MDI341

Introduction aux modèles graphiques : réseaux bayésiens et modèles de Markov cachés

Mars 2017

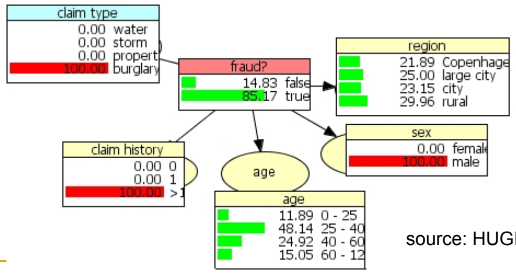
Laurence Likforman-Sulem
Telecom ParisTech/TSI
bureau : E 504
likforman@telecom-paristech.fr

Plan

- Modèles graphiques
- Réseaux bayesiens
- Réseaux bayesiens dynamiques (DBN)
- Lien avec les HMMs (Hidden Markov Models)

applications

- HMMs et réseaux bayésiens dynamiques:
  - reconnaissance de la parole,
  - de l'écriture, détection d'objets,
  - de visages dans les videos
- réseaux bayésiens statiques
  - Détection de fraudes, ex: HUGIN
  - diagnostic



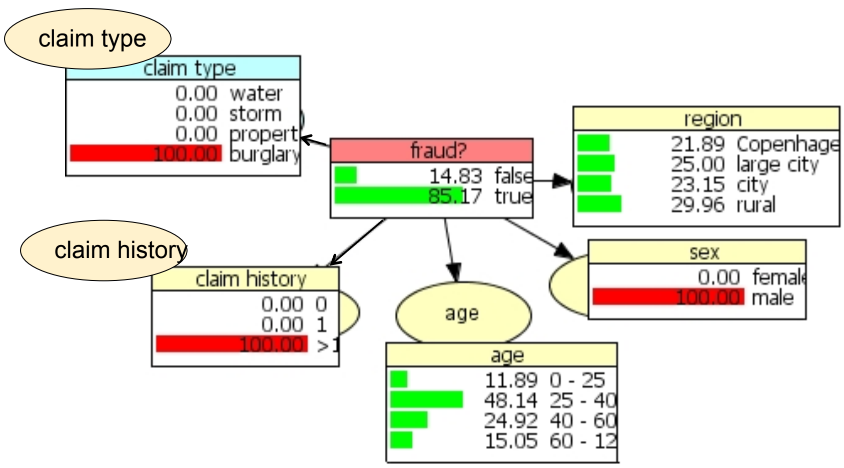
- observations

source: HUGINExpert

3

Application: fraud detection

inférer les valeurs de variables non observées, à partir de variables observées



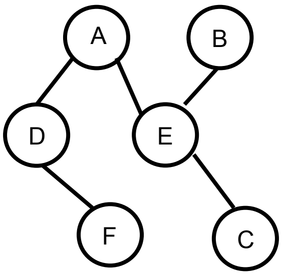
observations

source: HUGINExpert

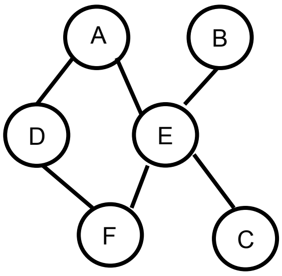
Modèles graphiques: graphes non dirigés

- arbre : graphe non dirigé + 1 seul chemin entre 2 nœuds
- graphe multi-connexe : non dirigé + plusieurs chemins entre noeuds

Arbre simple



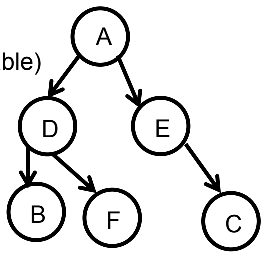
Graphe multi-connexe



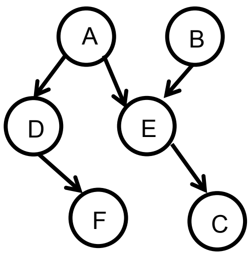
Modèles graphiques: graphes dirigés

- notion de fils, parents, descendants, ascendants, frères

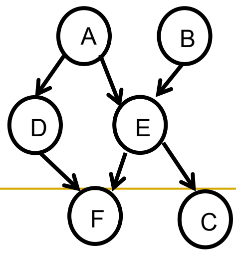
Arbre simple
(1 parent/variable)



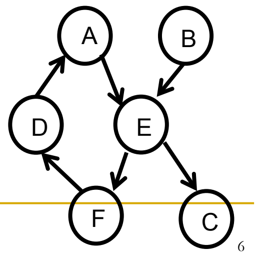
Poly-arbre
(1 seul chemin possible entre 2 nœuds, plusieurs parents possibles)



Graphe multi-connexe acyclique



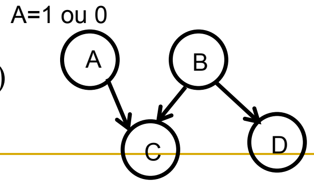
Graphe cyclique



réseau bayésien

- réseau bayésien: (G, θ)
- G=(V, E) graphe acyclique orienté
  - V : variables (nœuds),
  - E : arcs: relations entre variables (influence d'une variable sur une autre)
- θ : paramètres (tables de probabilité locales)
  - P(X|parents(X)), P(X) si X n'a pas de parent
- réseau: encode la probabilité jointe de manière factorisée
  - apprentissage: moins de paramètres: 1+1+4+2=8 au lieu de 2^4-1=15
  - inférence moins complexe

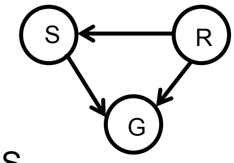
$$P(A,B,C,D) = P(A)P(B)P(C|A,B)P(D|B)$$



exemple réseau bayésien

[from Wikipedia BN]

- G: herbe du jardin humide (1 ou 0)
- S: arroseur en marche (1 ou 0)
- R: il a plu pendant la nuit (1 ou 0)



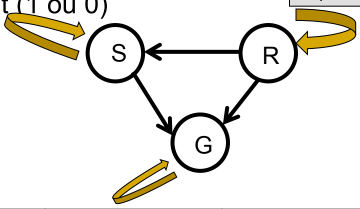
- S→ G : G est une conséquence de S
- R→ S: R a une influence sur S (s'il pleut, pas besoin de mettre en route l'arrosage)
- R→ G: G est une conséquence de R

- réciproquement: la connaissance de G modifie la « croyance » de S et de R : P(R=1|G=1) > P(R=1))

paramètres d'un réseau bayésien

- tables de probabilités locales : P(X| parents(X))
- Conditional Probability Tables or Distributions (CPT and CPD)
- G: herbe du jardin humide (1 ou 0)
- S: arroseur en marche (1 ou 0)
- R: il a plu pendant la nuit (1 ou 0)

P(R=0)=0.8
P(R=1)=0.2

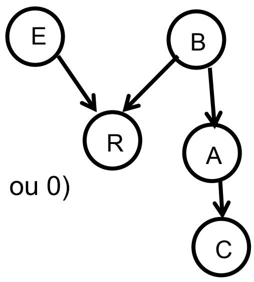


R	P(S=1|R)	P(S=0| R)
0	0.4	0.6
1	0.01	0.99

S	R	P(G=0| S,R)	P(G=1| S,R)
1	1	0.01	0.99
1	0	0.1	0.9
0	1	0.2	0.8
0	0	1.0	0.0

réseau bayésien: indépendance conditionnelle

- un nœud est indépendant de ses non-descendants connaissant ses parents
- C: le voisin appelle (1 ou 0)
- A: alarme maison s'est déclenchée (1 ou 0)
- B: il y a eu un cambriolage (1 ou 0)
- R: la radio annonce un tremblement de terre (1 ou 0)
- E: il y a un tremblement de terre (1 ou 0)

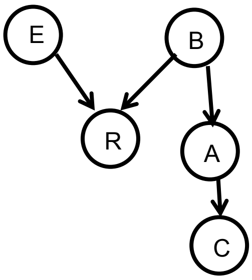


[d'après Nir Friedman]

- C est indépendant de R, B, E connaissant la valeur de A

Bayesian network: mini TD

- exercice : calculer le nombre de parametres de ce réseau

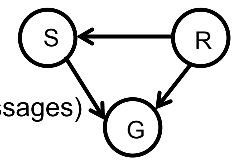


[from Nir Friedman]

11

Inférence dans un réseau bayésien

- calcul de P(variable(s) | variable(s) observée(s))
- variables observées: « evidence » ou observation
- inférence ou « propagation de l'observation »



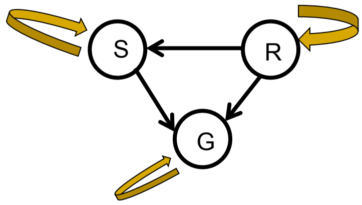
- algorithmes d'inférence:
  - exacts dans arbres, polyarbres (passages de messages)
  - stochastiques : échantillonnage

12

mini-TD : inférence

G observée : G=1, calculer P(R=1| G=1)

P(R=0)=0.8
P(R=1)=0.2

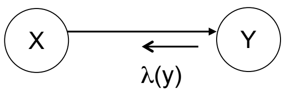


R	P(S=1|R)	P(S=0| R)
0	0.4	0.6
1	0.01	0.99

S	R	P(G=0| S,R)	P(G=1| S,R)
1	1	0.01	0.99
1	0	0.1	0.9
0	1	0.2	0.8
0	0	1.0	0.0

inférence dans une chaîne

- 2 nœuds dont 1 observé



$E = \{Y = y_2\}$

- paramètres: P(X=x) avec dom{X}={x1,x2}, et $P_{Y|X}$ avec dom{Y}={y1,y2, y3},

$$P_{Y|X} = \begin{bmatrix} P(Y = y_1 | X = x_1) & P(Y = y_2 | X = x_1) & P(Y = y_3 | X = x_1) \\ P(Y = y_1 | X = x_2) & P(Y = y_2 | X = x_2) & P(Y = y_3 | X = x_2) \end{bmatrix}$$

- variable backward λ:

$$\lambda(y) = P(E | Y = y) = [0 \quad 1 \quad 0]$$

$$\lambda(x) = P(E | X = x)$$

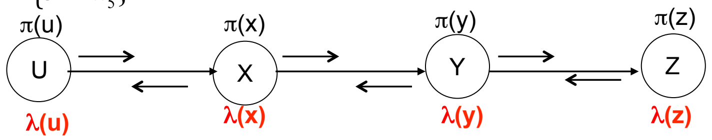
on calcule λ(x) à partir du « message » λ(y) envoyé par Y à X

$$\text{on a } \lambda(x) = P_{Y|X}\lambda(y) = \begin{bmatrix} P(Y = y_2 | X = x_1) \\ P(Y = y_2 | X = x_2) \end{bmatrix}$$

$$P(X = x, E) = P(E = e | X = x) P(X = x) = \lambda(x) P(X = x)$$

inférence dans une chaîne (backward)

$E^+ = \{U = u_5\}$ $\qquad$ $E^- = \{Z = z_4\}$



4 nœuds dont 2 observés (Z et U), observations amont E^+ et aval E^-:
dom{Z}={z_1, z_2, z_3, z_4}
dom{U}={u_1, u_2, u_3, u_4 , u_5}

on pose variable backward $\lambda(z) = P(E^- | Z = z) = [0 \quad 0 \quad 0 \quad 1]$

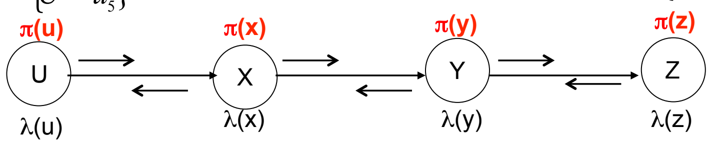
on a $\lambda(y) = P_{Z|Y}\lambda(z)$

et $\lambda(x) = P_{Y|X}\lambda(y)$

on calcule λ(y) à partir du message λ(z) envoyé par Z à Y puis λ(x) à partir du message λ(y) envoyé par Y à X

inférence dans une chaîne (forward)

$E^+ = \{U = u_5\}$ $\qquad$ $E^- = \{Z = z_4\}$



4 nœuds dont 2 observés

on calcule π(x) à partir du message π(u) envoyé par U à son enfant X

on pose variable forward $\pi(x) = P(X = x, E^+)$

on a $\pi(u) = [\ 0 \quad 0 \quad 0 \quad 0 \quad 1]$

on a $\pi(x) = \pi(u)P_{X|U}$

et $\pi(y) = \pi(x)P_{Y|X}$

$$\boxed{\begin{aligned} P(X = x, E) &= \lambda(x)\pi(x) \\ P(X = x|E) &\propto \lambda(x)\pi(x) \end{aligned}}$$

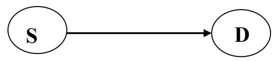
facteur normalisation 1/P(E) près

$$P(E) = \sum_{x \in dom\{X\}} \lambda(x)\pi(x)$$

exemple (chaîne)

extrait de P. Naïm et al. Réseaux Bayésiens

S: sexe : F ou M
D: daltonien: oui ou non



	D=oui	D=non
S=F	0.005	0.995
S=M	0,08	0.92

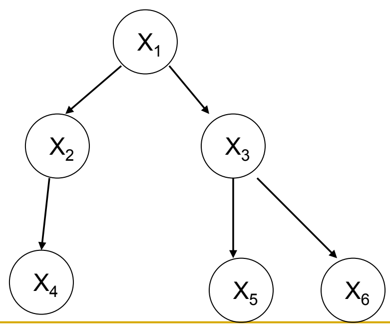
On observe :E={ D=oui}   calculer par algorithme d'inférence (passage de message) P(S|E)

quelle est la probabilité d'être une femme si on est daltonien ?

on considèrera que la proportion H/F est égale dans la population.

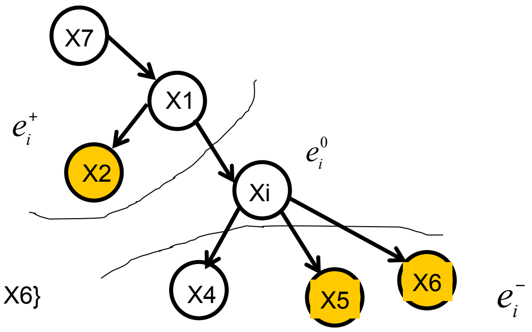
inference dans un arbre

- variables observées :X_1=2, X_4=1, X_5=1, X_6=2
- variables cachées (états) : X_2, dom{X_2}={3,4,5} X_3, dom{X_3} ={6,7}



inférence dans les arbres

- e_i^+: variables observées en amont de Xi
- e_i^-: variables observées en aval de Xi
- e_i^0: valeur observée de Xi, si Xi est observée
- $E = \{e_i^0, e_i^+, e_i^-\}$



observation: E={X2, X5, X6}

inférence dans les arbres

- on définit pour nœud Xi
  - variable λ (backward)
  - variable π forward

$$\lambda(x_i) = P(e_i^0, e_i^- | X_i = x_i)$$

$$\pi(x_i) = P(e_i^+, X_i = x_i)$$

$$P(E, X_i = x_i) = P(e_i^0, e_i^-, e_i^+, X_i = x_i)$$

$$\boxed{P(X_i = x_i, E) = \lambda(x_i)\pi(x_i)}$$

$$P(E) = \sum_{x_i \in dom\{X_i\}} \lambda(x_i)\pi(x_i)$$

$$P(X_i = x_i | E) = \frac{\lambda(x_i)\pi(x_i)}{\sum_{x_i \in dom\{X_i\}} \lambda(x_i)\pi(x_i)} \propto \lambda(x_i)\pi(x_i)$$

Algorithme d'inférence par passage de messages

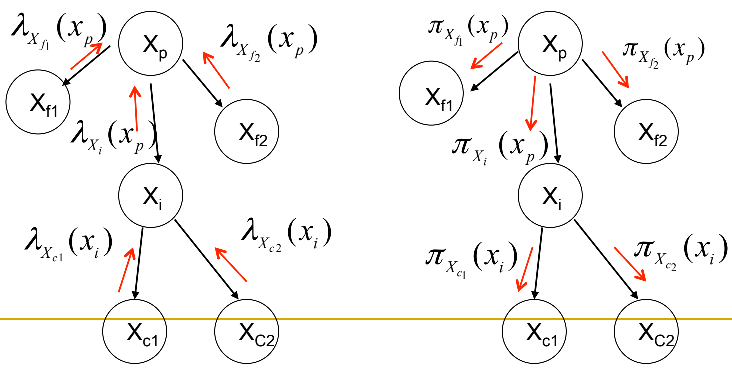
message λ des enfants de Xi, à Xi: $\lambda_{X_{c1}}(x_i)$

message λ de Xi à son parent Xp: $\lambda_{X_i}(x_p)$

message λ des frères de Xi à leur parent Xp: $\lambda_{X_{f2}}(x_p)$

message π envoyé par Xi à ses enfants: $\pi_{X_{c1}}(x_i)$

message π envoyé à Xi par son parent Xp $\pi_{X_i}(x_p)$



Algorithme d'inférence: calcul des λ

- les enfants X_{ci} envoient un message λ à leur parent Xi.

  $$\lambda_{X_{c1}}(x_i) = P_{X_{c1}|Xi} \bullet \lambda(x_{c1})$$

- si X_i feuille observée $\lambda(x_i) = [0 \quad 1 \quad 0]$ (position 1 dépend observation)
- si X_i feuille non observée: $\lambda(x_i) = [1 \quad 1 \quad 1]$
- si X_i nœud (pas feuille) caché :
- $$\lambda(x_i) = \prod_{Y:children(Xi)} \lambda_Y(x_i)$$
- avec $\lambda_Y(x_i)$ message de l'enfant Y de Xi, à Xi
- si X_i nœud (pas feuille) observé : $\lambda(x_i) = \prod_{Y:children(Xi)} \lambda_Y(x_i)$ pour xi=correspondant à l'observation, 0 sinon

Algorithme d'inférence: calcul des π

- de la racine aux feuilles
- si X_i racine est observée $\pi(x_i) = [0 \quad 1 \quad 0]$
- Si X_i est la racine, non observée $\pi(x_i) = P(X_i = x_i)$

- sinon: on utilise le message π : $\pi_{Xi}(x_p)$ envoyé à Xi par son unique parent Xp:

$$\pi(x_i) = \pi_{Xi}(x_p)\, P_{Xi|Xp}$$

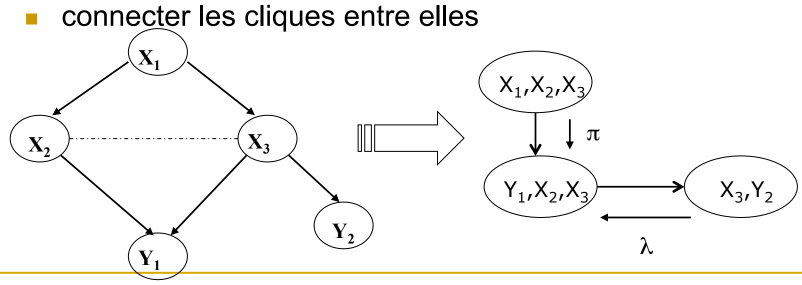
$$\pi_{X_i}(x_p) = \pi(x_p) \prod_{\text{Z, frères de Xi}} \lambda_Z(x_p)$$

- messages π envoyés par Xi à chacun de ses enfants (message λ du ou des autres enfants et π du parent de Xi)

$$\pi_{X_{c1}}(x_i) = \prod_{\text{X}_\text{f} \text{ freres de X}_\text{ci}} \lambda_{X_f}(x_i)\, \pi(x_i) = \prod_{X_f} \lambda_{X_f}(x_i)\, \pi_{X_i}(x_p)\, P_{X_i|X_p}$$

inférence dans un réseau quelconque (DAG)

- convertir le réseau dans une structure d'arbre (arbre de jonction)
- algorithme d'inférence de l'arbre de jonction [Jensen, 96] [Zweig, 2003]
  - moralisation (connecter les parents)
  - triangulation (former des cliques)
  - connecter les cliques entre elles

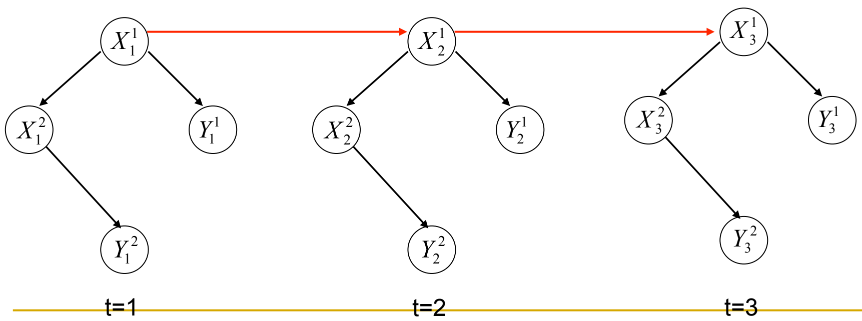


apprentissage des réseaux bayésiens

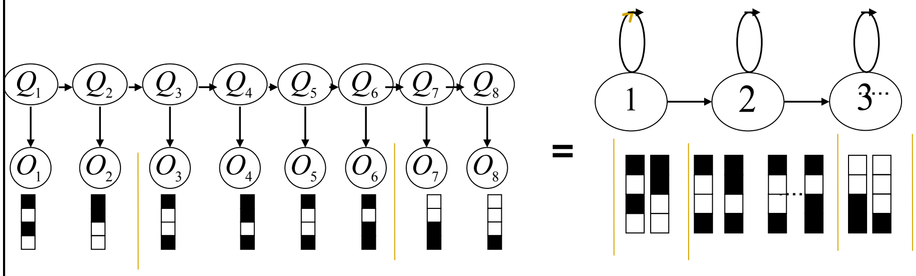
- apprentissage en données complètes, structure connue
  - estimation des paramètres $P(variable|Parents)$
  - par estimation au maximum de vraisemblance
- en données incomplètes, structure connue
  - algorithme EM ou descente de gradient ou méthodes stochastiques MCMC (échantillonnage de Gibbs)
- apprentissage de la structure, données complètes
  - algorithmes gloutons
- apprentissage de la structure, données incomplètes
  - EM + algorithmes gloutons

extension: réseaux bayésiens dynamiques (DBN)

- modélisent des processus stochastiques stationnaires
  - structure & paramètres ne varient pas au cours du temps
  - plusieurs variables d'états et d'observations pour chaque pas de temps



un modèle de Markov caché est un cas particulier de DBN



- HMM: Hidden Markov Model
- structure d'arbre
- 1 variable d'état + 1 variable d'observation à chaque instant t

$(Q_t)_{1 \le t \le T}$: variables d'état (cachées)

$(O_t)_{1 \le t \le T}$: variables d'observation (observées) générées par les états

références

- Wikipedia BN: http://en.wikipedia.org/wiki/Bayesian_network
- A. W. Moore, *Bayes Nets for representing and reasoning about uncertainty,* 2001
  - http://www.cs.cmu.edu/%7Eawm/tutorials
- S. Davis, A. Moore Bayesian networks: independencies and inference
  - http://www.cs.cmu.edu/~awm/tutorials
- K. Murphy, BayesNet Toolbox for matlab https://code.google.com/p/bnt/
- N. Friedman, D. Koller, Learning Bayesian Networks from data
- G. Zweig, Speech Recognition with Dynamic Bayesian Networks, Phd thesis, 1998, Univ. of California, Berkeley,
- P. Leray, Réseaux Bayésiens-Apprentissage de la structure
  - http://asi.insa-rouen.fr/enseignants/~pleray/RB2003/2-ApprentissageStructure.pdf
- P. Naïm , P-H. Wuillemin , P. Leray , O. Pourret , A. Becker, Les réseaux Bayésiens, Eyrolles, 2007.
- M. Sigelle, Bases de la Reconnaissance des Formes: Chaînes de Markov et Modèles de Markov Cachés, chapitre 7, Polycopié Telecom ParisTech, 2012.
- J. Pearl, Probabilistic Reasoning in intelligent systems: networks of plausible inference networks, 1988.
- L. Likforman-Sulem, E. Barney Smith, Reconnaissance des Formes: théorie et pratique sous matlab, Ellipses, TechnoSup, 2013.