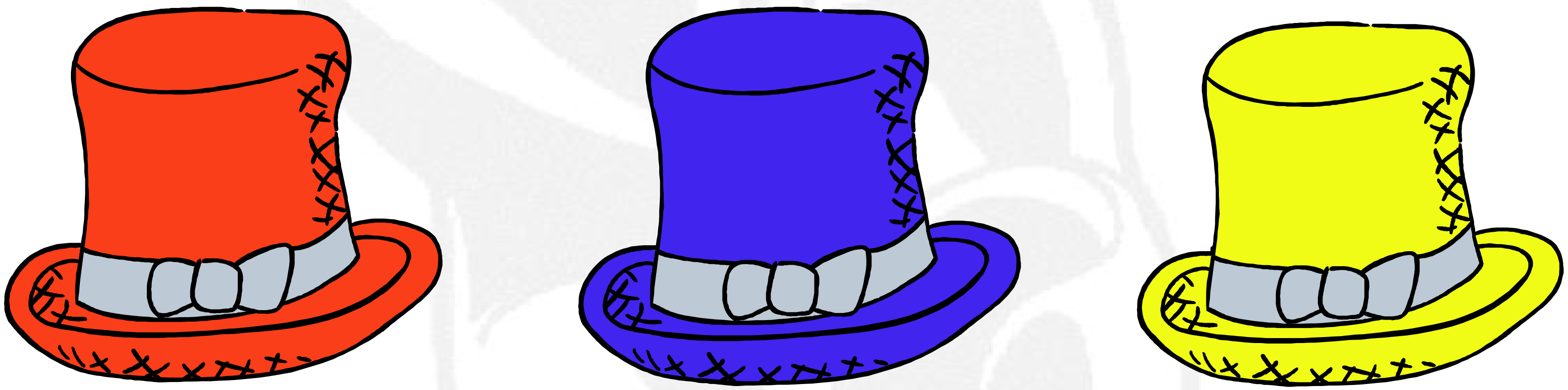
Les réseaux bayésiens

Un outil de modélisation des connaissances incertaines

- par apprentissage à partir des données
- par modélisation interactive

Petit exemple contre-intuitif

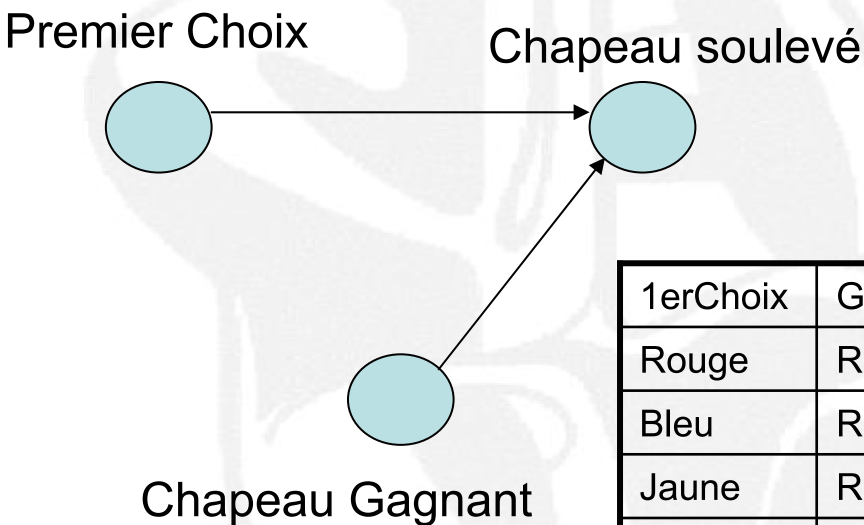
La voiture de vos rêves se cache sous un de ces chapeaux
Faites votre choix ?



Pour vous aider le présentateur va soulever un chapeau
Vous pouvez alors changer de choix ou non. Y a t'il une stratégie gagnante ?

Petit exemple contre-intuitif

Un réseau bayésien vous montrera que vous avez deux fois plus de chance de gagner si vous changez de choix



1erChoix	Gagnant	Rouge	Bleu	Jaune
Rouge	Rouge	0	0.5	0.5
Bleu	Rouge	0	0	1
Jaune	Rouge	0	1	0
Rouge	Bleu	0	0	1
Bleu	Bleu	0.5	0	0.5
Jaune	Bleu	1	0	0

Apport des réseaux bayésiens

- Ils conjuguent les avantages de diverses approches :
  - la compréhensibilité des modèles symboliques
  - les fondements probabilistes rigoureux des méthodes statistiques
  - la structure en réseau de composants simples des approches connexionistes

- Les réseaux bayésiens représentent **toutes** les relations entre les attributs (décrivant les exemples)
- Et permettent une utilisation **multidirectionnelle**

Axiomatique probabiliste (rappel)

Soit A,B et C des propositions

$$0 \leq P(A) \leq 1$$

$$P(Vrai) = 1, P(Faux) = 0$$

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

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

$$P(A|B) = P(A) \times \frac{P(B|A)}{P(B)}$$

$$P(A|BC) = P(A|C) \times \frac{P(B|AC)}{P(B|C)}$$

Interprétation probabiliste

- Premier point de vue
  - interprétation fréquentiste ou objectiviste
  - probabilités vues comme des propriétés réelles et mesurées des objets
- Second point de vue
  - interprétation bayésienne
  - probabilités vues comme des degrés de croyance d'un agent dans la vérité d'une proposition.
- Exemples :
  - « Mon dé a une chance sur 6 de tomber sur le 1 »
  - « Tous les éléphants sont gris »
  - Probabilité d'évènements jamais observés

Objectifs

- Le domaine étudié est représenté par un ensemble de variables (attributs) X_1, X_2,…, X_n
- On dispose des valeurs d'un sous-ensemble de ces variables (variables observées)
- On veut caractériser les valeurs possibles d'un autre sous-ensemble de variables (variables requises) sous la forme de leurs distributions de probabilités

Exemple simple

- Problème de diagnostic
  - on connaît les probabilités a priori des maladies
  - on connaît les probabilités des symptômes connaissant la maladie
  - on veut connaître la probabilité de la maladie en fonction des symptômes (probabilité conditionnelle)
- Exemple d'inférence : calcul de P(cavité|mal de dents)

$$P(cavit\acute{e}) = 0.1$$
$$P(mal\ de\ dents) = 0.05$$
$$P(mal\ de\ dents | cavit\acute{e}) = 0.4$$

$$P(cavit\acute{e} | mal\ de\ dents) = P(cavit\acute{e}) \times \frac{P(mal\ de\ dents | cavit\acute{e})}{P(mal\ de\ dents)}$$
$$= 0.1 \times \frac{0.4}{0.05} = 0.8$$

Intérêt de la Loi Conjointe

	cavité	¬cavité
mal de dents	0.04	0.01
¬mal de dents	0.06	0.89

Cette LC permet de calculer n'importe quelle valeur de probabilité du domaine

$$P(cavité \mid mal\ de\ dents) = \frac{P(cavité \wedge mal\ de\ dents)}{P(mal\ de\ dents)}$$

$$= \frac{0.04}{0.04 + 0.01} = 0.8$$

Cas général

- La connaissance de la LC est suffisante pour mener à bien toute inférence portant sur l'ensemble de variables $X_1, X_2, \ldots, X_n$
- La représentation de la LC nécessite un nombre de paramètres exponentiel par rapport au nombre de variables

Représentation compacte de la Loi Conjointe

- Si on exploite des indépendances conditionnelles,
- On peut écrire la LC sous la forme :

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i | Pa_i = pa_i)$$

- Au lieu d'un nombre de valeurs exponentiel par rapport au nombre de variables, on a besoin, pour chaque variable, d'un nombre de valeurs exponentiel par rapport au nombre de ses parents

Qu'est ce qu'un réseau bayésien ?

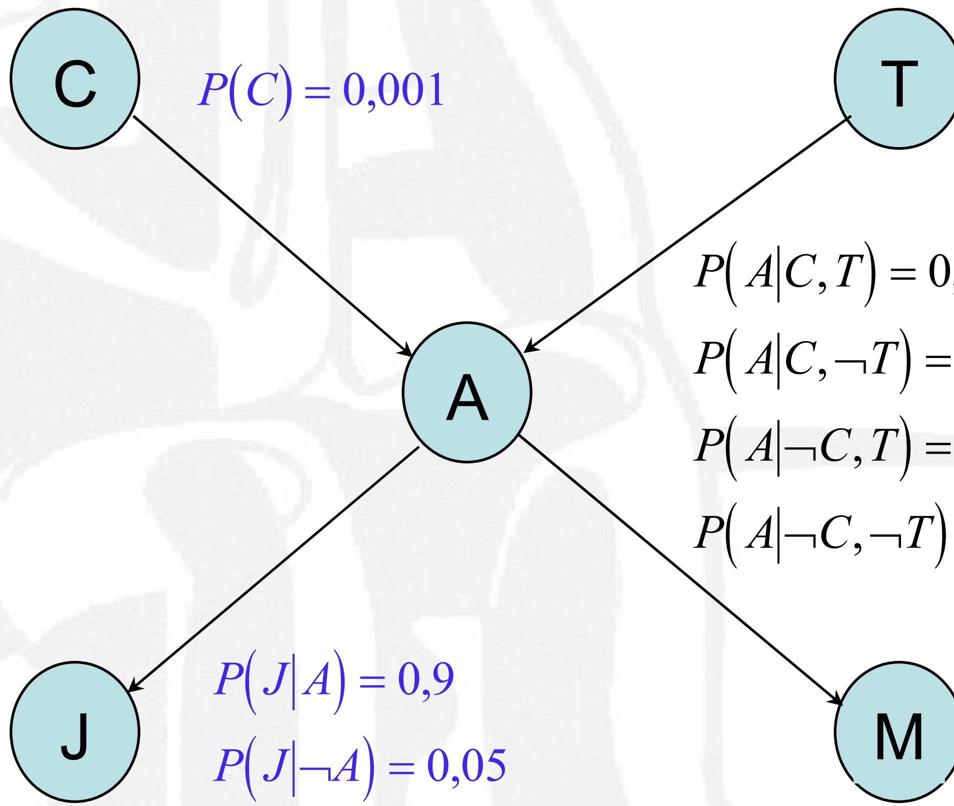
- Rôle
  - Représente la Loi Conjointe sous la forme compacte optimale
  - En exploitant les relations d'indépendance conditionnelle pour
    - réduire la quantité d'information nécessaire
    - simplifier les inférences
- Définition : graphe orienté sans cycle orienté dont
  - les nœuds sont étiquetés par les variables du domaine
  - les nœuds sont reliés entre eux par des arcs orientés
  - chaque nœud contient une table de probas conditionnelles

$$\overline{P}(X \mid parents(X)) \text{ où } parents(X) = \{U_1, U_2, \ldots, U_n\}$$

Exemple : contexte

- Système de sécurité doté d'une alarme
  - fonctionne correctement pour les cambriolages
  - se déclenche parfois à tort pour des tremblements de terre mineurs
- John et Mary se proposent d'appeler quand ils entendent l'alarme mais
  - John confond parfois la sonnerie du téléphone avec l'alarme
  - Mary n'entend pas toujours l'alarme

Exemple : réseau bayésien obtenu



C = cambriolage

T = Tremblement de terre

A = alarme

J = Appel de John

M = Appel de Mary

$$P(C) = 0{,}001$$

$$P(T) = 0{,}002$$

$$P(A|C,T) = 0{,}95$$

$$P(A|C,\neg T) = 0{,}94$$

$$P(A|\neg C,T) = 0{,}29$$

$$P(A|\neg C,\neg T) = 0{,}001$$

$$P(J|A) = 0{,}9$$

$$P(J|\neg A) = 0{,}05$$

$$P(M|A) = 0{,}7$$

$$P(M|\neg A) = 0{,}0$$

Inférence à partir des RB

- Problème général : cf inférence probabiliste
- Avantages par rapport aux autres modèles (AD, RN) :
  - rôle des variables non figé à l'avance
  - inférences possibles « dans tous les sens »
- L'inférence s'appuie sur
  - les tables de probabilités des nœuds
  - les indépendances conditionnelles (structure du réseau)
  - l'axiomatique probabiliste
- Complexité importante dans l'absolu et au delà des capacités humaines

Types d'inférences

- « Diagnostique »
  - des effets vers la cause : chaînage arrière
  - exemple : $P(C|J)$?
- Causale
  - des causes vers les effets : chaînage avant
  - exemple : $P(J|C)$?
- Inter-causale
  - entre les causes et un effet connu
  - exemple : $P(C|A \wedge T)$?
- Mixte : combine un ou plusieurs des types précédents
  - exemple : $P(A|J \wedge \neg T)$? (diagnostique + causale)

Apprentissage de réseaux bayésiens

- Apprentissage non-supervisé (associations et clustering), supervisé (caractérisation)
- Ses résultats peuvent être utilisés pour prédire la valeur de n'importe quelle variable à partir de n'importe quelles autres
- Perspectives nouvelles pour la découverte de relations causales
- Méthodes récentes et en plein essor

Méthodes d'apprentissage des RB

- Plusieurs méthodes selon que :
  - la structure du réseau est **connue** ou **non**
    - **apprentissage de paramètres**
    - **apprentissage de structure**
  - les valeurs de tous les attributs sont **disponibles** ou non
  - on dispose d'un réseau initial proposé par les experts ou **non**
  - on respecte à la lettre les principes de l'apprentissage bayésien ou **non**