

# Apprentissage de Réseaux Bayésiens Causaux à partir de données d'observation et d'expérimentation

Stijn Meganck, Sam Maes, Philippe Leray  
smeganck@vub.ac.be

Laboratoire d'Informatique, Traitement de l'Information, Systèmes.

JFRB  
Université de Valenciennes

18 mai Valenciennes

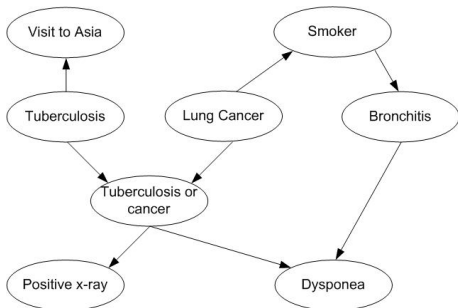


# Plan

- 1 Introduction
- 2 Apprentissage de Réseaux Bayésiens
- 3 Algorithmes
- 4 Conclusions

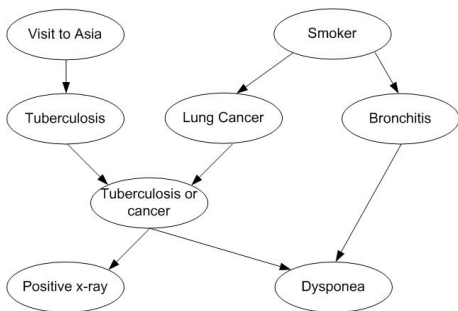
# Réseaux Bayésiens

- Modélisation de systèmes complexes
- Modèle graphique probabiliste
  - Modèle d'indépendance
    - Markov
  - Loi de probabilité
    - $P(V) = \prod_{X \in V} P(X|Pa(X))$
- Inférence probabiliste
  - $P(\text{cancer} | \text{smoker}, x - \text{ray})$



# Réseaux Bayésiens causaux

- Réseaux Bayésiens où toute relation entre variable correspond à une relation causale [Pearl, 2000]
  - $X \rightarrow Y$ :  $X$  est une cause de  $Y$
- Inférence causale
  - Manipulation
  - $P(\text{cancer} | \text{do}(TBC = \text{true}))$



# Apprentissage de Réseaux Bayésiens

- Structure connue
  - Apprentissage des paramètres
- Apprentissage de structure
  - Basé sur tests d'indépendance conditionnelle
  - Basé sur un score

# Apprentissage de Réseaux Bayésiens causaux

- Suffisance causale
  - Pas de variables latentes
- Données d'observation
  - Classe d'équivalence de Markov
  - Découvre tous les liens de dépendance
  - Ne découvre pas toutes les relations causales
- Expériences
  - Vérifier quelles variables agissent sur quelles autres

# Algorithme

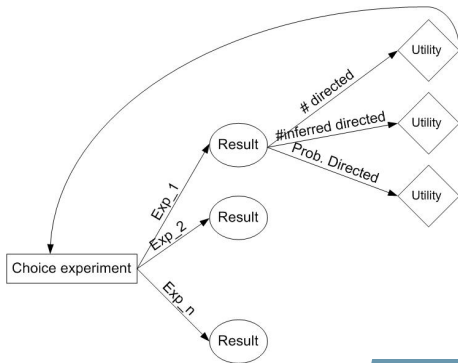
## Mécanisme

- Utilisation d'un algorithme traditionnel d'apprentissage de structure (SGS, PC [Spirtes et *al.*, 2000], IC [Pearl, 2000], ...)
  - Découverte du représentant de la classe d'équivalence de Markov
- Choisir l'expérience qui donne le plus d'information causale
- Exécuter l'expérience
- Interpréter les résultats
- Iteration

# Algorithme

## Choisir l'expérience

- Théorie de la Décision
- Maximaliser fonction d'utilité
  - $$U(A_{X_i}) = \frac{\alpha \text{gain}(A_{X_i})}{\beta \text{cost}(A_{X_i}) + \gamma \text{cost}(M_{X_i})}$$
- Critères de décision
  - Maximax
  - Maximin
  - Laplace
  - Utilité moyenne

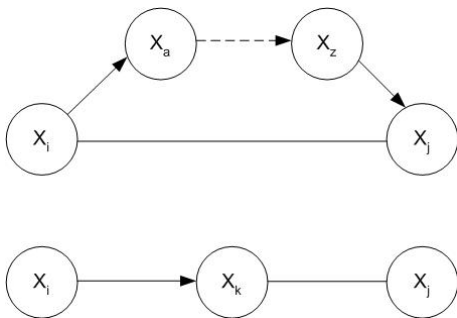




# Algorithme

## Inférer des arêtes

- la direction d'une arête peut en "forcer" d'autres
- Règles d'inférence
  - Si  $X_i - X_j$  et il existe un chemin dirigé reliant  $X_i$  et  $X_j$ , oriente  $X_i \rightarrow X_j$
  - Si  $X_i$  et  $X_j$  non adjacents,  $X_i \rightarrow X_k$  et  $X_k - X_j$ , oriente  $X_k \rightarrow X_j$
  - Répéter tant qu'il est possible d'orienter des arêtes



# Critères de décision

## Maximax

- Critère optimiste
- Choisir variable qui PEUT donner le plus d'information

- $$X_{best} = \underset{X_i}{\operatorname{argmax}} \left( \frac{Ne_U(X_i) + \max_{inst(A_{X_i})} (\#inferred(inst(A_{X_i})))}{cost(A_{X_i}) + cost(M_{X_i})} \right)$$

- Le nombre de voisins non dirigés + le plus haut nombre d'arêtes inféré pour une certaine instantiation

# Critères de décision

## Utilité moyenne

- Critère réaliste
- Choisir variable qui donne le plus d'information en moyenne

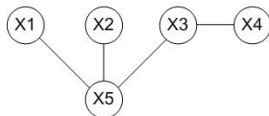
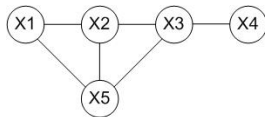
- $$\operatorname{argmax}_{X_i} \left( \frac{Ne_U(X_i) + \sum_{inst(A_{X_i})} \#inferred(inst(A_{X_i}))P(inst(A_{X_i}))}{cost(A_{X_i}) + cost(M_{X_i})} \right)$$

- **Problème**
  - Nombre de graphes dans classe d'équivalence de Markov est inconnu

# Critères de décision

## Approximation de l'utilité moyenne

- Nombre de graphes dans classe d'équivalence de Markov est inconnu
  - Cherche l'arbre "meilleur sous-graphe"
  - Applique MWST avec contraintes
    - Pas de nouveaux arêtes
    - Pas enlever des arêtes de  $X_i - Ne(X_i)$
  - Utilise arbre comme vrai structure
- $\#(\text{graphes dans classe d'équivalence Markov d'un arbre}) = N$



**Figure:** Exemple pour expérience sur  $X_5$

# Algorithme adaptatif

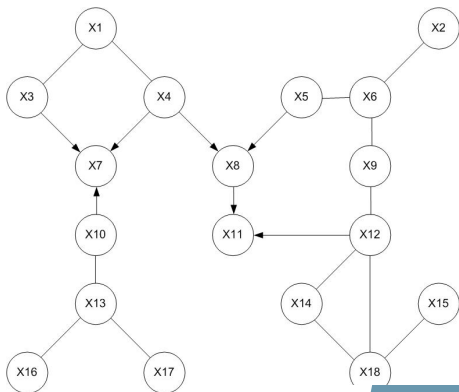
- 1 Calculer  $U(X)$  pour chaque nœud  $X$  tel que  $\#(Ne_U(x)) > 0$
- 2 Réaliser une intervention sur la variable  $X_{High}$  qui maximise  $U(X)$
- 3 Pour chaque  $Y \in Ne_U(X_{High})$ , si la distribution de  $Y$  change à cause de l'intervention, orienter  $X_{High} \rightarrow Y$ , sinon  $X_{High} \leftarrow Y$
- 4 Utiliser les règles d'inférence pour trouver des orientations
- 5 Retourner à (1) si toutes les arêtes n'ont pas été orientées

## Exemples

- Partager la structure en parties non-dirigés
- Dans chaque partie chercher la variable qui maximise  $U$ 
  - En parallèle
- Réaliser une intervention sur les variables "maximales"

Table de coût

$X_1$	$X_3$	$X_4$
$\infty$	2	3

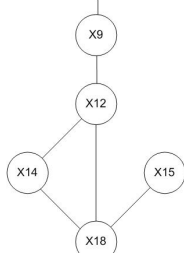
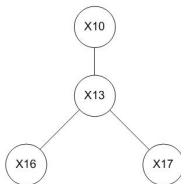
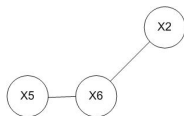
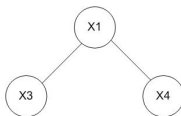


## Exemples

- Partager la structure en parties non-dirigés
- Dans chaque partie chercher la variable qui maximise  $U$ 
  - En parallèle
- Réaliser une intervention sur les variables "maximales"

Table de coût

	X1	X3	X4
X1	$\infty$	2	3



## Conclusions

- Les méthodes classiques donnent seulement la classe d'équivalence
- Besoin d'expériences pour compléter
- L'ordre des expériences:
  - Maximax
  - Maximin
  - Laplace
  - Utilité moyenne
- Algorithmes
  - Adaptatif



## Perspectives

- Des fautes dans le CPDAG
- Expériences multiples
- Apprentissage de réseaux bayésiens causaux dans le cas de problèmes qui n'ont pas la suffisance causale