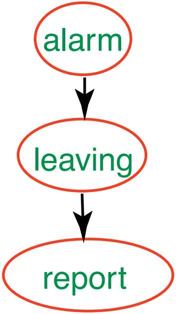
- $alarm$ and $report$ are independent:

Chain



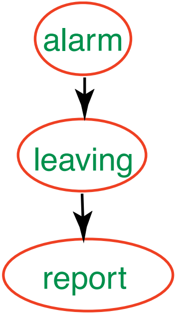
- $alarm$ and $report$ are independent: false.

Chain



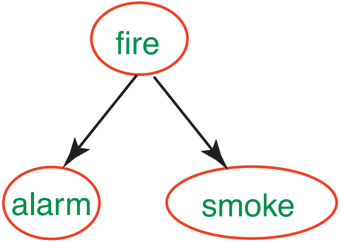
- $alarm$ and $report$ are independent: false.
- $alarm$ and $report$ are independent given $leaving$:

Chain



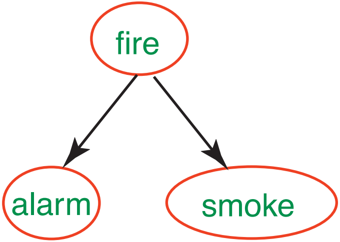
- $alarm$ and $report$ are independent: false.
- $alarm$ and $report$ are independent given $leaving$: true.
- Intuitively, the only way that the $alarm$ affects $report$ is by affecting $leaving$.

Common ancestors



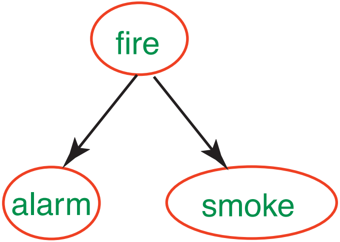
- *alarm* and *smoke* are independent:

Common ancestors



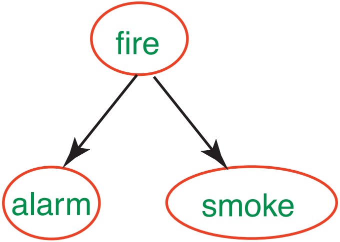
- $alarm$ and $smoke$ are independent: false.

Common ancestors



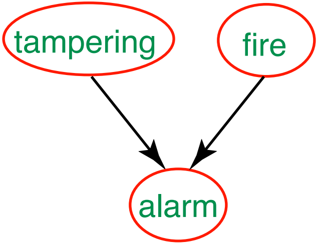
- $alarm$ and $smoke$ are independent: false.
- $alarm$ and $smoke$ are independent given $fire$:

Common ancestors



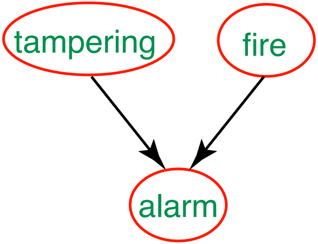
- $alarm$ and $smoke$ are independent: false.
- $alarm$ and $smoke$ are independent given $fire$: true.
- Intuitively, $fire$ can explain $alarm$ and $smoke$; learning one can affect the other by changing your belief in $fire$.

Common descendants



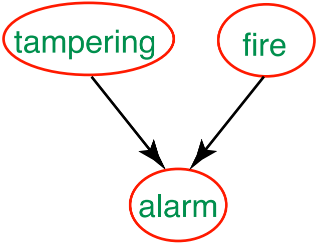
- $tampering$ and $fire$ are independent:

Common descendants



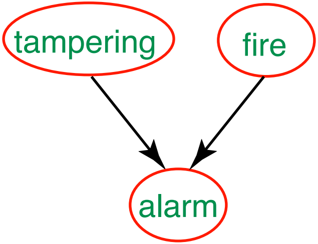
- $tampering$ and $fire$ are independent: true.

Common descendants



- $tampering$ and $fire$ are independent: true.
- $tampering$ and $fire$ are independent given $alarm$:

Common descendants



- $tampering$ and $fire$ are independent: true.
- $tampering$ and $fire$ are independent given $alarm$: false.
- Intuitively, $tampering$ can explain away $fire$

Lecture Overview

1. Recap
2. Belief Network Examples
3. Observing Variables
4. Belief Network Inference
5. Factors

Belief Network Inference

- Our goal: compute probabilities of variables in a belief network
- Two cases:
  1. the unconditional (prior) distribution over one or more variables
  2. the posterior distribution over one or more variables, conditioned on one or more observed variables

Evidence

- If we want to compute the posterior probability of Z given evidence $Y_1 = v_1 \wedge \ldots \wedge Y_j = v_j$:

$$
\begin{aligned}
& P(Z|Y_1 = v_1, \ldots, Y_j = v_j) \\
= \quad & \frac{P(Z, Y_1 = v_1, \ldots, Y_j = v_j)}{P(Y_1 = v_1, \ldots, Y_j = v_j)} \\
= \quad & \frac{P(Z, Y_1 = v_1, \ldots, Y_j = v_j)}{\sum_Z P(Z, Y_1 = v_1, \ldots, Y_j = v_j).}
\end{aligned}
$$

- So the computation reduces to the probability of $P(Z, Y_1 = v_1, \ldots, Y_j = v_j)$.

Belief Network Inference

- Our goal: compute probabilities of variables in a belief network
- Two cases:
    1. the unconditional (prior) distribution over one or more variables
    2. the posterior distribution over one or more variables, conditioned on one or more observed variables
- To address both cases, we only need a computational solution to case 1
- Our method: exploiting the structure of the network to efficiently eliminate (sum out) the non-observed, non-query variables one at a time.

Lecture Overview

1. Recap
2. Belief Network Examples
3. Observing Variables
4. Belief Network Inference
5. Factors

Factors

- A factor is a representation of a function from a tuple of random variables into a number.
- We will write factor f on variables $X_1, \ldots, X_j$ as $f(X_1, \ldots, X_j)$.
- A factor denotes a distribution over the given tuple of variables in some (unspecified) context
  - e.g., $P(X_1, X_2)$ is a factor $f(X_1, X_2)$
  - e.g., $P(X_1, X_2, X_3 = v_3)$ is a factor $f(X_1, X_2)$
  - e.g., $P(X_1, X_3 = v_3 | X_2)$ is a factor $f(X_1, X_2)$

Manipulating Factors

- We can make new factors out of an existing factor
- Our first operation: we can assign some or all of the variables of a factor.
  - $f(X_1 = v_1, X_2, \ldots, X_j)$, where $v_1 \in dom(X_1)$, is a factor on $X_2, \ldots, X_j$.
  - $f(X_1 = v_1, X_2 = v_2, \ldots, X_j = v_j)$ is a number that is the value of f when each X_i has value v_i.
- The former is also written as $f(X_1, X_2, \ldots, X_j)_{X_1 = v_1, \ldots, X_j = v_j}$

Example factors

$r(X, Y, Z)$:

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

$r(X{=}t, Y, Z)$:

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

$r(X{=}t, Y, Z{=}f)$:

Y	val
t	0.9
f	0.8

$r(X{=}t, Y{=}f, Z{=}f) = 0.8$

Summing out variables

Our second operation: we can sum out a variable, say X_1 with domain $\{v_1, \ldots, v_k\}$, from factor $f(X_1, \ldots, X_j)$, resulting in a factor on $X_2, \ldots, X_j$ defined by:

$$
\begin{aligned}
&(\sum_{X_1} f)(X_2, \ldots, X_j) \\
&\quad = \quad f(X_1 = v_1, \ldots, X_j) + \cdots + f(X_1 = v_k, \ldots, X_j)
\end{aligned}
$$

Summing out a variable example

f_3:

A	B	C	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

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

Multiplying factors

- Our third operation: factors can be multiplied together.
- The product of factor $f_1(\overline{X}, \overline{Y})$ and $f_2(\overline{Y}, \overline{Z})$, where $\overline{Y}$ are the variables in common, is the factor $(f_1 \times f_2)(\overline{X}, \overline{Y}, \overline{Z})$ defined by:

$$(f_1 \times f_2)(\overline{X}, \overline{Y}, \overline{Z}) \quad = \quad f_1(\overline{X}, \overline{Y}) f_2(\overline{Y}, \overline{Z}).$$

- Note: it's defined on all $\overline{X}, \overline{Y}, \overline{Z}$ triples, obtained by multiplying together the appropriate pair of entries from f_1 and f_2.

Multiplying factors example

f_1:

A	B	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

f_2:

B	C	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$f_1 \times f_2$:

A	B	C	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