Review of Expert Systems, Probabilities and Bayesian Networks

Yuqing Tang

Doctoral Program in Computer Science
The Graduate Center
City University of New York
ytang@cs.gc.cuny.edu

December 14, 2010

Outline

Expert systems

$$\text{Expert system} = \text{Knowledge base} + \text{Inference engine}$$

- Knowledge base contains facts about objects in the chosen domain and their relationships
  - Knowledge base can also contains concepts, theories, practical procedures, and their associations
- The inference mechanism is a set of procedures that are used to examine the knowledge based in an orderly manner to answer questions, solve problems, or make decisions within the domain

Overview of knowledge representation and methods of inference

Knowledge representation

- Logic
  - Propositional logic
  - Predicate logic
- Production rules
- Semantic networks/web
- Frames
- Probability

Methods of inference

- Reasoning with logic
- Inference with rules
  - Forward chaining
  - Backward chaining
- The inference tree
- Inference with frames
- Probabilistic inferences

Outline

Event Space

- Let U be the universe of all possible events
- For any possible event $X, Y \subseteq U$
- Set-theoretical operators
  - $X \cap Y \stackrel{\text{def}}{=} \{z|x \in X \text{ and } z \in Y\}$
  - $X \cup Y \stackrel{\text{def}}{=} \{z|x \in X \text{ or } z \in Y\}$
  - $X \setminus Y \stackrel{\text{def}}{=} \{z|x \in X \text{ but } z \notin Y\}$
  - $\bar{X} = U \setminus X$

Kolmogorov Axioms

1. $P(U) = 1$
2. For any $X \subseteq U$, $P(X) \geq 0$
3. For any two events $X, Y \subseteq U$
   if $X \cap Y = \emptyset$
   then $P(X \cup Y) = P(X) + P(Y)$

Conditional Probability

Definition (Conditional Probability)

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

Bayes' Rule

Definition

Bayes' Rule [Reverend Thomas Bayes (1764)]

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

- Read $P(E|H)$ as the likelihood of the event E given hypothesis H
- Read $P(H)$ as the prior of the hypothesis H
- Read $P(E)$ as the prior of the evidence E
- Read $P(H|E)$ as the posterior belief $Bel(H|E)$ of H given evidence E

Independence

Definition

Event X is said be independent of event Y, denoted by $X \perp\!\!\!\perp Y$,

- iff $P(X|Y) = P(X)$

- An equivalent definition is $P(X \cap Y) = P(X) \cdot P(Y)$
- Independence can be input knowledge
- Independence can arise from the probability

Conditional independence

Definition

Event X is said to be independent of event Y given Z, denoted $X \perp\!\!\!\perp Y|Z$

- iff $P(X|Y,Z) = P(X|Z)$

- An equivalent definition is $P(X \cap Y|Z) = P(X|Z) \cdot P(Y|Z)$

Outline

Multivariate

- A vector of variables $\mathbf{X} = \langle X_1, \ldots, X_k \rangle$ where $X_1, \ldots, X_k \in \mathcal{V}$
- A vector of values $\mathbf{x} = \langle x_1, \ldots, x_k \rangle$ where $x_i \in Domain(X_k)$ $(1 \leq i \leq k)$
- Multivariate notion of an event: $\mathbf{X} = \mathbf{x}$ means

$$X_1 = x_1 \wedge X_2 = x_2 \wedge \ldots \wedge X_k = x_k$$

  which is usually abbreviated by $\mathbf{x}$ if the variables $\mathbf{X}$ is clear in the context
- $P(\mathbf{X})$ corresponds to a table of probabilities with each assignment $\mathbf{x}$ to $\mathbf{X}$ having an entry in the table
- $P(\mathbf{X}, \mathbf{Y})$ corresponds to a table of probabilities with each assignment $\langle \mathbf{x}, \mathbf{y} \rangle$ to $\langle \mathbf{X}, \mathbf{Y} \rangle$ having an entry in the table

Examples I

- $\mathcal{V} = \{P, S, C\}$ where P for pollution, S for smoking, and C for having cancer
- The corresponding domains are $Domain(P) = \{low, high\}$, $Domain(S) = \{T, F\}$, $Doman(C) = \{T, F\}$
- All possible worlds are

$\langle P, S, C\rangle$
$\langle low, T, T\rangle$
$\langle low, F, T\rangle$
$\langle high, T, T\rangle$
$\langle high, F, T\rangle$
$\langle low, T, F\rangle$
$\langle low, F, F\rangle$
$\langle high, T, F\rangle$
$\langle high, F, F\rangle$

Examples II

- Event $P = low$ corresponds to

$$\{\langle P = low, S = T, C = T\rangle, \langle P = low, S = F, C = T\rangle,$$
$$\langle P = low, S = T, C = F\rangle, \langle P = low, S = F, C = F\rangle\}$$

- Event $S = T$ corresponds to

$$\{\langle P = low, S = T, C = T\rangle, \langle P = high, S = T, C = T\rangle,$$
$$\langle P = low, S = T, C = T\rangle, \langle P = high, S = T, C = T\rangle\}$$

- Event $P = low, S = T$ corresponds to

$$\{\langle P = low, S = T, C = T\rangle, \langle P = low, S = T, C = T\rangle\}$$

Multivariate version of conditional probability

$$P(\mathbf{X}|\mathbf{Y}) = \frac{P(\mathbf{X}, \mathbf{Y})}{P(\mathbf{Y})}$$

means for every assignment $\langle \mathbf{x}, \mathbf{y} \rangle$ to $\langle \mathbf{X}, \mathbf{Y} \rangle$, we have

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{x}, \mathbf{y})}{P(\mathbf{y})}$$

Multivariate version of conditional independence

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$$

means

$$P(\mathbf{X}|\mathbf{Y}, \mathbf{Z}) = P(\mathbf{X}|\mathbf{Z})$$

means for every assignment $\langle \mathbf{x}, \mathbf{y}, \mathbf{z} \rangle$ to $\langle \mathbf{X}, \mathbf{Y}, \mathbf{Z} \rangle$, we have

$$P(\mathbf{x}|\mathbf{y}, \mathbf{z}) = P(\mathbf{x}|\mathbf{z})$$

Multivariate version of Bayes' Rule

Bayes rule:

$$P(\mathbf{H}|\mathbf{E}) = \frac{P(\mathbf{E}|\mathbf{H}) \cdot P(\mathbf{H})}{P(\mathbf{E})}$$

For every assignment $\langle \mathbf{h}, \mathbf{e} \rangle$ to $\langle \mathbf{H}, \mathbf{E} \rangle$, we have

- If **e** is the only known evidence in the context
  - Read the likelihood of **e** given **h** simply as likelihood of **h**:

    $$\lambda(\mathbf{h}) = P(\mathbf{e}|\mathbf{h})$$

  - Read the belief of h given e simply as belief:

    $$Bel(\mathbf{h}) = Bel(\mathbf{h}|\mathbf{e}) = P(\mathbf{h}|\mathbf{e})$$

- Bayes' rule can then be read as

$$Posterior \;=\; \frac{Likelihood \times Prior}{Prob\ of\ \ evidence}$$

Marginalization

$$P(\mathbf{X} = \mathbf{x}) = \Sigma_{\mathbf{y} \in Domain(\mathbf{Y})} P(\mathbf{X} = \mathbf{x}, \mathbf{Y})$$

Example

P	S	$P(P, S)$
H	T	0.03
H	F	0.07
L	T	0.27
L	F	0.63

$$\begin{aligned} & P(P = low) \\ & = P(P = low, S = T) \\ & \quad + P(P = low, S = F) \\ & = 0.9 \end{aligned}$$

The chain rule

Each assignment $\langle \mathbf{x_1}, \ldots, \mathbf{x_n} \rangle$ to $\langle \mathbf{X}_1, \ldots, \mathbf{X_n} \rangle$ satisfies

$$
\begin{aligned}
P(\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}) = & \\
& P(\mathbf{X_1}) \\
& \times P(\mathbf{X_2} | \mathbf{X_1}) \\
& \times P(\mathbf{X_3} | \mathbf{X_1}, \mathbf{X_2}) \\
& \times \ldots \times P(\mathbf{X}_n | \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}) \\
= \quad & \Pi_{i=1,\ldots,n} P(\mathbf{X}_i | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})
\end{aligned}
$$

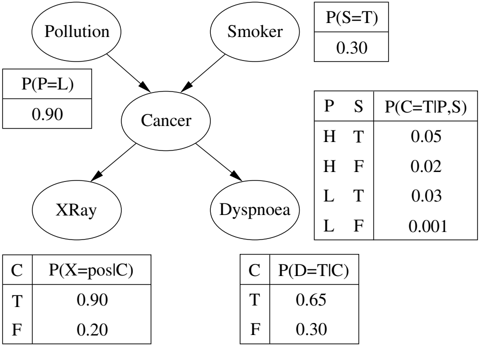
Example

Pollution-Smoking-Cancer

$$P(P, S, C) \quad = \quad P(P) \times P(S|P) \times P(C|P, S)$$

Outline

Bayesian networks



P(S=T)
0.30

P(P=L)
0.90

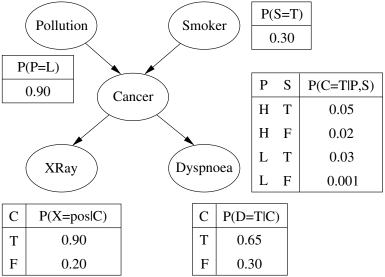
P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

C	P(X=pos|C)
T	0.90
F	0.20

C	P(D=T|C)
T	0.65
F	0.30

- A Bayesian Network is a directed acyclic graph (DAG)
  - Random variables makes up the nodes
  - Directed links or arrows connects pairs of nodes representing the dependence between variables.
  - Each node has a conditional probability table that quantifies the effects the parents have on the node.
- Gives a concise specification of the joint probability distribution.

Structure terminology and layout



P(P=L)
0.90

P(S=T)
0.30

P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

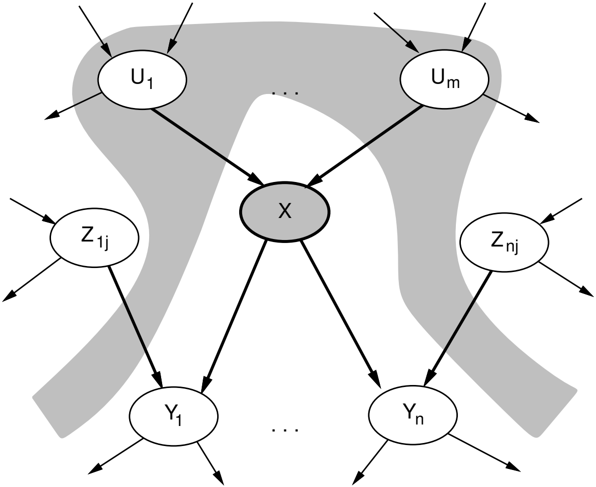
C	P(X=pos|C)
T	0.90
F	0.20

C	P(D=T|C)
T	0.65
F	0.30

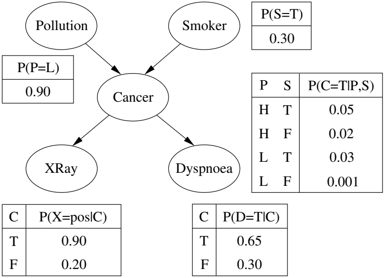
- Family metaphor: *Parent* $\Rightarrow$ *Child*
  Ancestor $\Rightarrow \ldots \Rightarrow$ *Descendant*
- Tree analogy:
  - root node: no parents
  - leaf node: no children
  - intermediate node: non-leaf, non-root
- Layout convention: root notes at top, leaf nodes at bottom, arcs point down the page.

Local semantics

Local semantics: each node is conditionally independent of its nondescendants given its parents.



Conditional probability tables



P(S=T)
0.30

P(P=L)
0.90

P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

C	P(X=pos|C)
T	0.90
F	0.20

C	P(D=T|C)
T	0.65
F	0.30

- One *conditional probability table (CPT)* for each node.
- Each row contains the conditional probability of every node value for a combination of its parents' values nodes.
- Each row sums to 1.
- A table for a Boolean var with n Boolean parents contain 2^{n+1} probabilities.
- A node with no parents has one row (the prior probabilities)

Bayesian network: A compact representation of joint probabilities

- Bayesian Network implies that the probability of a node is only conditional dependence of its parents

$$
\begin{aligned}
P(X_i|X_1,\ldots,X_{i-1}) &= P(X_i|Parents(X_i)) \\
P(X_1,\ldots,X_n) &= \Pi_{i=1,\ldots n} P(X_i|Parents(X_i))
\end{aligned}
$$

- Bayesian network regulates an ordering of variables
  - $\langle X_1, X_2, \ldots, X_N \rangle$
  - $Parents(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$
- Factoraization of joint probability with Bayesian network

$$
\begin{aligned}
P(X_1,\ldots,X_n) &= P(X_1) \times \ldots \times P(X_n|X_1,\ldots,X_{n-1}) \\
&= \Pi_{i=1,\ldots n} P(X_i|Parents(X_i))
\end{aligned}
$$

Example

Example

$$
\begin{aligned}
P(S = F, P = low, C = T, D = T, X = pos)& \\
= \quad & P(S = F) \\
& \times P(P = low | S = F) \\
& \times P(C = T | P = low, S = F) \\
& \times P(D = T | C = T, P = low, S = F) \\
& \times P(X = pos | D = T, C = T, P = low, S = F) \\
= \quad & P(S = F) \\
& \times P(P = low) \\
& \times P(C = T | P = low, S = F) \\
& \times P(D = T | C = T) \\
& \times P(X = pos | C = T)
\end{aligned}
$$

Pearl's network construction algorithm

1. Choose the set of relevant variables $\{X_i\}$ that describe the domain.
2. Choose an ordering for the variables, $\langle X_1, \ldots, X_n \rangle$.
3. While there are variables left:
    1. Add the next variable X_i to the network.
    2. Add arcs to the X_i nodes from some minimal set of nodes already in the net, $Parents(X_i)$, such that the following conditional independence property is satisfied:

       $$P(X_i | X_1', \ldots, X_m') = P(X_i | Parents(X_i))$$

       where $X_1', \ldots, X_m'$ are all the variables preceding X_i, including $Parents(X_i)$.
    3. Define the CPT for X_i

Outline

The query

$$
\begin{aligned}
&P(X|\mathbf{E}) \\
&= \alpha\Sigma_{U_1,\ldots,U_m}\{ \\
&\quad P(U_1,\mathbf{E}_{U_1\backslash X},\ldots,U_m,\mathbf{E}_{U_m\backslash X},X,\mathbf{E}_{Y_1\backslash X},\ldots,\mathbf{E}_{Y_n\backslash X})\} \\
&= \alpha\Sigma_{U_1,\ldots,U_m}\{P(X|U_1,U_2,\ldots,U_m) \\
&\quad \Pi_{i=1,\ldots,m}P(\mathbf{E}_{U_i\backslash X})\cdot\Pi_{i=1,\ldots,m}P(U_i|\mathbf{E}_{U_i\backslash X})\cdot\Pi_{k=1,\ldots,n}P(\mathbf{E}_{Y_k\backslash X}|X)\} \\
&= \alpha\cdot\Pi_{i=1,\ldots,m}P(\mathbf{E}_{U_i\backslash X})\cdot\Sigma_{U_1,\ldots,U_m}\{P(X|U_1,U_2,\ldots,U_m) \\
&\quad \Pi_{i=1,\ldots,m}P(U_i|\mathbf{E}_{U_i\backslash X})\cdot\Pi_{k=1,\ldots,n}P(\mathbf{E}_{Y_k\backslash X}|X)\}
\end{aligned}
$$

As $\Pi_{i=1,\ldots,m}P(\mathbf{E}_{U_i\backslash X})$ doesn't change when X takes different values, we can put it into the normalizing constant α:

$$
\begin{aligned}
P(X|\mathbf{E}) = \alpha\Sigma_{U_1,\ldots,U_m}\{&P(X|U_1,U_2,\ldots,U_m) \\
&\Pi_{i=1,\ldots,m}P(U_i|\mathbf{E}_{U_i\backslash X})\cdot\Pi_{k=1,\ldots,n}P(\mathbf{E}_{Y_k\backslash X}|X)\}
\end{aligned}
$$

Recursively compute $P(U_i|\mathbf{E}_{U_i\backslash X})$ and $P(\mathbf{E}_{Y_k}|X)$

$$
\begin{aligned}
P(X|\mathbf{E}) = \alpha\Sigma_{U_1,\ldots,U_m} \{ \\
& P(X|U_1, U_2, \ldots, U_m) \\
& \Pi_{i=1,\ldots,m} P(U_i|\mathbf{E}_{U_i\backslash X}) \\
& \Pi_{k=1,\ldots,n} P(\mathbf{E}_{Y_k\backslash X}|X)\}
\end{aligned}
$$

- $P(X|U_1, U_2, \ldots, U_m)$ can be looked from the CPTs
- Let $BN^{\pi}_{X}(U_i)$ be a sub-BN composed of U_i and all the nodes connecting to X through U_i.
  - Within $BN^{\pi}_{X}(U_i)$, $P(U_i|\mathbf{E}_{U_i\backslash X})$ can be computed recursively in the same manner
- Let $BN^{\lambda}_{Y_k}(X)$ be a sub-BN composed of Y_k and all the nodes connecting to X through Y_k.
  - Within $BN^{\lambda}_{Y_k}(X)$, $P(\mathbf{E}_{Y_k\backslash X}|X) = \Sigma_{Y_k} P(Y_k, \mathbf{E}_{Y_k\backslash X}|X)$ can be computed recursively in the same manner

A shorter equation: $\pi_X(U_i)$

Let

$$\begin{aligned} \pi_X(U_i) &= P(U_i|\mathbf{E}_{U_i\backslash X})) \\ \lambda_{Y_k}(X) &= P(\mathbf{E}_{Y_k\backslash X}|X) \end{aligned}$$

$$\begin{aligned} &\pi_X(U_i) \\ &\quad= P(U_i|\mathbf{E}_{U_i\backslash X})) \\ &\quad= \alpha\Sigma_{Parents(U_i)}\{ \\ &\qquad P(U_i|Parents(U_i))\cdot \\ &\qquad \Pi_{Z_j\in Parents(U_i)}\pi_{U_i}(Z_j)\cdot \\ &\qquad \Pi_{Y_k\in Children(U_i)\backslash\{X\}}\lambda_{Y_k}(U_i)\} \end{aligned}$$

When U_i is one of the evidence input

- $\pi_X(U_i = u_{i,e}) = 1$ if $u_{i,e}$ is the evidence value entered
- $\pi_X(U_i = u_{i,e}) = 0$ if $u_{i,e}$ is not the evidence value entered

A shorter equation: $\lambda_{Y_k}(X)$ I

$$\begin{aligned}\pi_X(U_i) &= P(U_i|\mathbf{E}_{U_i\setminus X})) \\ \lambda_{Y_k}(X) &= P(\mathbf{E}_{Y_k\setminus X}|X)\end{aligned}$$

$$\begin{aligned}&\lambda_{Y_k}(X) \\ &= \Sigma_{Y_k} P(Y_k, \mathbf{E}_{Y_k\setminus X}|X) \\ &= \Sigma_{Y_k} \Sigma_{Parents(Y_k)\setminus\{X\}} \{ \\ &\quad P(Y_k|Parent(Y_k)) \cdot \\ &\quad \Pi_{Z_{k,j}\in Parents(Y_k)\setminus\{X\}} \pi_{Y_k}(Z_{k,j}) \cdot \\ &\quad \Pi_{W_{k,l}\in Children(Y_k)} \lambda_{W_{k,l}}(Y_k) \}\end{aligned}$$

A shorter equation: $\lambda_{Y_k}(X)$ II

When Y_k is one of the evidence input, let $y_{k,e}$ be the evidence value entered, the marginalization over Y_k is replaced by setting $Y_k = y_{k,e}$ in the calculation.

$$
\begin{aligned}
&\lambda_{Y_k}(X) \\
&\quad = \quad P(Y_k = y_{k,e}, \mathbf{E}_{Y_k \setminus X} | X) \\
&\quad = \quad \Sigma_{Parents(Y_k) \setminus \{X\}} \Big\{ \\
&\qquad\quad P(Y_k = y_{k,e} | Parent(Y_k)) \cdot \\
&\qquad\quad \Pi_{Z_{k,j} \in Parents(Y_k) \setminus \{X\}} \pi_{Y_k}(Z_{k,j}) \cdot \\
&\qquad\quad \Pi_{W_{k,l} \in Children(Y_k)} \lambda_{W_{k,l}}(Y_k) \Big\}
\end{aligned}
$$

Message passing (bottom-up instead of recursion) I

- Query decomposition

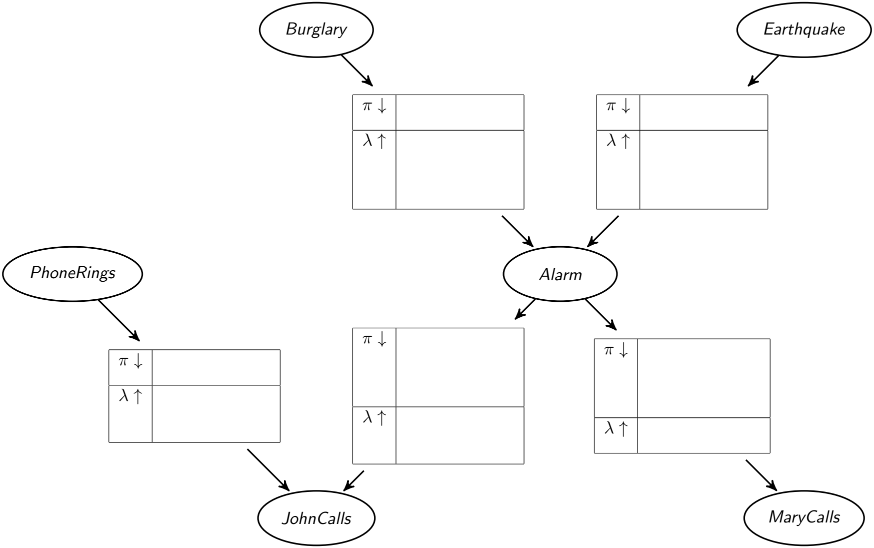
$$
\begin{aligned}
P(X|\mathbf{E}) = \alpha \Sigma_{U_1,\ldots,U_m} \{ \\
& P(X|U_1, U_2, \ldots, U_m) \cdot \\
& \Pi_{i=1,\ldots,m} \pi_{U_i}(X) \cdot \\
& \Pi_{k=1,\ldots,n} \lambda_{Y_k}(X) \}
\end{aligned}
$$

- Message passing:
  - Start with the nodes that don't need any messages to compute π
  - Start with the nodes that don't need any messages to compute λ
  - Once a node gets the π and λ messages required to compute its own messages sent to its children or parents, compute the messages and send them out. Notice that
    - ★ U_i send to its child X: $\pi_X(U_i)$ — computing $\pi_X(U_i)$ only requires messages from nodes in $BN^{\pi}_{X}(U_i)$
    - ★ Y_k send to its parent X: $\lambda_{Y_k}(X)$ — computing $\lambda_{Y_k}(X)$ only requires messages from nodes in $BN^{\lambda}_{Y_k}(X)$

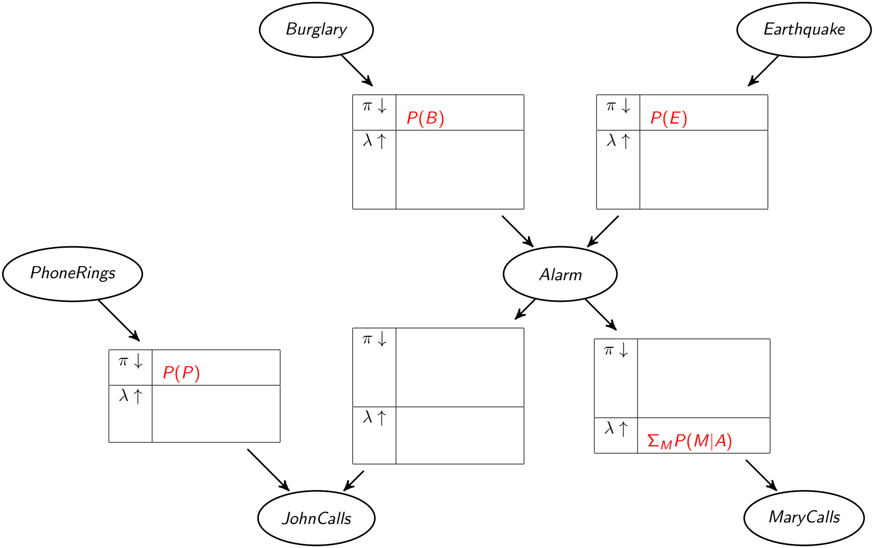
Message passing (bottom-up instead of recursion) II

- X computes query $P(X|\mathbf{E})$ using its conditional table $P(X|Parents(X))$, and the message $\pi_{U_i}(X)$ received from its parents, the message $\lambda_{Y_k}(X)$ received from its children
- For queries other than $P(X|\mathbf{E})$, X send messages
  - X send message $\pi_{Y_k}(X)$ to its child Y_k
  - X send message $\lambda_X(U_i)$ to its parent U_i

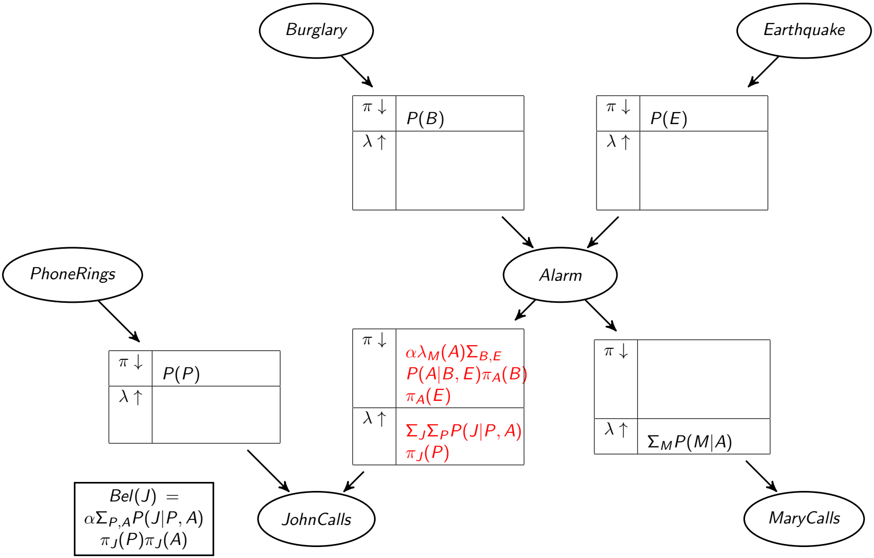
Message passing – No Evidence



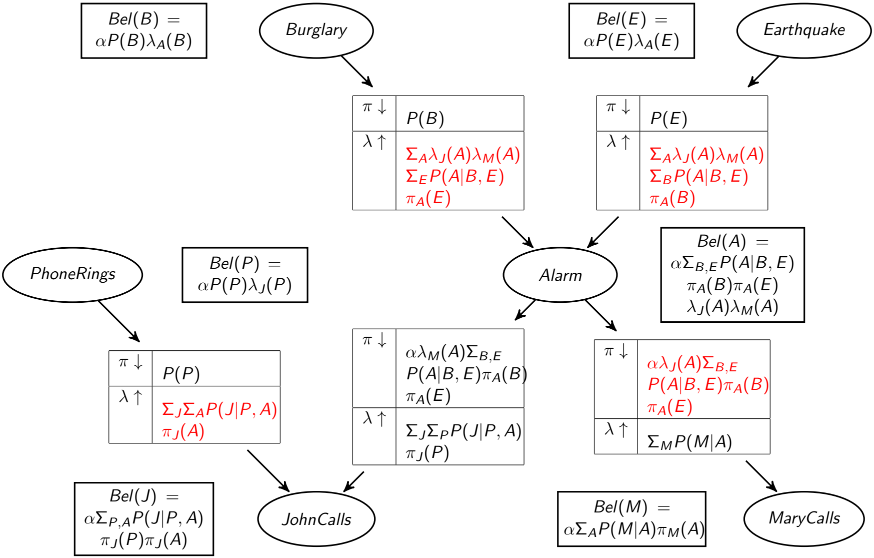
Message passing – No Evidence



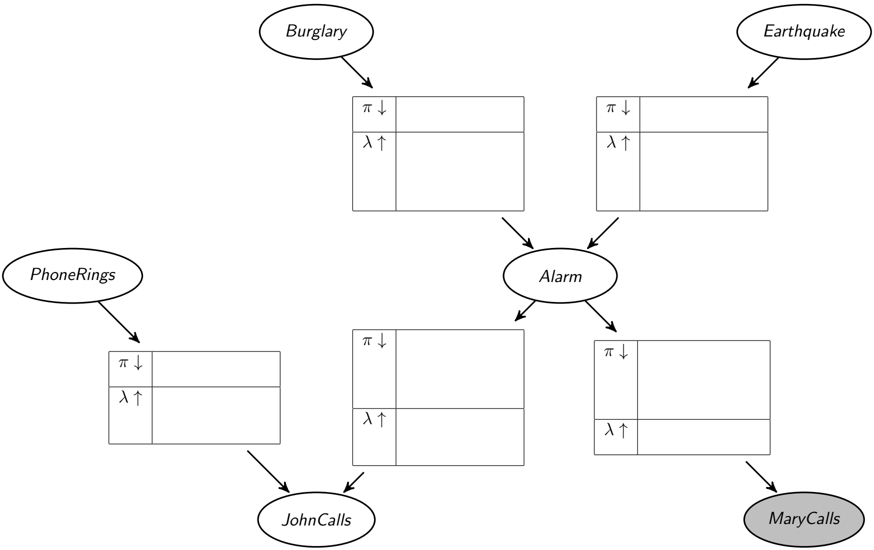
Message passing – No Evidence



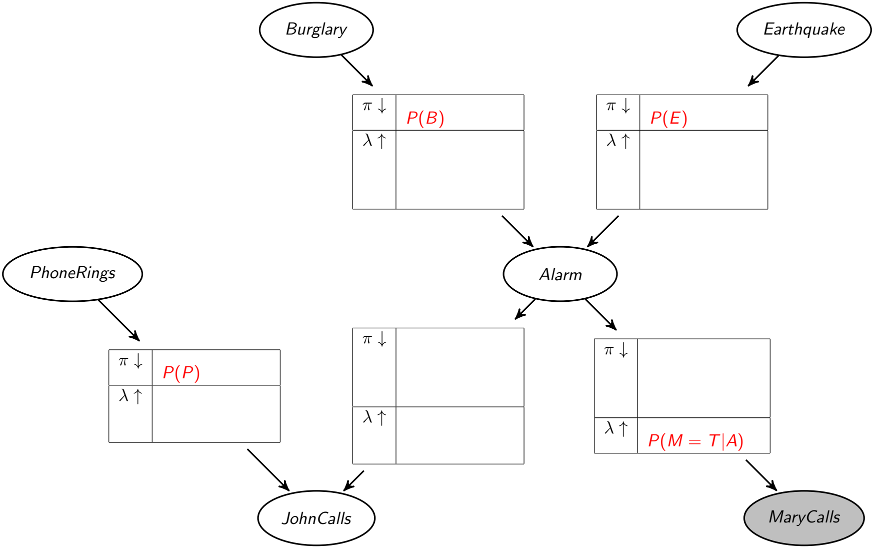
Message passing – No Evidence



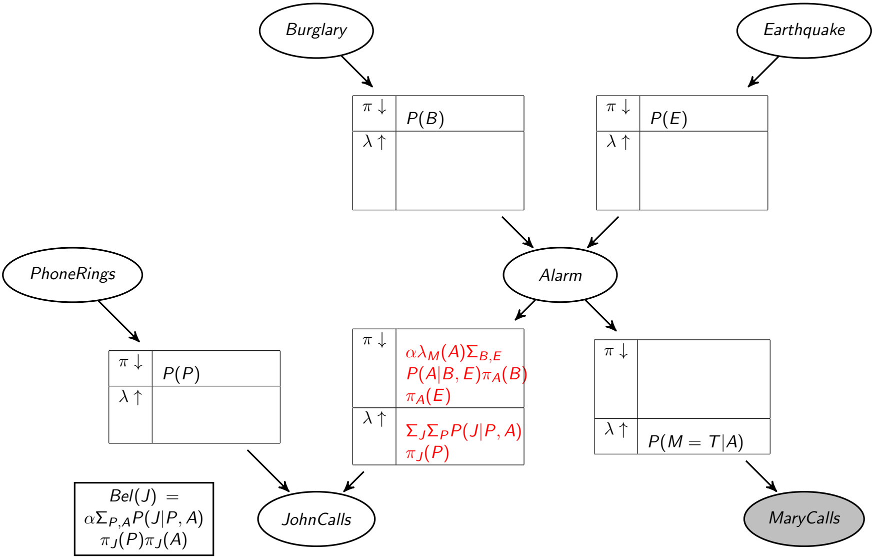
Message passing – Evidence Example



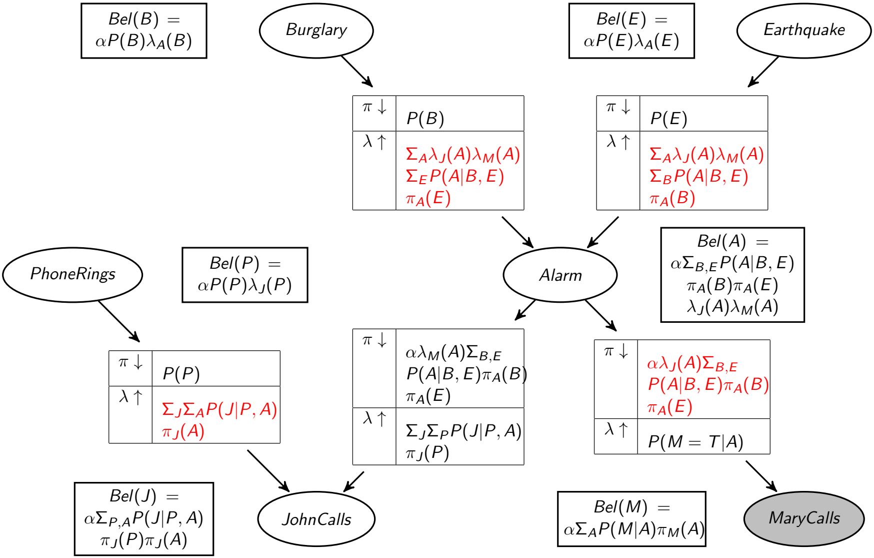
Message passing – Evidence Example



Message passing – Evidence Example

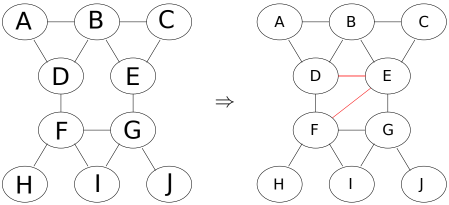


Message passing – Evidence Example



Outline

Triangulation ordering



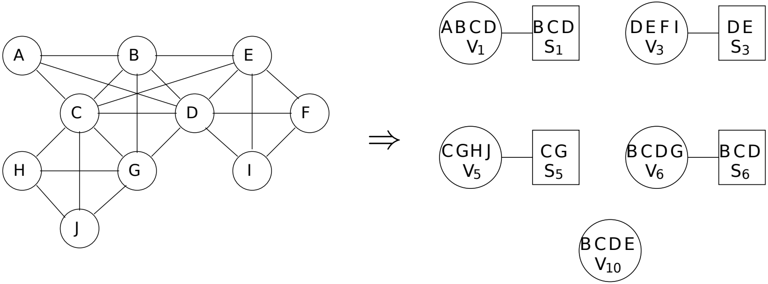
- Heuristics: Eliminate a node X with minimum number of family states, ie. $sz(fa(X))$
  - recall that $fa(X) = \{X\} \cup nb(X)$
  - $sz(V) = |Domain(V)|$
  - $sz(\{V_1, ..., V_n\}) = \Pi_i sz(V_i)$
- Let A, B, C, H, I, J have 2 states, D have 4 states, E have 5 states, F have 6 states, G have 7 states. Assume we have removed A, C, H, I, J already, then $sz(fa(B)) = 40$, $sz(fa(D)) = 48$, $sz(fa(E)) = 70$, $sz(fa(F)) = 168$ and $sz(fa(G)) = 210$.

Triangulation Algorithm

Given a graph G

1. For each node $X \in G$
   - If there is simplicial node X in G, eliminate X
2. Compute the family size of each node G, select a node X with minimum family size and eliminate X
3. If there are no nodes left in G then terminate; otherwise go to step 1

Transformation of a triangulated graph into a join tree I

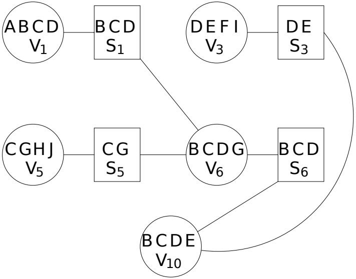


1. Establish an elimination sequence for the nodes in G and add fill-ins to G if neccessary (e.g. according to a triangulation algorithm)
2. $i \leftarrow 0$
3. For each node $X \in G$, if X is a simplicial node then $fa(X)$ is a clique
   - Remove nodes $Y_{x,1}, \ldots Y_{x,K}$ from $fa(X)$ with $nb(Y_{x,k}) \subseteq fa(X)$,

Transformation of a triangulated graph into a join tree II

- $i \leftarrow i + K$ (i is the number nodes removed so far)
- Construct a clique $V_i \leftarrow fa(X)$
- Construct a separator $S_i \leftarrow \{Y_{x,1}, \ldots Y_{x,K}\}$
- Connect V_i and S_i in the join tree T

4. For each separator S_i
   - Select a V_j such that $j > i$ and $S_i \subset V_j$
   - Connect S_i to V_j in the joint tree T



Outline

Other formalisms of uncertainty reasoning

- Default logic
- Certainty factor
- Dempster-Shafer theory
- Fuzzy set

Default logic

[Reiter, 1980]

- Default logic combines classical logic with domain-specific inference rules

$$\frac{Bird(x) : M\neg fly(X)}{fly(X)}$$

where Mp stands for "p is consistent" meaning that $\neg p$ can not be derived.

Definition

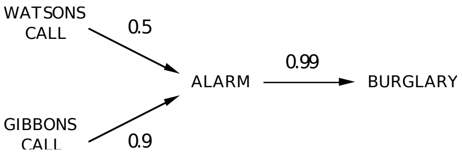
A default theory is a pair (W, D) where

- $W \subseteq L$ is a set of input knowledge represented in a language L
- D is a set of default rules of the form

$$\frac{\alpha : M\beta_1, \ldots, M\beta_n}{\gamma}$$

where $a, \beta_1, \ldots, \beta_n, \gamma \in L$, and $n \geq 0$.

The meaning of certainty factors



A rule with certainty factor: IF e THEN h, CF

- A ceraintfy factor (CF) represents a person's change in belief in the hypothesis (h) given the evidence (e)
- CF between 0 and 1 means that the person's belief in h given e increases
- CF between -1 and 0 means that the person's belief in h given e decreases
- Parallel-combination, serial-combination, conjunction-combination, disjunction-combination

Dempster-Shafer Theory

[Shafer, 1976]

- Frame of Discernment: Θ (set of possible events, one of them is true)
- Multi-Variable Frames:

$$D = \{x_1, \ldots, x_n\} \Rightarrow \Theta_D = \Theta_{x_1} \times \ldots \times \Theta_{x_n}$$

- Belief mass function: $m : 2^{\Theta_D} \rightarrow [0, 1]$ with $\Sigma_{A \subseteq \Theta_D} m(A) = 1$
- Belief: $Bel(A) = \Sigma_{B \subseteq A} m(A)$
- Focal Sets: $A \subseteq \Theta_D$ s.t. $m(A) \neq 0$
- Dempsters rule of combination:

$$m_{1,2}(A) = \frac{\Sigma_{B \cap C = A} m_1(B) \cdot m_2(C)}{\Sigma_{B \cap C \neq \emptyset} m_1(B) \cdot m_2(C)}$$

Fuzzy sets

[Zadeh, 1965]

- A set A is in terms of its characteristic function

$$\mu_A(x) : U \to [0, 1]$$

  A point x belongs to set A with possibility $_A(x)$.

- Union

$$\mu_{A \cup B}(x) = max(\mu_A(x), \mu_B(x))$$

- Intersection

$$\mu_{A \cap B}(x) = min(\mu_A(x), \mu_B(x))$$

- Complement

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

References I

Raymond Reiter.
A logic for default reasoning.
Artificial Intelligence, 13(1-2):81–132, 1980.

G. Shafer.
A mathematical theory of evidence.
Princeton university press, 1976.

L.A. Zadeh.
Fuzzy sets.
Information Control, 8:338–353, 1965.