Probabilistic Reasoning

Ch 14

22C:145
University of Iowa

Probability theory for representing uncertainty

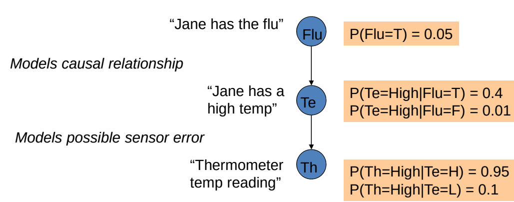
- Assigns a numerical degree of belief between 0 and 1 to facts
  - e.g. “it will rain today” is T/F.
  - P(“it will rain today”) = 0.2 prior probability (unconditional)
- Conditional probability (Posterior)
  - P(“it wil rain today” | “rain is forecast”) = 0.8
- Bayes’ Rule: $P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$

Bayesian networks

- Directed acyclic graphs
- Nodes: random variables,
  - R: “it is raining”, discrete values T/F
  - T: temperature, continuous *or* discrete variable
  - C: color, discrete values {red, blue, green}
- Arcs indicate dependencies (can have causal interpretation)

Bayesian networks

- Conditional Probability Distribution (CPD)
  - Associated with each variable
  - probability of each state given parent states

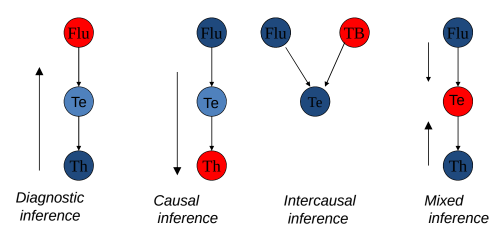


Inference in Belief Networks

- Main task of a belief network: Compute the conditional probability of a set of query variables given exact values for some evidence variables: P(query | evidence).
- Belief networks are flexible enough so that any node can serve as either a query or an evidence variable.

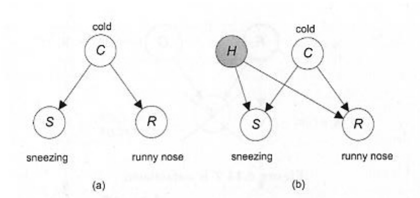
BN inference

- Evidence: observation of specific state
- Task: compute the posterior probabilities for query node(s) given evidence.



Building a BN

- Choose a set of random variables X_i that describe the domain.
  - Missing variables may cause the BN unreliable.



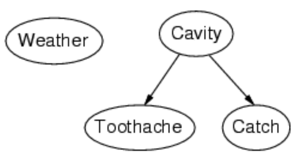
Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions
- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

    P (X_i | Parents (X_i))
- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

Example

- Topology of network encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* (steel probe catches in teeth) are conditionally independent given *Cavity*

Example

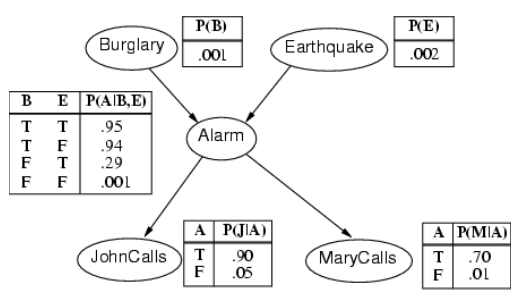
- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

The Alarm Example

- Mr. Holmes' security alarm at home may be triggered by either burglar or earthquake. When the alarm sounds, his two nice neighbors, Mary and John, may call him.

causal DAG

Example contd.



P(B)
.001

P(E)
.002

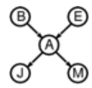
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness

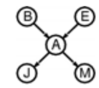
- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values
- Each row requires one number p for $X_i = true$ (the number for $X_i = false$ is just $1\text{-}p$)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. 2^5-1 = 31)



Semantics

The full joint distribution is defined as the product of the local conditional distributions:

$$\mathbf{P}(X_1, \dots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$



e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)\, \mathbf{P}(a \mid \neg b, \neg e)\, \mathbf{P}(\neg b)\, \mathbf{P}(\neg e)$

Building a BN

- Choose a set of random variables X_i that describe the domain.
- Order the variables into a list L
- Start with an empty BN.
- For each variable X in L do
  - Add X into the BN
  - Choose a minimal set of nodes already in the BN which satisfy the conditional dependence property with X
  - Make these nodes the parents of X.
  - Fill in the CPT for X.

Constructing Bayesian networks

- 1. Choose an ordering of variables $X_1, \dots, X_n$
- 2. For i = 1 to n
  - add X_i to the network
  - select parents from $X_1, \dots, X_{i-1}$ such that $\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \dots X_{i-1})$

This choice of parents guarantees:

$\mathbf{P}(X_1, \dots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \dots, X_{i-1})$
(chain rule)

$= \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$
(by construction)

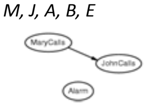
Example

- Suppose we choose the ordering *M, J, A, B, E*

$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

Example

- Suppose we choose the ordering *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$?

Example

- Suppose we choose the ordering *M, J, A, B, E*

P*(J | M) =* ***P****(J)?*
No
P*(A | J, M) =* ***P****(A | J)*? ***P****(A | J, M) =* ***P****(A)*? **No**
P*(B | A, J, M) =* ***P****(B | A)*?
P*(B | A, J, M) =* ***P****(B)*?

Example

- Suppose we choose the ordering M, J, A, B, E

P*(J | M) =* ***P****(J)?*
No
P*(A | J, M) =* ***P****(A | J)*? ***P****(A | J, M) =* ***P****(A)*? **No**
P*(B | A, J, M) =* ***P****(B | A)*? **Yes**
P*(B | A, J, M) =* ***P****(B)*? **No**
P*(E | B, A ,J, M) =* ***P****(E | A)*?
P*(E | B, A, J, M) =* ***P****(E | A, B)*?

Example

- Suppose we choose the ordering M, J, A, B, E

P*(J | M) =* ***P****(J)?*
No
P*(A | J, M) =* ***P****(A | J)*? ***P****(A | J, M) =* ***P****(A)*? **No**
P*(B | A, J, M) =* ***P****(B | A)*? **Yes**
P*(B | A, J, M) =* ***P****(B)*? **No**
P*(E | B, A ,J, M) =* ***P****(E | A)*? **No**
P*(E | B, A, J, M) =* ***P****(E | A, B)*? **Yes**

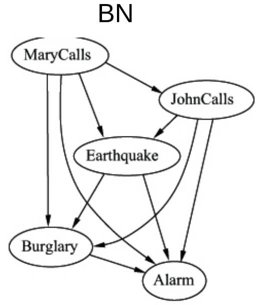
Example contd.

- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

The Alarm Example

- Variable order:
  – MaryCalls
  – JohnCalls
  – Earthquake
  – Burglary
  – Alarm

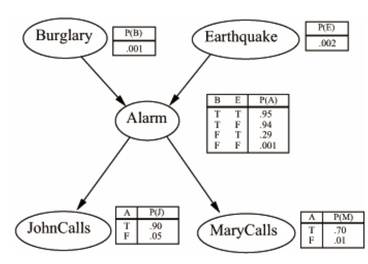
BN



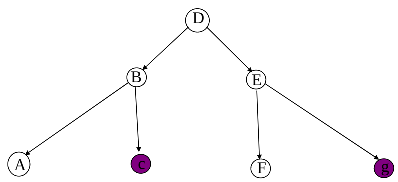
The Alarm Example

- Variable order:
  – Burglary
  – Earthquake
  – Alarm
  – JohnCalls
  – MaryCalls

BN

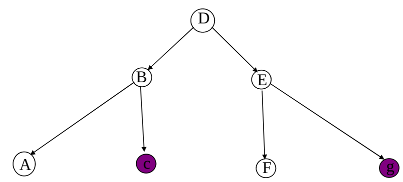


Example



Say we want to compute p(a | c, g)

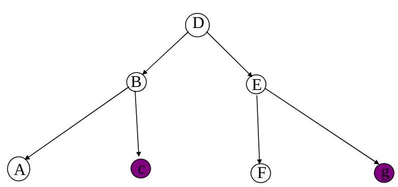
Example



Direct calculation: $p(a|c,g) = \sum_{bdef} p(a,b,d,e,f \mid c,g)$

Complexity of the sum is $O(m^4)$

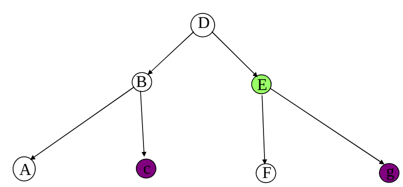
Example



Reordering:

$$\sum_b p(a|b) \sum_d p(b|d,c) \sum_e p(d|e) \sum_f p(e,f|g)$$

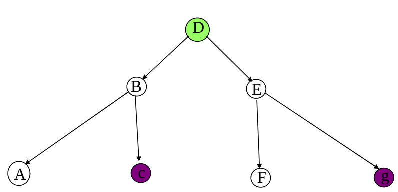
Example



Reordering:

$$\sum_b p(a|b) \sum_d p(b|d,c) \sum_e p(d|e) \underbrace{\sum_f p(e,f|g)}_{p(e|g)}$$

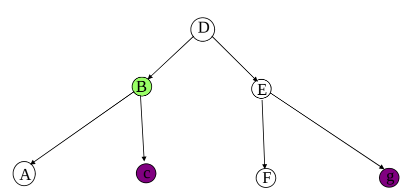
Example



Reordering:

$$\sum_b p(a|b) \sum_d p(b|d,c) \underbrace{\sum_e p(d|e)\, p(e|g)}_{p(d|g)}$$

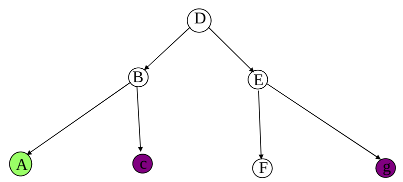
Example



Reordering:

$$\sum_b p(a|b) \underbrace{\sum_d p(b|d,c)\, p(d|g)}_{p(b|c,g)}$$

Example



Reordering:

$\sum_b p(a|b)\, p(b|c,g)$

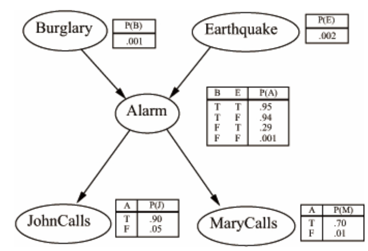
p(a|c,g)

Complexity is O(m), compared to O(m^4)

The Alarm Example

- P(A | B) = ?
- P(B | A) = ?
- P(B | E) = ?
- P(M | B) = ?
- P(B | M) = ?
- P(M | J) = ?
- P(B | M, J) = ?

BN



General Strategy for inference

- Want to compute P(q | e)

Step 1:

P(q | e) = P(q,e)/P(e) = α P(q,e), since P(e) is constant wrt Q

Step 2:

P(q,e) = $\Sigma_{a..z}$ P(q, e, a, b, …. z), by the law of total probability

Step 3:

$\Sigma_{a..z}$ P(q, e, a, b, …. z) = $\Sigma_{a..z}$ Π_i P(variable i | parents i)

(using Bayesian network factoring)

Step 4:

Distribute summations across product terms for efficient computation

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct

Weakness of BN

- Hard to obtain JPD (joint probability distribution)
  - Relative Frequency Approach: counting outcomes of repeated experiments
  - Subjective Approach: an individual's personal judgment about whether a specific outcome is likely to occur.
- Worst time complexity is NP-hard.

BN software

- Commerical packages: Netica, Hugin, Analytica (all with demo versions)
- Free software: Smile, Genie, JavaBayes, …

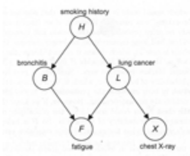
http://HTTP.CS.Berkeley.EDU/~murphyk/Bayes/bnsoft.html

Extensions of BN

- Weaker requirement in a DAG: Instead of

$I(X, ND_X \mid PA_X)$, ask $I(X, ND_X \mid MB_X)$, where

MB_X is called Markov Blanket of X, which is the set of neighboring nodes: its parents (PA_X), its children, and any other parents of X's children.



$PA_B = \{ H \}$
$MB_B = \{ H, L, F \}$
$ND_B = \{ L, X \}$

Open Research Questions

- Methodology for combining expert elicitation and automated methods
  - expert knowledge used to guide search
  - automated methods provide alternatives to be presented to experts
- Evaluation measures and methods
  - may be domain depended
- Improved tools to support elicitation
  - e.g. visualisation of d-separation
- Industry adoption of BN technology