Spezielle Themen der Künstlichen Intelligenz

8. Termin:

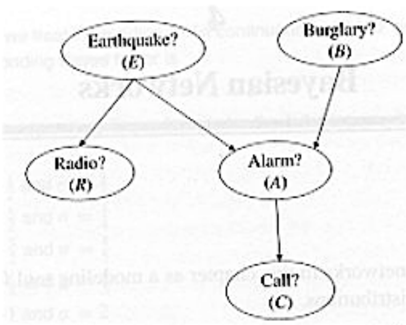
Bayesian Networks: Building & Inferencing

CITEC Sociable Agents

Bayesian network

Definition: A Bayesian network for variables Z is a pair (G, Θ) with

- Structure G: a directed acyclic graph with
  - a set of nodes, one per random variable
  - a set of edges representing *direct causal influence* between variables
- Parametrization Θ: a conditional probability table (CPT) for each variable
  - probability distribution for each node given its parents: *Pr(X$_i$ | Parents(X$_i$))* or *Pr(X$_i$)* if there are no parents
  - parameterizes the independence structure



Bayesian networks

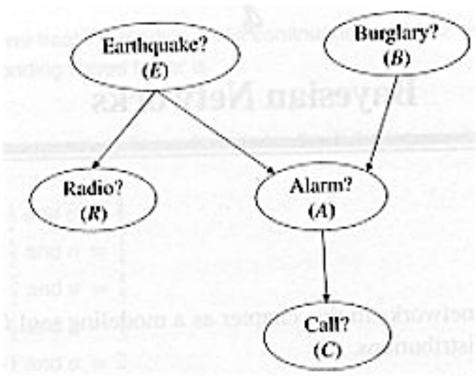
Each Bayesian network defines a set of cond. indep. statements:

$I(V, Parents(V), NonDescendants(V))$

- every variable is conditionally indep. of its nondescendants given its parents
  - Markovian assumptions: $Markov(G)$
- $Parents(V)$ are direct causes, $Descendants(V)$ are effects of V
  - given direct causes of *V*, beliefs in *V* are no longer influenced by any other variable, except possibly by its effects

Compact representation of a full joint distribution: network structure and parametrization are satisfied by *one and only one* prob. distribution given by the chain rule for Bayesian networks

$$Pr(\mathbf{z}) = \prod_{\theta_{x|\mathbf{u}} \sim \mathbf{z}} \theta_{x|\mathbf{u}} = \prod_{Pr(x|\mathbf{u}) \sim \mathbf{z}} Pr(x|\mathbf{u}), \text{ with } \mathbf{u} \text{ parents of } x$$



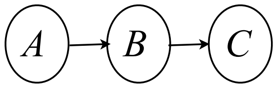
(chain rule of prob. calculus / repeated Bayesian cond.) $Pr(c, a, r, b, e)$

$$= Pr(c|a, r, b, e)Pr(a|r, b, e)Pr(r|b, e)Pr(b|e)Pr(e)$$

(decomposition / independencies)

$$= Pr(c|a)Pr(a|b, e)Pr(r|e)Pr(b)Pr(e)$$

$$= \theta_{c|a}\theta_{a|b,e}\theta_{r|e}\theta_b\theta_e$$



$$Pr(a,b,c) \quad = Pr(c|b,a)Pr(b|a)Pr(a) \quad = Pr(c|b)Pr(b|a)Pr(a)$$

requires 8 rows
(exponential)

a	b	c	Pr(.)
T	T	T	…
T	T	F	…
T	F	T	…
T	F	F	…
F	T	T	…
F	T	F	…
F	F	T	…
F	F	F	…

requires 2 rows
(+ normalization)

a	b	c	Pr(a)	Pr(b|a)	Pr(c|b)
T	T	T	…	…	…
F	F	T	…	…	…

Probabilistic independence

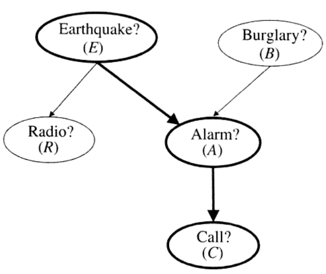
Bayesian network induces a belief state/prob distribution *Pr* satisfying the indep. relations $I(V, Parents(V), NonDescendants(V))$

Plus *further* implied independencies, can be derived graphically

- 3 types of „valves" with var *W*, either open or closed given var's **z**
- closed valves block a path in the graph, create d-separation

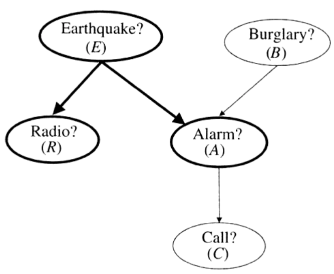
Sequential valve

closed if *W* in **z**



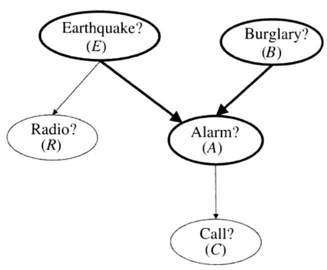
Divergent valve

closed if *W* in **z**



Convergent valve

closed if *W* not in **z**, nor any descendant of *W*



d-separation

__Definition:__

Variable sets ***X*** and ***Y*** are d-separated by ***Z*** iff every path between a node in ***X*** and a node in ***Y*** is blocked by ***Z*** (at least one valve on the path is closed given ***Z***).

$$dsep_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$$

__Theorem:__

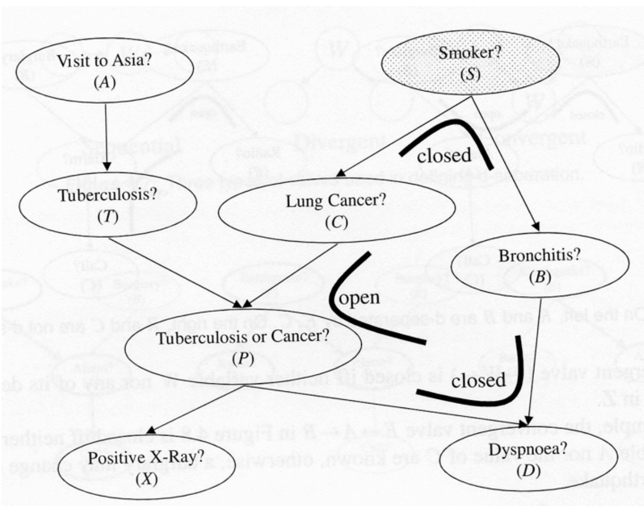
For every network graph G there is a parametrization Θ such that

$$I_{Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \leftrightarrow dsep_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$$

- *dsep* is always correct (sound)
- *dsep* is complete for a suitable parametrization (but not for every!)

d-separation

__Examples__:



Are *B* and *C* d-separeted by S?

Two paths:

- 1st one closed valve (C<-S->B) because S given
- 2nd one closed valve (B->D<-P) because D not given

➔ *B* and *C* are d-separated by *S*

➔ *B* and *C* are cond. indep. given *S*

Bayesian networks & independence

Terminology:

Graph *G* is an independence map (I-MAP) of a prob. distribution *Pr* over the same variables iff $dsep_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ only if $I_{Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

- d-separation in *G* implies independence in *Pr*
  (by definition for every *Pr* induced by the Bayesian network)
- minimal if *G* is no longer an I-MAP when removing any edge

Graph *G* is an dependence map (D-MAP) of a prob. distribution *Pr* over the same variables iff $I_{Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ only if $dsep_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

- lack of d-separation in *G* implies dependence in *Pr*
- *G* is not necessarily a D-MAP of any *Pr* induced by the Bayesian network, but of at least one with appropriate parameters Θ

If *G* is both I-MAP and D-MAP of *Pr*, then *G* is called perfect map (P-MAP)

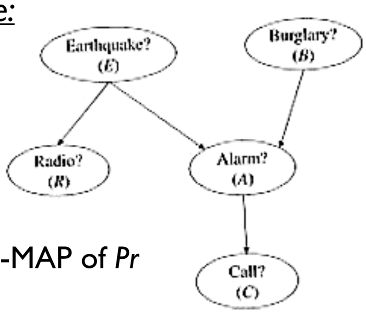
Bayesian networks & independence

Is there always a P-MAP for any distribution *Pr*?

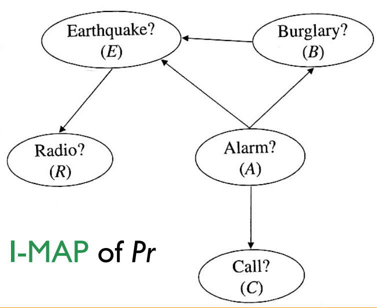
- **No**, there are distributions for which there are no P-MAPs!
  - reasonable since indep. captured by *G* satisfies properties (see above) not satisfied by any distribution

Given a distribution *Pr*, can we construct a minimal I-MAP?

- **Yes**, for variable X_i select parents **P** with $I_{Pr}(X_i, \mathbf{P}, \{X_1, ..., X_{i-1}\} \backslash \mathbf{P})$
- not unique, depending on order in which variables are considered
- *Example:*



supposed P-MAP of *Pr*



minimal I-MAP of *Pr*

not necessarily reasonable w.r.t causal relationships, yet with indep. relations!

Reasoning with Bayesian networks

How can a Bayesian network be used for answering queries about a domain?

There are (at least) four general types of queries one can pose:

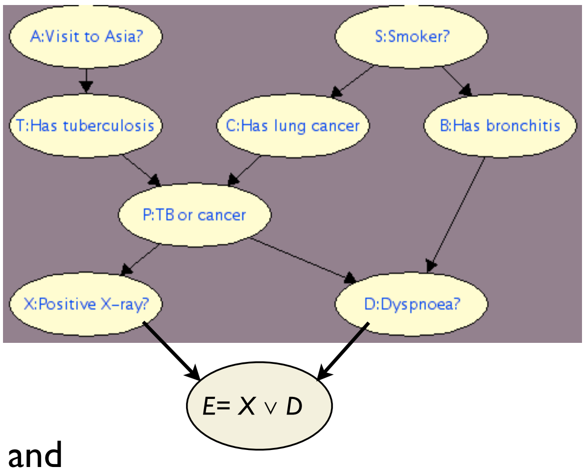
- probability of evidence:
  how likely is a variable instantiation **e** ➔ $Pr(\mathbf{e})=?$
- prior and posterior marginals: how probable is an instantiation of a limited set of variables ➔ $Pr(x_1,...,x_m)=?$ or $Pr(x_1,...,x_m|\mathbf{e})=?$
- most probable explanation (MPE): what is the most probable instantiation of all network var‘s given some evidence **e** ➔ **x** *with* $Pr(x_1,...,x_n|\mathbf{e})=max?$
- maximum a posteriori hypothesis (MAP): what is the most probable instantiation of a subset of var‘s given some evidence **e** ➔ **x** *with* $Pr(x_1,...,x_m|\mathbf{e})=max?$

Probability of evidence

Query: How likely is some variable instantiation **e** ➔ $Pr(\mathbf{e})=?$

Example: $Pr(X=yes, D=no)=?$

Example: $Pr(X=yes \vee D=yes)=?$



can be computed indirectly with the auxiliary-node technique:

- add node *E* with *X,D* as parents and $Pr(e|x,d)=1$ *iff* $e=1$ *and* $(d=1$ *or* $x=1)$
- possible when not too many evidence var‘s

Prior and posterior marginals

Query: How probable is an instantiation of a limited set of variables
➔ *Pr(x1,...,xm)=? or Pr(x1,...,xm|**e**)=?*

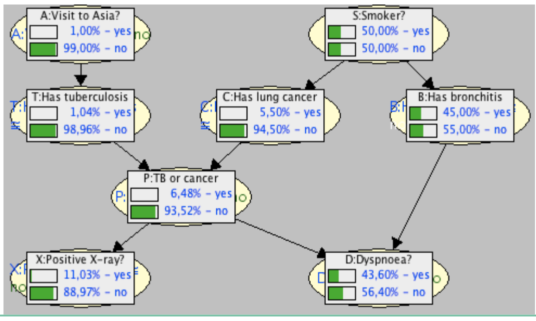
Definition: Given a joint distribution $Pr(x_1,...x_n)$ and a limited number *m* of variables,

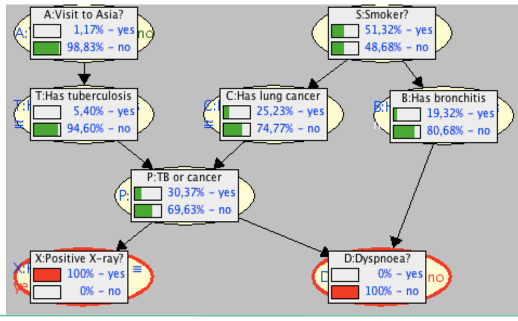
- prior marginal :

$$Pr(x_1, ..., x_m) = \sum_{x_{m+1},...x_n} Pr(x_1, ..., x_n)$$

- posterior marginal given **e** :

$$Pr(x_1, ..., x_m|\mathbf{e}) = \sum_{x_{m+1},...x_n} Pr(x_1, ..., x_n|\mathbf{e})$$

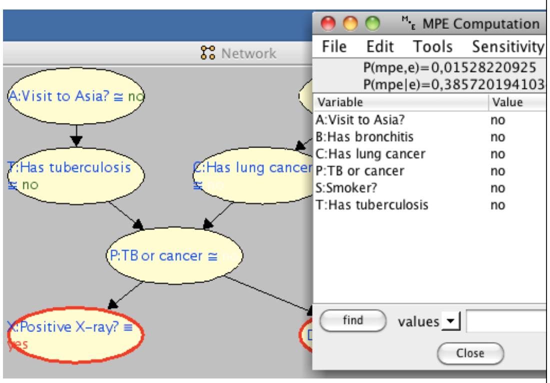




Most probable explanation (MPE)

Query: What is the most probable instantiation of all network var's given some evidence **e** ➔ ***x*** *with Pr(x1,...,xn|**e**)=max?*

Example: MPE for positive x-ray and not dyspnoea?



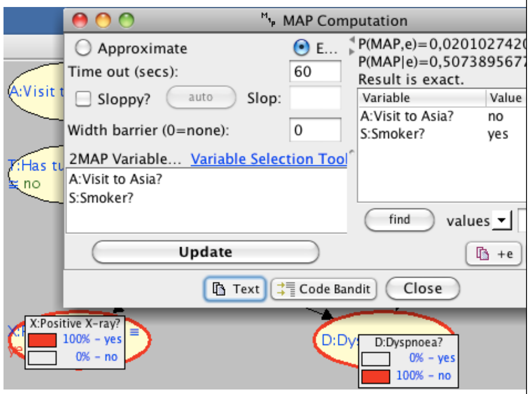
Note: cannot be computed directly from the maximal posterior marginals

- choosing x_i such that *Pr(*x_i*|**e**)=max* yields expl. *p* with *smoker=true and Pr(p|e)=20.03%* whereas *Pr(mpe|e)=38.57%*

Maximum a posteriori hypothesis (MAP)

Query: What is the most probable instantiation of a subset of var's $\mathbf{M}=X_1,...,X_m$ given some evidence $\mathbf{e} \rightarrow \mathbf{m}$ *with* $Pr(\mathbf{m}|\mathbf{e})=max$?

- MPE is a special case of MAP, easier to compute algorithmically



Example: Given X=*yes*, D=*no*, what is the most probable instantiation of $\mathbf{M}$={A,S}?

Approximative method to find MAP:

- compute MPE and return values for MAP variables (projecting MPE on MAP var's)
- but, here, leads to A=*no*, S=*yes* with prob ~48%, while A=*no*, S=*no* is MAP with prob ~50%

Modeling with Bayesian networks

Using Bayesian networks for real-world problems requires two steps:

- constructing an appropriate Bayesian network
- solve the problems by applying one of the previous queries

How to construct a Bayesian network?

1. define network variables and their values
   - distinguish between *query*, *evidence*, and *intermediary* variables
   - query and evidence var's usually determined from problem statement, intermediary var's less obvious
2. define network structure
   - for each var X, answer the question: what set of var's are direct causes of X?
3. define network parameters (CPTs)
   - difficulty and objectivity depend on problem

Example I: diagnosis model from expert

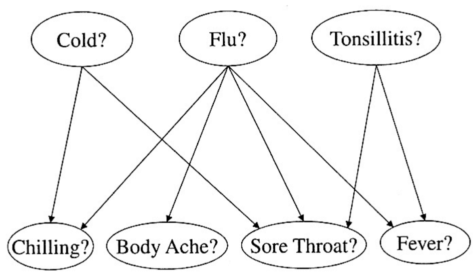
„Flu is an acute disease characterized by fever, body aches, and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a soar throat. Tonsillitis is an inflammation of the tonsils that leads to a soar throat and can be associated with fever.“

Variables:

- *query*: *flu, cold, tonsillitis*
- *evidence*: *chilling, body ache and pain, sore throat, fever*
- *intermediary*: -
- values: {true,false}

Structure?

Example I: diagnosis model from expert

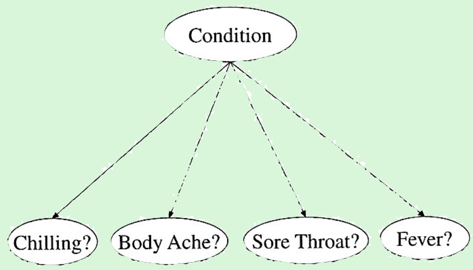


CPTs normally obtained from experts (subjective beliefs, empirical data)

- problem of parameter estimation
- *Example:* Given *N* patient records $\mathbf{d}_i$, find parametrization Θ such that

$$\prod_{i=1}^{N} Pr(\mathbf{d}_i) = max$$

Naive Bayes structure



- class variable *Condition* ∈ {*normal, cold, flu, tonsillitis*}
- attributes *Chilling, Body Ache, ...*
- single-fault assumption: only one cond. can hold at any time
- inconsistent with info given: given *Cond.=cold*, *Fever* and *Sore Throat* become independent

Example II: diagnosis model from expert

„Few weeks after inseminating a cow, we have three possible tests to confirm pregnancy. The first is scanning with a false positive of 1% and a false negative of 10%. The second is a blood test of progesterone with a false positive of 10% and a false negative of 30%. The third is a urine test of progesterone with false positive of 10% and a false negative of 20%. The prob. of a detectable progesterone level is 90% given pregnancy and 1% given no pregnany. The prob. that insemination will impregnate a cow is 87%.“

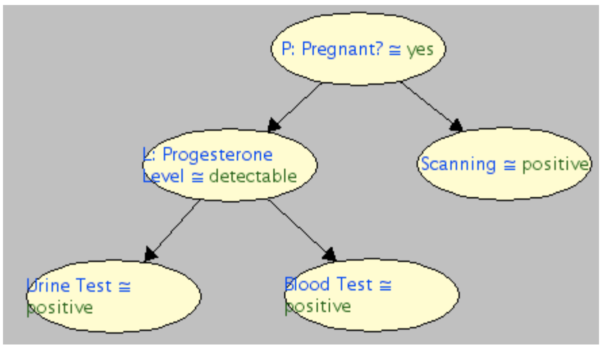
Goal: Build network to compute prob of pregnany given some test results

Variables:

- <u>*query*</u>: pregnancy? (*P*)
- <u>*evidence*</u>: scanning (*S*), blood test (*B*), urine test (*U*)
- <u>*intermediary*</u>: progesterone level (*L*)

Example II: diagnosis model from expert

Structure:

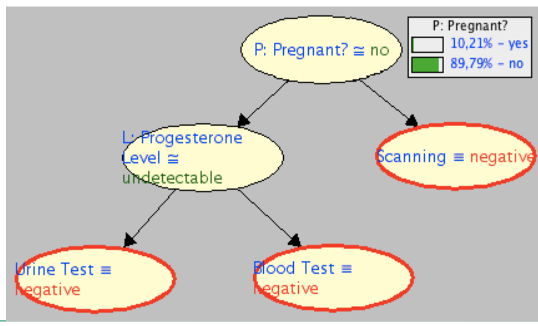


CPTs directly given by problem statement, e.g.

P	*L*	*P(l|p)*
yes	*undetect.*	*10*
no	*detectable*	*1*

<u>*Example:*</u> After insemination, all three tests are negative.

- *Pr(P|**e**)=?*
  Still 10,21%



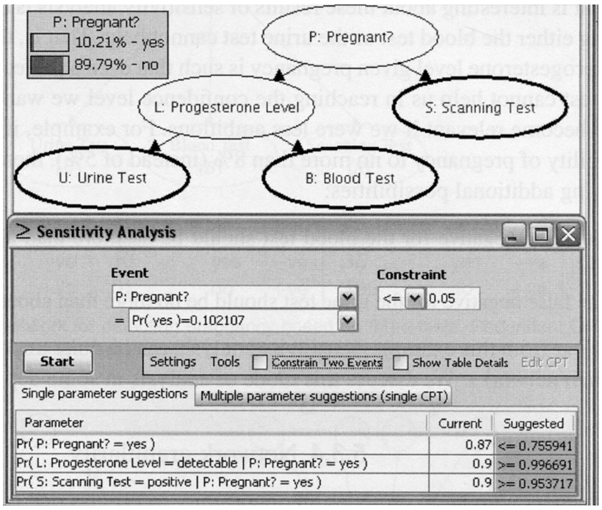
Example: sensitivity analysis

Q: What kind of a test is needed to get this error prob. down to ~5%?

- acceptable false positive/false negative?

„sensitivity analysis“:
which network parameters do we have to change, and how much, in oder to ensure that *Pr(P|L=neg.,B=neg.,U=neg.)≤5% ?*

- only improving the scanning test to a false negative of 4,63% helps



Further examples

See Darwiche (chap. 5) for further examples on

- diagnosis: model from design
- reliability analysis: model from design
  - depending on lifetime
- noisy channel coding
- commonsense knowledge
- how to deal with large CPTs

Next: main algorithms for drawing exact inferences

- by variable elimination / marginalization
- by factor elimination
- by (recursive) conditioning