Inteligență Artificială

Universitatea Politehnica Bucuresti
Anul universitar 2021-2022

Adina Magda Florea

Curs nr. 5 si 6

Reprezentarea cunostintelor incerte

- Teoria probabilitatilor
- Retele Bayesiene
- Inferente exacte si aproximative in retele Bayesiene

1. Teoria probabilitatilor

1.1 Cunostinte incerte

$\forall$p simpt(p, Dur_d) $\rightarrow$ factor(p,carie)

$\forall$p simpt(p, Dur_d) $\rightarrow$ factor(p,carie) $\vee$ factor(p,infl_ging) $\vee$…

- LP

- dificultate (« lene »)

- ignoranta teoretica

- ignoranta practica

- Teoria probabilitatilor ➔ un grad numeric de **incredere** sau **plauzibilitate** a afirmatiilor in [0,1]
- Gradul de adevar (fuzzy logic) $\neq$ gradul de incredere

1.2 Definitii TP

- ***Probabilitatea unui eveniment incert A*** este masura gradului de incredere sau plauzibilitatea produceri unui eveniment
- Camp de probabilitate, S
- ***Probabilitate neconditionata*** (apriori) - inaintea obtinerii de probe pt o ipoteza / eveniment
- ***Probabilitate conditionata*** (aposteriori) - dupa obtinerea de probe

Exemple

P(Carie) = 0.1

P(Vreme = Soare) = 0.7

P(Vreme = Ploaie) = 0.2   P(Vreme = Nor) = 0.1

Vreme - **variabila aleatoare**

- ***Distributie de probabilitate***

Definitii TP - cont

- ***Probabilitate conditionata*** (aposteriori) - P(A|B)

  P(Carie | Dur_d) = 0.8

  Masura probabilitatii producerii unui eveniment A este o functie $P{:}S \to R$ care satisface **axiomele**:
- $0 \leq P(A) \leq 1$
- P(S) = 1 (sau P(adev) = 1 si P(fals) = 0)
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

$P(A \vee \sim A) = P(A)+P(\sim A) - P(fals) = P(adev)$

$\Rightarrow P(\sim A) = 1 - P(A)$

Definitii TP - cont

Evenimente mutual exclusive

Moneda – cap/pajura – mutual exclusive

Zar – 1, 2, 3 – mutual exclusive

Evenimente exhaustive

Moneda – cap/pajura – mutual exhaustive

Zar – 1, 2, 3, 4, 5, 6 – mutual exhaustive

Definitii TP - cont

A si B mutual exclusive ➔

$P(A \vee B) = P(A) + P(B)$

$P(e_1 \vee e_2 \vee e_3 \vee \ldots e_n) =$

$$P(e_1) + P(e_2) + P(e_3) + \ldots + P(e_n)$$

e(a) – multimea de evenimente atomice mutual exclusive si exhaustive in care apare **a**

$$\boxed{P(a) = \sum_{e_i \in e(a)} P(e_i)}$$

1.3 Regula produsului

Probabilitatea conditionata de producere a evenimentului A ***in conditiile producerii evenimentului B***

- **$P(A|B) = P(A \wedge B) / P(B)$**

$$P(A \wedge B) = P(A|B) * P(B)$$

1.4 Teorema lui Bayes

$P(A|B) = P(A \wedge B) / P(B)$ – regula produsului

$$P(A|B) = P(A \wedge B) / P(B)$$

$$P(B|A) = P(A \wedge B) / P(A)$$

$$\boxed{P(B|A) = P(A|B) * P(B) / P(A)}$$

APosteriori = Plauzibilitate x APriori / Evidenta

Teorema lui Bayes

$P(B|A) = P(A|B) * P(B) / P(A)$

- Daca **B si ~B sunt mutual exclusive si exhaustive**, probabilitatea de producere a lui A in conditiile producerii lui B se poate scrie

$$\mathbf{P(A)} = P(A \wedge B) + P(A \wedge \sim B) = \mathbf{P(A|B)*P(B) + P(A|\sim B)*P(\sim B)}$$

$$\boxed{\begin{aligned} &\mathbf{P(B|A) =} \\ &\quad \mathbf{P(A \mid B) * P(B) / [P(A|B)*P(B) + P(A|\sim B)*P(\sim B)]} \end{aligned}}$$

Teorema lui Bayes

Generalizarea la mai multe ipoteze

B - h, A - e

P(h|e) = P(e | h) * P(h) / [P(e|h)*P(h) +P(e| ~h)*P(~h)]

Daca h_i mutual exclusive si exhaustive, i=1,k

$$\boxed{P(h_i|e) = \frac{P(e|h_i)\cdot P(h_i)}{\sum_{j=1}^{k} P(e|h_j)\cdot P(h_j)},\ i = 1,k}$$

Teorema lui Bayes

Generalizarea la mai multe ipoteze si probe

h_i – evenimente / ipoteze probabile (i=1,k);

$e_1,\ldots,e_n$ – probe (evenimente)

$P(h_i)$

$P(h_i \mid e_1,\ldots,e_n)$

$P(e_1,\ldots,e_n \mid h_i)$

$$\boxed{P(h_i|e_1,e_2,\ldots,e_n) = \frac{P(e_1,e_2,\ldots,e_n|h_i)\cdot P(h_i)}{\sum_{j=1}^{k} P(e_1,e_2,\ldots,e_n|h_j)\cdot P(h_j)},\ i = \overline{1,k}}$$

Teorema lui Bayes - cont

$$P(h_i|e_1,e_2,\ldots,e_n) = \frac{P(e_1,e_2,\ldots,e_n|h_i)\cdot P(h_i)}{\sum_{j=1}^{k} P(e_1,e_2,\ldots,e_n|h_j)\cdot P(h_j)},\ i=\overline{1,k}$$

Daca $e_1,\ldots,e_n$ sunt ipoteze independente atunci

$$\boxed{P(e|h_j) = P(e_1,e_2,\ldots,e_n|h_j) = P(e_1|h_j)\cdot P(e_2|h_j)\cdot\ldots\cdot P(e_n|h_j),\ j=\overline{1,k}}$$

PROSPECTOR – sistem expert pentru consultari privind exploatari miniere si evaluarea resurselor

1.5 Inferente din DP si TB

Distributie de probabilitate **P(Carie, Dur_d)**

	Dur_d	~Dur_d
Carie	0.04	0.06
~Carie	0.01	0.89

$P(\text{Carie}) = 0.04 + 0.06 = 0.1$

$P(\text{Carie} \vee \text{Dur\_d}) = 0.04 + 0.01 + 0.06 = 0.11$

$P(\text{Carie}|\text{Dur\_d}) = P(\text{Carie} \wedge \text{Dur\_d}) / P(\text{Dur\_d}) = 0.04 / 0.05$

Inferente din DP si TB

	Dur_d		~Dur_d	
	Evid	~Evid	Evid	~Evid
Carie	0.108	0.012	0.072	0.008
~Carie	0.016	0.064	0.144	0.576

Distributie de probabilitate **P(Carie, Dur_d, Evid)**

P(Carie) = 0.108 + 0.012 + 0.72 + 0.008 = 0.2
P(Carie $\vee$ Dur_d) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28

P(Carie, Dur_d, Vreme) – o tabela cu 2 x 2 x 3 = 12 intrari
Distributie de probablitate completa

Inferente din DP si TB

	Dur_d		~Dur_d	
	Evid	~Evid	Evid	~Evid
Carie	0.108	0.012	0.072	0.008
~Carie	0.016	0.064	0.144	0.576

P(Carie | Dur_d) = P(Carie $\wedge$ Dur_d) / P(Dur_d)
P(~Carie | Dur_d) = P(~Carie $\wedge$ Dur_d) / P(Dur_d)
α = 1 / P(Dur_d) – constanta de normalizare a distributiei

P(Carie | Dur_d) = **α** **P(Carie $\wedge$ Dur_d)** =
α [P(Carie $\wedge$ Dur_d $\wedge$ Evid) + P(Carie $\wedge$ Dur_d $\wedge$ ~Evid)] =
α [$\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle$] = **$\alpha$** $\langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle$

Chiar daca nu cunoastem α, adica P(Dur_d), putem calcula α
α = 1/(0.12+0.08) = 1/0.2

Inferente din DP si TB

Generalizare – procedura generala de inferenta bazata pe DPC
Interogare asupra lui X

X – variabila de interogat (Carie)
E – lista de variabile probe (Dur_d)
e – lista valorilor observate pt aceste variabile E
Y – lista variabilelor neobservate (restul) (Evid)

P(Carie | Dur_d) = $\boldsymbol{\alpha}$ [P(Carie $\wedge$ Dur_d $\wedge$ Evid) + P(Carie $\wedge$ Dur_d $\wedge$ ~Evid)]

Insumarea se face peste toate combinatiile de valori ale variab. neobservate Y

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

Inferente din DP si TB

$$\mathbf{P}(X \mid e) = \alpha\, \mathbf{P}(X , e) = \alpha \Sigma_{Y=y}\, \mathbf{P}(X, e, Y)$$

Avand DPC ecuatia poate da raspuns la interogari cu variabile discrete

Complex computational

n var Bool – tabela $O(2^n)$

- timp $O(2^n)$

1.6 Independenta conditionala

- Evenimentele **X1 si X2 sunt independente conditional fiind dat un eveniment Y** daca

  Stiind ca Y apare, aparitia lui X1 nu influenteaza aparitia lui X2 si aparitia lui X2 nu influenteaza aparitia lui X1

altfel spus

X1 si X2 sunt independente conditional fiind dat un eveniment Y daca, pentru orice valoare a lui Y:

- distributia de probabilitate a lui X1 este aceeasi pentru orice valoare ar lua X2
- distributia de probabilitate a lui X2 este aceeasi pentru orice valoare ar lua X1

$$\boxed{\mathbf{P}(X1,X2|Y) = \mathbf{P}(X1|Y)*\mathbf{P}(X2|Y)}$$

Independenta conditionala

$\mathbf{P}(\text{cauza} \mid \text{efect}) = \mathbf{P}(\text{efect} \mid \text{cauza}) * \mathbf{P}(\text{cauza}) / \mathbf{P}(\text{efect})$

$\mathbf{P}(y \mid x_1,..,x_n) = P(x_1,..,x_n \mid y) * \mathbf{P}(y) / \mathbf{P}(x_1,..x_n)$

$\mathbf{P}(y \mid x_1,..,x_n) = \alpha * P(x_1,..,x_n \mid y) * \mathbf{P}(y)$

Daca $x_1,..,x_n$ sunt independente conditional fiind dat y atunci

$P(x_1,..x_n \mid y) = \Pi_i P(x_i \mid y)$

$$\mathbf{P}(\text{Cauza} \mid \text{Efect}_1, \text{Efect}_2, \ldots) = \alpha\, \mathbf{P}(\text{Cauza}) * \Pi_i\, \mathbf{P}(\text{Efect}_i \mid \text{Cauza})$$

1.7 Model Bayesian naiv

$\mathbf{P}(\text{Cauza}_y \mid \text{Efect}_1, \text{Efect}_2, \ldots) = \alpha\, \mathbf{P}(\text{Cauza}_y) * \Pi_i\, \mathbf{P}(\text{Efect}_i|\text{Cauza}_y)$

Ipoteza de independenta conditionala / naiva

$$\text{Cauza\_MAP} = \text{argmax}_y\, \mathbf{P}(\text{Cauza}_y) * \Pi_i\, \mathbf{P}(\text{Efect}_i|\text{Cauza}_y)$$

Model Bayesian naiv

Model Bayesian naiv

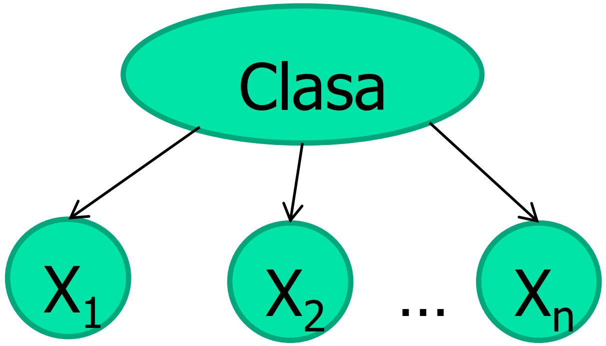
Vector de caracteristici $x_1,..,x_n$ si o serie de clase C_k

$$\mathbf{P}(C_k \mid x_1, x_2, \ldots) = \alpha\, \mathbf{P}(C_k) * \Pi_i\, \mathbf{P}(x_i|C_k)$$

$$C\_MAP = \text{argmax}_k\, \mathbf{P}(C_k) * \Pi_i\, \mathbf{P}(x_i|C_k)$$

Model Bayesian naiv

MAP – Maximum a Posteriori



1.8 Modele grafice probabiliste

- Fiecare nod reprezinta o variabila aleatoare si fiecare legatura reprezinta o relatie probabilistica
  - Retele Bayesiene – modele grafice orientate - permit reprezentare compacta a DP si punerea in evidenta a independentei conditionale
  - Modele Markov – lanturi Markov – stari si tranzitii probabilistice

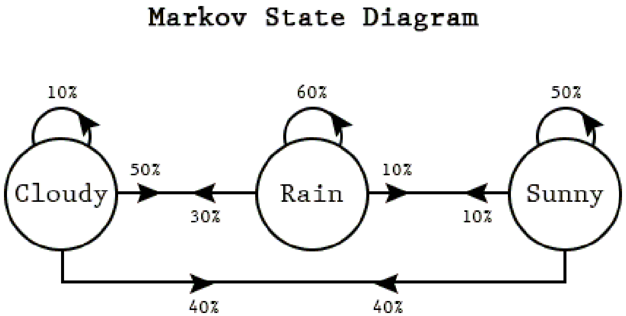
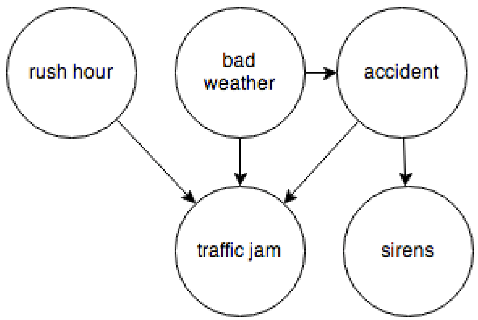


Figure 2



2 Retele Bayesiene

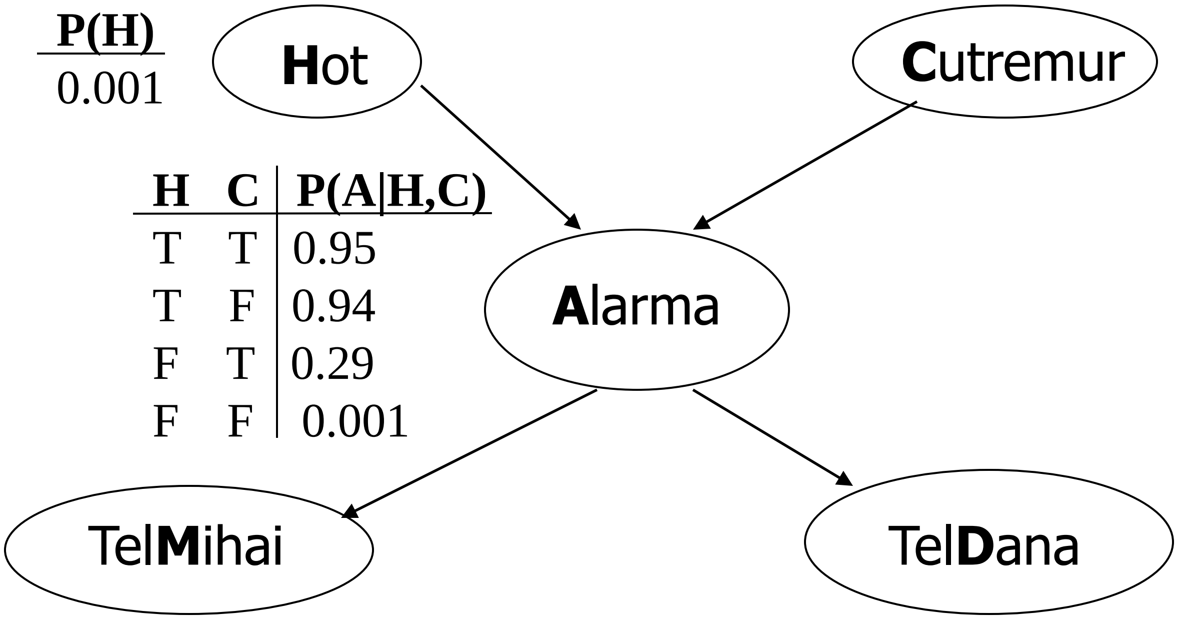
- Reprezinta dependente intre variabile aleatoare
- Specifica distributiei de probabilitate
- Simplifica calculele
- Au asociata o reprezentare grafica convenabila
- DAG care reprezinta relatiile cauzale intre variabile
- Pe baza structurii retelei se pot realiza diverse tipuri de inferente
- Calcule complexe in general dar se pot simplifica pentru structuri particulare

2.1 Structura retelelor Bayesiene

O RB este un DAG in care:

- Nodurile reprezinta variabilele aleatoare
- Legaturile orientate **X→Y**: X are o influenta directa asupra lui Y, **X=Parinte(Y)**
- Fiecare nod are asociata o tabela de probabilitati conditionate care cuantifica efectul parintilor asupra nodului
    - **P(X_i | Parinti(X_i))**

Structura retelelor Bayesiene - cont



P(H)
0.001

P(C)
0.002

H	C	P(A|H,C)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(M|A)
T	0.9
F	0.05

A	P(D|A)
T	0.7
F	0.01

Tabela de probabilitati conditionate

H	C	P(A | H, C)	
		T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

Structura retelelor Bayesiene

In general X (Cauza) $\rightarrow$ Y (Efect)

- Stabilesc topologia
- Specifica distributia de probabilitati conditionate
- Combinarea topologiei si distributia de probabilitati conditionate este suficienta pentru a specifica (implicit) intreaga DPC
- DPC poate raspunde la interogari
- Si RB la fel, mai eficient

2.2 Semantica retelelor Bayesiene

- Reprezentare a distributiei de probabilitate
- Specificare a independentei conditionale – constructia retelei

- Fiecare valoare din distributia de probabilitate poate fi calculata ca:

  $$P(X_1=x_1 \wedge \ldots X_n=x_n) = P(x_1,\ldots, x_n) = \Pi_{i=1,n} P(x_i \mid parinti(x_i))$$

  unde *parinti(x_i)* reprezinat valorile specifice ale variabilelor *Parinti(X_i)*

2.3 Construirea retelei

Cum sa construim o retea a.i. RB/DPC sa fie o buna reprezentare?

Ecuatia $\mathbf{P(x_1,\ldots,x_n) = \Pi_{i=1,n} P(x_i \mid parinti(x_i))}$ implica anumite relatii de independenta conditionala care pot ghida construirea retelei

$\mathbf{P(X_1=x_1 \wedge \ldots X_n=x_n) = P(x_1,\ldots,x_n) =}$

$\mathbf{P(x_n \mid x_{n-1},\ldots,x_1) * P(x_{n-1},\ldots,x_1) = \ldots =}$

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

$\mathbf{\Pi_{i=1,n} P(x_i \mid x_{i-1},\ldots,x_1)}$ – valabila in general

- DPC daca, pt fiecare variabila X_i din RB

$\mathbf{P(X_i \mid X_{i-1},\ldots,X_1) = P(x_i \mid Parinti(X_i))}$ cu conditia ca

$\mathbf{Parinti(X_i) \subseteq \{ X_{i-1},\ldots,X_1\}}$

Construirea retelei

Pt fiecare variabila X_i din RB

$\mathbf{P(X_i \mid X_{i-1},\dots, X_1) = P(x_i \mid Parinti(X_i))}$ cu conditia ca

$$\mathbf{Parinti(X_i) \subseteq \{ X_{i-1},\dots, X_1\}}$$

- O RB este o reprezentare corecta a domeniului cu conditia ca fiecare nod sa fie independent conditional de nondescendenti, fiind dati parintii lui

Construirea retelei

- Conditia poate fi satisfactuta prin etichetarea nodurilor intr-o ordine consitenta cu DAG

- Intuitiv, parintii unui nod X_i trebuie sa fie toate acele noduri
  $X_{i-1},\ldots, X_1$ care influenteaza direct X_i.

Construirea retelei - cont

- Alege o multime de variabile aleatoare relevante care descriu problema
- Alege o ordonare a acestor variabile
- **cat timp** mai sunt variabile **repeta**

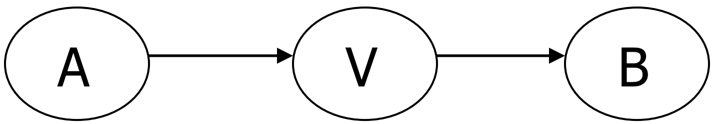
  (a) alege o variabila X_i si adauga un nod corespunzator lui X_i

  (b) atribuie Parinti(X_i) $\leftarrow$ un set minim de noduri deja existente in retea a.i. proprietatea de independenta conditionala este satisfacuta

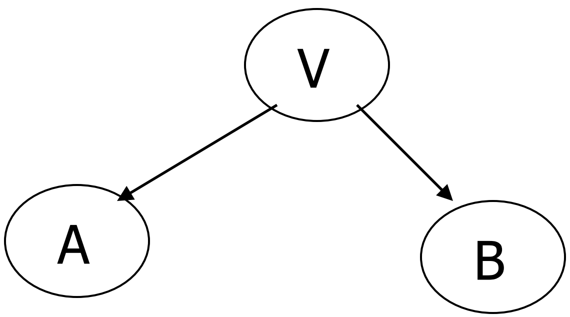
  (c) defineste tabela de probabilitati conditionate pentru X_i

Deoarece fiecare nod este legat numai la noduri anterioare $\rightarrow$ DAG

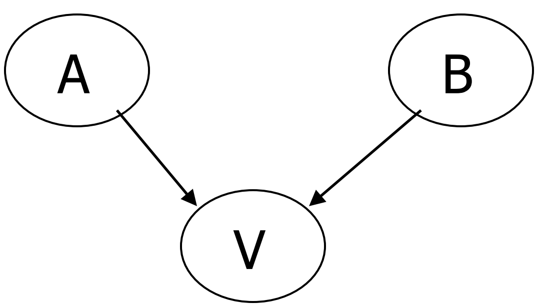
2.4 Inferente probabilistice



$$P(A \wedge V \wedge B) = P(A) * P(V|A) * P(B|V)$$

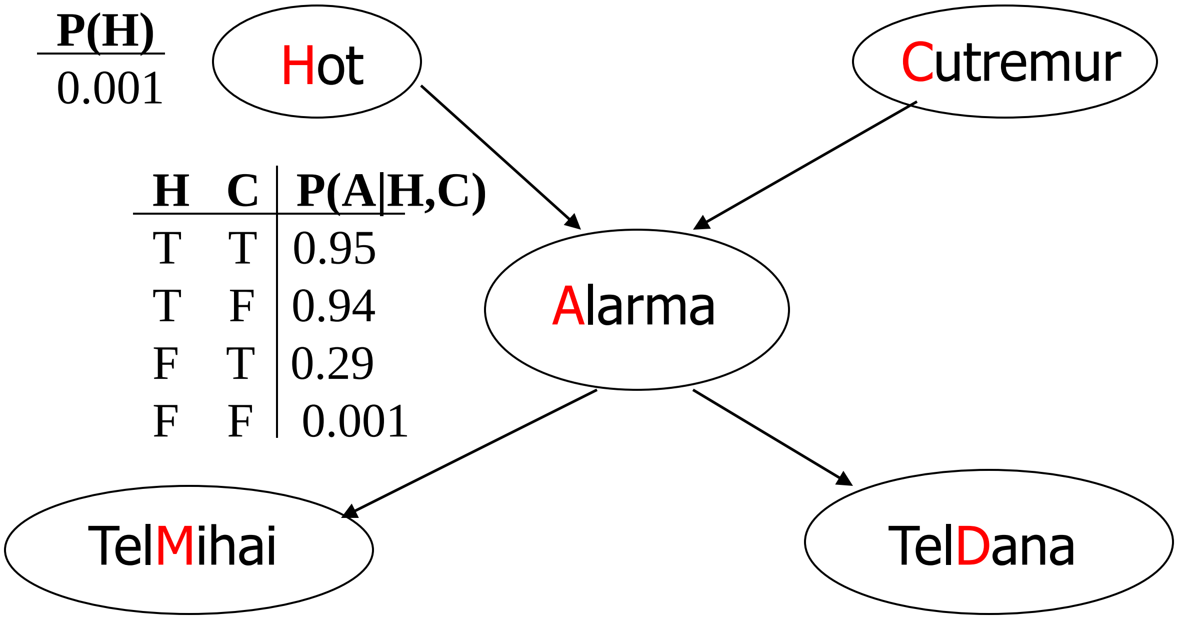


$$P(A \wedge V \wedge B) = P(V) * P(A|V) * P(B|V)$$



$$P(A \wedge V \wedge B) = P(A) * P(B) * P(V|A,B)$$

Inferente probabilistice



P(H)
0.001

P(C)
0.002

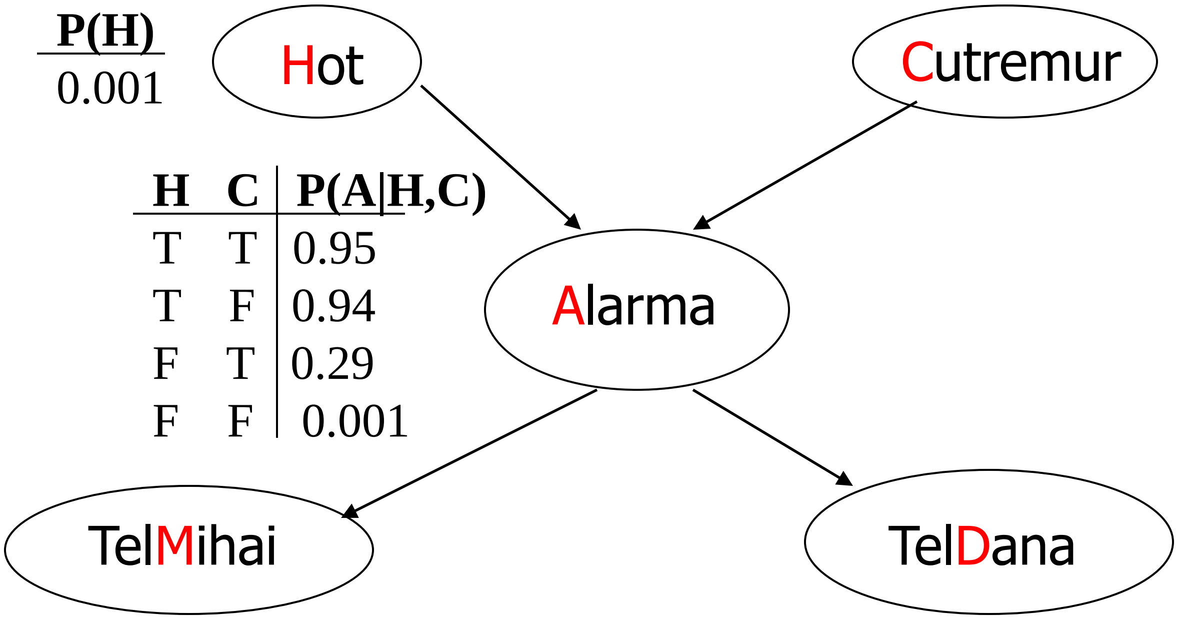
H	C	P(A|H,C)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(M|A)
T	0.9
F	0.05

A	P(D|A)
T	0.7
F	0.01

$$P(M \wedge D \wedge A \wedge \sim H \wedge \sim C) =$$
$$P(M|A) * P(D|A) * P(A|\sim H \wedge \sim C) * P(\sim H) \wedge P(\sim C) =$$
$$0.9 * 0.7 * 0.001 * 0.999 * 0.998 = 0.00062$$

Inferente probabilistice



P(H)
0.001

P(C)
0.002

H	C	P(A|H,C)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(M|A)
T	0.9
F	0.05

A	P(D|A)
T	0.7
F	0.01

$$P(A|H) = P(A|H,C) * P(C|H) + P(A|H,\sim C) * P(\sim C|H)$$

$$= P(A|H,C) * P(C) + P(A|H,\sim C) * P(\sim C)$$

$$= 0.95 * 0.002 + 0.94 * 0.998 = 0.94002$$