Introduction to Artificial Intelligence

Lecture: Bayesian Networks

Outline

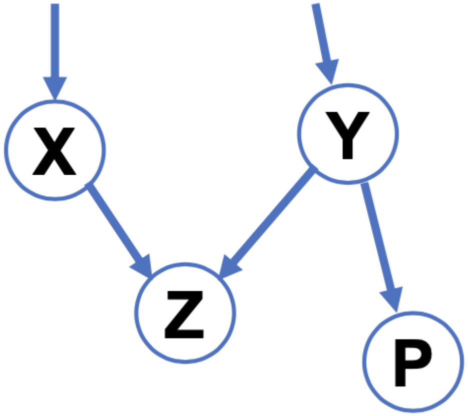
- Representing Knowledge in an Uncertain Domain
- Exact Inference in Bayesian Networks
- Constructing Bayesian Networks

Full joint probability distribution

- The full joint probability distribution (JPD) can answer any question about the domain.
- (Conditional) independence relationships among variables can greatly reduce the number of probabilities required.
- Bayesian networks can represent essentially, and in many cases very concisely, any full JPD.
  - Belief network, probabilistic network, causal network, knowledge map

Bayesian networks

- A Bayesian network is a directed graph in which each node is annotated with quantitative probability information.
- Each node presents a random discrete/continuous variable.
- A set of directed links or arrows connects pairs of nodes.
- Each node X_i has a probability distribution $\boldsymbol{P}(X_i \mid Parent(X_i))$ that quantifies the effect of the parents on the node.



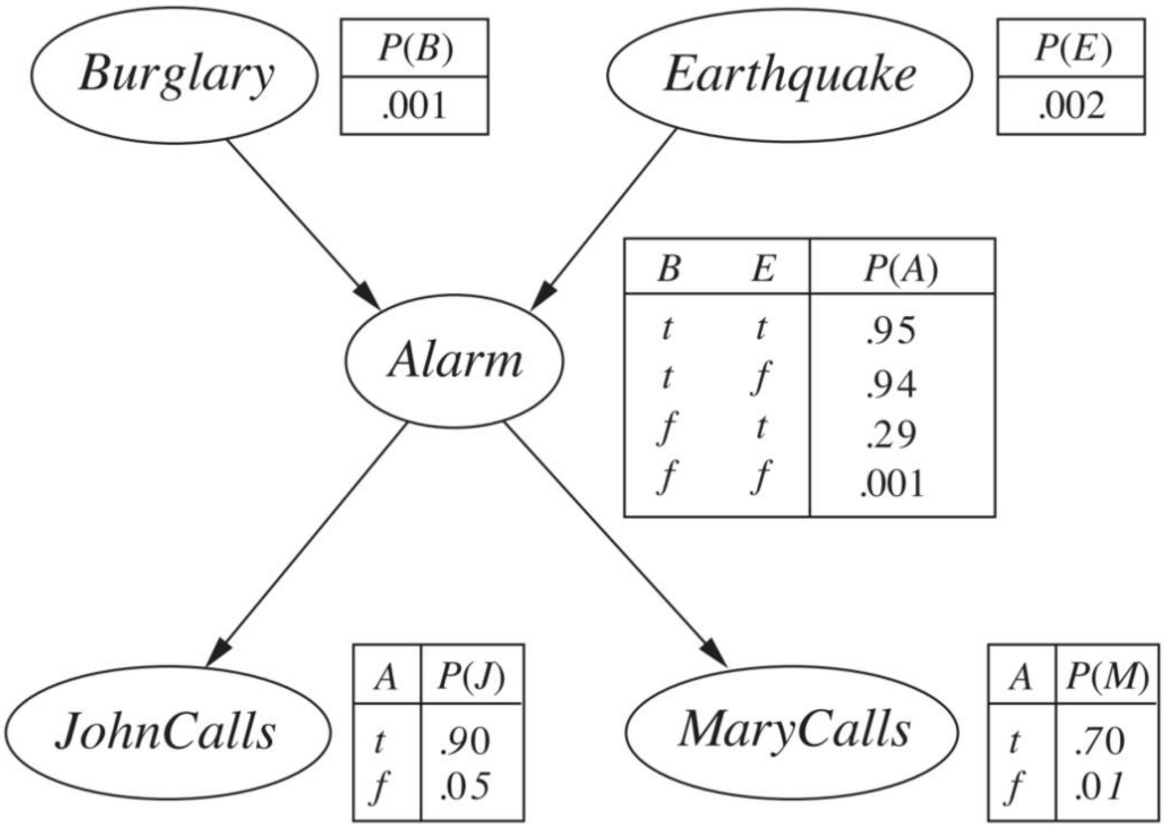
- DAG
- X and Y are parents of Z
- Y is the parent of P

Bayesian network topology

- The network topology defines the conditional independence relationships that hold in the domain.
  - An arrow means that X has a direct influence on Y, which suggests that causes should be parents of effects.
  - A domain expert decides what direct influences exist.
- Then, specify a conditional probability distribution for each variable, given its parents.

BN Example

- Burglary and earthquakes directly affect the probability of the alarm's going off
- Whether John and Mary call depends only on the alarm.



Conditional probability table (CPT)

- The probabilities summarizes a potentially infinite set of circumstances in which an event does (not) happen.
- In this way, a small agent can cope with a very large world, at least approximately.

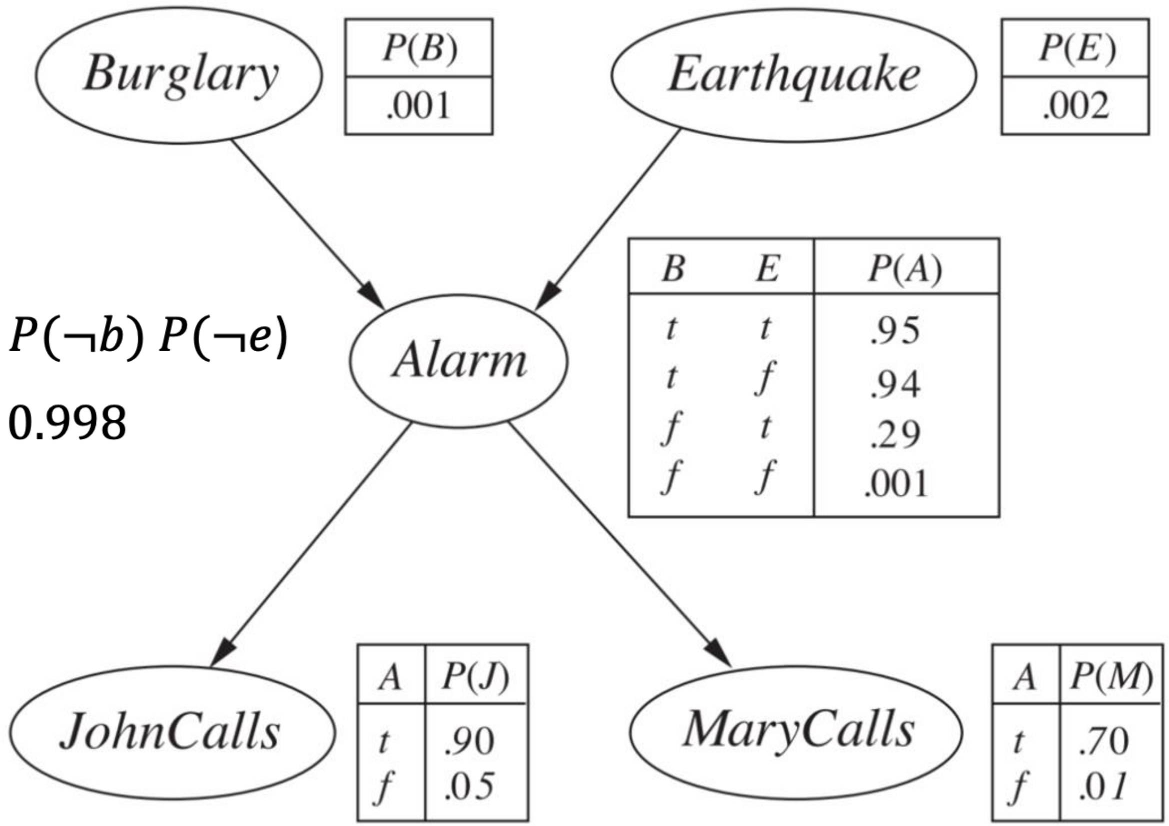
Represent the full joint distribution

- An entry in the joint distribution is the probability of a variable assignment, such as $\boldsymbol{P}(X_1 = x_1 \wedge \cdots \wedge X_n = x_n)$.

$$\boldsymbol{P}(x_1, \dots, x_n) = \prod_{i=1}^{n} \boldsymbol{P}(X_i \mid parent(X_i))$$

- Thus, it is the product of the appropriate elements of the CPTs in the Bayesian network.

Represent the full joint distribution



$P(j, m, a, \neg b, \neg e)$

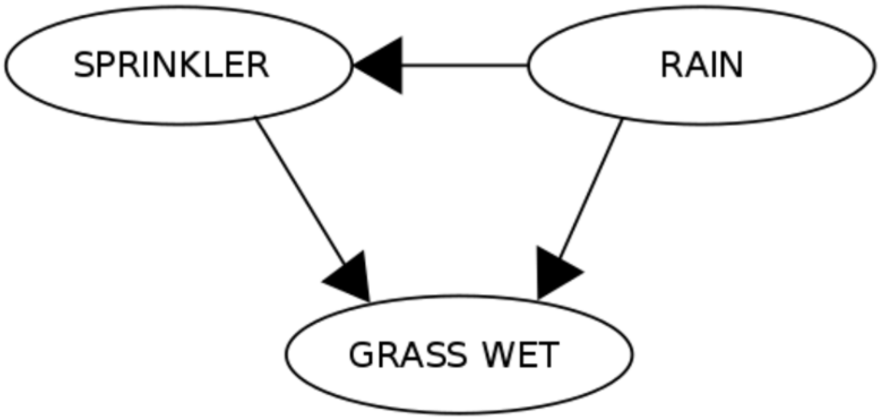
$= P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b \wedge \neg e)\, P(\neg b)\, P(\neg e)$

$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998$

$= 0.000628$

Represent the full joint distribution

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Notations

- X denotes the query variable.
- $\boldsymbol{E}$ denotes the set of evidence variables $E_1, \ldots, E_m$, and $\boldsymbol{e}$ is a particular observed event.
- Y denotes the non-evidence, non-query variables $Y_1, \ldots, Y_n$ (called the hidden variables).
- Thus, the complete set of variables is $\boldsymbol{X} = \{X\} \cup E \cup Y$.
- A typical query asks for the posterior probability $\boldsymbol{P}(X \mid \boldsymbol{e})$

Represent the full joint distribution

$$\boldsymbol{P}(R = T|G = T) = \frac{\boldsymbol{P}(G{=}T,R{=}T)}{\boldsymbol{P}(G{=}T)} = \frac{\sum_{S\in\{T,F\}} \boldsymbol{P}(G{=}T,S,R{=}T)}{\sum_{S,R\in\{T,F\}} \boldsymbol{P}(G{=}T,S,R)}$$

Using the expansion for the joint probability function $P(G, S, R)$ and the conditional probabilities from the CPTs stated in the diagram

$$\boldsymbol{P}(G = T, S = T, R = T) = \boldsymbol{P}(G = T \mid S = T, R = T)\,\boldsymbol{P}(S = T \mid R = T)P(R = T)$$

$$= 0.99 \times 0.01 \times 0.2 = 0.00198$$

The numerical results (subscripted by the associated variable values) are

$$\boldsymbol{P}(R = T \mid G = T) = \frac{0.00198_{TTT} + 0.1584_{TFT}}{0.00198_{TTT} + 0.288_{TTF} + 0.1584_{TFT} + 0.0_{TFF}}$$

$$= \frac{891}{2491} \approx 35.77\%$$

Inference by enumeration

- A query can be answered by computing sums of products of conditional probabilities from the Bayesian network.

$$\boldsymbol{P}(X \mid \boldsymbol{e}) = \alpha \boldsymbol{P}(X, \boldsymbol{e}) = \alpha \sum_{\boldsymbol{y}} \boldsymbol{P}(X, \boldsymbol{e}, \boldsymbol{y})$$

  - where α stands for the constant denominator term, which is usually simplified during calculation.

Inference by enumeration

- Consider the following query

$$P(Burglary \mid JohnCalls = true, MaryCalls = true)$$

- The hidden variables are $Earthquake$ and $Alarm$.
- Using initial letters for the variables, we have

$$P(B \mid j, m) = \alpha\, P(B, j, m) = \alpha \sum_{e} \sum_{a} P(B, j, m, e, a)$$

- For simplicity, we do this for $Burglary = true$.

$$P(b \mid j, m) = \alpha \sum_{e} \sum_{a} P(b)\, P(e)\, P(a \mid b, e) P(j \mid a)\, P(m \mid a)$$

- Complexity: $\boldsymbol{O(n2^n)}$ for a network with n Boolean variables

Inference by enumeration

function ENUMERATION-ASK(X, $\mathbf{e}$, bn) **returns** a distribution over X

    inputs: X, the query variable

        $\mathbf{e}$, observed values for variables $\mathbf{E}$

        bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* $\mathbf{Y}$ = *hidden variables* */

    $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty

    for each value x_i of X **do**

        $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{xi}$)

            where $\mathbf{e}_{xi}$ is $\mathbf{e}$ extended with $X = x_i$

    return NORMALIZE($\mathbf{Q}(X)$)

function ENUMERATE-ALL($vars$, $\mathbf{e}$) **returns** a real number

    if EMPTY?($vars$) **then return** 1.0

    $Y \leftarrow$ FIRST($vars$)

    if Y has value y in $\mathbf{e}$

        then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}$)

        else return $\Sigma_y\, P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)

            where $\mathbf{e}_y$ is $\mathbf{e}$ extended with $Y = y$

Construct a Bayesian network

- Scenario 1: Network structure known and all variables observable
  - Compute only the CPT entries
- Scenario 2: Network structure known while some variables hidden
  - Gradient descent (greedy hill-climbing) method, i.e., search for a solution along the steepest descent of a criterion function
- Scenario 3: Network structure unknown, all variables observable
  - Search through the model space to reconstruct network topology
- Scenario 4: Network structure unknown and all variables hidden
  - No good algorithms known for this purpose

Construct a Bayesian network

- The Chain Rule holds for any set of random variables.

$$\boldsymbol{P}(x_1, \dots, x_n) = \prod_{i=1}^{n} \boldsymbol{P}(x_i \,|x_{i-1}, \dots, x_1)$$

$$= P(x_n \,|x_{n-1}, \dots, x_1) P(x_{n-1}, \dots, x_1)$$

$$= P(x_n \,|x_{n-1}, \dots, x_1) P(x_2 \,|x_{n-2}, \dots, x_1) \dots P(x_2 \mid x_1) P(x_1)$$

- We generally assert that, for every variable X_i in the network

$$\boldsymbol{P}(X_i \,|X_{i-1}, \dots, X_1) = \boldsymbol{P}(X_i \mid Parent(X_i)) \; *$$

Construct a Bayesian network

- Each node must be conditionally independent of its other predecessors in the node ordering, given its parents.
- Nodes: Identify the set of variables required to model the domain and order them $\{X_1, ..., X_n\}$.
- Links: For i = 1 to n do
    - Choose, from $\{X_1, ..., X_{i-1}\}$, a minimal set of parents for X such that Equation (*) is satisfied.
    - For each parent insert a link from the parent to X_i.
    - CPTs: Write down the conditional probability table,

$$P(X_i | Parent(X_i))$$

- Intuitively, the parents of node X_i should contain all those nodes in $\{X_1, ..., X_{i-1}\}$ that directly influence X_i.

Construct a Bayesian network

- The network is guaranteed to be acyclic.
    - Each node is connected only to earlier nodes.
- Bayesian networks contain no redundant probability values.
    - If there is no redundancy, then there is no chance for inconsistency.
- It is impossible for the domain expert to create a Bayesian network that violates the axioms of probability.

Example

- You want to diagnose whether there is a fire in a building
- You can receive reports (possibly noisy) 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

- Start by choosing the random Boolean variables for this domain
- $Tampering$ (T) : the alarm has been tampered with
- $Fire$ (F): there is a fire
- $Alarm$ (A): there is an alarm
- $Smoke$ (S): there is smoke
- $Leaving$ (L): there are lots of people leaving the building
- $Report$ (R): the sensor reports that everyone are leaving the building.

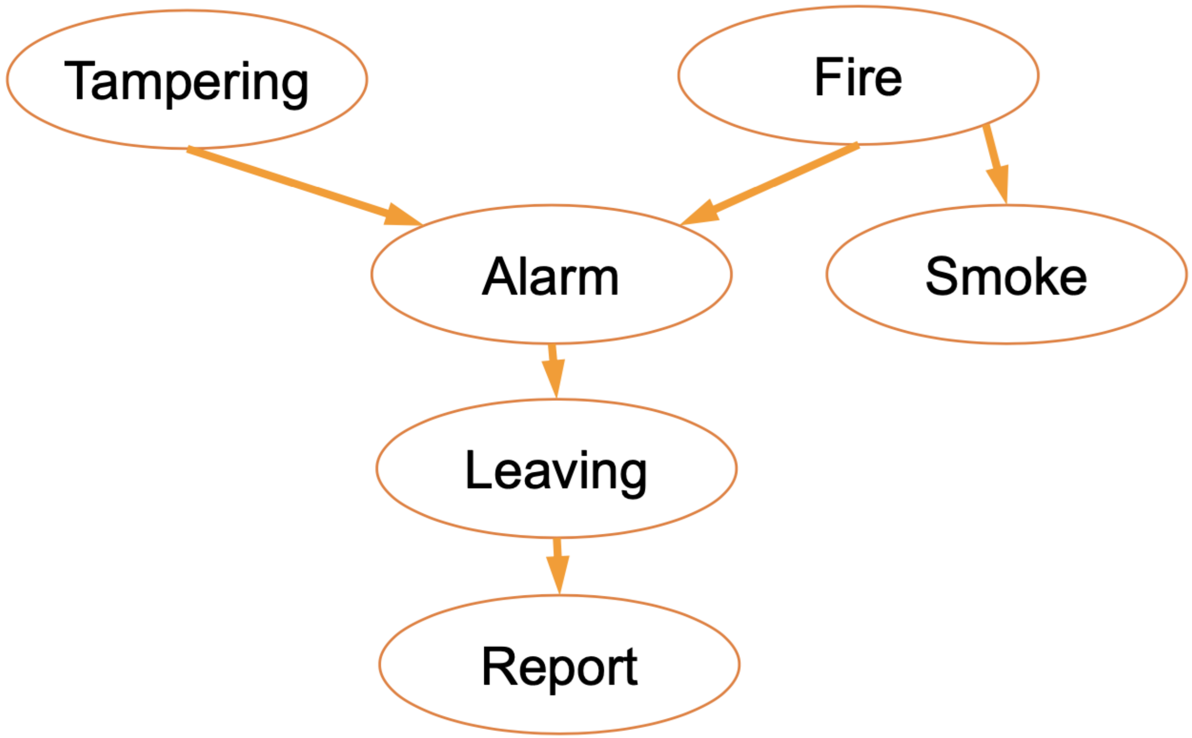
Example

- Define a total ordering of variables
  - Choose an order that follows the causal sequence of events
  - Fire (F) Tampering (T) Alarm (A) Smoke (S) Leaving (L) Report (R)
- Consider the following chain rule and use given clues to simplify it

$$\boldsymbol{P}(F,T,A,S,L,R) = P(F)\,P(T \mid F)\,P(A \mid F,T)\,P(S \mid F,T,A)$$

$$P(L \mid F,T,A,S)\,P(R \mid F,T,A,S,L)$$

Example

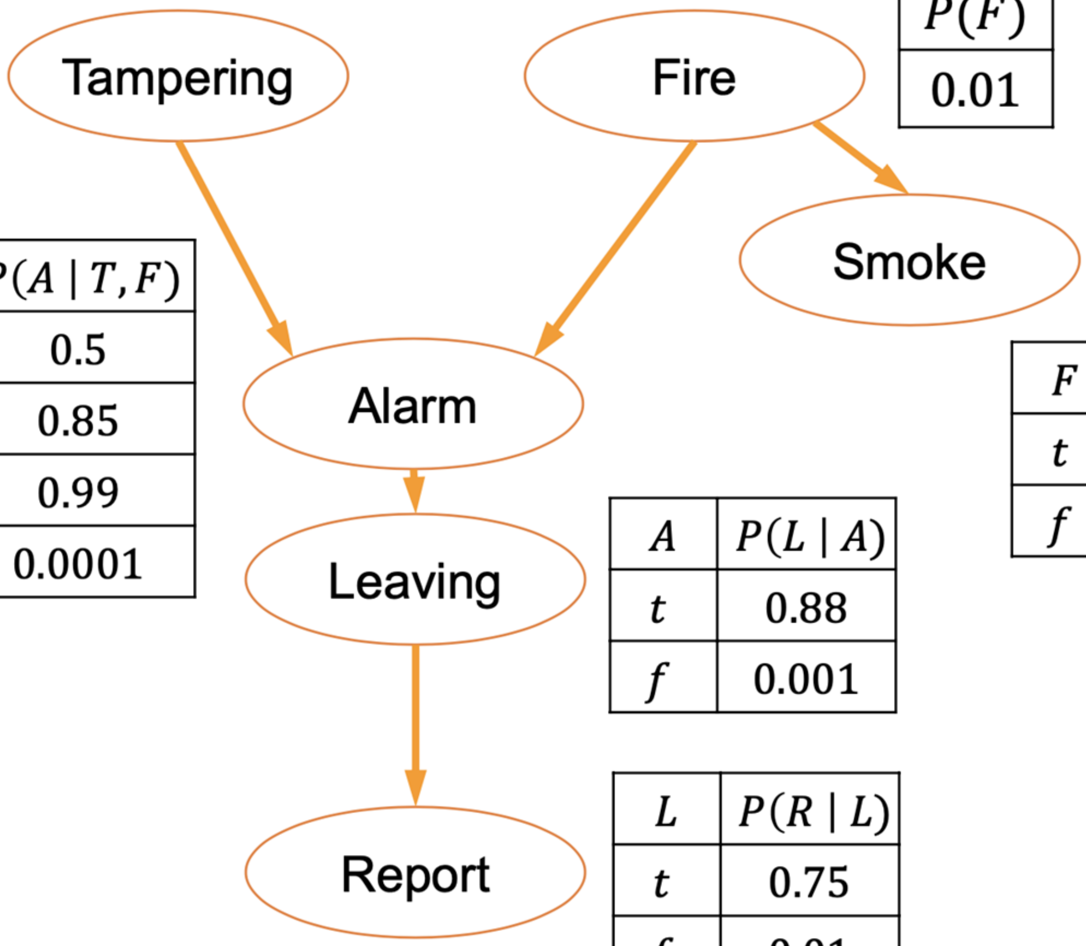


$$\boldsymbol{P}(F, T, A, S, L, R) = P(F)\, P(T)\, P(A \mid F, T)\, P(S \mid F)\, P(L \mid A)\, P(R \mid L)$$

Example

$P(T)$
0.02

$P(F)$
0.01



T	F	$P(A \mid T,F)$
t	t	0.5
t	f	0.85
f	t	0.99
f	f	0.0001

F	$P(S \mid F)$
t	0.9
f	0.01

A	$P(L \mid A)$
t	0.88
f	0.001

L	$P(R \mid L)$
t	0.75
f	0.01

Example

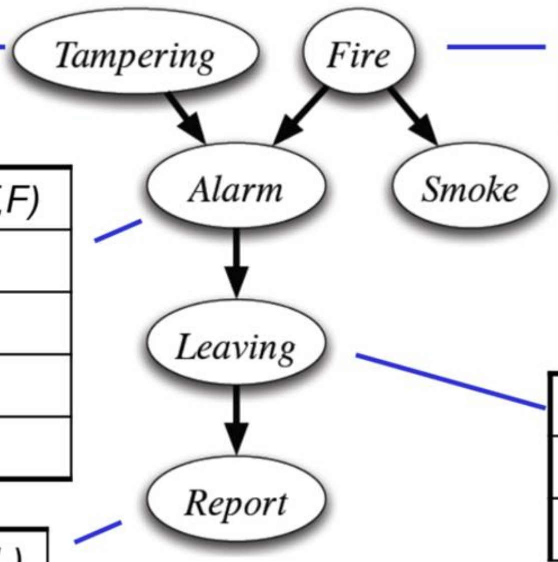
$$\boldsymbol{P}(T = t, F = f, A = t, S = f, L = t, R = t) = ?$$

Example

$$P(T = t, F = f, A = t, S = f, L = t, R = t) = ?$$

$$P(T = t) \times P(F = f) \times P(A = t \mid T = t, F = f) \times P(S = f \mid F = f)$$
$$\times P(L = t \mid A = t) \times P(R = t \mid L = t)$$

P(Tampering=t)
0.02



P(Fire=t)
0.01

Tampering T	Fire F	P(Alarm=t|T,F)
t	t	0.5
t	f	0.85
f	t	0.99
f	f	0.0001

Fire F	P(Smoke=t |F)
t	0.9
f	0.01

Alarm	P(Leaving=t|A)
t	0.88
f	0.001

Leaving	P(Report=t|L)
t	0.75
f	0.01

$= 0.02 \times (1 - 0.01) \times 0.85 \times (1 - 0.01) \times 0.88 \times 0.75$

$\mathbf{= 0.126}$

References

- Stuart Russell and Peter Norvig. 2009. Artificial Intelligence: A Modern Approach (3rd ed.). Prentice Hall Press, Upper Saddle River, NJ, USA.
- Lê Hoài Bắc, Tô Hoài Việt. 2014. Giáo trình Cơ sở Trí tuệ nhân tạo. Khoa Công nghệ Thông tin. Trường ĐH Khoa học Tự nhiên, ĐHQG-HCM.
- Nguyễn Ngọc Thảo, Nguyễn Hải Minh. 2020. Bài giảng Cơ sở Trí tuệ Nhân tạo. Khoa Công nghệ Thông tin. Trường ĐH Khoa học Tự nhiên, ĐHQG-HCM.