

Apprentissage de structure des réseaux bayésiens et données incomplètes

Olivier François et Philippe Leray

INSA Rouen - Laboratoire PSI - FRE CNRS 2645
BP 08 - Av. de l'Université, 76801 St-Etienne du Rouvray Cedex
{Olivier.Francois, Philippe.Leray}@insa-rouen.fr
<http://bnt.insa-rouen.fr>

Résumé. Le formalisme des modèles graphiques connaît actuellement un essor dans les domaines du *machine learning*. En particulier, les réseaux bayésiens sont capables d'effectuer des raisonnements probabilistes à partir de données incomplètes alors que peu de méthodes sont actuellement capables d'utiliser les bases d'exemples incomplètes pour leur apprentissage. En s'inspirant du principe de AMS-EM proposé par (Friedman, 1997) et des travaux de (Chow & Liu, 1968), nous proposons une méthode permettant de faire l'apprentissage de réseaux bayésiens particuliers, de structure arborescente, à partir de données incomplètes. Une étude expérimentale expose ensuite des résultats préliminaires qu'il est possible d'attendre d'une telle méthode, puis montre le gain potentiel apporté lorsque nous utilisons les arbres obtenus comme initialisation d'une méthode de recherche gloutonne comme AMS-EM.

1 Introduction

La détermination d'un réseau bayésien $\mathcal{B} = (\mathcal{G}, \theta)$ nécessite la définition d'un graphe acyclique dirigé (DAG) \mathcal{G} dont les sommets représentent un ensemble de variables aléatoires $X = \{X_1, \dots, X_n\}$ (la structure), et de matrices de probabilités conditionnelles du nœud i connaissant l'état de ses parents $Pa(X_i)$ dans \mathcal{G} , $\theta_i = [\mathbb{P}(X_i/X_{Pa(X_i)})]$ (les paramètres).

De nombreuses méthodes d'apprentissage de structure de réseaux bayésiens ont vu le jour ces dernières années. Alors qu'il est possible de faire de l'apprentissage de paramètres de réseaux bayésiens à partir de données incomplètes et que l'inférence dans les réseaux bayésiens est possible même lorsque peu d'attributs sont observés (Jensen, 1996, Pearl, 1998, Naïm *et al.*, 2004), les algorithmes d'apprentissage de structure avec des données incomplètes restent rares.

Il est possible de différencier trois types de données manquantes selon le mécanisme qui les a générées. Le premier type représente les données manquantes au hasard (MAR, *missing at random*). Dans ce cas, la probabilité qu'une variable ne soit pas mesurée ne dépend que de l'état de certaines autres variables observées. Lorsque cette probabilité ne dépend plus des variables observées, les données manquantes sont dites MCAR (*missing completely at random*). Par contre lorsque la probabilité qu'une variable soit manquante dépend à la fois de l'état de certaines autres variables observées mais également de phénomènes extérieurs, les données sont dites NMAR.

Par la suite, nous supposons que nous sommes en présence d'une base de données incomplètes suivant un mécanisme MAR ou MCAR. Ainsi, nous possédons toute l'information nécessaire pour estimer la distribution des données manquantes dans la base d'exemples.

Lorsque les données sont incomplètes, il est possible de déterminer les paramètres et la structure du réseau bayésien à partir des entrées complètes de la base. Comme les données manquantes sont supposées l'être aléatoirement, nous construisons ainsi un estimateur sans biais. Néanmoins, dans l'exemple d'une base de 2000 cas sur 20 attributs, avec une probabilité de 20% qu'une mesure soit manquante, nous ne disposerons en moyenne que de 23 cas complets. Les autres données à notre disposition ne sont donc pas négligeables et il serait donc préférable de faire l'apprentissage en utilisant toute l'information à laquelle nous avons accès.

Un avantage des réseaux bayésiens est qu'il suffit que seules les variables X_i et $Pa(X_i)$ soient observées pour estimer la table de probabilité conditionnelle correspondante. Dans ce cas, il est alors possible d'utiliser tous les exemples (même incomplets) où ces variables sont observées (dans

l'exemple précédent, en supposant que X_i possède trois parents, nous aurions 819 exemples en moyenne pour estimer les paramètres correspondants).

La recherche de structure de réseaux bayésiens peut utiliser des bases de données incomplètes, par exemple par le biais d'un échantillonnage de Gibbs (Myers *et al.*, 1999) ou encore en utilisant une approche comme l'algorithme EM (Friedman, 1997, Friedman, 1998). D'autres travaux utilisent des techniques plus originales, comme (Sebastiani & Ramoni, 2001) qui effectue l'apprentissage de sous-structures locales ou encore (Dash & Druzdzel, 2003) qui utilise une méthode à base de recherche d'indépendances conditionnelles.

Nous proposons de nous inspirer de la méthode de (Friedman, 1997) en associant un score à une structure arborescente à partir de données incomplètes, puis en choisissant le meilleur arbre selon ce score.

Dans la section suivante, nous rappelons comment associer un score à un modèle et une manière d'estimer ce score lorsque la base d'exemples est incomplète. Puis nous introduisons la méthode MWST-EM, qui est une généralisation de la recherche d'arbre de recouvrement maximal adaptée aux réseaux bayésiens (Chow & Liu, 1968, Heckerman *et al.*, 1994). Nous finissons par exposer quelques résultats préliminaires et apports éventuels d'une telle technique, notamment avec l'utilisation de l'arbre rendu par MWST-EM pour initialiser une recherche gloutonne comme AMS-EM (Friedman, 1997).

2 Score d'un réseau bayésien et données incomplètes

Soient $X = \{X_1, \dots, X_n\}$ un ensemble des variables aléatoires et \mathbf{D}_c une base de m tirages de X indépendants et identiquement distribués. Supposons par ailleurs que seule une version incomplète \mathbf{D} de la base \mathbf{D}_c soit disponible, celle-ci peut alors se décomposer en

$$\mathbf{D} = [[\mathbf{X}_i^l]]_{\substack{1 \leq i \leq n \\ 1 \leq l \leq m}} = [\mathbf{O}, \mathbf{H}]$$

où \mathbf{O} est l'ensemble des variables \mathbf{X}_i^l mesurées et \mathbf{H} l'ensemble des variables \mathbf{X}_i^l manquantes.

Score bayésien et données incomplètes

Le score bayésien est défini par $BD(\mathcal{G}, \mathbf{D}) = \mathbb{P}(\mathcal{G}, \mathbf{D}) = \mathbb{P}(\mathcal{G})\mathbb{P}(\mathbf{D}|\mathcal{G})$ où

$$\mathbb{P}(\mathbf{D}|\mathcal{G}) = \int_{\Theta} \mathbb{P}(\mathbf{D}|\mathcal{G}, \Theta)\mathbb{P}(\Theta|\mathcal{G})d\Theta$$

Pour une base d'exemples complète \mathbf{D}_c et en supposant que toutes les variables sont discrètes, (Cooper & Hersovits, 1992) ont donné le résultat suivant :

$$\mathbb{P}(\mathbf{D}_c|\mathcal{G}) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{((\sum_{k=1}^{r_i} N_{ijk}) + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (1)$$

où N_{ijk} est le nombre d'instances où $\{X_i = x_{i,k}$ et $Pa(X_i) = pa_{i,j}\}$ dans la base \mathbf{D}_c .

Or, en présence de données incomplètes, nous avons

$$\mathbb{P}(\mathbf{D}|\mathcal{G}) = \sum_{\mathbf{H}} \int_{\Theta} \mathbb{P}(\mathbf{O}, \mathbf{H}|\mathcal{G}, \Theta)\mathbb{P}(\Theta|\mathcal{G})d\Theta \quad (2)$$

L'équation 2 peut alors être évaluée par application multiple de l'équation 1 pour toutes les complétions possibles des variables manquantes \mathbf{H} . La complexité d'un tel calcul est alors exponentielle en fonction du nombre de valeurs manquantes dans la base d'exemples. En pratique, ceci n'est pas utilisable, il faut donc utiliser une méthode d'approximation pour $\mathbb{P}(\mathbf{D}|\mathcal{G})$.

Approximation d'un score avec des données incomplètes

Soit $S(\mathcal{B}|\mathbf{D}_c)$ une fonction de score quelconque pour un modèle $\mathcal{B} = (\mathcal{G}, \Theta)$ et pour une base complète \mathbf{D}_c . Le score S peut être le score bayésien ou tout autre score (pour plus d'information

sur les fonctions de score, voir (Naïm *et al.*, 2004)). Il est possible d'estimer le score de ce modèle avec des données incomplètes en calculant

$$Q^S(\mathcal{B}|\mathbf{D}) = \mathbb{E}_{\mathbf{H} \sim \mathbb{P}(\mathbf{H})} [S(\mathcal{B}|\mathbf{O}, \mathbf{H})] \quad (3)$$

Malheureusement, nous n'avons pas accès à la loi $\mathbb{P}(\mathbf{H})$. Il va donc falloir approcher cette loi à partir d'un modèle supposé générateur de \mathbf{D} . Soit \mathcal{B}^0 un tel modèle, alors il est possible d'écrire

$$Q^S(\mathcal{B} : \mathcal{B}^0|\mathbf{D}) = \mathbb{E}_{\mathbf{H} \sim \mathbb{P}(\mathbf{H}|\mathcal{B}^0)} [S(\mathcal{B}|\mathbf{O}, \mathbf{H})] = \sum_{\mathbf{H}} S(\mathcal{B}|\mathbf{O}, \mathbf{H}) \mathbb{P}(\mathbf{H}|\mathcal{B}^0) \quad (4)$$

La loi $\mathbb{P}(\mathbf{H}|\mathcal{B}^0)$ peut être obtenue par inférence dans le réseau bayésien \mathcal{B}^0 . Cette méthode permet donc, à partir d'une fonction de score $S(\mathcal{B}|\mathbf{D}_c)$ quelconque, de créer une fonction de score $Q^S(\mathcal{B} : \mathcal{B}^0|\mathbf{D})$ qui donne un résultat (approché) sur des bases d'exemples incomplètes. De plus, ce score à la particularité de conserver la propriété de *décomposabilité* :

si $S(\mathcal{B}|\mathbf{D}_c) = \sum_{i=1}^n s(X_i, Pa(X_i)|\mathbf{D}_c)$ alors $Q^S(\mathcal{B} : \mathcal{B}^0|\mathbf{D}) = \sum_{i=1}^n q^s(X_i, Pa(X_i) : \mathcal{B}^0|\mathbf{D})$.

Par exemple pour le score BIC cela donne :

$$q_i^{BIC}(X_i, Pa(X_i) : \mathcal{B}^0|\mathbf{D}) = \sum_{X_i} \sum_{Pa(X_i)} \log(\hat{\theta}_{X_i|Pa(X_i)}) \mathbb{E}_{\mathbf{H} \sim \mathbb{P}(\mathbf{H}|\mathcal{B}^0)} (N_{X_i, Pa(X_i)}) - \frac{1}{2} Dim(X_i, Pa(X_i)) \log N \quad (5)$$

3 Algorithme EM et apprentissage de structure

Structural-EM

(Friedman, 1997) a été un des premiers à proposer une méthode déterministe efficace pour faire de la recherche de structure à partir de données incomplètes. Le principe général de la méthode AMS-EM proposée par (Friedman, 1997) est rappelé dans l'algorithme 1.

Algorithme 1 Algorithme EM pour l'apprentissage de structure

- 1: $k = 0$
 - 2: Choisir le graphe \mathcal{G}^k et les paramètres $\Theta^{k,0}$ aléatoirement ou en utilisant une heuristique
 - 3: **Tant que** l'on a pas convergence et $k \leq k_{max}$ **Faire**
 - 4: $l = 0$
 - 5: **Tant que** l'on a pas convergence et $l \leq l_{max}$ **Faire**
 - 6: $l = l + 1$
 - 7: $\Theta^{k,l} = \underset{\Theta}{\operatorname{argmax}} Q(\mathcal{G}^n, \Theta : \mathcal{G}^k, \Theta^{k,l-1})$
 - 8: **Fin Tant que**
 - 9: $k = k + 1$
 - 10: $\mathcal{G}^k = \underset{\mathcal{G}}{\operatorname{argmax}} Q(\mathcal{G}, \cdot : \mathcal{G}^{k-1}, \Theta^{k,l})$
 - 11: $\Theta^{k,0} = \underset{\Theta}{\operatorname{argmax}} Q(\mathcal{G}^k, \Theta : \mathcal{G}^{k-1}, \Theta^{k-1,l})$
 - 12: **Fin Tant que**
-

Nous ne pouvons pas déterminer l'élément qui maximise le score de l'étape 10 parmi tous les DAG (Robinson, 1977). Heureusement la version généralisée de l'algorithme EM (Dempster *et al.*, 1977) montre qu'il suffit de trouver 'une' meilleure solution plutôt que 'la' meilleure sans perdre la convergence de la méthode. Dans le cadre de la méthode AMS-EM, l'*argmax* est ainsi calculé sur l'ensemble des voisins du graphe \mathcal{G}^{k-1} , c'est-à-dire les graphes obtenus par *suppression*, *ajout* ou *inversion* d'un arc.

3.1 MWST-EM, une nouvelle méthode dans l'espace des arbres

En s'inspirant des recommandations de (Heckerman *et al.*, 1994), (François & Leray, 2004) ont montré que dans le cas de données complètes, la méthode MWST (*maximum weight spanning tree*) de recherche de l'arbre de recouvrement maximal adaptée aux réseaux bayésiens permet d'obtenir

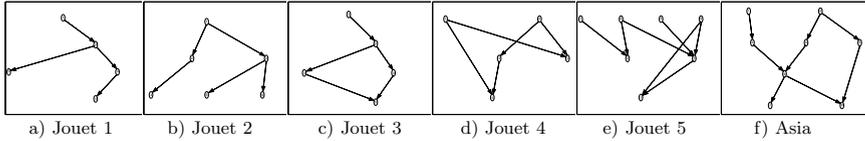


FIG. 1 – Structures utilisées pour générer les bases d'exemples

rapidement une structure simple, ayant un bon score et pouvant également servir d'initialisation pour une méthode de recherche gloutonne. Nous proposons ici d'adapter cette méthode aux bases de données incomplètes.

L'algorithme MWST-EM est très proche de celui présenté précédemment. La différence principale réside dans la recherche de l'*argmax* de l'étape 10. Nous proposons de rechercher le meilleur graphe dans l'ensemble des arbres plutôt que le meilleur dans le voisinage. Cette étape nécessite le calcul d'une matrice de similarité pour laquelle nous devons utiliser une fonction de score *décomposable* localement. Le terme générique de cette matrice $((M_{ij}^k))_{1 \leq i, j \leq n}$ est donné par :

$$M_{ij}^k = \begin{cases} \mathbb{E}_{\mathbf{H} \sim \mathbb{P}(\mathbf{H}|\mathcal{B}^{k-1})} [s(X_i, X_j | \mathbf{O}, \mathbf{H}) - s(X_i, \emptyset | \mathbf{O}, \mathbf{H})] & \text{si } i \neq j \\ 0 & \text{sinon} \end{cases} \quad (6)$$

L'algorithme de Kruskal permet ensuite d'obtenir, grâce à cette matrice, l'arbre qui maximise $Q(\mathcal{G}, \dots, \mathcal{G}^{k-1}, \Theta^{k,l})$.

L'algorithme MWST-EM, tout comme l'algorithme AMS-EM, est une méthode itérative. Il doit donc être initialisé avec un modèle de départ. (Friedman, 1997) propose de choisir une chaîne reliant toutes les variables comme initialisation de l'algorithme AMS-EM. Nous allons faire de même pour MWST-EM.

MWST-EM comme initialisation de AMS-EM

Nous proposons ensuite d'initialiser AMS-EM avec l'arbre rendu par MWST-EM pour étudier les apports éventuels d'une telle initialisation.

Variante pour les problèmes de classification

Le *réseau bayésien naïf* est souvent utilisé pour les tâches de classification. La structure de ce réseau correspond à une hypothèse d'indépendance des variables conditionnellement à la variable classe. Il est possible d'assouplir cette hypothèse d'indépendance en ajoutant des arcs entre les variables. En particulier, la méthode TANB (*tree augmented naive Bayes*) ajoute à la structure de Bayes naïve le meilleur arbre reliant les variables autres que la classe.

Nous pouvons donc maintenant utiliser la même approche avec des données incomplètes et notre algorithme MWST-EM pour obtenir une variante pour les problèmes de classification que nous appellerons TANB-EM.

4 Expérimentations

Nous avons testé les algorithmes suivants : arbre de recouvrement maximal avec données manquantes (MWST-EM), Structural EM (AMS-EM) et Structural EM initialisé avec l'arbre issu de MWST-EM (AMS-EM+T). Les bases de données incomplètes utilisées ont été générées à partir des réseaux bayésiens dont les structures sont décrites sur la figure 1, avec des données manquantes de type MCAR pour différentes probabilités (20%, 30%). Durant l'apprentissage, nous avons utilisé l'approximation du score BIC décrit dans la section 2. Des bases de test complètes ont également été générées. La table 1 contient la moyenne des scores BIC sur les bases de test des DAG obtenus à partir de 5 bases d'apprentissage différentes ainsi que les moyennes des temps de calculs (en secondes), données à titre indicatif, qui ont été obtenu avec une machine de 1GHz.

Interprétations des résultats

Pour des structures initialement arborescentes (jouet 1 et jouet 2), MWST-EM trouve un meilleur graphe que AMS-EM et plus rapidement.

	Jouet 1	Jouet 2	Jouet 3	Jouet 4	Jouet 5	Asia
napp, ntest	300, 500	500, 1000	1000, 1000	1000,1000	2000, 1000	2000, 1000
DAG init,	-1339,1	-2950,8	-2676,0	-2581,3	-3869,5	-2281,1
20 MWST-EM	-1375,6 (63,9)	-3047,9 (101,1)	-2862,8 (89,1)	-2711,9 (216,3)	-4134,2 (534,3)	-2903,6 (772,8)
AMS-EM	-1375,3 (136,6)	-3083,6 (539,5)	-2846,0 (282,0)	-2713,7 (660,4)	-4041,2 (1898)	-2831,9 (12393)
AMS-EM+T	-1421 (187,4)	-3097,1 (550,2)	-2729,8 (356,0)	-2698,2 (885,2)	-4003,2 (2855)	-2713,0 (9633)
30 MWST-EM	-1374,4 (69,5)	-3034,1 (135,0)	-2852,7 (103,4)	-2708,1 (217,3)	-4121,5 (923,5)	-2885,3 (796)
AMS-EM	-1404,5 (173,6)	-3098,7 (461,4)	-2846,9 (325,0)	-2694,6 (565,2)	-3997,6 (3123)	-2638,7 (16476)
AMS-EM+T	-1421,1 (191,0)	-3112,9 (465,6)	-2821,2 (594,8)	-2721,9 (736,4)	-4035,2 (4482)	-2552,5 (11266)

TAB. 1 – Scores BIC calculés sur des données de test pour les réseaux initiaux et pour les réseaux appris par les méthodes MWST-EM, AMS-EM et AMS-EM+T calculés avec des bases d'apprentissage vidées aléatoirement de 20% puis 30%. Les temps de calculs sont précises entre parenthèses. Les résultats sont une moyenne sur 5 exécutions pour différentes bases d'apprentissage.

Pour des réseaux plus complexes, MWST-EM trouve une structure moins bonne que AMS-EM, mais avec un score souvent proche, et toujours beaucoup plus rapidement.

De plus, les résultats de AMS-EM+T sont souvent meilleurs que ceux de AMS-EM. L'initialisation que nous proposons semble permettre de converger plus rapidement vers la même solution, ou de converger vers une meilleure solution.

Nous remarquons également que cette méthode est plus stable que la méthode AMS-EM. En effet, la recherche gloutonne possède l'inconvénient de s'arrêter dans les optimum locaux, alors qu'une initialisation par MWST-EM permet de trouver un meilleur graphe.

5 Conclusions et perspectives

Nous avons proposé une méthode originale d'apprentissage de structure arborescente de réseaux bayésiens à partir de bases de données incomplètes. Puis nous avons comparé empiriquement cette méthode à l'algorithme AMS-EM (Friedman, 1997).

Les premiers résultats nous permettent de dire que MWST-EM est une méthode assez efficace et peu complexe. Elle permet de retrouver des structures ayant un bon score et ceci plus rapidement que AMS-EM. Néanmoins notre méthode a le désavantage de travailler dans l'espace des arbres, ce qui est restrictif.

Ensuite nous avons utilisé notre méthode pour initialiser l'algorithme AMS-EM. Cette initialisation permet souvent d'obtenir de meilleurs résultats qu'en utilisant l'initialisation proposée par (Friedman, 1997).

Nous envisageons maintenant de tester et d'évaluer ces algorithmes sur un éventail de problèmes plus large, notamment avec des données MAR et pour des problèmes de classification.

Les méthodes MWST-EM et AMS-EM sont des adaptations d'algorithmes de recherche de structure dans l'espace des graphes acycliques dirigés, mais (Chickering, 2002) a proposé récemment GES, un algorithme optimal de recherche dans un autre espace, l'espace des représentants des équivalents de Markov. La suite logique de nos travaux est donc l'adaptation de GES pour le cas des bases de données incomplètes.

Remerciements

Ce travail a été partiellement financé par le programme de la communauté européenne IST, sous le réseau d'excellence PASCAL, IST-2002-506778. Cette publication ne reflète que les opinions de ses auteurs.

Références

(Chickering, 2002) CHICKERING D. M. (2002). Learning equivalence classes of bayesian-network structures. *Journal of machine learning research*, **2**, pages 445-498, 2002.

- (Chow & Liu, 1968) CHOW C. & LIU C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, **14**(3), pages 462-467, 1968.
- (Cooper & Hersovits, 1992) COOPER G. & HERSOVITS E. (1992). A bayesian method for the induction of probabilistic networks from data. *Maching Learning*, **9**, pages 309-347, 1992.
- (Dash & Druzdzel, 2003) DASH D. & DRUZDZEL M. (2003). Robust independence testing for constraint-based learning of causal structure. Proceedings of The Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI03), 2003.
- (Dempster *et al.*, 1977) DEMPSTER A., LAIRD N. & RUBIN D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, **B 39**, pages 1-38, 1977.
- (François & Leray, 2004) FRANÇOIS O. & LERAY P. (2004). Evaluation d'algorithmes d'apprentissage de structure dans les réseaux bayésiens. In *14ième Congrès francophone de Reconnaissance des formes et d'Intelligence artificielle*, pages 1453-1460, 2004.
- (Friedman, 1997) FRIEDMAN N. (1997). Learning belief networks in the presence of missing values and hidden variables. In *in the Proceedings of the 14th International Conference on Machine Learning*, Morgan Kaufmann, pages 125-133, 1997.
- (Friedman, 1998) FRIEDMAN N. (1998). The Bayesian structural EM algorithm. In G. F. COOPER & S. MORAL, Eds., *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, San Francisco : Morgan Kaufmann, pages 129-138, 1998.
- (Heckerman *et al.*, 1994) HECKERMAN D., GEIGER D. & CHICKERING M. (1994). Learning Bayesian networks : The combination of knowledge and statistical data. In R. L. DE MANTARAS & D. POOLE, Eds., *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, USA : Morgan Kaufmann Publishers, pages 293-301, 1994.
- (Jensen, 1996) JENSEN F. V. (1996). *An introduction to Bayesian Networks*. Taylor and Francis, London, United Kingdom, 1996.
- (Myers *et al.*, 1999) MYERS J. W., LASKEY K. B. & LEWITT T. S. (1999). Learning bayesian network from incomplete data with stochastic search algorithms. In *In the Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI99)*, 1999.
- (Naïm *et al.*, 2004) NAÏM P., WUILLEMIN P.-H., LERAY P., POURRET O. & BECKER A. (2004). *Réseaux bayésiens*. Eyrolles. ISBN : 2-212-11137-1, 1999.
- (Pearl, 1998) PEARL J. (1998). Graphical models for probabilistic and causal reasoning. In D. M. GABBAY & P. SMETS, Eds., *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1 : Quantified Representation of Uncertainty and Imprecision*, pages 367-389. Dordrecht : Kluwer Academic Publishers, 1998.
- (Robinson, 1977) ROBINSON R. W. (1977). Counting unlabeled acyclic digraphs. In C. H. C. LITTLE, Ed., *Combinatorial Mathematics V*, volume 622 of *Lecture Notes in Mathematics*, Springer, Berlin pages 28-43, 1977.
- (Sebastiani & Ramoni, 2001) SEBASTIANI P. & RAMONI M. (2001). Bayesian selection of decomposable models with incomplete data. *Journal of the American Statistical Association*, Vol. 96, No. 456, 2001.

Summary

The framework of graphical models is more and more used in the area of Machine Learning. Specially, Bayesian Networks allow to carry out probabilistic reasoning from incomplete datasets but few methods can use incomplete datasets to learn their structure. Using the principle of the Structural-EM (Friedman, 1997) and the work of (Chow & Liu, 1968), we propose a new method to perform a structural search of bayesian network in the space of tree graphs from incomplete datasets. An experimental study gives first results of this method and then shows the potential improvement if we use trees that it gives to initialise a greedy method like the Structural EM.