CIS587 - Artificial Intellgence

Bayesian Networks

KB for medical diagnosis. Example.

We want to build a KB system for the **diagnosis of pneumonia**.

Problem description:

- **Disease:** pneumonia
- **Patient symptoms (findings, lab tests)**:
  - Fever, Cought, Paleness, WBC (white blood cells) count, Chest pain, etc.

Representation of a patient case:

- Statements that hold (are true) for that patient.

  E.g: Fever =*True*
  Cough =*False*
  WBCcount=*High*

Diagnostic task: we want to infer whether the patient suffers from the pneumonia or not given the symptoms

Uncertainty

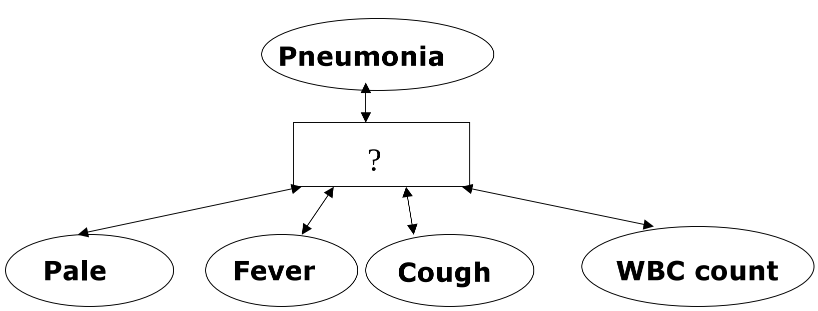
To make diagnostic inference possible we need to represent rules or axioms that relate symptoms and diagnosis

Problem: disease/symptoms relation is not deterministic (things may vary from patient to patient) – it is **uncertain**

- **Disease → Symptoms uncertainty**
  - A patient suffering from pneumonia may not have fever all the times, may or may not have a cough, white blood cell test can be in a normal range.
- **Symptoms → Disease uncertainty**
  - High fever is typical for many diseases (e.g. bacterial diseases) and does not point specifically to pneumonia
  - Fever, cough, paleness, high WBC count combined do not always point to pneumonia

Modeling the uncertainty.

- Relation between the disease and symptoms is not deterministic. **Key issues:**
- How to describe, represent the relations in the presence of uncertainty?
- How to manipulate such knowledge to make inferences?
  - **Humans can reason with uncertainty.**



Methods for representing uncertainty

KB systems based on propositional and first-order logic often represent uncertain statements, axioms of the domain in terms of

- rules with various **certainty factors**

Very popular in 70-80s (MYCIN)

If	1. The stain of the organism is gram-positive, and 2. The morphology of the organism is coccus, and 3. The growth conformation of the organism is chains
Then	**with certainty 0.7** the identity of the organism is streptococcus

Problems:

- Chaining of multiple inference rules (propagation of uncertainty)
- Combinations of rules with the same conclusions
- After some number of combinations results not intuitive.

Probability theory

a well-defined coherent theory for representing uncertainty and for reasoning with it

Representation:

Propositional statements – assignment of values to random variables

$Pneumonia = True \qquad WBCcount = high$

Probabilities over statements model the degree of belief in these statements

$P(Pneumonia = True) = 0.001$

$P(WBCcount = high) = 0.005$

$P(Pneumonia = True, Fever = True) = 0.0009$

$P(Pneumonia = False, WBCcount = normal, Cough = False) = 0.97$

Joint probability distribution

Joint probability distribution (for a set variables)

- Defines probabilities for all possible assignments to values of variables in the set

$\mathbf{P}(pneumonia, WBCcount)$ 2×3 table

$\mathbf{P}(Pneumonia)$

		WBCcount			
		high	*normal*	*low*	$\mathbf{P}(Pneumonia)$
Pneumonia	*True*	0.0008	0.0001	0.0001	0.001
	False	0.0042	0.9929	0.0019	0.999
$\mathbf{P}(WBCcount)$		0.005	0.993	0.002	

$\mathbf{P}(WBCcount)$

Marginalization - summing out variables

Conditional probability distribution

Conditional probability distribution:

- Probability distribution of A given (fixed B)

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

- Conditional probability is defined in terms of joint probabilities
- Joint probabilities can be expressed in terms of conditional probabilities

$$P(A,B) = P(A \mid B)P(B)$$

- Conditional probability – is useful for **diagnostic reasoning**
  - the effect of a symptoms (findings) on the disease

$$P(Pneumonia = True \mid Fever = True, WBCcount = high, Cough = True)$$

Modeling uncertainty with probabilities

- **Full joint distribution:** joint distribution over all random variables defining the domain
  - it is sufficient to represent the complete domain and to do any type of probabilistic reasoning

Problems:

- **Space complexity.** To store full joint distribution requires to remember $O(d^n)$ numbers.

  n – number of random variables, d – number of values
- **Inference complexity.** To compute some queries requires $O(d^n)$ steps.
- **Acquisition problem.** Who is going to define all of the probability entries?

Pneumonia example. Complexities.

- **Space complexity**.
  - Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBCcount (3: high, normal, low), paleness (2: T,F)
  - Number of assignments: 2*2*2*3*2=48
  - We need to define at least 47 probabilities.
- **Time complexity**.
  - Assume we need to compute the probability of Pneumonia=T from the full joint

$$P(Pneumonia = T) =$$
$$= \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k = h,n,l} \sum_{u \in T,F} P(Fever = i, Cough = j, WBCcount = k, Pale = u)$$

  - Sum over 2*2*3*2=24 combinations

Modeling uncertainty with probabilities

- Knowledge based system era (70s – early 80's)
  - Extensional non-probabilistic models
  - Probability techniques avoided because of space, time and acquisition bottlenecks in defining full joint distributions
  - Negative effect on the advancement of KB systems and AI in 80s in general

- Breakthrough (late 80s, beginning of 90s)
  - **Bayesian belief networks**
    - Give solutions to the space, acquisition bottlenecks
    - Partial solutions for time complexities

Bayesian belief networks (BBNs)

Bayesian belief networks.

- Represent the full joint distribution more compactly with a smaller number of parameters.
- Take advantage of conditional and marginal independences among components in the distribution

- **A and B are independent**

$$P(A,B) = P(A)P(B)$$

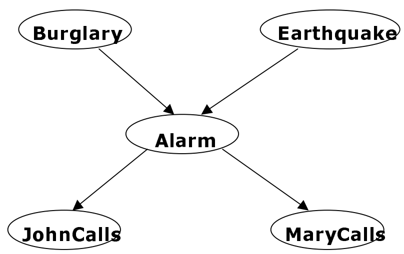
- **A and B are conditionally independent given C**

$$P(A,B \mid C) = P(A \mid C)P(B \mid C)$$
$$P(A \mid C,B) = P(A \mid C)$$

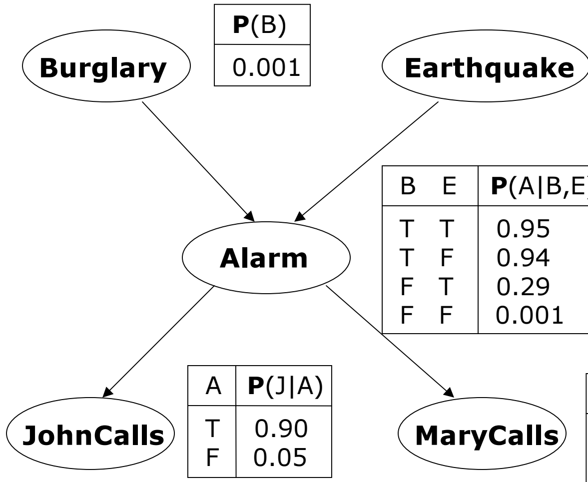
Alarm system example.

- Assume your house has an **alarm system** against **burglary**. You live in the seismically active area and the alarm system can get occasionally set off by an **earthquake**. You have two neighbors, **Mary** and **John**, who do not know each other. If they hear the alarm they call you, but this is not guaranteed.
- We want to represent the probability distribution of events:
  - Burglary, Earthquake, Alarm, Mary calls and John calls

Causal relations



Bayesian belief network example.



P(B)
0.001

P(E)
0.002

B	E	**P**(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

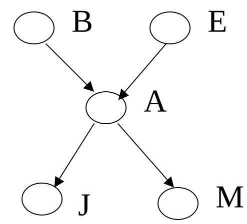
A	**P**(J|A)
T	0.90
F	0.05

A	**P**(M|A)
T	0.70
F	0.01

Bayesian belief networks (general)

Two components: $B = (S, \Theta_S)$

- **Directed acyclic graph**
  - Nodes correspond to random variables
  - (Missing) links encode independences



- **Parameters**
  - Local conditional probability distributions
  - for every variable-parent configuration

$$\mathbf{P}(X_i \mid pa(X_i))$$

Where:

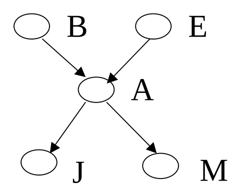
$pa(X_i)$ - stand for parents of X_i

B	E	**P**(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

Joint distribution in Bayesian networks

Full joint distribution is defined in terms of local conditional distributions (via the chain rule):

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$



Example:

Probability for one possible assignments of values:

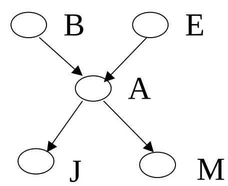
$$P(B = T, E = T, A = T, J = T, M = F) =$$

?

Full joint distribution in BBNs

Full joint distribution is defined in terms of local conditional distributions (via the chain rule):

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$



Example:

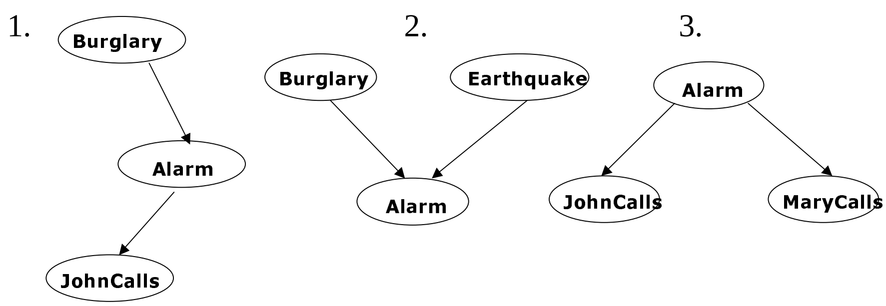
Probability for one possible assignments of values:

$$P(B=T, E=T, A=T, J=T, M=F) =$$

$$P(B=T)P(E=T)P(A=T \mid B=T, E=T)P(J=T \mid A=T)P(M=F \mid A=T)$$

Independences in BBNs

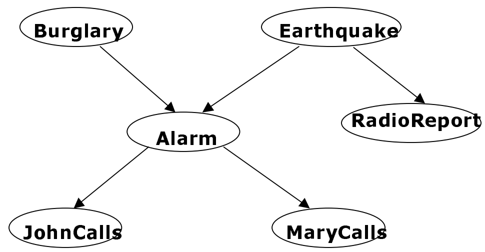
- **3 basic independence structures**



1. JohnCalls **is independent** of Burglary given Alarm
2. Burglary **is independent** of Earthquake (not knowing Alarm) Burglary and Earthquake **are not independent** given Alarm !!
3. MaryCalls **is independent** of JohnCalls given Alarm

Independences in BBNs

- Other dependences/independences in the network

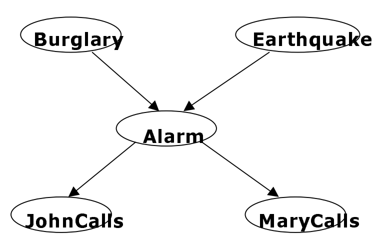


- Earthquake and Burglary are **not independent** given MaryCalls
- Burglary and MaryCalls **are not independent** (not knowing Alarm)
- Burglary and RadioReport **are independent** given Earthquake
- Burglary and RadioReport **are not independent** given MaryCalls

Parameter complexity problem

- In the BBN the full joint distribution is expressed as a product of conditionals (of smaller) complexity

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$



Parameters:

full joint: $2^5 = 32$

BBN: $2^3 + 2(2^2) + 2(2) = 20$

Parameters to be defined:

full joint: $2^5 - 1 = 31$

BBN: $2^2 + 2(2) + 2(1) = 10$

Model acquisition problem

The structure of the BBN typically reflects causal relations

- BBNs are also sometime referred to as **causal networks**
- Causal structure is very intuitive in many applications domain and it is relatively easy to obtain from the domain expert

Probability parameters of BBN correspond to conditional distributions relating a random variable and its parents only

- Their complexity much smaller than the full joint
- Easier to come up (estimate) the probabilities from expert or automatically by learning from data

Inference in Bayesian networks

- BBN models compactly the full joint distribution by taking advantage of existing independences between variables
- Simplifies the acquisition of a probabilistic model
- But we are interested in solving various **inference tasks**:
  - **Diagnosis**
  - **Prediction**

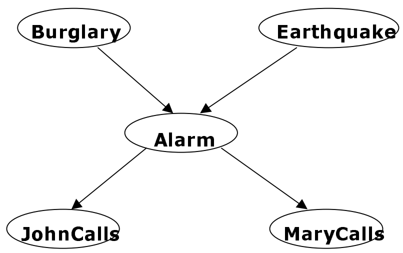
  Require to compute a variety of probabilistic queries:

  $$\mathbf{P}(Burglary \mid JohnCalls = T)$$
  $$\mathbf{P}(JohnCalls \mid Burglary = T)$$
  $$\mathbf{P}(Alarm)$$

- **Question:** Can we take advantage of independences to construct special algorithms and speeding up the inference?

Inference in Bayesian network

- **Bad news:**
  - Exact inference problem in BBNs is NP-hard (Cooper)
  - Approximate inference is NP-hard (Dagum, Luby)
- **But** very often we can achieve significant improvements
- Assume our Alarm network



- Assume we want to compute: $P(J = T)$

Inference in Bayesian networks

Approach 1. Blind approach.

- Sum over the joint distribution for all uninstantiated variables, express the joint distribution as a product of conditionals

$$P(J=T) = \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k \in T,F} \sum_{l \in T.F} P(B=i, E=j, A=k, J=T, M=l)$$

$$= \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k \in T,F} \sum_{l \in T,F} P(J=T \mid A=k) P(M=l \mid A=k) P(A=k \mid B=i, E=j) P(B=i) P(E=j)$$

Computational cost:

Number of additions: **16**

Number of products: 16*5=**80**

Inference in Bayesian networks

Approach 2. Interleave sums and products

- Combines sums and product in a smart way (multiplications by constants can be taken out of the sum)

$$P(J=T)=$$
$$=\sum_{i\in T,F}\sum_{j\in T,F}\sum_{k\in T,F}\sum_{l\in T,F}P(J=T\mid A=k)P(M=l\mid A=k)P(A=k|B=i,E=j)P(B=i)P(E=j)$$
$$=\sum_{i\in T,F}\sum_{k\in T.F}\sum_{l\in T,F}P(J=T\mid A=k)P(M=l\mid A=k)P(B=i)[\sum_{j\in T,F}P(A=k\mid B=i,E=j)P(E=j)]$$
$$=\sum_{k\in T,F}P(J=T\mid A=k)[\sum_{l\in T,F}P(M=l\mid A=k)][\sum_{i\in T,F}P(B=i)[\sum_{j\in T,F}P(A=k\mid B=i,E=j)P(E=j)]]$$

Computational cost:

Number of additions: 2*(4+2)=**12**

Number of products: 2*8+2*4+3*2= 2*(15)=**30**

Inference in Bayesian network

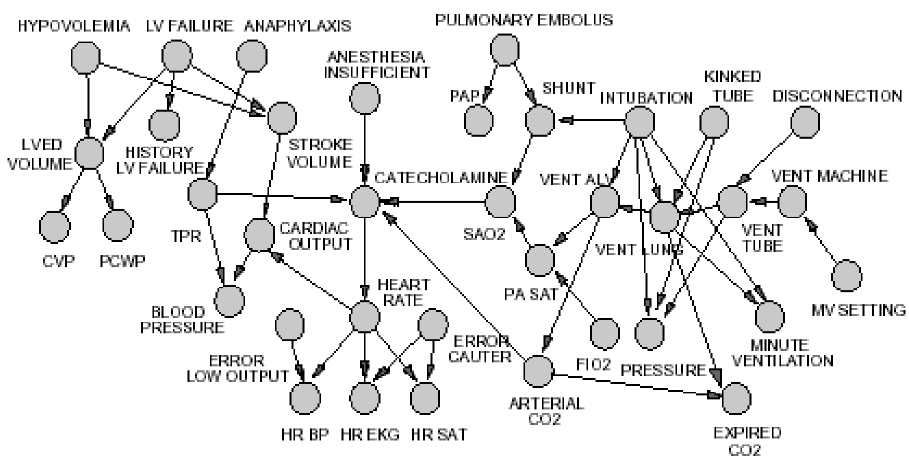
- **Exact inference algorithms:**
  - Symbolic inference (D'Ambrosio)
  - Pearl's message passing algorithm (Pearl)
  - Clustering and Join tree approach (Lauritzen, Spiegelhalter)
  - Arc reversal (Olmsted, Schachter)

- **Approximate inference algorithms:**
  - Monte Carlo methods:
    - Forward sampling, Likelihood sampling
  - Variational methods

BBNs built in practice

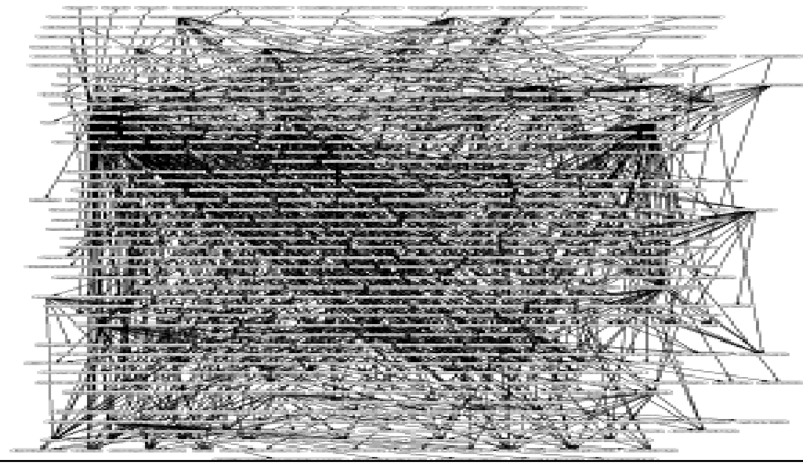
- **In various areas:**
  - Intelligent user interfaces (Microsoft)
  - Troubleshooting, diagnosis of a technical device
  - Medical diagnosis:
    - Pathfinder (Intellipath)
    - CPSC
    - Munin
    - QMR-DT
  - Collaborative filtering
  - Military applications
  - Insurance, credit applications

(ICU) Alarm network



CPCS

- **C**omputer-based **P**atient **C**ase **S**imulation system (CPCS-PM) developed by Parker and Miller (at University of Pittsburgh)
- 422 nodes and 867 arcs



QMR-DT

- **Medical diagnosis in internal medicine**

Bipartite network of disease/findings relations