Uncertainty and Probabilistic Reasoning

1 The Need of Uncertainty Reasoning

- Humans easily reason with uncertain (not 100% correct) statements and knowledge.
- Classical logic needs 100% correct statements.
- Probability theory needs lots of data to compute conditional probabilities.
- Rules with certainty factors are oversimplified model.

2 Uncertainty Reasoning – Basic Terms

- *Random variable*: term in predicate calculus that may take a number of values (including continous variables).
- *Variable domain*: $dom(x)$ = the set of possible values of x.
- *Statement*: Boolean expression of variable assignments ($x_i = v_j$). For example, *A=(outlook = rain)* $\vee$ *(temperature = cool)* $\vee$ $\neg$ *(wind = strong)*.
- *Probability*: measure of confidence in a statement A. If $P(A) = 0 \rightarrow 100\%$ confidence that A is false, $P(A) = 1 \rightarrow 100\%$ that A is true. If $P(A)$ is between 0 and 1 this does not mean that A is "partially true", rather this reflects our uncertainty about the actual truth value of A.
- *Probability distribution*: the probabilities of each value of a random variable. If $dom(x) = \{v_1, ..., v_n\}$, then $\sum_{i=1}^{n} P(x = v_i) = 1$.
- *Prior probability*: probability without any additional information about a statement.
- *Conditional probability*: probability when the values of other random variables are known. For example, $P(temperature = cool|humidity = low)$.

3 Uncertainty Reasoning – Basic Formulas

- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- A, B – *independent* (knowing the one does not change the probabilty of the other), $P(A \wedge B) = P(A)P(B)$.
- *Disjoint events*: never happen together, $P(A \wedge B) = 0$.
- Conditional probability: $P(A|B) = \frac{P(A \wedge B)}{P(B)}$
- Bayes Theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

4 Probabilistic Model

- *Atomic event*: $(x_1 = v_1) \wedge (x_2 = v_2) \wedge ... \wedge (x_n = v_n)$, where $x_1, ..., x_n$ are random variables. Describes a state of the world.
- *Joint probability distribution* – n-dimensional table with m_i cells $(i = 1, ..., n)$ along each dimension, if x_i has m_i possible values. Each cell shows the probability of the corresponding atomic event. For example:

	toothache=true	*toothache=false*
cavity= true	0.04	0.06
cavity=false	0.01	0.89

- Given the joint probability distribution the probability of any even can be found. For example:

$$P(cavity = true) = 0.04 + 0.06 = 0.1$$

$$P(cavity = true \vee toothache = true) = 0.04 + 0.01 + 0.06 = 0.11$$

$$P(cavity = true|toothache = true) = \frac{P(cavity = true \wedge toothache = true)}{P(toothache = true)} =$$
$$0.04/(0.04 + 0.01) = 0.8$$

5 Probabilistic Inference

- Using the joint distribution
- Using Bayes theorem
  - Geven: e – a set of symptoms, $d_1, ..., d_n$ – diagnoses, $P(d_i)$ and $P(e|d_i)$, $i = 1, ..., n$ are also known.
  - Compute $P(d_i|e)$ for each i and find the most probable diagnosis given the symptoms e.

$$P(d_i|e) = \frac{P(d_i)P(e|d_i)}{P(e)}$$

  - $P(e)$ can be computed (if needed) by using the formulas:

$$\sum_{i=1}^{n} P(d_i|e) = \sum_{i=1}^{n} \frac{P(d_i)P(e|d_i)}{P(e)} = 1$$

$$P(e) = \sum_{i=1}^{n} P(d_i)P(e|d_i)$$

  - Computing $P(e|d_i)$ requires enumerating all possible combinations of values of the atomic symptoms contained in e (*exponential complexity*).
- Inference with the independence assumption ("Naive" Bayes).
  - Assuming that the atomic symptoms in $e = \{e_1, ..., e_k\}$ are independent:

$$P(e|d_i) = \prod_{j=1}^{k} P(e_j|d_i)$$

  - e_j $(j = 1, ..., k)$ can be easily computed from observations or determined by an expert.
  - Example:

Probability	*healthy*	*flu*	*alergy*
$P(d)$	0.9	0.05	0.05
$P(sneeze\|d)$	0.1	0.9	0.9
$P(cough\|d)$	0.1	0.8	0.7
$P(fever\|d)$	0.01	0.7	0.4

Assume the symptoms e are sneezing and cough, but no fever. Then:
$P(healthy|e) = (0.9)(0.1)(0.1)(0.99)/P(e) = 0.0089/P(e)$
$P(flu|e) = (0.05)(0.9)(0.8)(0.3)/P(e) = 0.01/P(e)$
$P(\mathbf{alergy}|e) = (0.05)(0.9)(0.7)(0.6)/P(e) = 0.019/P(e)$
Normalization: $P(e) = 0.0089 + 0.01 + 0.019 = 0.0379$
$P(healthy|e) = 0.23$
$P(flu|e) = 0.26$
$P(\mathbf{alergy}|e) = 0.50$

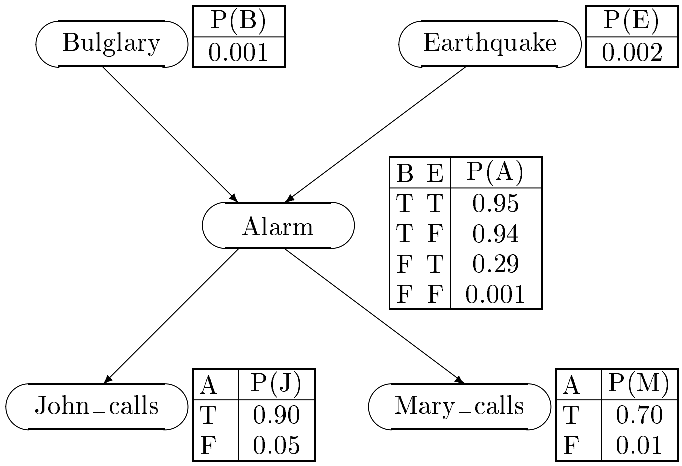

Figure 1: A Bayesian Net

6 Bayesian (Believe) Nets

- The independence assumption is too strong and almost never present in practice.
- Bayesian (Believe) Nets: using an acyclic graph to represent the dependencies between variables (a concise representation of the complete joint distribution).
- Each random variable is represented by a node.
- Links describe direct causal relations.
- For each node there is a conditional probability table (CPT). See the example in figure 1.
- Bayes Nets define implicitly the joint distribution of the variables in the nodes. For example, if $x_1, ..., x_n$ are random variables and $P(v_1, ..., v_n)$ is the joint probability that they get the values $v_1, ..., v_n$ respectively, then:

$$P(v_1, ..., v_n) = \prod_{i=1}^{n} P(v_i | Parents(x_i)),$$

where $P(v_i|Parents(x_i))$ is the conditional probability of $x_i = v_i$ given the values of the parent variables of x_i, $Parents(x_i)$. For example:

$$P(J, M, A, \neg B, \neg E) = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.000628$$

7 Inference with BN

Given the values of a subset of variables (evidence variables) find the probabilities of another subset of variables (query variables).

- Diagnosis: inference from evidence to cause. For example, $P(B|J) =?$
- Prediction: inference from cause to evidence. For example, $P(J|B) =?$
- Intercausal: among causes of an evidence. For example, $P(B|E) =?$
- Mixed, for example: $P(A|J \wedge \neg E) =?$

8 Algorithms for implementing the inference in BN

- Exact and approximate algorithms.
- Computing the full joint distribution (exact).
- Propagating evidence – computing the probabilities of all variables according to the evidence. Iterative modification of the probabilities by message passing between nodes: successors $\rightarrow$ parents and parents $\rightarrow$ successors, beginning with the evidence variables.
- Algorithms led by the query variables (do not compute all variables, approximate).
- In general the problem is exponential with the number of variables.
- For special cases of BN's there exists polynomial complexity algorithms.