Introduction to Artificial Intelligence

Belief networks

Chapter 15.1–2

Dieter Fox

Outline

◇ Bayesian networks: syntax and semantics

◇ Inference tasks

Belief 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:

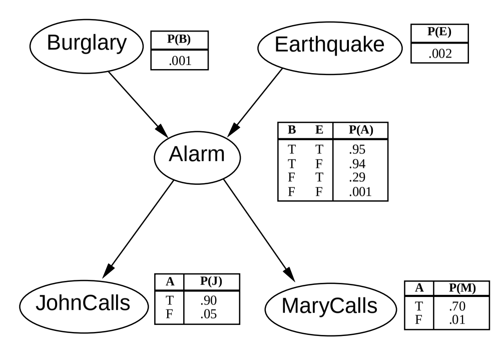
$$\mathbf{P}(X_i | Parents(X_i))$$

In the simplest case, conditional distribution represented as a **conditional probability table** (CPT)

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: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
Network topology reflects "causal" knowledge:



Semantics

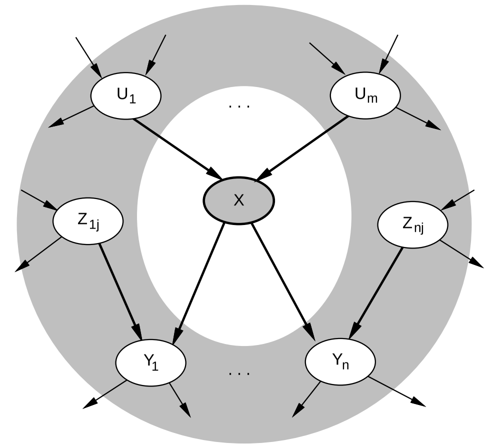
"Global" semantics defines the full joint distribution as
the product of the local conditional distributions:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

e.g., $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$ **is given by**??
=

Markov blanket

Each node is conditionally independent of all others given its
Markov blanket: parents + children + children's parents



Constructing belief networks

Need a method such that a series of locally testable assertions of
conditional independence guarantees the required global semantics

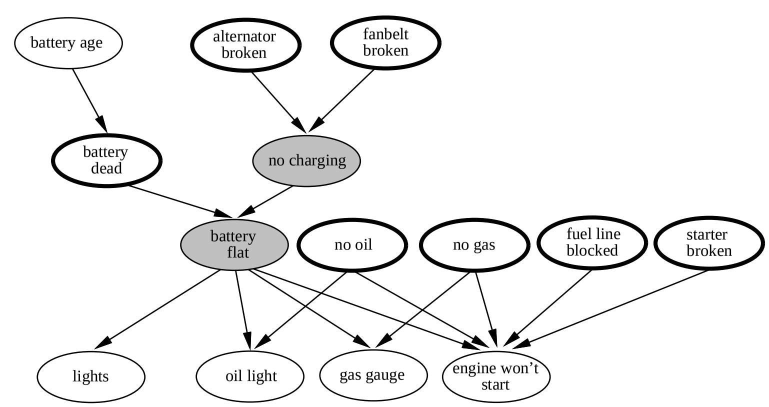
1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For i = 1 to n
   add X_i to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1, \ldots, X_{i-1})$

This choice of parents guarantees the global semantics:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1}) \text{ (chain rule)}$$
$$= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) \text{ by construction}$$

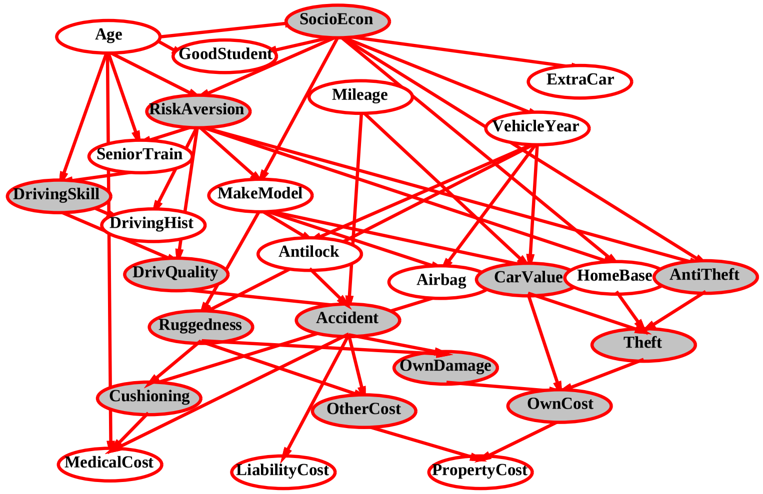
Example: Car diagnosis

Initial evidence: engine won't start
Testable variables (thin ovals), diagnosis variables (thick ovals)
Hidden variables (shaded) ensure sparse structure, reduce parameters



Example: Car insurance

Predict claim costs (medical, liability, property)
given data on application form (other unshaded nodes)



Compact conditional distributions

CPT grows exponentially with no. of parents
CPT becomes infinite with continuous-valued parent or child

Solution: **canonical** distributions that are defined compactly

Deterministic nodes are the simplest case:
$X = f(Parents(X))$ for some function f

E.g., Boolean functions
$NorthAmerican \Leftrightarrow Canadian \vee US \vee Mexican$

E.g., numerical relationships among continuous variables

$$\frac{\partial Level}{\partial t} = \text{inflow + precipation - outflow - evaporation}$$

Compact conditional distributions contd.

Noisy-OR distributions model multiple noninteracting causes
1) Parents $U_1 \ldots U_k$ include all causes (can add **leak node**)
2) Independent failure probability q_i for each cause alone

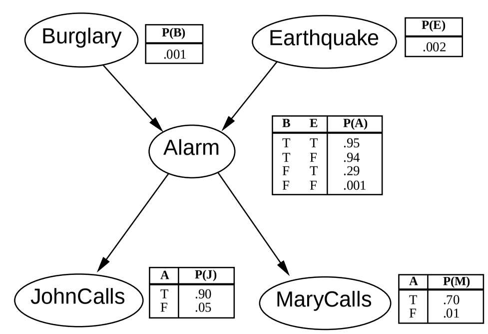
$$\Rightarrow P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \prod_{i=1}^{j} q_i$$

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Number of parameters **linear** in number of parents

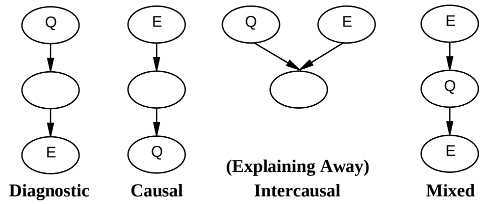
Inference in Bayesian networks

Instantiate some nodes (evidence nodes) and query other nodes.



$P($**Burglary** $|$ **JohnCalls**$)$??

Types of inference



1. **Diagnostic**: From effects to causes
   $P(\textit{Burglary} \mid \textit{JohnCalls}) = 0.016$
2. **Causal**: From causes to effects
   $P(\textit{JohnCalls} \mid \textit{Burglary}) = 0.86$
3. **Intercausal**: between causes of common effect
   $P(\textit{Burglary} \mid \textit{Alarm}) = 0.376$, but $P(\textit{Burglary} \mid \textit{Alarm}, \textit{Earthquake}) = 0.003$.
4. **Mixed:** Combinations of 1.-3.
   $P(\textit{Alarm} \mid \textit{JohnCalls}, \neg \textit{Earthquake}) = 0.03$

Inference tasks

Queries: compute posterior marginal $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$
e.g., $P(NoGas|Gauge=empty, Lights=on, Starts=false)$

Optimal decisions: decision networks include utility information;
probabilistic inference required for $P(outcome|action, evidence)$

Value of information: which evidence to seek next?

Sensitivity analysis: which probability values are most critical?

Explanation: why do I need a new starter motor?