Uncertain Knowledge and Reasoning

Outline

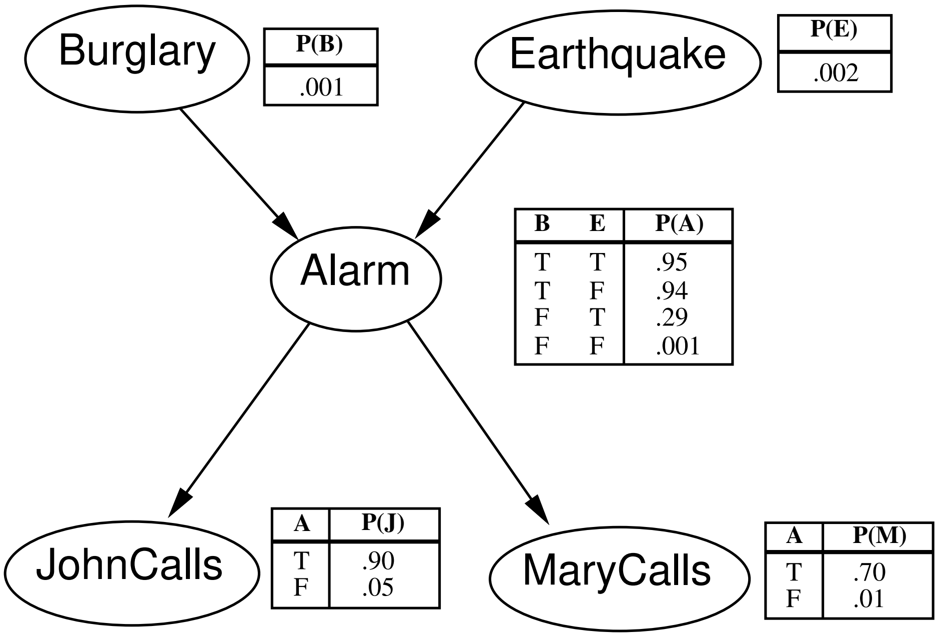
- Belief networks
- Inference in belief networks
- Other approaches

Belief networks

(Bayesian network, probabilistic network, causal network, knowledge map)

Compact specification of the joint probability distribution.

- Nodes: random variables
- Arrows: direct influence
- Each nodes has a conditional probability table (CPT): effects of the parents on the node.
- The graph is acyclic.



Semantics of belief networks

➤ Network represents the joint probability distribution.
$(P(X_1 = x_1 \wedge \cdots \wedge X_n = x_n) = P(x_1, \ldots, x_n))$

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

Example.

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) =$$
$$= P(J|A)P(M|A)P(A|\neg B \wedge \neg E)P(\neg E)P(\neg B)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.998 \times 0.999$$
$$= 0.00063$$

A method for constructing belief networks

➤ Network encodes conditional independence statements.

$P(x_1, \ldots, x_n) = P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1)$
$= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1} | x_{n-2}, \ldots, x_1) \cdots P(x_2 | x_1) P(x_1)$
$= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$

If nodes labelled such that $Parents(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$

$$\mathbf{P}(X_i | X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i | Parents(X_i)) \qquad (1)$$

➤ A belief network is a correct representation if each node is conditionally independent of its predecessors given its parents.

➤ Parents of X_i should contain all those $X_1, \ldots, X_{i-1}$ that directly influence X_i.

Example. Only $Alarm$ directly influences $MaryCalls$.

$\mathbf{P}(MaryCalls | JohnCalls, Alarm, Earthquake, Burglary) =$
$\mathbf{P}(MaryCalls | Alarm)$

Incremental network construction:

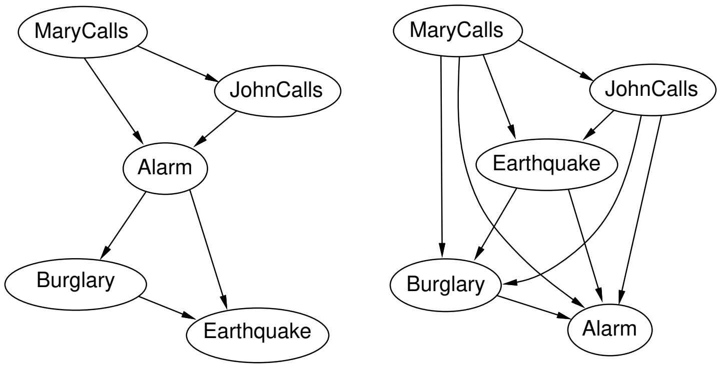
1. Choose a set of relevant variables X_i.
2. Choose an ordering for the variables.
3. While there are variables left:
   (a) Pick a variable X_i and add a node.
   (b) Set $Parents(X_i)$ to some minimal set of nodes already in the net such that conditional independence property (1) holds.
   (c) Define the conditional probability table for X_i.

☞ Network acyclic and axioms of probability satisfied.

Compactness and node ordering

➤ A belief network can be a compact representation of the joint probability distribution.
Locally structured / sparse system
If each Boolean variable directly influences at most k others, conditional probability tables: $n2^k$ numbers (instead of 2^n)

➤ Choosing a good node ordering non-trivial:
The correct order to add nodes is to add the "root causes" first, then variables they influence and so on.



$P(JohnCalls|MaryCalls) \neq P(JohnCalls)$

$P(Burglary|Alarm, JohnCalls, MaryCalls) =$

$P(Burglary|Alarm)$

$P(Earthquake|Burglary, Alarm, JohnCalls, MaryCalls) =$

$P(Earthquake|Burglary, Alarm)$

Representation of conditional probability tables

➤ Filling conditional probability tables needs a lot of work.
☞ Canonical distributions

➤ Deterministic nodes
Values specified exactly by the values of the parents (no uncertainty).

➤ "Noisy" logical relationships

Noisy-OR

(i) Each cause has an independent chance of causing the effect.
(ii) All possible causes are listed.
(iii) Whatever inhibits some cause from causing an effect is independent of whatever inhibits some other cause from causing the effect.

Inhibitors summarized as noise parameters.

Noisy relationships in which a variable depends on k parents can be described using $O(k)$ parameters (instead of $O(2^k)$ entries in the full table).

Example.

$P(Fever|Cold) = 0.4$
$P(Fever|Flu) = 0.8$
$P(Fever|Malaria) = 0.9$
Noise parameters $0.6, 0.2, 0.1$, respectively.

$Cold$	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	$0.012 = 0.6 \times 0.2 \times 0.1$

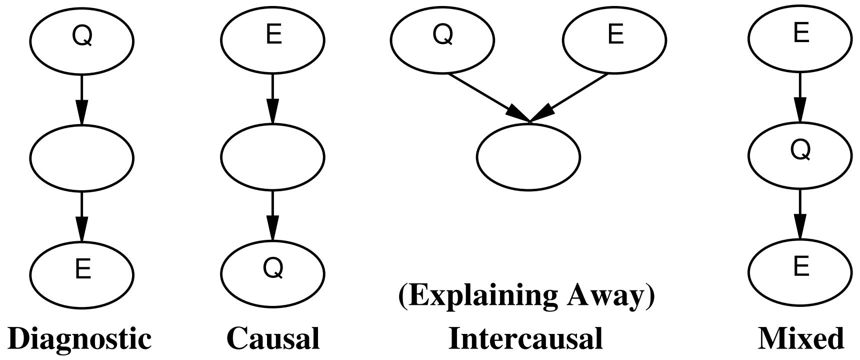
Inference in Belief Networks

- An agent gets values for evidence variables from its percepts and asks about the possible values of other variables so that it can decide what action to take.
- BELIEF-NET-TELL: adding evidence in the network
- BELIEF-NET-ASK: computing the posterior probability distribution for a given query variable $\mathbf{P}(Query|Evidence)$?

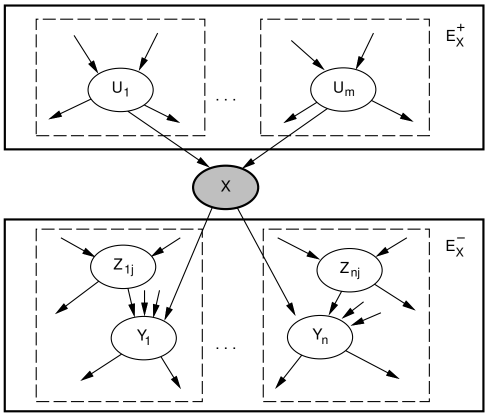
Example. $P(Burglary|JohnCalls, MaryCalls)$? $P(Alarm|JohnCalls, Earthquake)$?

The nature of probabilistic inference

- Diagnostic inferences (from effects to causes)
  $P(Burglary|JohnCalls) = 0.016$
- Causal inferences (from causes to effects)
  $P(JohnCalls|Burglary) = 0.86$
- Intercausal inferences (between causes/common effect)
  $P(Burglary|Alarm) = 0.376$
  $P(Burglary|Alarm \wedge Earthquake) = 0.003$
- Mixed inferences (combining two or more of the above)
  $P(Alarm|JohnCalls \wedge \neg Earthquake) = 0.03$
  $P(Burglary|JohnCalls \wedge \neg Earthquake) = 0.017$



If the network is a **polytree** (singly connected; at most one undirected path between any two nodes), then the probability distribution can be computed in **linear time**.



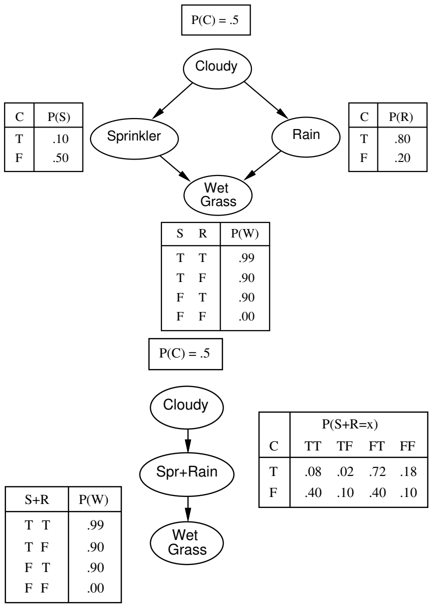
$$\mathbf{P}(X|E) = \alpha \mathbf{P}(E_X^-|X)\mathbf{P}(X|E_X^+)$$

Inference in Belief Networks

- Algorithms for polytrees are used as subroutines in algorithms for general (multiply connected) networks.
- Clustering
- Conditioning
- Stochastic simulation

☞ Exact inference in belief networks is NP-hard.

Clustering methods



C	P(S)
T	.10
F	.50

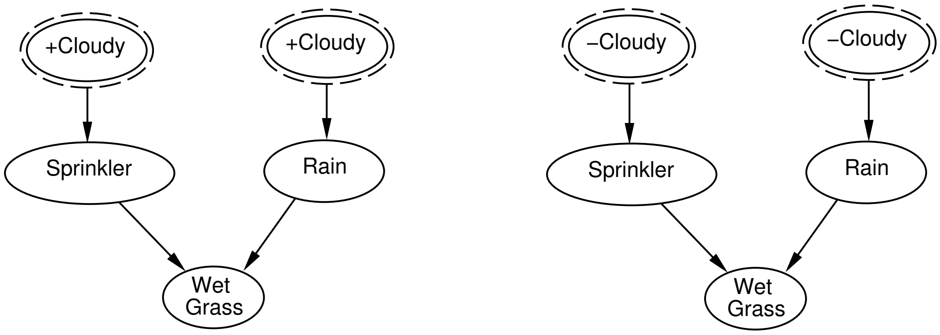
C	P(R)
T	.80
F	.20

S	R	P(W)
T	T	.99
T	F	.90
F	T	.90
F	F	.00

	P(S+R=x)			
C	TT	TF	FT	FF
T	.08	.02	.72	.18
F	.40	.10	.40	.10

S+R	P(W)
T T	.99
T F	.90
F T	.90
F F	.00

Cutset conditioning methods



Stochastic simulation methods

- Logic sampling

  Simulate the world described by the belief network and estimate the probabilities by counting the frequencies with which events occur.

  Starting from the root nodes of the graph values are randomly chosen and the choices are weighted by the prior/conditional probabilities.

  ☞ Problems when interested in a case E that rarely occurs. (e.g. $P(WetGrass|Sprinkler \wedge Rain)$)

- Likelihood weighting

  For an evidence variable, instead of randomly choosing a value, the given value is taken but the conditional probability is used to see how likely that is.

  Likelihood weighting converges considerably faster than logic sampling.

  ☞ Takes still a long time to reach accurate probabilities for unlikely events.

Example. P(WetGrass|Rain)?

1. Choose a value for $Cloudy$ ($P(Cloudy) = 0.5$).
$Cloudy := False$.
2. Choose a value for $Sprinkler$
($P(Sprinkler|\neg Cloudy) = 0.5$). $Sprinkler := True$.
3. Choose a value for $Rain$. $Rain := True$ (evidence;
$P(Rain|\neg Cloudy) = 0.2$)
4. Choose a value for $WetGrass$
($P(WetGrass|Sprinkler \wedge Rain) = 0.99$).
$WetGrass := True$.
5. We have completed a run with likelihood 0.2 that says
$WetGrass = True$ given $Rain = True$.

Knowledge Engineering for Uncertainty

- Decide what to talk about.
- Decide on a vocabulary of random variables.
- Encode general knowledge about the dependences between variables.
- Encode a description of the specific problem instance.
- Pose queries to the inference procedure and get answers.

Other Approaches

- Default reasoning
- Dempster-Shafer theory
- Fuzzy logic

Default reasoning

- Default reasoning: infer in absence of any information to the contrary
- A qualitative approach
- Compact knowledge representation
- Improving implementation techniques

$$
\begin{aligned}
&do(A) \leftarrow preconds(A) \wedge \mathrm{not}(exceptions(A)) \\
&exceptions(A) \leftarrow ex_1(A) \\
&exceptions(A) \leftarrow ex_2(A) \\
&\vdots
\end{aligned}
$$

Rule-based methods with certainty factors

- Truth-functionality and chaining problematic

  $Sprinkler \mapsto WetGrass$ and $WetGrass \mapsto Rain$

Dempster-Shafer theory

- $Bel(X)$: probability that the evidence supports X.

  $Bel(Heads) = 0$, $Bel(\neg Heads) = 0$

  0.9 certainty that the coin is fair:

  $Bel(Heads) = 0.5 \times 0.9 = 0.45$, $Bel(\neg Heads) = 0.45$

Fuzzy logic

- Vagueness (rather than uncertainty)

  "Mika Myllylä is tall" (truth value between 0 and 1)