



L'intégration des connaissances ontologiques pour l'apprentissage des réseaux bayésiens causaux

Montassar Ben Messaoud, Nahla Ben Amor, Philippe Leray

► To cite this version:

Montassar Ben Messaoud, Nahla Ben Amor, Philippe Leray. L'intégration des connaissances ontologiques pour l'apprentissage des réseaux bayésiens causaux. 5èmes Journées Francophones sur les Réseaux Bayésiens (JFRB2010), May 2010, Nantes, France. <hal-00474395>

HAL Id: hal-00474395

<https://hal.archives-ouvertes.fr/hal-00474395>

Submitted on 20 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'intégration des connaissances ontologiques pour l'apprentissage des réseaux bayesiens causaux

Montassar Ben Messaoud* — Nahla Ben Amor* — Philippe Leray**

* LARODEC, ISG de Tunis, 41 rue de la liberté 2000, Le Bardo, Tunisie

** LINA, UMR 6241, Ecole Polytechnique de l'Université de Nantes, France

RÉSUMÉ. L'apprentissage de la vraie structure causale d'un Réseau Bayésien (RB) est un sujet vaste et complexe. Dans certains nouveaux domaines, l'apprentissage des Réseaux Bayésiens Causaux (RBC) s'avère indispensable. Pour apprendre de tels modèles, nous avons besoin de données d'intervention, qui sont, en général, très coûteuses à exécuter. Cet article propose une nouvelle méthode pour apprendre les RBCs à partir de données observationnelles et des interventions. En effet, notre nouvelle approche se base essentiellement sur l'algorithme d'apprentissage causal MyCaDo (Meganck et al., 2006) en lui rajoutant une nouvelle étape d'intégration de connaissances ontologiques. Ceci va nous permettre de choisir efficacement les interventions à effectuer pour obtenir un CBN complet.

ABSTRACT. Learning the causal structure of Bayesian Networks (BN) is a really complex research field. Some new domains appear where the model will only be learnt in order to discover these causal relationships. To learn such models, interventional data are required. These interventions are usually very expensive to perform, therefore the choice of variables to experiment on can be vital when the number of experimentations is restricted. This article proposes a new method for learning Causal Bayesian Networks (CBNs) from observational data and interventions. We first extend the greedy approach for perfect observational and experimental data proposed in (Meganck et al., 2006), by adding a new step based on the integration of ontological knowledge, which will allow us to choose efficiently the interventions to perform in order to obtain the complete CBN.

MOTS-CLÉS : RBCs, données d'interventions, découverte causale, connaissances ontologiques.

KEYWORDS: CBNs, interventional data, causal discovery, ontological knowledge

1. Introduction

Tout au long de ces dernières années, nous avons assisté à une utilisation accrue des ontologies dans le domaine informatique. La tâche essentielle de ces représentations est de dégager une sémantique de domaine et de permettre ainsi à la machine de résonner à base de relations entre concepts (Cannataro *et al.*, 2004). On note que différentes solutions à base d'ontologies ont été mises en places dans différents domaines de recherche tels que la traduction du langage naturel, la médecine, le commerce électronique et la bioinformatique. Parallèlement, la découverte causale constitue encore un vrai challenge pour les informaticiens. L'une des techniques les plus répandues est sans doute les Réseaux Bayésiens Causaux (RBC), qui sont des modèles graphiques probabilistes, exigeant que chaque relation causale soit représentée par un arc dirigé (Pearl, 2000). Tout à fait comme leurs ancêtres les réseaux bayésiens, les RBC utilisent des données observationnelles pour apprendre un graphe partiellement dirigé et des interventions en complément, pour diriger le reste de modèle. Dans ce contexte, différentes recherches ont aboutit à des algorithmes d'apprentissage de structure de RBC à partir de données observationnelles et d'interventions. Le présent article va plus loin en abordant la problématique d'intégration des connaissances ontologiques dans la découverte causale (Ben Messaoud *et al.*, 2009). Les travaux se rapportant à cette combinaison de réseaux bayésiens et ontologies sont très rares. Et la majorité d'entre eux se concentrent sur le développement d'outils de mapping d'ontologies basés sur des modèles probabilistes (Pan *et al.*, 2005; Svátek, 2006). Le plan de cet article est comme suit : La section 2 introduit les réseaux bayésiens, les réseaux bayésiens ainsi que les ontologies. La section 3 décrit notre nouvelle approche pour la découverte causale à base de connaissances ontologiques. Nous finissons avec la section 4 qui présente un exemple illustratif de la méthode proposée.

2. Notations

Dans cette section, nous allons définir progressivement les Réseaux Bayésiens, les Réseaux Bayésiens Causaux et les ontologies. Les notations et conventions suivantes seront utilisées tout au long de l'article. Soit $X = \{X_1, X_2, \dots, X_N\}$ un ensemble de variables. On note par x toute instance de X . Pour chaque noeud $X_i \in X$, nous notons par $\text{Pa}(X_i)$ (resp. $\text{Ne}(X_i)$) les parents (resp. voisinage) de la variable X_i . $P(X_i)$ représente la distribution de probabilités relative à X_i , et $P(X_i=x_k)$ dénote la probabilité que la variable X_i soit égale à x_k .

2.1. Réseaux Bayésiens

Un Réseau Bayésien (RB) (Garey *et al.*, n.d.) consiste en un Graphe Dirigé Acyclique (DAG) et un ensemble de tables de probabilités conditionnelles (CPTs), de façon que chaque noeud dans le DAG corresponde à une variable, et la table qui lui est associée (i.e. $P(X_i|\text{Pa}(X_i))$) contienne les probabilités de chaque état de cette variable

dans le contexte des combinaisons possibles des états de ses parents. Les réseaux bayesiens sont très utiles pour l'inférence probabiliste puisqu'ils satisfont la *propriété Markovienne*, qui suppose que chaque noeud est indépendant de ses non-descendants dans le contexte de ses parents. Ainsi la distribution de probabilité jointe relative à $X=(X_1, X_2, \dots, X_n)$ peut être calculée grâce à la règle de chaînage suivante :

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1..n} P(x_i | Pa(X_i)). \quad [1]$$

Notons qu'une même distribution de probabilité jointe peut être représentée par plusieurs RBs. De tels réseaux seront appelés *équivalents* (Verma *et al.*, 1990). Les structures équivalentes peuvent être représentées à travers un DAG partiellement dirigé (CPDAG). Ce CPDAG garde le même squelette que le DAG original en possédant à la fois des arcs dirigés et non dirigés.

Dans le cadre de la suffisance causale (i.e. le système étudié n'est influencé par aucune variable latente), un grand nombre de techniques d'apprentissage de structures est disponible pour apprendre un CPDAG et choisir ensuite une instantiation dans l'espace des graphes équivalents. Ces techniques peuvent être classées en deux groupes, les algorithmes à base de scores et les algorithmes à base de contraintes. Les algorithmes à base de score (Chickering *et al.*, 2002; Cooper *et al.*, 1992) essaient d'identifier le réseau qui maximise une certaine fonction de score. Ceux à base de contraintes (Cooper *et al.*, 1999; Spirtes *et al.*, 2000) recherchent les (in)dépendances au niveau des données et modélisent directement cette information dans la structure graphique résultante.

2.2. Réseaux Bayesiens Causaux

Un Réseau Bayésien Causal (RBC) (Pearl, 2000) est essentiellement un réseau bayésien avec la propriété que chaque arc dirigé représente une relation causale autonome. En utilisant des RBCs, nous pouvons aller au delà de l'inférence probabiliste pour effectuer une inférence causale. Pearl a introduit l'opérateur *do* comme standard de notation pour les interventions externes au niveau des modèles causaux. Ainsi, l'effet d'une certaine action " $do(X_j=x_k)$ " dans le modèle causal correspond à une perturbation du système en forçant la variable X_j à prendre la valeur x_k . En d'autres termes, l'inférence causale est le processus via lequel nous calculons l'effet de manipuler une certaine variable X_j en la forçant à prendre la valeur x_k , ceci peut être traduit par $P(X_i|do(X_j=x_k))$.

Parmi les algorithmes dédiés à l'apprentissage des RBCs nous pouvons citer les travaux de Cooper et Yoo (Cooper *et al.*, 1999) et ceux de Tong et Koller (Tong *et al.*, 2001), qui se placent dans le cadre de l'apprentissage actif, où les seules données seront celles obtenues par expérimentation, et où le modèle va être construit au fur et à mesure de ces expériences.

Par ailleurs, Meganck et al. (Meganck *et al.*, 2006) ont proposé l'algorithme MYCADO (My Causal DiscOverY) en partant d'une hypothèse différente. Ils supposent qu'un ensemble de données d'observation est déjà disponible, et a permis d'obtenir le représentant de la classe d'équivalence de Markov après application de l'algorithme PC de Spirtes et al. (Spirtes *et al.*, 2000). Il reste donc à finir d'orienter cette structure à partir d'expérimentations sur le système. Ainsi MYCADO propose à l'utilisateur l'expérience à réaliser qui pourrait lui permettre d'orienter potentiellement le plus d'arêtes, en maximisant une fonction d'utilité U . Trois variables sont mises en jeu le gain d'information, le coût d'expérience et le coût de mesure de l'impact de l'expérience sur les autres variables. Soient A_{X_i} (resp. M_{X_i}) la réalisation d'une expérience sur la variable X_i (resp. la mesure de l'impact de l'expérience sur les variables du voisinage). La fonction d'utilité relative à une expérience sur un noeud donné est exprimée par :

$$U(A_{X_i}) = \frac{\alpha \text{gain}(A_{X_i})}{\beta \text{cost}(A_{X_i}) + \gamma \text{cost}(M_{X_i})} \quad [2]$$

où α , β et γ sont les mesures d'importance pour chaque terme.

Selon la situation de l'expérimentation, nous pouvons utiliser le *Maximax*, *Maximin* ou l'*Expected Utility* comme critères de décision. Une fois que les résultats de l'expérience sont pris en compte dans le graphe, il faut réévaluer la situation pour choisir l'expérience suivante. Ce processus sera réitéré jusqu'à l'obtention de la structure correcte. Cette approche permet aussi de tenir compte des coûts éventuels d'expérimentation ou d'observation de chaque variable.

Borchani et al. (Borchani *et al.*, 2007) ont également élaboré une approche pour l'apprentissage de RBCs à partir de données d'observations incomplètes et d'interventions simultanées. D'autres approches ont été proposées pour le cas de modèles graphiques avec variables latentes (Maes *et al.*, 2007a; Maes *et al.*, 2007b).

2.3. Ontologie

L'ontologie est définie comme un ensemble structuré de termes et concepts représentant la sémantique d'un champ d'informations. Les relations dans une ontologie sont soit taxonomiques (e.g. is-a, part of) ou non taxonomiques (défini par l'utilisateur). Les relations taxonomiques sont les plus utilisées. Dans sa forme la plus simple, l'ontologie est assimilée à un arbre ou une hiérarchie représentant une taxonomie de concepts. Formellement :

Definition 1. Une *Taxonomie de Concepts* $H = (C, E, Rt)$ est un arbre où $C = \{c_1, c_2, \dots, c_n\}$ est l'ensemble de tous les concepts c_i , E l'ensemble de tous les liens de Subsumption (is-a) et Rt la racine. Les éléments structurants une taxonomie de

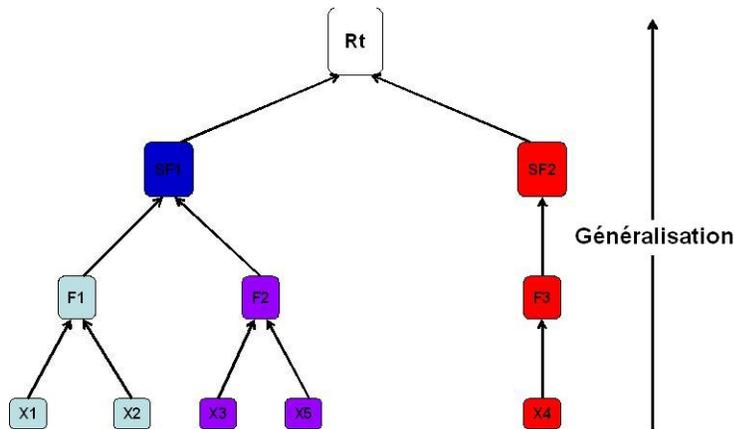


Figure 1. Exemple de hiérarchie représentant une taxonomie de concepts.

concepts sont les relations de subsumption, qui supposent que si le concept c_i est le fils du concept c_j alors toutes les propriétés de c_j seront transférées à c_i et c_j sera considéré comme le subsumant de c_i .

Dans la suite, nous allons considérer les notations suivantes :

- $pths(c_i, c_j)$: est l'ensemble de chemins entre les concepts c_i et c_j ,
- $len_e(e)$: la longueur en nombre d'arcs du chemin e ,
- $mcs(c_i, c_j)$: est le subsumant commun le plus spécifique de c_i et c_j ,

La taxonomie de concepts constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine. Elle est employée pour raisonner sur des objets du domaine concerné.

Exemple 1. Figure 1 montre un exemple d'ontologie où plus précisément, un arbre is-a, où les nœuds feuilles ($X1, X2, \dots, X5$) peuvent être assimilées à des gènes. Leurs subsumants directs ($F1, F2, F3$) représentent des fonctions biologiques. Ces fonctions sont également subsumées par des super-fonctions biologiques ($SF1, SF2$). Pour l'arbre étudié, nous retrouvons qu'un seul chemin entre deux différents concepts. Pour illustrer cette idée, nous allons considérer les deux concepts $X1$ and $SF2$. Ici, $pths(X1, SF2)$ représente l'ensemble des arcs $X1-F1, F1-SF1, SF1-RT$ et $RT-SF2$, relatif à l'unique chemin entre $X1$ et $SF2$. Par conséquent, la longueur en nombre d'arcs de ce chemin est égale à 4.

Récemment, des travaux de recherche ont mis l'accent sur l'importance d'évaluer la force des liens sémantiques au niveau des ontologies de domaine. On peut distinguer 3 classes de mesures : le rapprochement sémantique, la similarité sémantique et la distance sémantique, évaluant, respectivement, le rapprochement, la ressemblance et

l'éloignement entre deux concepts bien déterminés. Les approches de mesure des distances sémantiques sont *Rada et al.'s distance*, *Sussna's distance* et *Jiang and Conrath's distance*. Pour la similarité sémantique, nous citons *Leacock and Chodorow's similarity*, *Wu and Palmer's similarity* et *Lin similarity*, par contre pour le rapprochement sémantique, la mesure la plus utilisée est *Hirst and St Onge's relatedness*. Voir (Blanchard *et al.*, 2005) pour une étude comparative entre ces trois mesures.

Dans ce qui suit, nous allons nous intéresser à la distance sémantique et plus particulièrement à *Rada et al.'s distance*. Cette distance se base sur le fait que plus le chemin est court, plus les objets comparés sont similaires. En cas de multiples chemins, nous choisissons le chemin le plus court. Formellement, pour deux concepts c_i et c_j , la distance Rada et al.'s est définie comme suit :

$$dist_{rmbb}(c_i, c_j) = \min_{p \in pths(c_i, c_j)} len_e(p) \quad [3]$$

3. Nouvelle approche pour la découverte causale à base de connaissances ontologiques (SemCaDo)

Dans cette section, nous proposons une nouvelle approche SemCaDo (Semantical Causal Discovery) pour l'apprentissage des RBCs. Notre idée est d'étendre l'algorithme MyCaDo (Meganck *et al.*, 2006) en intégrant les connaissances ontologiques (voir algorithme 1). En effet, les ontologies disponibles nous fournissent des connaissances du domaine qui peuvent être efficacement utilisées pour enrichir le processus d'apprentissage et optimiser la découverte causale. Nous allons travailler sous l'hypothèse que les feuilles de la taxonomie de concepts sont également les noeuds du RBC. Il est à noter que les relations causales à découvrir dans le RBC n'existent pas forcément au niveau de l'ontologie, ce qui explique la complémentarité des deux modèles.

Dans la suite, nous considérons que, suite à chaque expérimentation sur X_i , nous pouvons mesurer l'impact sur toutes les variables du voisinage de X_i ($Ne_U(X_i)$).

Cette mesure se fera à travers une fonction d'utilité qui représente une généralisation de la fonction d'utilité relative à MyCaDo. Nous remplacerons ainsi $Ne_U(X_i)$ par l'inertie sémantique correspondante $SemIn(Ne_U(X_i))$. Cette fonction sera exprimée comme suit :

$$SemIn(M) = \frac{\sum_{X_i \in M} \min_{p \in pths(X_i, mscs(M))} (len_e(p))}{card(M)} \quad [4]$$

C'est à travers cette nouvelle notion que nous allons intégrer les connaissances ontologiques extraites des ontologies et guider le choix des expériences à réaliser. Dans

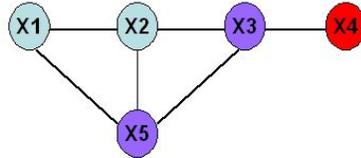


Figure 2. *Un exemple de CPDAG*

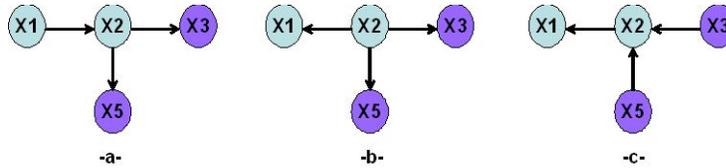


Figure 3. *Toutes les instantiations possibles de $X_2 - Ne_U(X_2)$*

beaucoup de situations, un noeud X_i peut avoir un nombre important de voisins qui sont très rapprochés au niveau de l'ontologie. Ceci implique que nous aurons, exactement, deux alternatives (i.e. X_i est la cause directe de tous ses voisins où l'inverse). Or le fait qu'un noeud ait un voisinage très rapproché sémantiquement va réduire le nombre de fonctions étudiées et par conséquent la contribution de l'expérience sera minimale. Par contre, si les voisins sont sémantiquement distants, il y aura encore plus de relations causales inter fonctions à découvrir. Les grandes lignes de notre nouvelle approche sont résumées par l'algorithme 1.

4. Exemple Illustratif

Afin d'illustrer notre approche et la comparer à MyCaDo, nous allons considérer le CPDAG de la figure 2 et la taxonomie de concepts de la figure 1. Pour une expérimentation potentielle sur X_2 , la Figure 3 dénombre toutes les instanciions possibles des arcs $X_i - Ne_U(X_i)$.

Il est clair que les arcs $X_2 - X_3$ et $X_2 - X_5$ ont la même orientation vu que leurs deux voisins X_3 et X_5 partagent la même fonction dans l'ontologie correspondante. Dans notre cas, nous avons trois instanciions possibles alors qu'avec MYCADO, nous aurons eu 6 cas possibles. Par conséquent, nous avons considérablement réduit le champ des instanciions possibles dans le PDAG (i.e. CPDAG avec des arcs dirigés). Ici, le subsumant commun le plus spécifique de $Ne_U(X_2) = X_1, X_3, X_5$ est le

concept SF1. Par référence à l'ontologie représentée au niveau de la figure 1, l'inertie sémantique des noeuds X_1 , X_3 et X_5 peut être calculée comme suit :

$$\begin{aligned} SemIn(X_1, X_3, X_5) &= \frac{\sum_{X_j \in Ne_U(X_2)} \min_{p \in pths(X_j, mscs(Ne_U(X_2)))} (len_e(p))}{\#Ne_U} \\ &= \frac{2+2+2}{3} = 2 \end{aligned}$$

Dans le cas où les voisins se partagent le même subsumant au niveau de l'ontologie, l'inertie sémantique du voisinage sera égale à un.

Cette inertie sémantique dépend également du nombre de voisinage. Par exemple, si nous éliminons X_1 du voisinage de X_2 , SemIn va décroître automatiquement et nous aurons :

$$SemIn(X_3, X_5) = \frac{1+1}{2} = 1$$

Et, finalement, plus les variables voisines sont distantes au niveau de l'ontologie, plus l'inertie sémantique est importante et l'utilité est maximisée. Ainsi, si nous remplaçons le noeud X_3 par X_4 , qui est encore plus distant du voisinage de X_2 , SemIn va croître considérablement, avec :

$$SemIn(X_1, X_4, X_5) = \frac{3+3+3}{3} = 3$$

D_O : Données observationnelles ;

H : Taxonomie de concepts ;

1) Appliquer un algorithme d'apprentissage de structures sur D_O pour obtenir notre CPDAG G .

2) Calculer $X_{best} = \text{argmax}(U(A_{X_i}))$

3) Réaliser une expérience sur la variable X_{best} .

4) Instancier l'orientation de tous les arcs non dirigés reliés à X_{best}

Pour ($X_j \in Ne_U(X_{best})$) **faire**

Si (la distribution de X_j change suite à l'expérience) **Alors**

 orienter $X_{best} - X_j$ comme $X_{best} \rightarrow X_j$

Sinon

 orienter $X_{best} - X_j$ comme $X_{best} \leftarrow X_j$

Fin Si

Fin Pour

5) Inferer les arcs (Règles d'orientation relatives à l'algorithme PC)

Répéter

 Appliquer les règles-PC ;

jusqu'à ce que (aucun arc ne peut être orienté)

6) Retour à l'étape (2) jusqu'à ce que tous les arcs soient dirigés.

7) Retourner le CBN G .

Algorithme 1: Algorithme SEMCADO

5. Conclusion

Dans cet article, nous avons démontré comment les connaissances ontologiques peuvent être utiles pour la découverte causale. Plus précisément, nous avons étendu l'algorithme MYCADO via l'introduction de la notion d'inertie sémantique. En intégrant les coûts des expériences et de mesures, notre approche guide le choix vers les meilleures expériences pour obtenir un réseau bayésien causal totalement orienté. Les travaux futurs à venir vont se focaliser sur l'intégration des connaissances sémantiques pour l'apprentissage de modèles causaux en présence de variables latentes, à la fois non observables et non mesurables (Cannataro *et al.*, 2004).

6. Bibliographie

- Ben Messaoud M., Leray P., Ben Amor N., « Integrating Ontological Knowledge for Iterative Causal Discovery and Visualization. », in C. Sossai, G. Chemello (eds), *ECSQARU*, vol. 5590 of *Lecture Notes in Computer Science*, Springer, p. 168-179, 2009.
- Blanchard E., Harzallah M., Briand H., Kuntz P., « A typology of ontology-based semantic measures », *2nd INTEROP-EMOI Open Workshop on Enterprise Models and Ontologies for Interoperability at the 17th Conference on Advanced Information Systems Engineering (CAISE'05)*, vol. 160, p. 407-412, 2005.
- Borchani H., Chaouachi M., Ben Amor N., « Learning Causal Bayesian Networks from Incomplete Observational Data and Interventions. », in K. Mellouli (ed.), *ECSQARU*, vol. 4724 of *Lecture Notes in Computer Science*, Springer, p. 17-29, 2007.
- Cannataro M., Massara A., Veltri P., « The OnBrowser ontology manager : Managing ontologies on the Grid », In : *International Workshop on Semantic Intelligent Middleware for the Web and the Grid*, 2004.
- Chickering D. M., Boutilier C., « Optimal Structure Identification with Greedy Search », 2002.
- Cooper G. F., Dietterich T., « A Bayesian method for the induction of probabilistic networks from data », *Machine Learning*, p. 309-347, 1992.
- Cooper G. F., Yoo C., « Causal discovery from a mixture of experimental and observational data », In *UAI*, Morgan Kaufmann, p. 116-125, 1999.
- Garey M. R., Computers D. S. J., Theory A. G. T. T., Np-completeness O., Freeman P. W. H., Geeman S., Relaxation D. G. S., Distributions G., Poole D., Rojas-guzman C., « [7] Judea Pearl. Probabilistic Reasoning in Intelligent Systems : Networks of Plausible In- », n.d.
- Maes S., Leray P., Meganck S., « Causal Graphical Models with latent variables : learning and inference », in D. E. Holmes, L. Jain (eds), *Innovations in Bayesian Networks : Theory and Applications*, Springer, Germany, 38 pages, 2007a.
- Maes S., Meganck S., Leray P., « An integral approach to causal inference with latent variables », in F. Russo, J. Williamson (eds), *Causality and Probability in the Sciences*, Texts In Philosophy series, London College Publications, 23 pages, 2007b.
- Meganck S., Manderick B., Leray P., « Learning causal bayesian networks from observations and experiments : A decision theoretic approach », In *Modeling Decisions in Artificial Intelligence, LNCS*, p. 58-69, 2006.
- Pan R., Ding Z., Yu Y., Peng Y., « A Bayesian Network Approach to Ontology Mapping », In : *Proceedings ISWC 2005*, Springer, p. 563-577, 2005.
- Pearl J., *Causality : models, reasoning, and inference*, 2000.
- Spirtes P., Glymour C., Scheines R., *Causation, Prediction, and Search*, 2nd edn, MIT press, 2000.
- Svátek V., « Combining ontology mapping methods using bayesian networks », In : *Workshop on Ontology Matching at ISWC-2006*, 2006.
- Tong S., Koller D., « Active Learning for Structure in Bayesian Networks. », in B. Nebel (ed.), *IJCAI*, Morgan Kaufmann, p. 863-869, 2001.
- Verma T., Pearl J., « Equivalence and synthesis of causal models. », in P. P. Bonissone, M. Henrion, L. N. Kanal, J. F. Lemmer (eds), *UAI*, Elsevier, p. 255-270, 1990.