Zpracování neurčitosti
– vybrané teoretické přístupy

14. listopadu 2023

Na úvod krátké historické shrnutí

- Pro modelování neurčitosti použijeme teorii pravděpodobnosti (podobně jako ve fyzice či ekonomii), ale bylo tomu tak vždy?
- První expertní systémy (kolem roku 1970) používaly čistě logický přístup bez práce s neurčitostí, což se ukázalo nepraktické.
- Další generace expertních systémů zkusila pravděpodobnostní techniky, ale ty měly problém se škálovatelností (efektivní odvozování na Bayesovských sítích ještě nebylo známé).
- Proto byly zkoušeny alternativní přístupy (1975 – 1988) pro modelování neurčitosti, ignorance a vágnosti – pravidlově orientované systémy, Dempster-Shaferova teorie, fuzzy logika.

Opakování – podmíněná pravděpodobnost

Podmíněná pravděpodobnost

$$P(x|y) = P(x \wedge y)/P(y)$$

Příklad: Jaká je pravděpodobnost, že na dvou kostkách hodíme pár, pokud víme, že na kostce 1 padla 5?

- $P(dvojice|Hod1 = 5) = \frac{1/36}{1/6} = 1/6$

Podmíněnou pravděpodobnost často zapisujeme v podobě součinového pravidla:

$$P(x \wedge y) = P(x|y) \times P(y) \qquad \mathbf{P}(X, Y) = \mathbf{P}(X|Y) \times \mathbf{P}(Y)$$

- $\mathbf{P}$ znamená, že výsledek je vektor čísel
- $\mathbf{P}(X|Y)$ dává hodnoty $P(X = x_i|Y = y_j)$ pro každý možný pár i, j
- $P(x, y)$ odpovídá $P(X = x \wedge Y = y)$
- $\mathbf{P}(X, Y)$ označuje pravděpodobnosti pro všechny kombinace hodnot X, Y

(Úplná) nezávislost: náhodné proměnné jsou na sobě nezávislé

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

Opakování – úplná sdružená distribuce, marginalizace

Pravděpodobnosti elementárních jevů můžeme popsat tabulkou,
tzv. úplnou sdruženou distribucí $\mathbf{P}(Tootchache, Cavity, Catch)$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Chceme-li znát pravděpodobnost nějakého tvrzení, sečteme pravděpodobnosti všech „světů", kde tvrzení platí (marginalizace)

$$\mathbf{P}(\mathbf{Y}) = \sum_{z \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, z)$$

- př. $P(tootchache = true) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$
- pozn. $\mathbf{Y}$ značí množinu proměnných vs. Y značí jednu proměnnou

Opakování – Bayesova věta

Víme, že platí

$$P(a \wedge b) = P(a|b)\,P(b) = P(b|a)\,P(a)$$

Můžeme odvodit Bayesovu větu v obecné podobě:

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

- α je normalizační konstanta, zajišťující, že položky v $\mathbf{P}(Y|X)$ jsou sumarizovány na 1
  - př. $\alpha\,\langle 0.2, 0.3\rangle = \langle 0.4, 0.6\rangle$
- $1/\mathbf{P}(X)$ vlastně dočasně zanedbáme
- podobně u podmíněné pravděpodobnosti

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

Bayesovské sítě

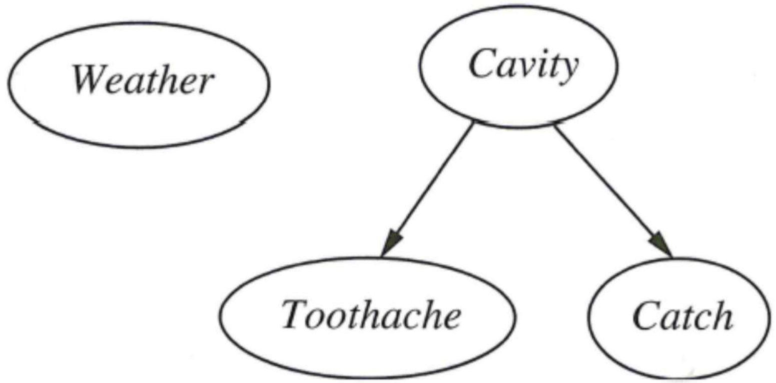
Bayesovská síť (Bayesian belief network) je orientovaný acyklický graf, jehož uzlům odpovídají náhodné proměnné a vazby reprezentují kauzální závislosti mezi těmito proměnnými.

Hrana $X \rightarrow Y$ znamená, že X kauzálně ovlivňuje Y (pozorování X poskytuje kauzální podporu Y, pozorování Y poskytuje diagnostickou podporu pro X). Bayesovská síť umožňuje provádět prediktivní i diagnostické inference.

Každému uzlu je přiřazena tabulka rozdělení pravděpodobnosti. Jestliže uzel nemá žádné předchůdce (rodiče), jedná se o nepodmíněnou pravděpodobnost, v opačném případě jde o podmíněnou pravděpodobnost.

Bayesovská síť

- Zachycuje závislosti mezi náhodnými proměnnými
- Orientovaný acyklický graf (DAG)
  - uzel odpovídá náhodné proměnné
  - předchůdci uzlu v grafu se nazývají rodiče
  - každý uzel má přiřazenu tabulku podmíněné pravděpodobnostní distribuce $\mathbf{P}(X|Parents(X))$



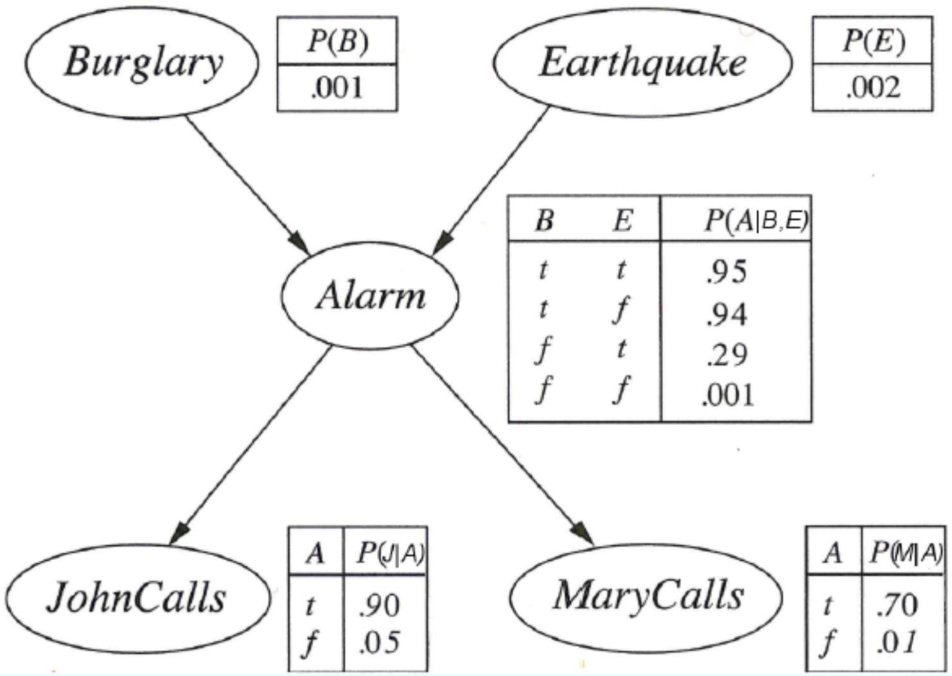
- Jiné názvy
  - belief network, probabilistic network, causal network (speciální případ BS), knowledge map

Bayesovská síť – modelová situace

- Máme v domě zabudovaný alarm, který se spustí při vloupání ale někdy také při zemětřesení
- Sousedi Mary a John slíbili, že vždy, když alarm uslyší, tak nám zavolají
  - John volá skoro vždy, když slyší alarm, ale někdy si ho splete s telefonním zvoněním
  - Mary poslouchá hlasitou hudbu a někdy alarm přeslechne
- Zajímá nás pravděpodobnost vloupání, pokud John i Mary volají
- Další předpoklady
  - sousedi přímo nevidí vloupání ani necítí zemětřesení
  - sousedi se nedomlouvají (volají nezávisle na sobě)

Bayesovská síť – příklad:

- Náhodné boolovské proměnné reprezentují možné události
  - některé události (zvonění telefonu, přelet letadla, vadu alarmu, ...) ingnorujeme
- Pravděpodobnostní tabulky reprezentují vztah podmíněné pravděpodobnosti
  - stačí reprezentovat hodnoty true



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J|A)
t	.90
f	.05

A	P(M|A)
t	.70
f	.01

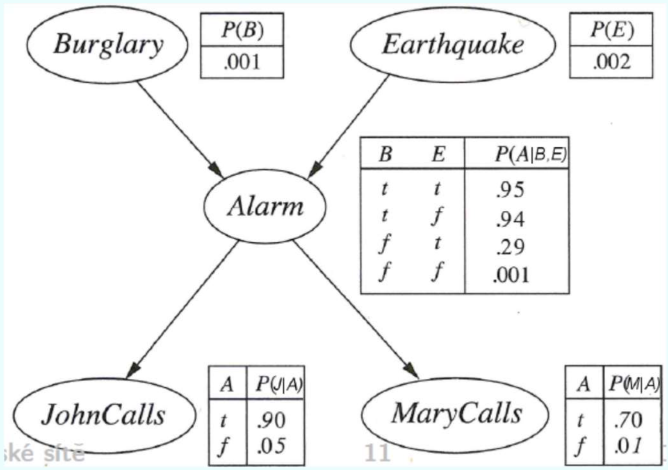
Sémantika sítě

- Bayesovská síť kompaktním způsobem reprezentuje úplnou sdruženou distribuci

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | parents(X_i))$$

$$\begin{aligned} \text{př. } P(j, m, a, \neg b, \neg e) &= P(j|a)\, P(m|a)\, P(a|\neg b \wedge \neg e)\, P(\neg b)\, P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628 \end{aligned}$$

- Zpětně lze ukázat, že tabulky $\mathbf{P}(X|Parents(X))$ jsou podmíněné pravděpodobnosti podle výše uvedené sdružené distribuce
- Protože úplnou sdruženou distribuci lze použít pro odpověď na libovolnou otázku v dané doméně, lze stejnou odpověď získat z Bayesovské sítě marginalizací



Jak konstruovat Bayesovské sítě

Rozepíšeme $P(x_1, \dots, x_n)$ pomocí tzv. řetězcového pravidla

- $n = 4 : \mathbf{P}(X_4, X_3, X_2, X_1) = \mathbf{P}(X_4|X_3, X_2, X_1)\, \mathbf{P}(X_3|X_2, X_1)\, \mathbf{P}(X_2|X_1)\, \mathbf{P}(X_1)$

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

Dále dostaneme

$$\mathbf{P}(X_i|X_{i-1}, \dots X_1) = \mathbf{P}(X_i|Parents(X_i))$$

za předpokladu $Parents(X_i) \subseteq \{X_{i-1}, \dots, X_1\}$, což platí, pokud je očíslování uzlů konzistentní s uspořádáním uzlů v síti

- pozn. $\mathbf{P}(X_4|X_3, X_2, X_1) = \mathbf{P}(X_4|X_3, X_2)$ pokud X_4 nezávisí na X_1

Tj. sdružená distribuce

$$P(x_1, \dots, x_n) = \prod_i P(x_i|parents(X_i))$$

odpovídá přidání $\mathbf{P}(X_i|Parents(X_i))$ pro každý uzel sítě X_i.

Konstrukce sítě – algoritmus:

Uzly: rozhodněte, jaké náhodné proměnné jsou potřeba a uspořádejte je

- funguje libovolné uspořádání, ale pro různá uspořádání dostaneme různě kompaktní sítě
- doporučené uspořádání je takové, kdy příčiny předcházejí efekty

Hrany: bereme proměnné X_i v daném pořadí od 1 do n

- v množině $\{X_1, \ldots, X_{i-1}\}$ vybereme nejmenší množinu rodičů X_i tak, že platí $\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_{i-1}, \ldots X_1)$
- z rodičů vedeme hranu do X_i
- vypočteme podmíněné pravděpodobnostní tabulky $\mathbf{P}(X_i | Parents(X_i))$

Vlastnosti

- síť je z principu konstrukce acyklická
- síť neobsahuje redundantní informaci a tudíž je vždy konzistentní (splňuje axiomy pravděpodobnosti)

Konstrukce sítě – poznámky:

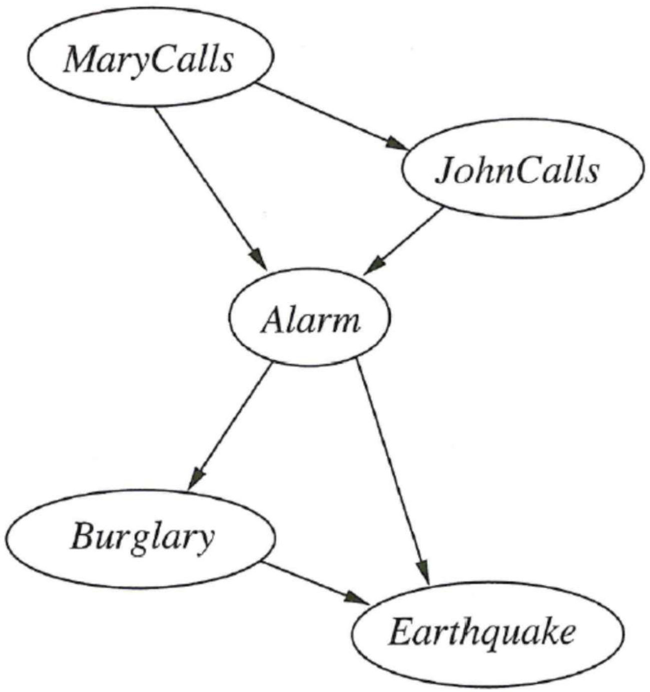
Bayesovská síť může být mnohem kompaktnější než úplná sdružená distribuce, pokud je síť řídká (je lokálně strukturovaná)

- náhodná proměnná často závisí jen na omezeném počtu jiných proměnných
- nechť je takových proměnných k a celkem máme n proměnných, pak potřebujeme prostor
  - $n.2^k$ pro Bayesovskou síť
  - 2^n pro uplnou sdruženou distribuci
- můžeme ignorovat „slabé vazby", čímž budeme mít menší přesnost reprezentace, ale reprezentace bude kompaknější
  - př. nebudeme brát v úvahu, zda Mary nebo John volají kvůli zemětřesení
- samozřejmě kompaktnost sítě hodně závisí na vhodném uspořádání proměnných

Konstrukce sítě – příklad:

Nechť jsme zvolili pořadí MarryCalls, JohnCalls, Alarm, Burglary, Earthquake

- MarryCalls nemá rodiče
- pokud volá Marry, je zřejmě aktivní alarm, což ovlivňuje Johnovo zavolání
- Alarm asi zní, pokud volá Marry nebo John
- pokud známe stav Alarmu, tak vloupání nezávisí na tom, zda volá Marry nebo John
  - $P(Burglary|Alarm, JohnCalls, MarryCalls) = P(Burglary|Alarm)$
- Alarm je svým způsobem detektor zemětřesení, ale pokud došlo k vloupání, tak pravděpodobně nebylo zemětřesení



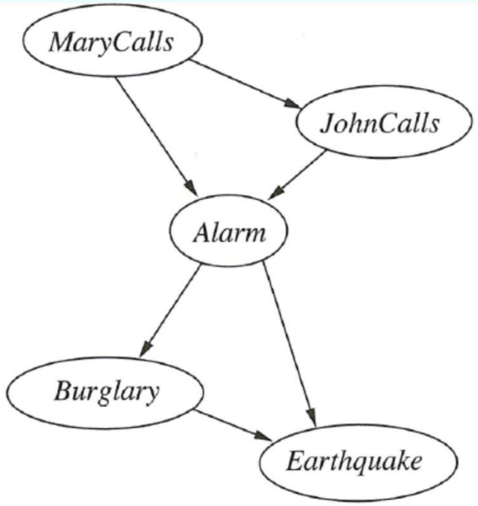
Konstrukce sítě – uspořádání proměnných

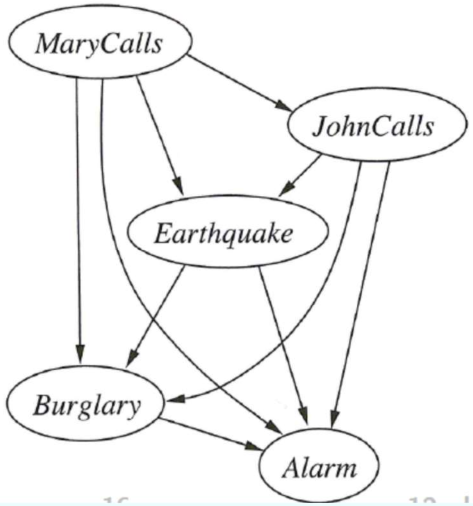
Máme sice jen dvě hrany navíc oproti původnímu návrhu, ale problém je vyplnění tabulek podmíněných závislostí

- stejný problém jako u kauzální vs. diagnostické vazby
- diagnostickou vazbu $P(pricina|nasledek)$ často neznáme
- je lepší držet se kauzální vazby (příčina před následkem) $P(nasledek|pricina)$
  - dává menší sítě a je snažší vyplnit tabulky podmíněných závislostí

Při „špatném" uspořádání proměnných nemusíme nic uspořit vzhledem k úplné sdružené distribuci

- MarryCalls, JohnCalls, Eartquake, Burglary, Alarm





Odvozování v Bayesovských sítích enumerací

- Připomeňme, k čemu mají Bayesovské sítě sloužit – zjistit pravděpodobnostní distribuce náhodných proměnných X v dotazu za předpokladu znalosti hodnot e proměnných z pozorování (ostatní proměnné jsou skryté).

$$\mathbf{P}(X|\mathbf{e}) = \alpha\,\mathbf{P}(X,\mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X,\mathbf{e},\mathbf{y})$$

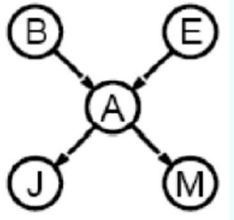
- Hodnotu $P(X, e, y)$ zjistíme z Bayesovské sítě
  $P(x_1, \ldots, x_n) = \prod_i P(x_i|parents(X_i))$

- Můžeme ještě vhodně přesunout
  některé členy $P(x_i|parents(X_i))$ před součty
  - viz příklad na další straně

Odvozování enumerací – příklad:

- Nechť máme dotaz, zda došlo k vloupání, pokud Marry i John volají

$$
\begin{aligned}
P(b|j,m) &= \alpha \sum_e \sum_a P(b)\,P(e)\,P(a|b,e)\,P(j|a)\,P(m|a) \\
&= \alpha P(b) \sum_e P(e) \sum_a P(a|b,e)\,P(j|a)\,P(m|a)
\end{aligned}
$$

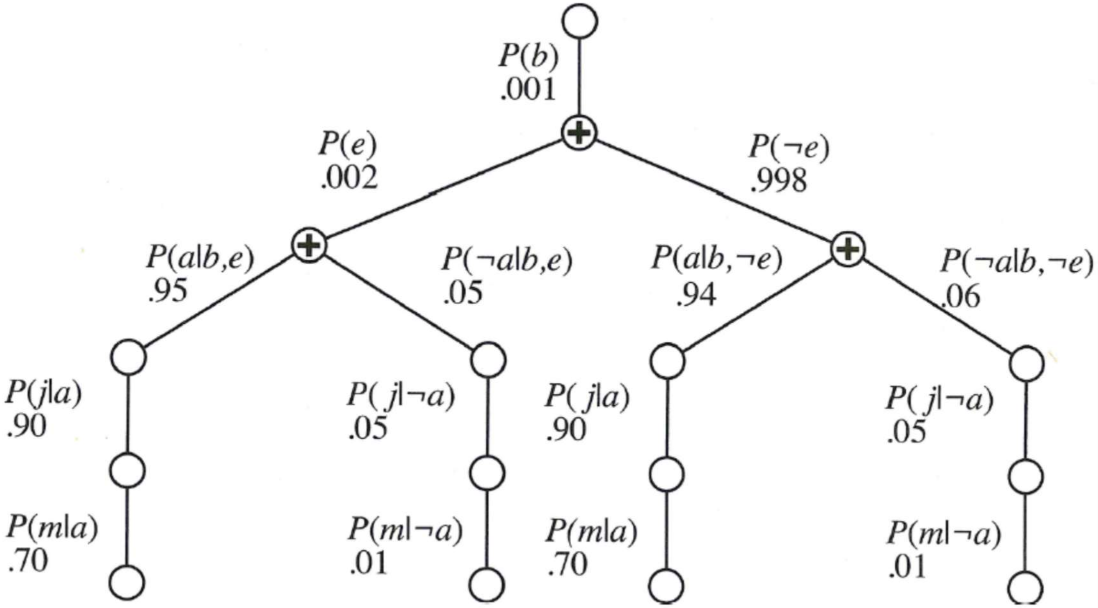


  - Earthquake a Alarm jsou skryté proměnné

  Strukturu výpočtu můžeme zachytit stromovou strukturou

  - je to hodně podobné řešení CSP a SAT

- Všimněme si, že některé části výpočtu se opakují!
- Evaluace se provádí shora dolů násobením hodnot po každé cestě a sečítáním hodnot v „+" uzlech



Odvozování enumerací – algoritmus

function enumerace(X, **e**, bn) **returns** distribuci pro X
  (X: náhodná proměnná
   e**: hodnoty proměnných z pozorování pro **E
   *bn: BS s proměnnými $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$, kde **Y** jsou skryté proměnné)*
 $\mathbf{Q}(X)$ = distribuce pro X, iniciálně prázdná;
 forall hodnoty x_i proměnné X **do**
  $\mathbf{Q}(x_i)$ = enumerace-hodnot($bn.VARS$, $\mathbf{e}_{x_i}$), kde $\mathbf{e}_{x_i}$ je **e** rozšířeno o $X = x_i$;
 return normalizace($\mathbf{Q}(X)$)

function enumerace-hodnot($vars$, **e**) **returns** reálné číslo
 if $vars = \emptyset$ **then return** 1.0;
 Y = first($vars$);
 if Y má hodnotu y v **e**
 then return $P(y|parents(Y)) \times$ enumerace-hodnot(rest($vars$), **e**)
 else return $\sum_y P(y|parents(Y)) \times$ enumerace-hodnot(rest($vars$), $\mathbf{e}_y$)
  kde $\mathbf{e}_y$ je **e** rozšířeno o $Y = y$;

Eliminace proměnných

- Enumerační metoda zbytečně opakuje některé výpočty
- Stačí si výsledek zapamatovat a následně použít

$$\begin{aligned}\mathbf{P}(B|j,m) &= \alpha\,\mathbf{P}(B)\textstyle\sum_e P(e)\sum_a \mathbf{P}(a|B,e)\,P(j|a)\,P(m|a)\\ &= \alpha\,\mathbf{f}_1(B)\textstyle\sum_e \mathbf{f}_2(E)\sum_a \mathbf{f}_3(A,B,E)\,\mathbf{f}_4(A)\,\mathbf{f}_5(A)\end{aligned}$$

- Činitelé f_i jsou matice (tabulky) pro dané proměnné
- Vyhodnocení provedeme zprava doleva (viz příklad dále)
  - násobení činitelů je násobení po prvcích (ne násobení matic)
  - vysčítáním činitelů eliminujeme příslušnou proměnnou
  - na závěr provedeme normalizaci

Eliminace proměnných – operace s tabulkami

- Máme-li dvě tabulky, potom jejich součin je tabulka nad sjednocením proměnných z obou tabulek

$$\mathbf{f}(A_1,\ldots,A_j,B_1,\ldots,B_k,C_1,\ldots,C_l)=\mathbf{f}_1(A_1,\ldots,A_j,B_1,\ldots,B_k)\times\mathbf{f}_2(B_1,\ldots,B_k,C_1,\ldots,C_l)$$

A	B	f_1(A,B)	B	C	f_2(B,C)	A	B	C	f_3(A,B,C)
T	T	0.3	T	T	0.2	T	T	T	0.06 = 0.3*0.2
T	F	0.7	T	F	0.8	T	T	F	0.24 = 0.3*0.8
F	T	0.9	F	T	0.6	T	F	T	0.42 = 0.7*0.6
F	F	0.1	F	F	0.4	T	F	F	0.28 = 0.7*0.4
						F	T	T	0.18 = 0.9*0.2
						F	T	F	0.72 = 0.9*0.8
						F	F	T	0.06 = 0.1*0.6
						F	F	F	0.04 = 0.1*0.4

- Při vyščítání dojde k eliminaci proměnné

$$\mathbf{f}(B,C)=\sum_a \mathbf{f}_3(A,B,C)=\mathbf{f}_3(a,B,C)+\mathbf{f}_3(\neg a,B,C)$$

$$\begin{bmatrix}0.06 & 0.24\\ 0.42 & 0.28\end{bmatrix}+\begin{bmatrix}0.18 & 0.72\\ 0.06 & 0.04\end{bmatrix}=\begin{bmatrix}0.24 & 0.96\\ 0.48 & 0.32\end{bmatrix}$$

Eliminace proměnných – algoritmus

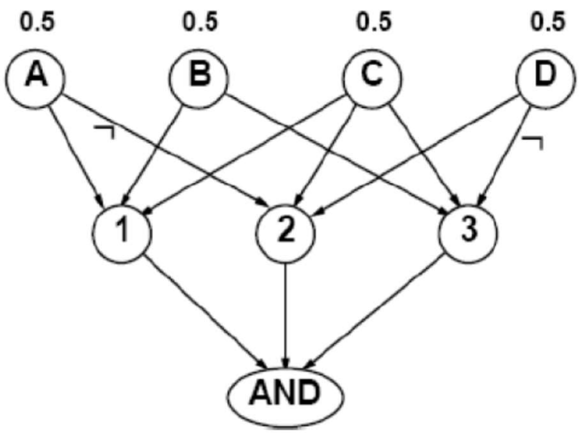
function eliminace(X, $\mathbf{e}$, bn) **returns** distribuci pro X
(X: náhodná proměnná
$\mathbf{e}$: hodnoty proměnných z pozorování pro $\mathbf{E}$
bn: BS s úplnou sdruženou distribucí $\mathbf{P}(X_1, \ldots, X_n)$)
tabulky = [];
forall *var* $\in$ order($bn.VARS$) **do**
tabulky = [vytvoř-tabulku(*var*, $\mathbf{e}$)|*tabulky*];
if *var* je skrytá proměnná **then** *tabulky* = marginalizace(*var*, *tabulky*);
return normalizace(součin-po-prvcích(*tabulky*))

- Algoritmus funguje pro libovolné uspořádání proměnných (order)
- Složitost je dána velikostí největšího činitele (tabulky) v průběhu výpočtu
- Vhodné je proto pro eliminaci vybrat proměnnou, jejíž eliminací vznikne nejmenší tabulka

Složitost úlohy (problému)

- Eliminace proměnných urychluje odvozování, ale jak moc?
- Pokud je Bayesovská síť poly-strom (mezi každými dvěma vrcholy vede maximálně jedna neorientovaná cesta), potom je časová a prostorová složitost odvozování lineární vzhledem k velikosti sítě (tj. velikosti tabulek)
- Pro více-propojené sítě je to horší
  - 3SAT lze redukovat na odvození v Bayesovské síti, takže odvození je NP-těžké

1. $A \vee B \vee C$
2. $C \vee D \vee \neg A$
3. $B \vee C \vee \neg D$



  - odvozování v Bayesovské síti je ekvivalentní zjištění počtu řešení SAT-formule, je tedy striktně těžší než NP-úplné problémy

Vzorkovací metody

- Exaktní odvozování je výpočetně náročné, můžeme ale použít aproximační techniky založené na metodě Monte Carlo
- Monte Carlo algoritmy slouží pro odhad hodnot, které je těžké spočítat exaktně
  - vygeneruje se množství vzorků
    - vzorek = ohocení náhodných proměnných
  - hledaná hodnota se zjistí statisticky
  - více vzorků = větší přesnost

- Pro Bayesovské sítě ukážeme přístupy
  - příme vzorkování
  - vzorkování se zamítáním
  - vzorkování s vážením věrohodností
- Další zajímavé metody (viz Russel & Norvig)
  - vzorkování s Markovovskými řetězci

Přímé vzorkování

- Vzorkem pro nás bude ohodnocení náhodných proměnných
- Vzorek je potřeba generovat tak, aby „odpovídal" tabulkám v Bayesovské síti
  - uzly (proměnné) bereme v topologickém uspořádání
  - ohodnocení rodičů nám dá pravděpodobnostní distribuci hodnot aktuální náhodné proměnné
  - náhodně vybereme hodnotu podle této distribuce
- Nechť N je počet vzorků a $N(x_1, \ldots, x_n)$ je počet výskytů jevu $x_1, \ldots, x_n$, potom $P(x_1, \ldots, x_n) = \lim_{N\to\infty}(N(x_1, \ldots, x_n)/N)$

---

function vzorky(bn) **returns** vzorek získaný dle bn
        (bn: BS specifikující úplnou sdruženou distribuci $\mathbf{P}(X_1, \ldots, X_n)$*)*
  forall proměnnou X_i z $X_1, \ldots, X_n$ **do**
    x_i = náhodný vzorek z $\mathbf{P}(X_i|parents(X_i))$;
  return x

Přímé vzorkování – příklad:

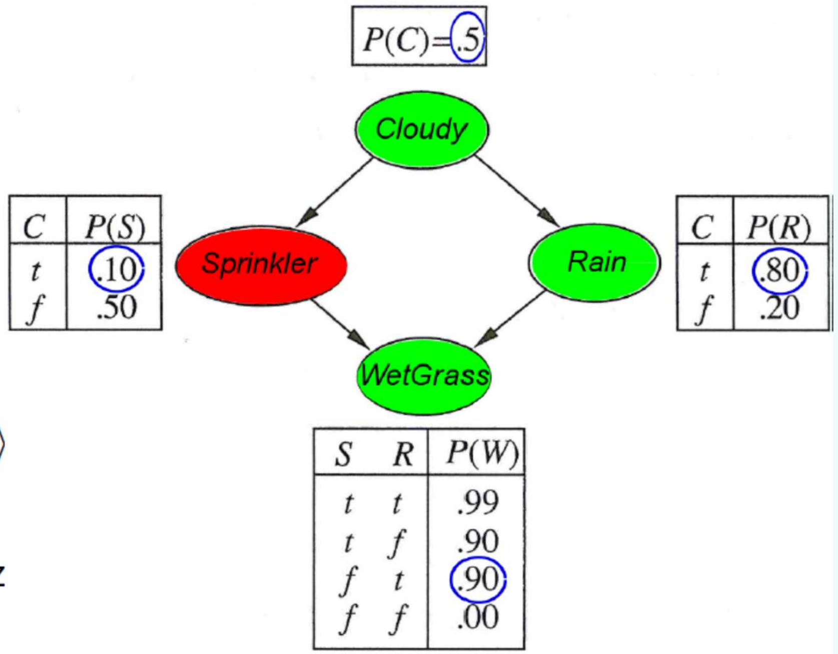
- Vybereme hodnotu pro *Cloudy* z distribuce $\mathbf{P}(Cloudy) = \langle 0.5, 0.5\rangle$

  nechť je true
- Vybereme hodnotu pro *Sprinkler* z distribuce $\mathbf{P}(Sprinkler|Cloudy = true) = \langle 0.1, 0.9\rangle$

  nechť je false
- Vybereme hodnotu pro *Rain* z distribuce $\mathbf{P}(Rain|Cloudy = true) = \langle 0.8, 0.2\rangle$

  nechť je true
- Vybereme hodnotu pro *WetGrass* z distribuce $\mathbf{P}(WetGrass|Sprinkler = false, Rain = true) = \langle 0.9, 0.1\rangle$

  nechť je true



Získali jsme vzorek $Cloudy = true, Sprinkler = false, Rain = true, WetGrass = true$

Pravděpodobnost jeho získání je zřejmě $0.5 * (1 - 0.1) * 0.8 * 0.9 = 0.324$

Vzorkování se zamítáním

- Nás ale zajímá $\mathbf{P}(X|e)$!
- Ze vzorků, které vygenerujeme, vezmeme jen ty, které jsou kompatibilní s e (ostatní zamítneme) $\mathbf{P}(X|e) \approx \mathbf{N}(X,e)/N(e)$

---

function vzorky-zamítání(X, $\mathbf{e}$, bn, N) **returns** odhad pro $\mathbf{P}(X|\mathbf{e})$
(X: náhodná proměnná
$\mathbf{e}$: hodnoty proměnných z pozorování pro $\boldsymbol{E}$
bn: BS s úplnou sdruženou distribucí $\mathbf{P}(X_1, \ldots, X_n)$
N: celkový počet vzorků, který máme generovat)
forall hodnotu x v X **do** $\mathbf{N}[x] = 0$; *(vektor počtu hodnot inicializujeme na 0)*
for $j = 1$ **to** N **do**
  $\mathbf{x}$ = vzorky(bn);
  if $\mathbf{x}$ je konzistentní s $\mathbf{e}$ **then**
    $\mathbf{N}[x] = \mathbf{N}[x] + 1$, kde x je hodnota X v $\mathbf{x}$;
return normalizace($\mathbf{N}$)

- Nechť v našem příkladu vygenerujeme 100 vzorků, z toho u 27 platí $Sprinkler = true$ a z nich u 8 je $Rain = true$ a u 19 je $Rain = false$. Potom $\mathbf{P}(Rain|Sprinkler = true) \approx$ normalizace$(\langle 8, 19\rangle) = \langle 0.296, 0.704\rangle$
- Hlavní nevýhoda metody je zamítání příliš mnoha vzorků!

Vážení věrohodností

Místo zamítání vzorků je efektivnější generovat pouze vzorky vyhovující pozorování e

- Zafixujeme hodnoty z pozorování e a vzorkujeme pouze ostatní proměnné
- Pravděpodobnost získání vzorku je (mezi rodiči máme i pozorování)

$$S_{WS}(\mathbf{z}, \mathbf{e}) = \prod_i P(z_i | parents(Z_i))$$

- To ale není to, co potřebujeme! Ještě nám chybí

$$w(\mathbf{z}, \mathbf{e}) = \prod_j P(e_j | parents(E_j))$$

- Tedy

$$S_{WS}(\mathbf{z}, \mathbf{e}) w(\mathbf{z}, \mathbf{e}) = \prod_i P(z_i | parents(Z_i)) \prod_j P(e_j | parents(E_j)) = P(\mathbf{z}, \mathbf{e})$$

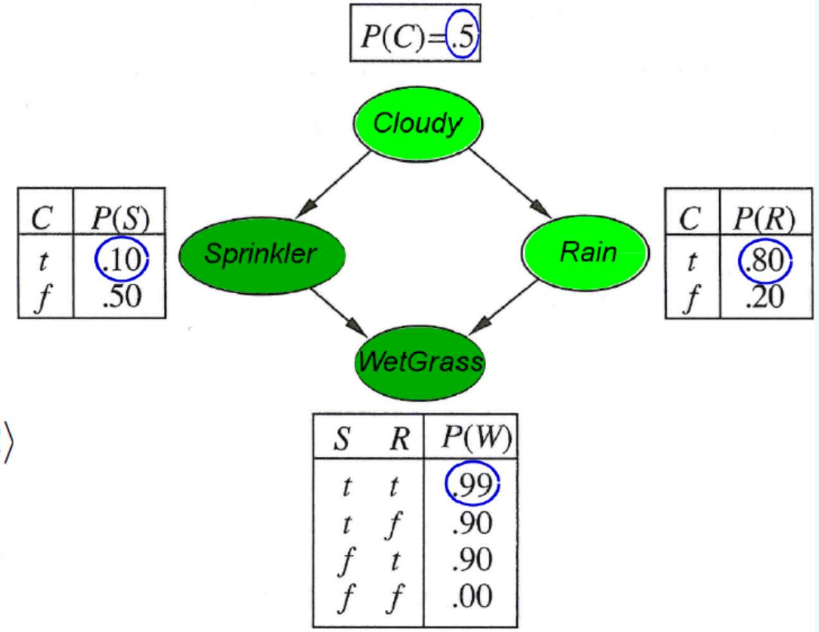
- Každý vzorek tedy doplníme o příslušnou váhu

$$\mathbf{P}(X | e) \approx \alpha \, \mathbf{N}(X, e) \, w(X, e)$$

Vážení věrohodností – příklad:

Nechť zpracováváme dotaz $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$

- Počáteční váha vzorku $w = 1.0$
- Vybereme hodnotu pro *Cloudy* z distribuce $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$ nechť je true
- Hodnotu $Sprinkler = true$ známe, ale upravíme váhu $w = w \times P(Sprinkler = true|Cloudy = true) = 0.1$
- Vybereme hodnotu pro *Rain* z distribuce $\mathbf{P}(Rain|Cloudy = true) = \langle 0.8, 0.2 \rangle$ nechť je true
- Hodnotu $WetGrass = true$ známe, ale upravíme váhu $w = w \times P(WetGrass = true|Sprinkler = true, Rain = true) = 0.099$



Získali jsme vzorek $Cloudy = true, Sprinkler = true, Rain = true, WetGrass = true$, který má váhu 0.099

Vážení věrohodností – algoritmus

function vážení-věrohodností($X, \mathbf{e}, bn, N$) **returns** odhad pro $\mathbf{P}(X|\mathbf{e})$
(X: náhodná proměnná
$\mathbf{e}$: hodnoty proměnných z pozorování pro $\mathbf{E}$
bn: BS s úplnou sdruženou distribucí $\mathbf{P}(X_1, \ldots, X_n)$
N: celkový počet vzorků, který máme generovat)
forall hodnotu x v X **do** $\mathbf{W}[x] = 0$; *(vektor vah inicializujeme na 0)*
for $j = 1$ **to** N **do**
$\mathbf{x}, w$ = vážený-vzorek($bn, \mathbf{e}$);
$\mathbf{W}[x] = \mathbf{W}[x] + w$, kde x je hodnota X v $\mathbf{x}$;
return normalizace($\mathbf{W}$)

---

function vážený-vzorek($bn, \mathbf{e}$) **returns** vzorek a váhu
forall X_i v $X_1, \ldots, X_n$ **do**
if X_i má v $\mathbf{e}$ hodnotu x_i
then $w = w \times P(X_i = x_i | parents(X_i))$
else x_i = náhodný vzorek z $\mathbf{P}(X_i | parents(X_i))$;
returns $\mathbf{x}, w$

Neurčitost prostřednictvím Bayesovských sítí – shrnutí

1. Opakování: pravděpodobnost
2. Bayesovská síť
3. Sémantika sítě
4. Konstrukce sítě
5. Odvozování v Bayesovských sítích
   - Exaktní metody
     - Odvozování enumerací
     - Eliminace proměnných
   - Aproximační metody
     - Přímé vzorkování
     - Vzorkování se zamítáním

Dempster-Shaferova teorie

Dempster-Shaferova teorie byla vyvinuta ve snaze o překonání některých nedostatků pravděpodobnostního přístupu, jako např. reprezentace neznalosti (ignorance) a požadavku, že součet měr důvěry v událost a její negaci musí být roven 1.

Zatímco pravděpodobnost představuje stupeň, ve kterém je tvrzení považováno za pravdivé, míra domnění v Dempster-Shaferově teorii představuje podporu tomuto tvrzení. Růst pravděpodobnosti hypotézy redukuje pravděpodobnost komplementu, kdežto v Dempster-Shaferově teorii růst podpory hypotézy nezpůsobuje změnu podpory komplementu.

Dempster-Shaferova teorie je obecnější než bayesovský přístup, jelikož důvěra ve tvrzení a důvěra v negaci tohoto tvrzení nemusí být v součtu rovna 1.

Základní přiřazení

Prostředí: úplný systém vzájemně disjunktních základních hypotéz

$$X = \{h_1, h_2, \dots, h_n\}$$

Základní přiřazení (*basic probability assignment*, *mass probability function*) je funkce definovaná na množině všech podmnožin množiny X a nabývající hodnot z intervalu $\langle 0, 1\rangle$, tj.

$$m\colon 2^X \rightarrow \langle 0, 1\rangle\,,$$

která má tyto vlastnosti:

$$m(\emptyset) = 0, \quad \sum_{A \subseteq X} m(A) = 1$$

Hodnota $m(A)$ představuje *míru důvěry*, že platí právě hypotéza A, přičemž nevypovídá nic o míře důvěry ve složky množiny A. Nemusí tedy platit $m(B) \leq m(A)$ pro $B \subseteq A$.

Řekneme, že A je *fokální element,* jestliže $m(A) > 0$.

Příklad:

Mějme nějaké (možné) hodnoty veličin, např. tři delikventy, kteří způsobili nějakou újmu, a víme jen, že z 50% to udělal A nebo B, ale nevíme, který z nich; na druhou stranu ale víme, že z 50% to udělal C. Pokud těmto tvrzením přiřadíme pravděpodobnosti a poté dostaneme další informaci, že A vinen není, tak se zvýší pravděpodobnost viny C.

V Dempster-Shaferově teorii vyloučíme A, tedy celých 50% přejde na B a 50% zůstane na C (zjednodušení).

Příklad:

Mějme nějaké (možné) hodnoty veličin, např. tři delikventy, kteří způsobili nějakou újmu, a víme jen, že z 50% to udělal A nebo B, ale nevíme, který z nich; na druhou stranu ale víme, že z 50% to udělal C. Pokud těmto tvrzením přiřadíme pravděpodobnosti a poté dostaneme další informaci, že A vinen není, tak se zvýší pravděpodobnost viny C.

V Dempster-Shaferově teorii vyloučíme A, tedy celých 50% přejde na B a 50% zůstane na C (zjednodušení).

Dále: Základní veličiny Dempster-Shaferovy teorie

Míra domnění

Míra domnění (*measure of belief*) v platnost hypotézy *A* je definována jako součet základních přiřazení všech podmnožin množiny *A*:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

Na rozdíl od základního přiřazení tedy $Bel(A)$ vyjadřuje míru domnění v hypotézu *A* nebo v jakoukoli z jejích podmnožin.

Platí:

$$Bel(A) + Bel(A^{c}) \leq 1, \text{ kde } A^{c} = X - A,$$

$$Bel(A) \leq Bel(B) \text{ pro } A \subseteq B,$$

$$Bel(X) = 1.$$

Míra věrohodnosti a interval domnění

Míra věrohodnosti, resp. připustitelnosti, (*measure of plausibility*):

$$Pl(A) = 1 - Bel(A^c)$$

Hodnota $Pl(A)$ vyjadřuje míru chyby při zamítnutí A.

Platí:

$$Bel(A) \leq Pl(A)$$

$$Pl(A) + Pl(A^c) \geq 1,$$

$$Pl(A) \leq Pl(B) \text{ pro } A \subseteq \text{B}.$$

Interval domnění:

$$\langle\, Bel(A), Pl(A) \,\rangle$$

Tento interval vyjadřuje rozsah naší jistoty o hypotéze A. Rozdíl $Pl(A) - Bel(A)$ se nazývá *nejistota* o hypotéze A nebo *ignorance*.

Poznámka: Míru věrohodnosti (připustitelnosti, plausibility) lze definovat rovněž takto:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

tj. součet přes všechna B mající s A neprázdný průnik.

Poznámka: Míru věrohodnosti (připustitelnosti, plausibility) lze definovat rovněž takto:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

tj. součet přes všechna B mající s A neprázdný průnik.

Shrnutí: $Bel(A)$ sumarizuje, nakolik evidence (známé) ukazuje na A.

$Pl(A)$ říká, jak bychom věřili A, kdyby vše neznámé ukazovalo na A.

→ pravdivá hodnota leží někde mezi.

Příklad:

Uvažujme univerzum $U = \{H, C, P\}$ a základní přiřazení

$$
\begin{aligned}
\mathrm{m}(\{\mathrm{H}\}) &= 0.3 \\
\mathrm{m}(\{\mathrm{H,C}\}) &= 0.2 \\
\mathrm{m}(\{\mathrm{H,C,P}\}) &= 0.5
\end{aligned}
$$

Pak

$\mathrm{Bel}(\{\mathrm{H}\})$	$= 0.3$	$\mathrm{Pl}(\{\mathrm{H}\})$	$= 1.0$
$\mathrm{Bel}(\{\mathrm{H,C}\})$	$= 0.5$	$\mathrm{Pl}(\{\mathrm{H,C}\})$	$= 1.0$
$\mathrm{Bel}(\{\mathrm{P}\})$	$= 0$	$\mathrm{Pl}(\{\mathrm{P}\})$	$= 0.5$
$\mathrm{Bel}(\{\mathrm{C}\})$	$= 0$	$\mathrm{Pl}(\{\mathrm{C}\})$	$= 0.7$

Kombinace základních přiřazení

Dempsterovo kombinační pravidlo:

$$m_3(C) = \frac{\sum\limits_{A,B\subseteq X,\ A\cap B=C} m_1(A)\cdot m_2(B)}{1-\sum\limits_{A,B\subseteq X,\ A\cap B=\emptyset} m_1(A)\cdot m_2(B)}$$

$$m_3(C) = m_1(A) + m_2(B)$$

Dvě základní přiřazení m_1 a m_2 mohou pocházet od různých expertů, případně m_1 může být výchozí základní přiřazení a m_2 základní přiřazení získané na základě nových skutečností.

Změna domnění

Jestliže pozorujeme *B*, pak se základní přiřazení, domnění a věrohodnost *A* mění takto:

$$m(A \mid B) = \frac{\sum\limits_{C \subseteq X,\, C \cap B = A} m(C)}{1 - \sum\limits_{C \subseteq X,\, C \cap B = \emptyset} m(C)}$$

pro $A \neq \emptyset$; $m(\emptyset|B) = 0$

$$Bel(A \mid B) = \frac{Bel(A \cup B^c) - Bel(B^c)}{1 - Bel(B^c)}$$

$$Pl(A \mid B) = \frac{Pl(A \cap B)}{Pl(B)}$$

Příklad:

Pro univerzum $U = \{D, D'\}$ a

$m_1(\{D\}) = 0.8$; $m_1(\{D'\}) = 0$; $m_1(\{D,D'\}) = 0.2$;

$m_2(\{D\}) = 0.9$; $m_2(\{D'\}) = 0$; $m_2(\{D,D'\}) = 0.1$;

vytvoříme tabulku:

		m_2					
	m_1	0.9	{D}	0	{D'}	0.1	{D,D'}
0.8	{D}	0.72	{D}	0	{}	0.08	{D}
0	{D'}	0	{}	0	{}	0	{D'}
0.2	{D,D'}	0.18	{D}	0	{}	0.02	{D,D'}

$$\begin{aligned} m_1 + m_2(\{D\}) &= 0.72 + 0.08 + 0.18 = 0.98 \\ m_1 + m_2(\{D'\}) &= 0 \\ m_1 + m_2(\{D, D'\}) &= 0.02 \end{aligned}$$

Bayesovské sítě precizně (pro zájemce)

Bayesovská síť (Bayesian belief network) je orientovaný acyklický graf, jehož uzlům odpovídají náhodné proměnné a vazby reprezentují kauzální závislosti mezi těmito proměnnými.

Hrana $X \rightarrow Y$ znamená, že X kauzálně ovlivňuje Y (pozorování X poskytuje kauzální podporu Y, pozorování Y poskytuje diagnostickou podporu pro X). Bayesovská síť umožňuje provádět prediktivní i diagnostické inference.

Každému uzlu je přiřazena tabulka rozdělení pravděpodobnosti. Jestliže uzel nemá žádné předchůdce (rodiče), jedná se o nepodmíněnou pravděpodobnost, v opačném případě jde o podmíněnou pravděpodobnost.

Pojmy potřebné pro definici bayesovské sítě

Nechť $G = (V, E)$ je orientovaný acyklický graf a nechť $v \in V$.

Definujme následující množiny:

$C(v) = \{ u \in V \mid (u, v) \in E \}$,

$D(v) = \{ w \in V \mid$ existuje cesta z v do $w \}$,

$A(v) = \{ x \in V \mid x \neq v$ a $x \notin C(v) \cup D(v) \}$.

Množina $C(v)$ je množinou bezprostředních předchůdců (rodičů) uzlu v, $D(v)$ je množinou všech následníků uzlu v (nejen bezprostředních). V případě bayesovské sítě prvky $C(v)$ nazýváme také *příčinami* .

Definice bayesovské sítě

Nechť (Ω, P) je pravděpodobnostní prostor, kde $\Omega = \Omega_1 \times \ldots \times \Omega_n$, a nechť X_i $(i = 1, \ldots, n)$ je projekce Ω na Ω_i (tj. $X_i : \Omega \to \Omega_i$ je náhodná proměnná). Nechť (V, E) je orientovaný acyklický graf, kde $V = \{X_1, \ldots, X_n\}$.

Řekneme, že (V, E, P) je *bayesovská síť*, jestliže pro všechna $X_i \in V$ a všechna $W \subseteq A(X_i)$ jsou X_i a W podmíněně nezávislé při daném $C(X_i)$. To znamená, že když $W = \{Y_1,\ldots,Y_k\}$, $C(X_i) = \{Z_1,\ldots, Z_m\}$,

$$P(Y_1 \wedge \ldots \wedge Y_k \wedge Z_1 \wedge \ldots \wedge Z_m) \neq 0, \quad \text{pak}$$

$$P(X_i \,|\, Y_1 \wedge \ldots \wedge Y_k \wedge Z_1 \wedge \ldots \wedge Z_m) = P(X_i \,|\, Z_1 \wedge \ldots \wedge Z_m).$$

Zjednodušeně řečeno, když známe příčiny X_i , nic jiného než X_i samotné nebo jeho následníci nám nemůže dát nějaké další informace o X_i . Místo pojmu bayesovská síť se někdy užívají pojmy *kauzální síť* či *influenční diagram*.

Poznámky k symbolice

Nechť X je náhodná proměnná s oborem hodnot $O(X)$ a P je pravděpodobnost. Symbolem $P(X)$ rozumíme funkci definovanou na $O(X)$ tak, že pro $x \in O(X)$ je $P(x) = P(X = x)$.

Zápisy pravděpodobností můžeme dále zjednodušit následujícím způsobem. Nechť $V = \{X_1, \dots , X_n\}$, $C(X_i) = \{Z_1, \dots , Z_m\}$. Pak:

$$P(V) = P(\{X_1\} \cup \dots \cup \{X_n\}) = P(X_1, \dots , X_n) = P(X_1 \wedge \dots \wedge X_n),$$

$$P(X_i \,|\, C(X_i)) = P(X_i \,|\, Z_1, \dots , Z_m) = P(X_i \,|\, Z_1 \wedge \dots \wedge Z_m).$$

Uvedenou symbolikou můžeme definici bayesovské sítě přepsat:

Řekneme, že (V, E, P) je bayesovská síť, jestliže pro všechna $X_i \in V$ a všechna $W \subseteq A(X_i)$, $P(W \cup C(X_i)) \neq 0$, platí, že

$$P(X_i \,|\, W \cup C(X_i)) = P(X_i \,|\, C(X_i)).$$

Vlastnosti bayesovské sítě

Nechť (V, E, P) je bayesovská síť. Pak platí:

$$P(V) = \prod_{\substack{X \in V \\ P(C(X)) \neq 0}} P\big(X \mid C(X)\big)$$

Nechť (V, E) je orientovaný acyklický graf, kde $V = \{X_1, \ldots, X_n\}$, přičemž X_i jsou proměnné s obory hodnot $O(X_i)$. Nechť $f(X \mid C(X))$ je nezáporná reálná funkce taková, že $\sum_{x \in O(X)} f\big(x \mid C(X)\big) = 1$ pro všechny kombinace hodnot proměnných z $C(X)$. Pak

$$\Omega = O(X_1) \times \ldots \times O(X_n) \quad \text{a} \quad P(V) = \prod_{X \in V} f\big(X \mid C(X)\big)$$

definují pravděpodobnostní prostor, pro nějž (V, E, P) je bayesovská síť. Přitom $P(X \mid C(X))$ je buď 0 nebo $f(X \mid C(X))$.

Konstrukce bayesovské sítě

1. Specifikujeme veličiny $X_1, \dots, X_n$ a jejich obory hodnot $O(X_i)$.
2. Zkonstruujeme orientovaný acyklický graf (V, E), kde $V = \{X_1, \dots, X_n\}$, vyjadřující kauzální závislosti mezi veličinami.
3. Odhadneme pravděpodobnost P tak, že odhadneme $P(X \mid C(X))$ pro všechna X, všechny hodnoty X a všechny kombinace hodnot proměnných z $C(X)$. Podle předchozí věty je nezbytné splnění pouze těchto podmínek:

$$0 \leq P\big(X \mid C(X)\big) \leq 1$$

$$\sum_{x \in O(X)} P\big(x \mid C(X)\big) = 1$$

Problém pro bayesovskou síť

Problém řešený pomocí bayesovské sítě je možno zjednodušeně formulovat takto:

Nechť je dána bayesovská síť (V, E, P) a množiny $U \subseteq V$, $W \subseteq V$, $U \cap W = \emptyset$. Jsou-li zadány hodnoty proměnných z množiny U, je třeba zjistit $P(W|U)$.

Po zadání hodnot některých proměnných se provádí *inference*, což znamená přepočet podmíněných pravděpodobností pro ostatní proměnné. Inference v bayesovské síti je založena na Bayesových vzorcích.

Obecně je výše zmíněný problém NP-složitý, což znamená, že pro něj neexistují algoritmy s polynomiální časovou složitostí. Pokud bayesovská síť nemá speciální strukturu, je nutné použít aproximační techniky, které jsou obvykle založeny na nějakých transformacích bayesovské sítě na jednodušší tvar.

Jednoduše souvislá bayesovská síť

Řekneme, že bayesovská síť je jednoduše souvislá, jestliže mezi každými dvěma uzly existuje právě jedna neorientovaná cesta.

Jednoduše souvislá síť se také nazývá polystrom nebo les. Zvláštním případem polystromu je strom, což je graf, kde každý uzel má nejvýše jednoho rodiče.

Pro polystromovou bayesovskou síť existují algoritmy pro inferenci, které mají polynomiální časovou složitost.

Příklady: viz následující http:// adresy:

http://www.uai.fme.vutbr.cz/~jdvorak/Vyuka/

http://ktiml.ms.mff.cuni.cz/~marta/SU.html

8. Zpracování neurčitosti – teoretické přístupy