

Olivier PARENT  
Julien EUSTACHE

# Les Réseaux Bayésiens

A la recherche de la vérité

Master 2 Recherche Connaissance et Raisonnement, 2006 – 2007  
Cours Cognition et connaissance - Alain MILLE

# TABLE DES MATIERES

Table des matières .....	2
Introduction.....	3
Réseau bayésien .....	4
Inférence bayésienne .....	6
Conclusion .....	8
Références.....	9

Les réseaux bayésiens s'appuient sur un théorème : **Le théorème de Bayes**. C'est un résultat de base en théorie des probabilités, issu des travaux du révérend Thomas Bayes (1702-1761), présenté à titre posthume en 1763. Voici ces résultats :

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

Le terme  $P(A)$  est la *probabilité a priori* de  $A$ . Elle est « antérieure » au sens qu'elle précède toute information sur  $B$ .  $P(A)$  est aussi appelée la *probabilité marginale* de  $A$ . Le terme  $P(A|B)$  est appelée la *probabilité a posteriori* de  $A$  sachant  $B$  (ou encore de  $A$  sachant  $B$ ). Elle est « postérieure », au sens qu'elle dépend directement de  $B$ . Le terme  $P(B|A)$ , pour un  $B$  connu, est appelée la *fonction de vraisemblance* de  $A$ . De même, le terme  $P(B)$  est appelé la *probabilité marginale* ou *a priori* de  $B$ .

Cependant, certains considèrent qu'il n'y a pas toujours de rapprochement entre un modèle de type "réseau bayésien" et la "statistique bayésienne". Le terme **diagramme d'influence**, plus général et dénué de connotation est donc parfois préféré à réseau bayésien.

# RESEAU BAYESIEN

## Définition

Un réseau bayésien est un système représentant la connaissance et permettant de calculer des probabilités conditionnelles apportant des solutions à différentes sortes de problématiques.

La structure de ce type de réseau est simple : un **graphe** dans lequel les **nœuds** représentent des variables aléatoires, et les **arcs** (le graphe est donc orienté) reliant ces dernières sont rattachées à des probabilités conditionnelles.

Notons que le graphe est **acyclique** : il ne contient pas de boucle. Les arcs représentent des relations entre variables qui sont soit déterministes, soit probabilistes.

Ainsi, l'observation d'une ou plusieurs causes n'entraîne pas systématiquement l'effet ou les effets qui en dépendent, mais modifie seulement la probabilité de les observer.

L'intérêt particulier des réseaux bayésiens est de tenir compte simultanément de connaissances a priori d'experts (dans le graphe) et de l'expérience contenue dans les données.

Domaines d'utilisation principaux : diagnostic (médical et industriel), analyse de risques, détection de spams, datamining, détection de fraudes, exploitation du retour d'expérience, modélisation et simulation de systèmes complexes, détection d'intrusions, TextMining, analyse de BioPuces, analyse de trajectoires de santé.

Pour résumer, un réseau bayésien est un modèle probabiliste graphique permettant d'acquérir, de capitaliser et d'exploiter des **connaissances**, né du besoin de créer des systèmes experts à base de probabilités.

## Construction d'un graphe

Construire un réseau bayésien c'est donc :

- Définir le graphe du modèle
- Définir les tables de probabilités de chaque variable, conditionnellement à ses causes.

Le graphe est aussi appelé la "structure" du modèle, et les tables de probabilités ses "paramètres". Généralement, la structure est définie par des experts et les tables de probabilités calculées à partir de données expérimentales. Il est possible d'utiliser des algorithmes tels que K2, le recuit simulé ou encore certains algorithmes génétiques pour construire le réseau.

## Espace de recherche

L'espace de recherche est relatif aux nombre de variables bien sûr, mais aussi au nombre d'arcs et de valeurs. Au pire des cas, cet espace peut mesurer  $2^{\text{nb variables}}$ , lorsque toutes les variables sont

binaires. De plus, il augmente si les variables ont de multiples valeurs (exemple : une variable fumeur peut avoir en valeurs : non, léger, gros).

Cependant, cela signifie que toutes les variables sont dépendantes les unes des autres. Or, couramment, les variables ne sont pas dépendantes de toutes les autres, ce qui réduit considérablement la taille de l'espace de recherche.

### Tables de probabilités

Les tables de probabilités sont définies par des statistiques relatives au problème à résoudre (peuvent aussi être déterminées par des experts).

Chacune des variables dispose d'une table de probabilités conditionnelles relatives aux variables causales dont elle dépend. Par exemple (cf. présentation ppt ci-jointe), l'alarme peut se déclencher soit à cause d'un cambriolage, soit à cause d'un séisme. Les probabilités conditionnelles alarme sachant cambriolage ou/et séisme sont déduites en fonction des probabilités que tel ou tel événement survienne.

$P(\text{Alarme}|\text{Cambriolage},\text{Séisme})$

	Cambriolage, Séisme =			
	O,O	O,N	N,O	N,N
Alarme=O	0.75	0.10	0.99	0.10
Alarme=N	0.25	0.90	0.01	0.90

Lorsqu'une variable possède plusieurs valeurs, pour chacune d'elles est calculé les probabilités conditionnelles en fonction des événements causaux.

# INFERENCE BAYESIENNE

L'inférence bayésienne est basée sur l'utilisation d'énoncés probabilistes, qui dans le cas général sont trouvés par des experts étudiant un système qui leurs ait connu. Ces énoncés doivent être clairs et précis afin d'éviter toute confusion dans les relations de dépendance qui en découleront. L'inférence bayésienne est particulièrement utile dans les problèmes d'induction, car se basant sur des cas particuliers et n'a de validité qu'en terme probabiliste. Les méthodes bayésiennes se distinguent des méthodes dites standard par l'application systématique de règles formelles de transformation des probabilités. On cherche à induire sur un système bayésien aussi bien par le haut que par le bas, aussi bien les conséquences que les causes, du graphe de dépendance.

Les règles de la logique des probabilités utilisées sont les suivantes:

La règle d'addition  $p(A \cup B|C) = p(A|C) + p(B|C) - p(A \cap B|C)$

La règle de multiplication  $p(A \cap B) = p(A|B)p(B) = p(B|A)p(A)$

Le théorème de Bayes peut être dérivé simplement en mettant à profit la symétrie de la règle de multiplication

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$

Le théorème de Bayes permet d'inverser les probabilités. C'est-à-dire que si l'on connaît les conséquences d'une cause, l'observation des effets permet de remonter aux causes, c'est l'effet d'induction « bottom-up ». Sachant aussi qu'une lecture littéral du théorème de Bayes permet une induction « top-down », c'est à dire à partir des causes en déduire les conséquences. Mais il existe aussi un troisième type d'induction dit « explaining away » ou comment réfuter une cause en en constatant une autre, autrement dit partir d'une conséquence pour remonter aux causes, constater la quelle est vrai et réfuter les conséquences sous jacentes des autres causes.

En résumé pour induire sur un réseau bayésien, il faut en premier lieu trouver les probabilités conditionnelles de chaque variable aléatoire avec lesquelles elles sont directement dépendantes. Ce que les experts font à partir des statistiques étudiées sur le système voulu. Puis partir de faits produits auxquels on appliquera une probabilité de 1 ou 0 sur le réseau bayésien suivant qu'ils identifient une variable à vrai ou faux dans celui ci. Et enfin par le biais de calcul respectant la règle d'addition et ou de multiplication précédemment décrite, on modifie les probabilités causales et ou conséquentes. La nouvelle probabilité obtenue est l'induction que l'on peut faire sur un réseau bayésien.

## Algorithmes employés

Les réseaux bayésiens ont été développés au début des années 1980 pour tenter de résoudre certains problèmes de prédiction et d'abduction, courants en intelligence artificielle (IA). Dans ce type de tâche, il est nécessaire de trouver une interprétation cohérente des observations avec les données

connues **a priori**. L'inférence probabiliste signifie donc le calcul de  $P(\mathbf{Y}|\mathbf{X})$  où  $\mathbf{X}$  est un ensemble d'observations et  $\mathbf{Y}$  un ensemble de variables décrivant le problème et qui sont jugées importantes pour la prédiction ou le diagnostic.

Le calcul d'inférence probabiliste étant en général NP-difficile, deux types de méthodes ont été développés, il y a les approches complète et approximative. On a constaté qu'il existait une classe de réseaux pour lesquels le calcul est réalisable : les polyarbres (Pour tous nœud  $x$  du graphe,  $x$  d-sépare ses voisins, et le fait qu'un nœud  $x$  d-séparé deux autres nœuds  $y$  et  $z$  implique que  $y$  et  $z$  sont indépendant si  $x$  est connu, [voir notion de dépendance et de d-séparation](#)). C'est pourquoi les méthodes exactes vont se rapprocher d'un polyarbre pour inférer sur le réseau.

### Les méthodes complètes

Les premiers algorithmes d'inférence pour les réseaux bayésiens sont basés sur une **architecture à passage de messages** et ils étaient limités aux arbres. Dans cette technique, à chaque nœud est associé un processeur qui peut envoyer des messages de façon asynchrone à ses voisins jusqu'à ce qu'un équilibre soit atteint, en un nombre fini d'étapes. Cette méthode a été depuis étendue aux réseaux quelconques pour donner l'algorithme de **l'arbre de jonction (JT)**. Cette algorithme s'applique en 4 étapes de transformation du graphe: **Moralisation, Absorbé les faits mesuré, Triangulation, Construire l'Arbre de Jonction**. Et une étape d'induction : Transmettre les messages pour réaliser la cohérence. Le coût de l'algorithme **JT** est déterminé par la taille de la plus grande clique et est exponentiel en espace, pour les graphes densément connectés l'inférence peut être impraticable.

Une autre méthode s'appelle le **cut-set conditioning** : elle consiste à instancier un certain nombre de variables de manière à ce que le graphe restant forme un arbre. On procède à une propagation par messages sur cet arbre. Puis une nouvelle instanciation est choisie. On réitère ce processus jusqu'à ce que toutes les instanciations possibles aient été utilisées. On fait alors la moyenne des résultats. L'avantage de cette méthode est une complexité en temps linéaire sur la taille du réseau, mais le calcul des probabilités conditionnelles est en général impraticable pour les réseaux assez grands car ayant une complexité dans le pire des cas exponentielle dans le nombre de variables et aussi à cause du problème des boucles dans le réseau.

### Les méthodes approximatives

Il existe trois approches pour réaliser des inférence approchées, faire comme si le graphe était un arbre : « loopy belief propagation », Markov chain Monte Carlo ( e.g . échantillonnage de Gibbs), Inférence variationnelle.

Markov chain Monte Carlo exploite la topologie du réseau et effectue un échantillonnage de Gibbs sur des sous-ensembles locaux de variables de façon séquentielle et concurrente.

L'inférence variationnelle est une méthode de plus en plus utilisée, elle est une sorte d'adaptation de l'algorithme EM (Expectation-Maximization).

## Quelques outils...

- Matlab : <http://www.mathworks.fr/>
- Association for Uncertainty in Artificial Intelligence : <http://www.auai.org/>
- Bayesia : <http://www.bayesia.fr>
- Hugin : <http://www.hugin.com>
- Netica : <http://www.norsys.com>
- Elvira: <http://leo.ugr.es/~elvira>

Les réseaux bayésiens sont donc un outil de choix dans la représentation de connaissances et dans l'exploitation de celles-ci. Nous l'avons vu, beaucoup de domaines sont intéressés par ce type de représentation. Comme on a pu le voir, l'inférence sur les réseaux bayésiens est un problème NP-difficile, c'est pourquoi il était convenable de le voir de façon complète pour des instances réalisables et incomplète dans les autres cas.

Après cette approche statique, pour aller plus loin, il pourrait être intéressant de se pencher sur les réseaux bayésiens dynamiques. Ceux ci sont une répétition du réseau classique dans lesquels on rajoute un lien causal d'un pas de temps à l'autre. Ils contiennent chacun un certain nombre de variables aléatoires représentant les observations et les états cachés du processus. Le temps ici est discret et chaque unité de temps représente une nouvelle observation, l'unité de temps n'a donc pas toujours la même valeur en temps réel, la complexité inférencielle des réseaux bayésiens dynamiques est évidemment bien plus élevée que celle vu précédemment.

Enfin, on retrouve les réseaux bayésiens dans beaucoup d'applications, sans même le savoir. Microsoft par exemple est un fervent utilisateur de cette structure (Answer Wizard, assistant Office par exemple), mais aussi Google et Mozilla via leurs filtres anti-spam. De nombreux travaux dans le domaine sont réalisés, preuve de l'intérêt porté par la communauté scientifique, et de la puissance de ces réseaux.

# REFERENCES

## Bibliographie

- « Automated filtering of network intrusion detection alarms. », Faour, A., Leray, P., and Eter, B. (2006).
- « Multi-agent causal models for dependability analysis. », Maes, S. and Leray, P. (2006).
- « Apprentissage de structure dans les réseaux bayésiens et données incomplètes. », Francois, O. and Leray, P. (2005).
- « Détection et localisation de dysfonctionnement par réseaux bayésiens - applications à un réacteur tubulaire. », Buvat, J., Leray, P., Estel, L., Mouhab, N., and Cosmao, J. (2005).
- « Réseaux bayésiens. », Naïm, P., Wuillemin, P.-H., Leray, P., Pourret, O., and Becker, A. (2004).
- « Etude comparative d'algorithmes d'apprentissage de structure dans les réseaux bayésiens », Francois, O. and Leray, P. (2003).
- « Finding optimal bayesian networks. », Chickering, D. and Meek, C. (2002).
- « Learning dynamic bayesian networks. », Ghahramani, Z. (1998).
- « Knowledge representation and inference in similarity networks and bayesian multinets. », Geiger, D. and Heckerman, D. (1996).
- « Learning Bayesian networks: The combination of knowledge and statistical data. », Heckerman, D., Geiger, D., and Chickering, M. (1994).

## Webographie

- <http://tisc.inrets.fr/Seminaires/seminaire2/doc/coursMangeas.pdf>
- [http://david.bellot.free.fr/These\\_David\\_Bellot.pdf](http://david.bellot.free.fr/These_David_Bellot.pdf)
- <http://perso-math.univ-mlv.fr/users/bouissou.marc/siteRB/Page0.htm>
- [http://www.univ-angers.fr/docs/etudquassi/methodes\\_bayesiennes.pdf](http://www.univ-angers.fr/docs/etudquassi/methodes_bayesiennes.pdf)
- <http://asi.insa-rouen.fr/enseignement/siteUV/rna/>
- <http://www.cs.ubc.ca/%7Emurphyk/Bayes/bnintro.html>
- <http://www.lri.fr/%7Eantoine/Courses/IIE/WEB-ISX/Tr-SE-3-4.pdf>
- <http://maia.loria.fr/GTMS/Cr/cr5.html>