MADI

Réseaux bayésiens, modèles graphiques

Pierre-Henri Wuillemin

LIP6 – Université Paris 6, France

Présentation des Modèles Graphiques dans MADI

1. Modèle graphique probabiliste : les réseaux Bayésiens
2. Inférence exacte dans les BNs
3. Inférence approchée dans les BNs
4. Apprentissages dans les BNs
5. Extension à la décision : les diagrammes d'influence
6. Extension pour la représentation des préférences
7. Extension dans le cas des probabilités imprécises

Organisation

- **3 intervenants** : PH.Wuillemin, P.Perny, P.Viappiani (dans l'ordre d'apparition)
- **site du cours** : Moodle
- **Évaluation** : 30% Projet, 35% Réparti1, 35% Réparti2
- **Projet** : Sujet courant octobre, rendu et mini-soutenance dernier cours.

Un peu d'histoire (rapide et approximative) du siècle dernier

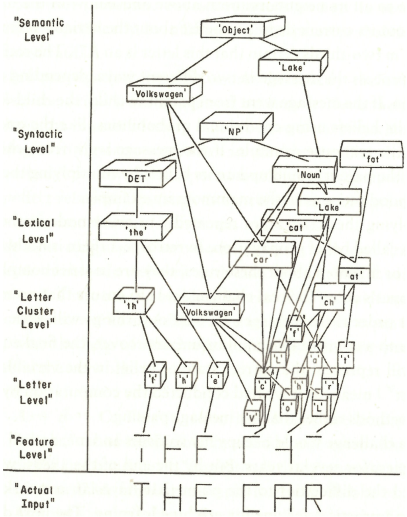
- Jusque 1970-80, raisonnement en IA = logique, rule-based, expert systems, etc.

$$\begin{cases} X \\ X \Rightarrow Y \end{cases} \models Y$$

  Chaînage avant, chaînage arrière, etc ...
- À partir de 1970, émergence de l'idée de raisonnement numérique. Par exemple :
  - Fuzzy sets (Zadeh, 1965)
  - Belief functions (Shafer, Dempter, 1967, ...)

  mais
  - perte d'information sur les séquences
  - pas d'utilisation simultanée en chaînage avant/arrière (prédiction/diagnostic)
  - modélisation de l'expert plutôt que du système étudié

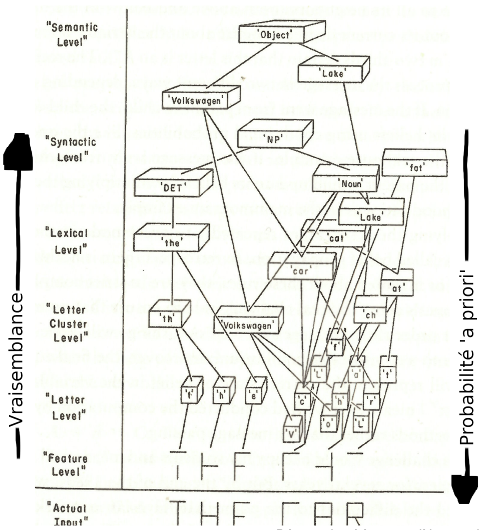
Flux d'information dans le modèle : Rumelhart, 1976



En général, les flux d'information vont de la prise en compte des données vers les paramètres du modèle mais aussi des paramètres du modèle vers la prise en compte des données.

Flux d'information dans le modèle : version probabiliste

$P(modele|data)$ **et** $P(data|modele)$

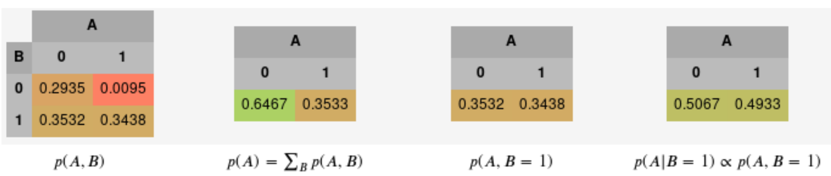


Inférence dans les modèles probabilistes discrets

Inférence

Soient A, B 2 variables aléatoires (discrètes),
L'inference consiste à calculer la distribution de A étant donné des observations sur B.

$$P(A|\color{red}{B = \epsilon_b}) = P(A|\epsilon_b) = \frac{P(\epsilon_b|A) \cdot P(A)}{P(\epsilon_b)} \propto P(\epsilon_b|A) \cdot P(A) = P(A, \epsilon_b)$$



	A	
B	0	1
0	0.2935	0.0095
1	0.3532	0.3438

$p(A, B)$

A	
0	1
0.6467	0.3533

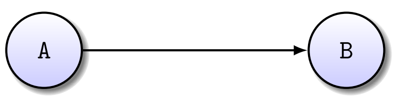
$p(A) = \sum_B p(A, B)$

A	
0	1
0.3532	0.3438

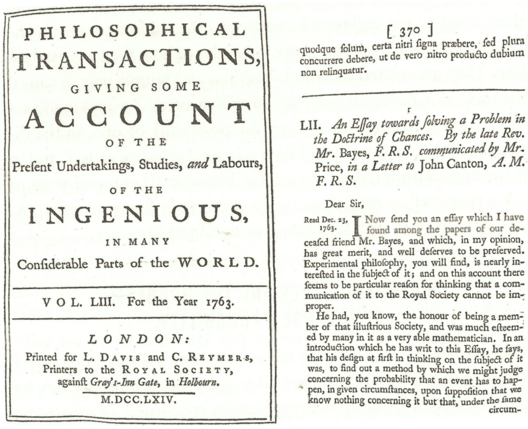
$p(A, B = 1)$

A	
0	1
0.5067	0.4933

$p(A|B = 1) \propto p(A, B = 1)$



Pourquoi ne pas utiliser les probabilités ?



PHILOSOPHICAL TRANSACTIONS, GIVING SOME ACCOUNT OF THE Preſent Undertakings, Studies, *and* Labours, OF THE INGENIOUS, IN MANY Conſiderable Parts of the WORLD.

VOL. LIII. For the Year 1763.

LONDON: Printed for L. DAVIS and C. REYMERS, Printers to the ROYAL SOCIETY, againſt *Gray's-Inn Gate*, in *Holbourn*.

M.DCC.LXIV.

[370]

quodque ſolum, certa nitri ſigna præbere, ſed plura concurrere debere, ut de vero nitro producto dubium non relinquatur.

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays, that his deſign at firſt in thinking on the ſubject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumſtances, upon ſuppoſition that we know nothing concerning it but that, under the ſame circum-

Complexités !

- Complexité spatiale
- Complexité temporelle
- Complexité en nombre de modèles

Inférence dans les modèles probabilistes discrets complexes

$$P(A|B=\epsilon_b) = P(A|\epsilon_b) = \frac{P(\epsilon_b|A)\cdot P(A)}{P(\epsilon_b)} \propto P(\epsilon_b|A)\cdot P(A) = P(A,\epsilon_b)$$

Systèmes complexes probabilistes : ⚠ $P(X_1,\cdots,X_n)$ 2^n paramètres !

Inference dans les systèmes complexes : $P(X_i|X_j=\epsilon_j) \propto P(X_i,\epsilon_j)$

$$P(X_i|\epsilon_j) = \frac{\displaystyle\sum_{k\notin\{i,j\}} P(x_1,\cdots,X_i,\cdots,\epsilon_j,\cdots,x_n)}{\displaystyle\sum_{k\neq j} P(x_1,\cdots,x_i,\cdots,\epsilon_j,\cdots,x_n)}$$

⚠ explosion combinatoire, curse of dimensionality !

FactorisationS de distribution jointe

On voudrait représenter une distribution jointe comme un produit de distributions marginales ou conditionnelles.

(on ne se pose pas de question de probabilité=0 : toutes les lois sont strictement positives)

$$\begin{aligned} P(A,B) &= P(A|B)\cdot P(B) \\ &= P(B|A)\cdot P(A) \\ &= P(A)\cdot P(B) \text{ si } A \perp\!\!\!\perp B \end{aligned}$$

Complexité en nombre de modèles ?

$$P(X_1,\cdots,X_n)?$$

Combien de modèles possibles ?

Un exemple : Altitude et Rain

Altitude et Rain

En regardant les villes d'un pays, on s'aperçoit qu'il y a une relation (hypothètique) entre Altitude (basse, moyenne, haute) de la ville et sa pluviométrie Rain (Low,High).

Comment représenter ce modèle probabiliste ?

$$P(A, R) = P(A|R) \cdot P(R) = P(R) \cdot P(R|A)$$

A		
0	1	2
0.5000	0.3000	0.2000

$p(A)^*$

	R	
A	0	1
0	0.2000	0.8000
1	0.4000	0.6000
2	0.6000	0.4000

$p(R|A)$

	A		
R	0	1	2
0	0.1000	0.1200	0.1200
1	0.4000	0.1800	0.0800

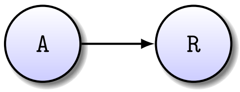
$= p(A, R) =$

R	
0	1
0.3400	0.6600

$p(R)*$

	A		
R	0	1	2
0	0.2941	0.3529	0.3529
1	0.6061	0.2727	0.1212

$p(A|R)$



exemple 2 : Altitude, Rain, Happy

Altitude, Rain, Happy

On s'aperçoit que les habitants d'une ville sont plus heureux quand il fait beau, quelle que soit l'altitude de la ville.

$$P(A, R, H) = P(A) \cdot P(R|A) \cdot P(H|R)$$

A		
0	1	2
0.5000	0.3000	0.2000

$p(A)$

	R	
A	0	1
0	0.2000	0.8000
1	0.4000	0.6000
2	0.6000	0.4000

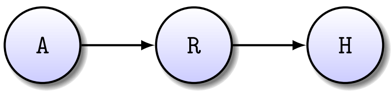
$p(R|A)$

	H	
R	0	1
0	0.8000	0.2000
1	0.4000	0.6000

$p(H|R)$

		R	
H	A	0	1
0	0	0.0800	0.1600
1	0	0.0200	0.2400
0	1	0.0960	0.0720
1	1	0.0240	0.1080
0	2	0.0960	0.0320
1	2	0.0240	0.0480

$p(A, R, H)$



Exemple 3 : Altitude, Rain et Demography

Altitude, Rain et Demography

Une variable Demography (haute ou basse) est ajouté dans l'exemple 1 : on considère qu'une ville est d'autant plus peuplée que son altitude est basse (quelle que soit sa pluviométrie).

$$P(A, R, D) = P(A) \cdot P(R|A) \cdot P(D|A)$$

	D	
A	0	1
0	0.1000	0.9000
1	0.3000	0.7000
2	0.6000	0.4000

$p(D|A)$

A		
0	1	2
0.5000	0.3000	0.2000

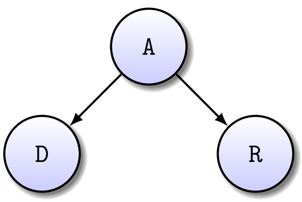
$p(A)$

	R	
A	0	1
0	0.2000	0.8000
1	0.4000	0.6000
2	0.6000	0.4000

$p(R|A)$

		A		
D	R	0	1	2
0	0	0.0100	0.0360	0.0720
1	0	0.0900	0.0840	0.0480
0	1	0.0400	0.0540	0.0480
1	1	0.3600	0.1260	0.0320

$p(A, R, D)$



Exemple 4 : Altitude, Rain et Sea

Altitude, Rain et Sea

Une variable Sea (proche ou éloigne) indique la proximité de la mer : on considère qu'une ville est d'autant plus pluvieuse qu'elle est proche de la mer.

$$P(A, R, S) = P(A) \cdot P(S) \cdot P(R|A, S)$$

A: 0	1	2
0.5000	0.3000	0.2000

$p(A)$

A	S	R: 0	1
0	0	0.3000	0.7000
1	0	0.5000	0.5000
2	0	0.5000	0.5000
0	1	0.4000	0.6000
1	1	0.2000	0.8000
2	1	0.6000	0.4000

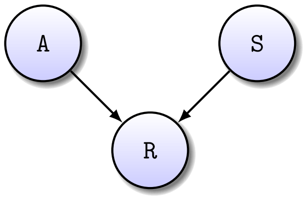
$p(R|A, S)$

S: 0	1
0.7000	0.3000

$p(S)$

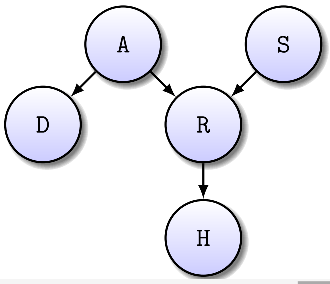
A	R	S: 0	1
0	0	0.1050	0.0600
1	0	0.1050	0.0180
2	0	0.0700	0.0360
0	1	0.2450	0.0900
1	1	0.1050	0.0720
2	1	0.0700	0.0240

$p(A, R, S)$



Exemple complet : Altitude, Rain, Happy, Demography et Sea

$$P(A, R, H, D, S) = P(A) \cdot P(S) \cdot P(R|A, S) \cdot p(H|R) \cdot P(D|A)$$



H	A	S	D	R: 0	1
0	0	0	0	0.0084	0.0098
1	0	0	0	0.0021	0.0147
0	1	0	0	0.0252	0.0126
1	1	0	0	0.0063	0.0189
0	2	0	0	0.0336	0.0168
1	2	0	0	0.0084	0.0252
0	0	1	0	0.0048	0.0036
1	0	1	0	0.0012	0.0054
0	1	1	0	0.0043	0.0086
1	1	1	0	0.0011	0.0130
0	2	1	0	0.0173	0.0058
1	2	1	0	0.0043	0.0086
0	0	0	1	0.0756	0.0882
1	0	0	1	0.0189	0.1323
0	1	0	1	0.0588	0.0294
1	1	0	1	0.0147	0.0441
0	2	0	1	0.0224	0.0112
1	2	0	1	0.0056	0.0168
0	0	1	1	0.0432	0.0324
1	0	1	1	0.0108	0.0486
0	1	1	1	0.0101	0.0202
1	1	1	1	0.0025	0.0302
0	2	1	1	0.0115	0.0038
1	2	1	1	0.0029	0.0058

A: 0	1	2
0.5000	0.3000	0.2000

$p(A)$

S: 0	1
0.7000	0.3000

$p(S)$

A	D: 0	1
0	0.1000	0.9000
1	0.3000	0.7000
2	0.6000	0.4000

$p(D|A)$

A	S	R: 0	1
0	0	0.3000	0.7000
1	0	0.5000	0.5000
2	0	0.5000	0.5000
0	1	0.4000	0.6000
1	1	0.2000	0.8000
2	1	0.6000	0.4000

$p(R|A, S)$

R	H: 0	1
0	0.8000	0.2000
1	0.4000	0.6000

$p(H|R)$

Bayesian networks : definition

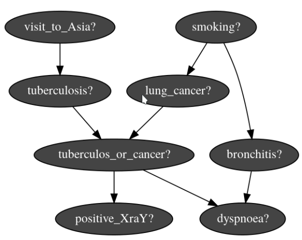
Définition (Bayesian Network (BN))

Un réseau bayésien (BN) est une distribution jointe sur un ensemble de variables aléatoires (discrètes). Un réseau bayésien est décrit par un graphe orienté sans circuit (DAG) et un ensemble de tables de probabilités conditionnelles (CPT) pour chaque variable aléatoire (noeuds du graphe) $P(X_i|\text{parents}_i)$

Factorization of the joint distribution in a BN

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P\left(X_i|\text{parents}(X_i)\right)$$

$P(A, S, T, L, O, B, X, D)$



$P(A) \cdot P(S) \cdot P(T|A) \cdot P(L|S) \cdot P(O|T, L) \cdot P(B|S) \cdot P(X|O) \cdot P(D|O, B)$

($2^8 = 256$ parameters)

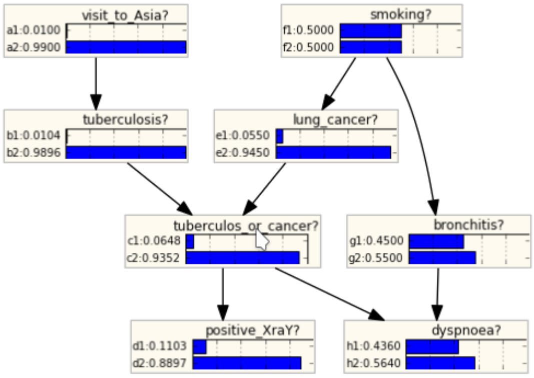
(2+2+4+4+8+4+4+8 = 32 parameters)

Bayesian networks : inférence

Définition (Bayesian Network (BN))

Un réseau bayésien est décrit par un graphe orienté sans circuit (DAG) et un ensemble de tables de probabilités conditionnelles (CPT) pour chaque variable aléatoire (noeuds du graphe) $P(X_i | parents_i)$

Inference : $P(\textbf{dyspnoea})$ **?**

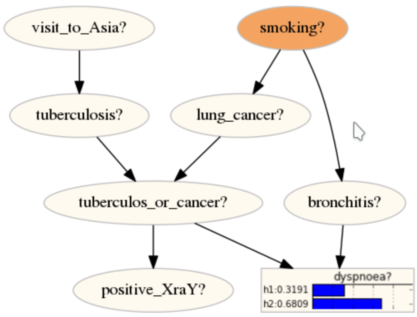


Bayesian networks : inférence

Définition (Bayesian Network (BN))

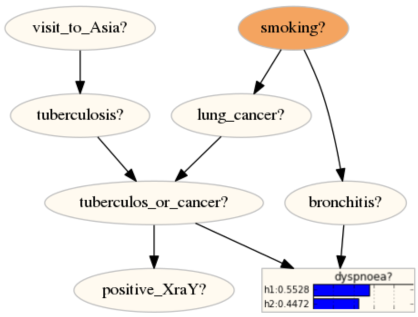
Un réseau bayésien est décrit par un graphe orienté sans circuit (DAG) et un ensemble de tables de probabilités conditionnelles (CPT) pour chaque variable aléatoire (noeuds du graphe) $P(X_i | parents_i)$

Inference : $P(\textbf{dyspnoea}|\textbf{smoking})$ **?**



Inference in 14.29ms

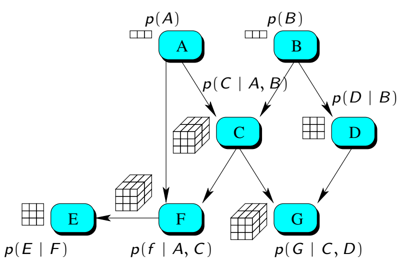
$P(dyspnoea|smoking = 1)$



Inference in 11.35ms

$P(dyspnoea|smoking = 0)$

Utilisations des BNs : inférence probabiliste



diagnostic : $P(A \mid F)$

- *diagnostic de panne*
- *sûreté de fonctionnement*
- *filtrage de spams*

prédiction $P(E \mid B, A)$

- *Simulation de process (industriels)*
- *prévisions boursières*
- *modélisation de joueurs*

Autres tâches

- *Cas le plus probable :* $\arg\max P(\mathfrak{X} \mid D)$
- *Analyse de sensibilité, information mutuelle, etc.*
- *Troubleshooting :* $\arg\max \frac{P(.)}{C(.)}$

```
https:
//www.lip6.fr/Pierre-Henri.Wuillemin/Docs/exoBN2019.tgz
```