Apprentissage des réseaux possibilistes à partir de données: un survol

Nahla Ben Amor, Maroua Haddad, Philippe Leray

▶ **To cite this version:**

Nahla Ben Amor, Maroua Haddad, Philippe Leray. Apprentissage des réseaux possibilistes à partir de données: un survol. Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle, 2015, 29 (2), pp.229-252. 10.3166/HSP.x.1-24 . hal-01166897

HAL Id: hal-01166897
https://hal.science/hal-01166897

Submitted on 15 Apr 2020

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.

Apprentissage des réseaux possibilistes à partir de données

Maroua Haddad [1,2], **Philippe Leray** [2], **Nahla Ben Amor** [1]

1. LARODEC
Institut Supérieur de Gestion, Université de Tunis
41 rue de la liberté 2000 Le Bardo
Tunisie
maroua.haddad@gmail.com,nahla.benamor@gmx.fr

2. LINA-UMR CNRS 6241
Equipe DUKe (Data User Knowledge)
Université de Nantes
France
philippe.leray@univ-nantes.fr

RÉSUMÉ. *Les réseaux possibilistes représentent des outils importants de modélisation et de raisonnement, en particulier, en présence d'informations imprécises et/ou incertaines. Ces modèles graphiques ont été efficacement utilisés dans plusieurs applications réelles. Vu que leur construction par des experts est une tâche complexe qui prend beaucoup de temps, plusieurs chercheurs ont essayé d'apprendre ces modèles à partir de données. Dans ce travail, nous essayons de présenter tous les travaux pertinents reliés à l'apprentissage des réseaux possibilistes à partir de données. En effet, nous présentons un aperçu des méthodes qui ont déjà été proposées dans ce sens et les limites de chaque méthode vis-à-vis des recherches récentes développées dans le cadre de la théorie des possibilités.*

ABSTRACT. *Possibilistic networks are important tools for modeling and reasoning, especially in the presence of imprecise and/or uncertain information. These graphical models have been successfully used in several real applications. Since their construction by experts is complex and time consuming, several researchers have tried to learn them from data. In this work, we try to present relevant works related to learning possibilistic networks from data. In fact, we give an overview of methods that have already been proposed in this context and limitations of each one of them towards recent researches developed in possibility theory framework.*

MOTS-CLÉS : *théorie des possibilités, modèles graphiques, réseaux possibilistes, apprentissage automatique.*

KEYWORDS: *possibility theory, graphical models, possibilistic networks, machine learning.*

1. Introduction

Au cours des dernières années, les modèles graphiques connaissent une utilisation accrue dans diverses applications du monde réel vu qu'ils représentent les données d'une manière très simple et intuitive. Les modèles probabilistes restent les plus connus et les plus utilisés. En particulier, les réseaux bayésiens (Pearl, 1988; Jensen, 1996) ont été largement étudiés et utilisés dans des applications réelles (Pourret *et al.*, 2008). Cependant, dans des domaines où l'existence des informations incertaines et imprécises est primordiale et inévitable ou lorsque les informations à disposition sont tout simplement des préférences, d'autres théories d'incertitude non classiques comme la théorie des possibilités (Zadeh, 1999), la théorie des fonctions de croyance (Shafer, 1976) et la théorie des probabilités imprécises (Dempster, 1967; Augustin *et al.*, 2014) s'imposent comme meilleures alternatives à la théorie des probabilités (Dubois, Prade, 2006). Par conséquent, d'autres modèles graphiques ont été proposés qui permettent de modéliser et de raisonner avec ce type de données ou de connaissances imparfaites. Parmi ces modèles graphiques, nous nous intéressons aux réseaux possibilistes qui représentent la contrepartie possibiliste des réseaux bayésiens. Le choix de la théorie des possibilités est motivé par sa capacité à offrir un cadre formel naturel et simple pour la représentation d'informations imprécises et incertaines et sa capacité à traiter l'aspect qualitatif et quantitatif au niveau de la description des états du monde (Dubois, Prade, 1998).

En théorie des possibilités, il existe deux façons de définir la contrepartie des réseaux bayésiens selon l'opérateur de conditionnement utilisé : min ou produit (Dubois, Prade, 2000). Ces modèles ont suscité l'intérêt de plusieurs chercheurs et ont été utilisés dans certains domaines comme la recherche d'information (Boughanem *et al.*, 2009), les systèmes tuteurs intelligents (Adina, 2006), l'industrie automobile (Kruse, Borgelt, 1995) et la fusion de données (Kruse, Borgelt, 1995). Malgré la multitude des travaux proposés concernant la propagation d'information dans ces modèles (Fonck, 1992; Ben Amor *et al.*, 2003; Benferhat, Smaoui, 2007; Ayachi *et al.*, 2010), les travaux liés à leur apprentissage (Sangüesa *et al.*, 1998a; Borgelt, Kruse, 2003) restent très limités et discutables.

Dans cet article, nous discutons du problème d'apprentissage des paramètres des réseaux possibilistes. Pour cela, nous proposons d'appliquer les méthodes d'estimation de distributions de possibilités existantes dans notre contexte. Ensuite, nous présentons un survol des méthodes d'apprentissage de structure à partir de données tout en redéfinissant quelques notions de base indispensables comme la notion de score possibiliste et en mettant en avant les limites de ces méthodes.

Cet article est organisé comme suit : la section 2 présente un aperçu concernant la théorie des possibilités et les réseaux possibilistes. La section 3 est dédiée à l'apprentissage des paramètres de ces réseaux. La section 4 présente l'apprentissage de la structure.

2. Concepts de base et réseaux possibilistes

Cette section présente un bref aperçu de la théorie des possibilités détaillée de manière plus précise dans (Dubois *et al.*, 1988; Dubois, Prade, 2000; Dubois, 2006). Ensuite, elle décrit les réseaux possibilistes (Fonck, 1992) tout en mettant en avant l'intérêt de ces modèles graphiques.

2.1. Concepts de base de la théorie des possibilités

2.1.1. Notations et définitions

Soient $V = \{X_1, ..., X_n\}$ un ensemble de variables définies dans leurs domaines respectifs finis $D_1, ..., D_n$ et x_{ik} la $k^{\text{ème}}$ instance d'une quelconque variable X_i. Chaque élément $x_{ik} \in D_i$ correspond à un état (une valeur possible) de X_i. L'ensemble des états d'une variable X_i peut être codé par une distribution de possibilité $\pi_i(X_i)$ qui correspond à une application de D_i vers l'intervalle $[0, 1]$. Pour un état $x_{ik} \in D_i$, $\pi(x_{ik}) = 1$ signifie que la réalisation de x_{ik} est totalement possible et $\pi(x_{ik}) = 0$ signifie que x_{ik} est un état impossible. Généralement, on suppose qu'au moins un état est totalement possible et par conséquent la distribution de possibilité π_i est dite normalisée. Etant donnée une variable X_i, les deux cas de connaissance extrêmes dans la théorie de possibilité sont : l'ignorance totale si $\forall x_{ik} \in D_i, \pi(x_{ik}) = 1$ (tous les états relatifs à X_i sont totalement possibles) et la précision si $\exists x_{ij} \in D_i$ t.q. $\pi(x_{ij}) = 1$ et $\forall x_{ik} \in D_i$ t.q. $x_{ik} \neq x_{ij}, \pi_i(X_i) = 0$. Soit $\pi_{(i)}$ le i^{eme} degré dans la distribution π_i ayant une valeur considérée dans un ordre décroissant de ses valeurs ($\pi_{(1)}$ est la plus grande valeur et $\pi_{(m)}$ est la plus petite). La définition d'une distribution de possibilité peut être généralisée pour un ensemble de variables V défini sur l'univers de discours $\Omega = D_1 \times ... \times D_n$ codé par π_i. Une distribution de possibilité π_i correspond alors à une application de Ω vers l'intervalle $[0, 1]$. $\omega \in \Omega$ est appelée interprétation ou élémentaire et est noté par un tuple $(x_{1k}, ..., x_{nl})$. $\omega[X_i]$ correspond à la valeur associée à la variable X_i dans ω.

La marginalisation d'une distribution de possibilité utilise l'opérateur maximum et est exprimée comme suit :

$$\pi(X_i) = \max_{X_j, j \neq i} \pi(X_1, \ldots, X_n) \tag{1}$$

La théorie des possibilités se base sur le principe de la spécificité (Yager, 1992) : soient deux distributions de possibilités π et π', π est dite plus spécifique que π', en d'autres termes π est plus informative que π' ssi :

$$\forall x_{ik} \in D_i, \pi(x_{ik}) \leq \pi'(x_{ik}) \tag{2}$$

Dans ce sens, une des mesures de non-spécificité qui permet de mesurer l'imprécision relative à une distribution de possibilité est notée nsp et exprimée comme suit (Klir, Mariano, 1987) :

$$\text{nsp}(\pi) = [\sum_{i=1}^{m} (\pi_{(i)} - \pi_{(i+1)}) \ \log_2 \ i] + (1 - \pi_{(1)}) \ \log_2 \ m \tag{3}$$

où $\pi_{m+1} = 0$ par convention.

Etant donnée une distribution de possibilité π_i, on peut définir pour tout sous-ensemble $A \subseteq D_i$, deux mesures duales : la mesure de possibilité $\Pi(A) = \max_{x_{ik} \in A} \pi(x_{ik})$ et la mesure de nécessité $N(A) = 1 - \Pi(\bar{A})$ évaluant à quel niveau $\bar{A}$ est impossible.

La particularité de l'échelle possibiliste est qu'elle peut être interprétée de deux manières : d'une manière ordinale si les degrés de possibilités ne reflètent que l'ordre entre les valeurs possibles, utilisant comme opérations d'aggrégation le minimum et le maximum. Dans l'interprétation numérique, nous nous intéressons aux valeurs prises par les degrés de possibilités possibiliste et les opérations d'aggrégation deviennent le produit et le maximum. Ces deux interprétations induisent deux définitions du conditionnement possibiliste (Dubois, Prade, 1998) qui correspond à la révision d'une distribution de possibilité initiale π_i par une nouvelle information certaine A. Le conditionnement numérique basé sur l'opérateur produit est défini comme suit :

$$\pi(\omega|_{\times} A) = \begin{cases} \frac{\pi(\omega)}{\Pi(A)} & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases} \tag{4}$$

Le conditionnement ordinal basé sur l'opérateur minimum est défini comme suit :

$$\pi(\omega \mid_m A) = \begin{cases} 1 & \text{si } \pi(\omega) = \Pi(A) \text{ et } \omega \in A \\ \pi(\omega) & \text{si } \pi(\omega) < \Pi(A) \text{ et } \omega \in A \\ 0 & \text{sinon.} \end{cases} \tag{5}$$

Tableau 1. Un exemple d'une distribution de possibilité jointe sur deux variables binaires

X_1	X_2	$\pi(X1, X_2)$
x_{11}	x_{21}	$0,8$
x_{11}	x_{22}	$0,4$
x_{12}	x_{21}	$0,6$
x_{12}	x_{22}	1

EXEMPLE 1. — Soit X_1 et X_2 deux variables binaires définies sur $D_1 = \{x_{11}, x_{12}\}$ et $D_2 = \{x_{21}, x_{22}\}$ dont la distribution jointe est décrite dans le tableau 1. Prenons ces exemples de distributions de possibilités conditionnelles calculées à partir de l'équation 4 :
$\pi(X_1 = x_{11}, X_2 = x_{21}|X_1 = x_{11}) = \frac{0,8}{\max(0,8,0,4)} = 1.$
$\pi(X_1 = x_{11}, X_2 = x_{22}|X_1 = x_{11}) = \frac{0,4}{\max(0,8,0,4)} = 0,5.$
$\pi(X_1 = x_{12}, X_2 = x_{21}|X_1 = x_{11}) = 0.$
$\pi(X_1 = x_{12}, X_2 = x_{22}|X_1 = x_{11}) = 0.$
$\pi(X_1 = x_{11}, X_2 = x_{21}|X_1 = x_{12}) = 0.$
$\pi(X_1 = x_{11}, X_2 = x_{22}|X_1 = x_{12}) = 0.$
$\pi(X_1 = x_{12}, X_2 = x_{21}|X_1 = x_{12}) = \frac{0,6}{\max(0,6,1)} = 0,6.$

$\pi(X_1 = x_{12}, X_2 = x_{22}|X_1 = x_{12}) = \frac{1}{\max(0,6,1)} = 1.$

Nous pouvons alors trouver $\pi(X_2|X_1)$ en marginalisant (équation 1) :
$\pi(X_2 = x_{21}|X_1 = x_{11}) = \max(1, 0) = 1.$
$\pi(X_2 = x_{22}|X_1 = x_{11}) = \max(0, 5, 0) = 0, 5.$
$\pi(X_2 = x_{21}|X_1 = x_{12}) = \max(0, 6, 0) = 0, 6.$
$\pi(X_2 = x_{22}|X_1 = x_{12}) = \max(1, 0) = 1.$

□

2.1.2. Théorie des possibilités et théorie des probabilités

La théorie des probabilités représente un cadre formel bien fondé pour représenter et raisonner avec des informations incertaines mais précises. Par conséquent, en présence de l'imprécision, cette théorie ne présente pas le cadre adéquat et la manipulation probabiliste de ces informations notamment la manipulation de l'ignorance totale et partielle aboutit à des paradoxes connus comme celui de Peter, Paul et Mary (Smets, Kennes, 1994) décrit dans l'exemple 2.

EXEMPLE 2. — Le big boss a décidé que Mr Jones devrait mourir. Le juge sait que les trois tueurs possibles sont Peter, Paul et Mary et il n'a aucune idée sur le sexe du tueur. Comment peut-on représenter les connaissances du juge sur l'auteur du meurtre?
Trois tueurs possibles Peter, Paul et Mary : P(Peter) = P(Paul) = P(Mary) = $\frac{1}{3}$ ce qui entrainerait P(homme) = $\frac{2}{3}$ et P(femme) = $\frac{1}{3}$. Or, le juge n'a aucune idée sur le sexe de l'auteur du meurtre ce qui devrait mener à P(homme) = $\frac{1}{2}$ et P(femme) = $\frac{1}{2}$ □

Une interprétation de la théorie des possibilités consiste à considérer une distribution de possibilité comme une famille de distributions de probabilités (Dubois, 2006) pour laquelle la mesure de probabilité de chaque sous-ensemble $A \subseteq D_i$ est bornée par sa nécessité et sa possibilité. Si $\mathcal{P}$ est l'ensemble de toutes les distributions de probabilités définies sur D_i, la famille des distributions de probabilités relative à π_i, notée $\mathcal{P}_\pi$, est définie comme suit :

$$\mathcal{P}_\pi = \{p \in \mathcal{P}, \forall A \subseteq D_i, N(A) \leq P(A) \leq \Pi(A)\} \quad (6)$$

Plusieurs chercheurs ont proposé de transformer une distribution de probabilités en une distribution de possibilité soit directement (Klir, Parviz, 1992; Dubois *et al.*, 1993, 2004; Mouchaweh *et al.*, 2006; Bouguelid, 2007; Dubois *et al.*, 2008) soit à travers d'intervalles de confiance (De Campos, Huete, 2001; Masson, Denœux, 2006; Hou, Yang, 2010). Nous détaillons maintenant une méthode de transformation de chaque catégorie c'est-à-dire la *transformation optimale* (Dubois *et al.*, 1993) et l'adaptation de la *normalisation par maximum* avec les intervalles de confiance (De Campos, Huete, 2001). Soit $P(X_i)$ une distribution de probabilités sur X_i et supposons que $\{p_{(1)}, p_{(2)}, ..., p_{(m)}\}$ est l'ordre décroissant de $P(X_i)$. La *transformation optimale*, c'est-à-dire qui donne la distribution de possibilité la plus spécifique tout en satisfai-

sant les propriétés de cohérence et de préservation de l'ordre, est exprimée comme suit :

$$\pi(x_{ik}) = \sum_{j/p_{(j)} \leq p_{(k)}} p_{(j)} \tag{7}$$

L'adaptation de la *normalisation par le maximum* permet de prendre en considération la notion de confiance qui augmente en augmentant la taille de la base de données. L'adaptation de la *normalisation par le maximum* avec un intervalle de confiance [pu_k,pl_j] et un degré de confiance ϵ est exprimée comme suit (De Campos, Huete, 2001) :

$$\pi(x_{ik}) = \min(\frac{pu_k}{\max_{j=1..m} pl_j}, 1) \tag{8}$$

où $pu_k = \min(p_{(k)} + \frac{c_\epsilon}{\sqrt{N}}\sqrt{p_{(k)}(1-p_{(k)})}, 1)$,
$pl_j = \max(p_{(j)} - \frac{c_\epsilon}{\sqrt{N}}\sqrt{p_{(j)}(1-p_{(j)})}, 0)$, $c_\epsilon = \frac{1}{100} \times \frac{1+\epsilon}{2}$ et N est le nombre d'observations dans la base de données.

Dans l'interprétation ordinale, il existe des travaux qui estiment une distribution de possibilité mais à partir des probabilités infinitésimales (Henrion *et al.*, 1994; Darwiche, Goldszmidt, 1994; Giang, Shenoy, 1999), pour plus de détails voir (Sabbadin, 2001). Notons que (De Campos, Huete, 2001) propose que les transformations décrites dans leur article s'appliquent dans le cas numérique et dans le cas ordinal, mais en se basant sur la définition considérant une distribution de possibilité comme une famille de distributions de probabilités, la définition que (Dubois, 2006) restreint de son côté au cas numérique.

2.1.3. Estimation d'une distribution de possibilité

Dans le cadre possibiliste numérique, (Joslyn, 1991) a proposé une méthode basée sur la notion d'histogrammes possibilistes qui permet d'estimer des distributions de possibilités directement à partir de données imprécises décrites par plus d'une valeur. L'auteur relie sa méthode avec la notion de spécificité d'une distribution de possibilité dans quelques cas particuliers comme le cas d'un ensemble de données certaines et le cas d'un ensemble de données consistant. Soit $\mathcal{D}_i = \{d_i^{(l)}\}$ un ensemble de données relatif à une variable X_i, par conséquent, $d_i^{(l)} \in D_i$ (resp. $d_i^{(l)} \subseteq D_i$) si les données sont précises (resp. imprécises). Le nombre d'occurrences d'une instance $x_{ik} \in D_i$, noté N_{ik}, est le nombre de fois où x_{ik} apparaît dans $\mathcal{D}_i$: $N_{ik} = N(\{x_{ik}\} \in \mathcal{D}_i)$. L'estimation non normalisée de $\hat{\pi}^{nn}(x_{ik})$ est donnée par la formule suivante :

$$\hat{\pi}^{nn}(x_{ik}) = \frac{N_{ik}}{N} \tag{9}$$

où N le nombre d'observations dans $\mathcal{D}_i$. N est égal (resp. inférieur ou égal) à la somme de N_{ik} si les données sont précises (resp. imprécises).

Joslyn normalise les degrés de possibilités en les divisant par leur valeur maximale. L'équation 9 devient :

$$\hat{\pi}^{n}(x_{ik}) = \frac{N_{ik}}{\max(N_{ik})} \tag{10}$$

L'équation 10 est généralisable à un ensemble de variables $X_i, X_j, ...X_w$. Dans ce cas N_{ik} devient $N_{ik,jl,...,wp} = N(\{x_{ik}x_{jl}...x_{wp}\} \subseteq \mathcal{D}_{ijw})$

Tableau 2. Exemple d'une base de données imprécises

X_1	X_2
x_{11}, x_{12}	x_{21}, x_{22}
x_{12}, x_{13}	x_{21}, x_{22}
x_{11}, x_{12}	x_{21}
x_{11}, x_{13}	x_{21}
x_{12}	x_{22}

EXEMPLE 3. — Le tableau 2 présente un exemple de données imprécises relatives à une variable ternaire X_1 t.q. $D_1 = \{x_{11}, x_{12}, x_{13}\}$ et une variable binaire X_2 t.q. $D_2 = \{x_{21}, x_{22}\}$. A partir de cet ensemble de données :

$N(x_{11}) = 3$, $N(x_{12}) = 4$ et $N(x_{13}) = 2$

L'estimation normalisée de π_i est :

$\hat{\pi}^n(x_{11}) = \frac{3}{4}$, $\hat{\pi}^n(x_{12}) = \frac{4}{4}$ et $\hat{\pi}^n(x_{13}) = \frac{2}{4}$

De la même manière,

$N(x_{21}) = 4$ et $N(x_{22}) = 3$

$\hat{\pi}^n(x_{21}) = \frac{4}{4}$ et $\hat{\pi}^n(x_{22}) = \frac{3}{4}$

Nous avons aussi :

$N(x_{11}, x_{21}) = 3$, $N(x_{11}, x_{22}) = 1$, $N(x_{12}, x_{21}) = 3$, $N(x_{12}, x_{22}) = 3$, $N(x_{13}, x_{21}) = 2$ et $N(x_{13}, x_{22}) = 1$

$\hat{\pi}^n(x_{11}, x_{21}) = \frac{3}{3}$, $\hat{\pi}^n(x_{11}, x_{22}) = \frac{1}{3}$, $\hat{\pi}^n(x_{12}, x_{21}) = \frac{3}{3}$, $\hat{\pi}^n(x_{12}, x_{22}) = \frac{3}{3}$, $\hat{\pi}^n(x_{13}, x_{21}) = \frac{2}{3}$ et $\hat{\pi}^n(x_{13}, x_{22}) = \frac{1}{3}$

□

2.1.4. Echantillonnage d'une variable

Le processus d'échantillonnage d'une variable correspond à la génération d'un ensemble de données représentatif d'une distribution. Dans le cadre possibiliste numérique, il existe à notre connaissance deux approches s'intéressant à l'échantillonnage d'une variable décrite par une distribution de possibilité (Yager, 1982; Guyonnet *et al.*, 2003). Ces méthodes se basent essentiellement sur la notion de α-cut proposée par Zadeh (Zadeh, 1975) : α-cut$_{X_i} = \{x_{ik} \in D_i$ t.q. $\pi(x_{ik}) \geq \alpha\}$ où $\alpha \in [0, 1]$ en générant aléatoirement α. La méthode proposée dans (Guyonnet *et al.*, 2003) s'intéresse à la génération de données imprécises en retournant toutes les valeurs de α-cut$_{X_i}$ pour une variable quelconque X_i. La méthode de (Yager, 1982) se consacre à la génération de données précises en retournant une seule valeur choisie uniformément parmi α-cut$_{X_i}$.

EXEMPLE 4. — Soit une variable X_1 dont la distribution de possibilité est $\pi(x_{11}) = 0,1$, $\pi(x_{12}) = 1$ et $\pi(x_{13}) = 0,8$. Pour une valeur de $\alpha = 0,7$ (resp. $\alpha = 0,1$, $\alpha = 0,9$ et $\alpha = 1$), α-cut$_{X_1} = \{x_{12}, x_{13}\}$ (resp. α-cut$_{X_1} = \{x_{11}, x_{12}, x_{13}\}$, α-cut$_{X_1} = \{x_{12}\}$ et α-cut$_{X_1} = \{x_{12}\}$ □

2.2. Réseaux possibilistes

2.2.1. Définitions

Les réseaux possibilistes (Fonck, 1992, 1994b; Borgelt, Kruse, 2003) représentent la contrepartie possibiliste des réseaux bayésiens (Pearl, 1988; Jensen, 1996). La première composante de ces modèles, qualitative ou graphique, est donnée par un graphe orienté sans circuit (DAG) qui code un ensemble de relations d'indépendance. Chaque variable $X_i \in V$ est conditionnellement indépendante de ses non-descendants sachant ses parents (Fonck, 1992). La seconde composante, quantitative ou numérique, correspond à l'ensemble des distributions de possibilités conditionnelles normalisées relatives à chaque nœud $X_i \in V$ sachant ses parents, notés Pa(X_i), c'est-à-dire $\pi(X_i|\text{Pa}(X_i))$.

Les deux formes de conditionnement possibilistes relatives aux interprétations de l'échelle possibiliste induisent naturellement deux manières différentes de définir les réseaux possibilistes : les réseaux possibilistes basés sur le conditionnement numérique (équation 4), nommés également réseaux possibilistes quantitatifs. Ces réseaux sont très semblables aux réseaux bayésiens théoriquement et algorithmiquement. En effet, les deux modèles partagent la même composante graphique (c'est-à-dire le DAG) et également l'opérateur produit dans le processus calculatoire. Cela n'est pas le cas des réseaux possibilistes basés sur le conditionnement minimum (équation 5), nommés réseaux possibilistes qualitatifs et qui représentent une sémantique bien différente. Notons que les réseaux possibilistes quantitatifs peuvent présenter une approximation des réseaux crédaux (Cozman, 2000) basés sur les probabilités imprécises.

Dans les deux cas, les réseaux possibilistes sont une représentation compacte des distributions de possibilités. Plus précisément, la distribution de possibilité jointe relative aux réseaux possibilistes est calculée avec la règle de chaînage exprimée par :

$$\pi_{\otimes}(X_1, ..., X_n) = \otimes_{i=1..n} \pi(X_i \mid_{\otimes} \text{Pa}(X_i)) \tag{11}$$

où $\otimes$ correspond à l'opérateur minimum (min) pour les réseaux possibilistes qualitatifs et à l'opérateur produit ($\times$) pour les réseaux quantitatifs et $|_{\otimes}$ correspond à l'opérateur de conditionnement associé (équations 4 et 5). Notons que la condition d'acyclicité garantit la normalisation de la distribution de possibilité jointe.

Les réseaux possibilistes permettent de modéliser des relations d'indépendance conditionnelle pour lesquelles il existe plusieurs définitions dans le cadre possibiliste (De Campos, Huete, 1999b,a; Ben Amor, Benferhat, 2005) contrairement au cadre probabiliste où une seule définition existe. Les définitions les plus communes et qui

satisfont les propriétés des semi graphoides sont exprimées comme suit (Dubois *et al.*, 1994) :

– Relation d'indépendance basée sur le produit, notée $\mathrm{I}_{\mathrm{prod}}$, utilisant le conditionnement numérique sur lequel se basent les réseaux possibilistes quantitatifs et exprimée par :

$$\forall X_i, X_j \in V \quad \mathrm{I}_{\mathrm{prod}}(X_i, X_j) \text{ ssi } \quad \pi(X_i, X_j) = \pi(X_i) \times \pi(X_j) \qquad (12)$$

– Relation de non-interactivité, notée I_{NI}, utilisant le conditionnement ordinal sur lequel se basent les réseaux possibilistes qualitatifs et exprimée :

$$\forall X_i, X_j \in V \quad \mathrm{I}_{\mathrm{NI}}(X_i, X_j) \text{ ssi } \quad \pi(X_i, X_j) = \min(\pi(X_i), \pi(X_j)) \qquad (13)$$

EXEMPLE 5. — Le DAG de la figure 1 et les distributions de possibilités conditionnelles du tableau 3 représentent un réseau possibiliste défini sur quatre variables binaires X_1, X_2, X_3 et X_4 définies sur $D_1 = \{x_{11}, x_{12}\}$ $D_2 = \{x_{21}, x_{22}\}$, $D_3 = \{x_{31}, x_{32}\}$ et $D_4 = \{x_{41}, x_{42}\}$. Le tableau 4 présente la distribution de possibilité jointe dans les interprétations numérique ($\pi_\times$) et ordinale (π_m).

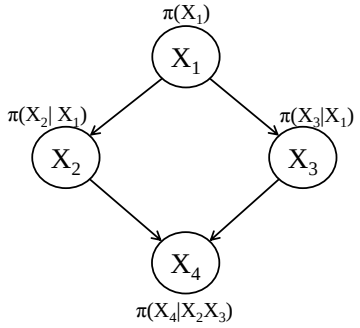


Figure 1. Composante graphique d'un réseau possibiliste

Tableau 3. Composante numérique du réseau possibiliste de la figure 1

X_1	$\pi(X_1)$
x_{11}	$0,4$
x_{12}	1

X_1	X_2	$\pi(X_2\|X_1)$
x_{11}	x_{21}	$0,3$
x_{11}	x_{22}	1
x_{12}	x_{21}	1
x_{12}	x_{22}	$0,7$

X_1	X_3	$\pi(X_3\|X_1)$
x_{11}	x_{31}	1
x_{11}	x_{32}	1
x_{12}	x_{31}	$0,4$
x_{12}	x_{32}	1

X_2	X_3	X_4	$\pi(X_4\|X_2, X_3)$
x_{21}	x_{31}	x_{41}	1
x_{21}	x_{31}	x_{42}	$0,3$
x_{21}	x_{32}	x_{41}	1
x_{21}	x_{32}	x_{42}	$0,2$

X_2	X_3	X_4	$\pi(X_4\|X_2, X_3)$
x_{22}	x_{31}	x_{41}	$0,5$
x_{22}	x_{31}	x_{42}	1
x_{22}	x_{32}	x_{41}	1
x_{22}	x_{32}	x_{42}	$0,1$

Notons que dans (Borgelt, Kruse, 2003), les auteurs définissent les réseaux possibilistes quantitatifs différemment avec une composante numérique dont les distributions de possibilités conditionnelles ne sont pas forcément normalisées. En outre, ils se basent sur la relation de non-interactivité qui n'est applicable que dans l'interprétation ordinale. Dans ce cadre, il est aussi possible d'appliquer la règle de chaînage (équation 11) sur des distributions de possibilités non normalisées.

Tableau 4. distribution de possibilité jointe du réseau possibiliste défini par la figure 1 et le tableau 3 dans les interprétations numérique ($\pi_\times$) et ordinale (π_m)

X_1	X_2	X_3	X_4	$\pi_\times$	π_m	X_1	X_2	X_3	X_4	$\pi_\times$	π_m
x_{11}	x_{21}	x_{31}	x_{41}	$0,12$	$0,3$	x_{12}	x_{21}	x_{31}	x_{41}	$0,4$	$0,4$
x_{11}	x_{21}	x_{31}	x_{42}	$0,036$	$0,3$	x_{12}	x_{21}	x_{31}	x_{42}	$0,12$	$0,3$
x_{11}	x_{21}	x_{32}	x_{41}	$0,12$	$0,3$	x_{12}	x_{21}	x_{32}	x_{41}	1	1
x_{11}	x_{21}	x_{32}	x_{42}	$0,024$	$0,2$	x_{12}	x_{21}	x_{32}	x_{42}	$0,2$	$0,2$
x_{11}	x_{22}	x_{31}	x_{41}	$0,2$	$0,4$	x_{12}	x_{22}	x_{31}	x_{41}	$0,14$	$0,4$
x_{11}	x_{22}	x_{31}	x_{42}	$0,4$	$0,4$	x_{12}	x_{22}	x_{31}	x_{42}	$0,28$	$0,4$
x_{11}	x_{22}	x_{32}	x_{41}	$0,4$	$0,4$	x_{12}	x_{22}	x_{32}	x_{41}	$0,7$	$0,7$
x_{11}	x_{22}	x_{32}	x_{42}	$0,04$	$0,1$	x_{12}	x_{22}	x_{32}	x_{42}	$0,07$	$0,1$

□

2.2.2. Propagation

Un des traitements les plus intéressants que nous pouvons appliquer sur les réseaux possibilistes est la propagation qui consiste à évaluer l'impact de la réalisation d'un évènement sur le reste des variables. Comme c'est le cas avec les réseaux bayésiens, ce problème est classé comme NP-difficile sauf pour les poly-arbres (graphes simplement connectés) où la propagation peut être réalisée en temps polynomial (Cooper, 1990). La plupart des algorithmes de propagation proposés dans la littérature (Fonck, 1994a; Gebhardt, Kruse, 1997) présentent une adaptation des méthodes initialement dédiées aux réseaux bayésiens et plus spécifiquement la méthode exacte de Pearl (Pearl, 1988) pour le cas particulier des DAG simplement connectés et l'algorithme de Jensen (Jensen, 1996) dans le cas plus général des DAG à connexions multiples. Ces travaux ont montré que les algorithmes concernant les réseaux possibilistes basés sur l'opérateur produit sont très semblables aux algorithmes probabilistes puisqu'ils utilisent le même opérateur produit. Ceci n'est pas le cas pour les réseaux basés sur le min. En effet, cet opérateur possède des propriétés particulières, telle que l'idempotence, ce qui permet la proposition d'algorithmes spécifiques. Nous citons en particulier l'algorithme approximatif anytime proposé dans (Ben Amor *et al.*, 2003) permettant d'éviter la transformation du DAG initial en arbre de jonction dans le cas des DAG à connexions multiples. Dans (Benferhat, Smaoui, 2007), les auteurs ont proposé un algorithme de propagation dans les réseaux possibilistes hybrides où la représentation de l'incertitude associée à chaque nœud et ses parents est faite avec la logique possibiliste. Les algorithmes de propagation proposés dans (Ayachi *et al.*, 2010) se basent sur les techniques de compilation des connaissances à l'instar des travaux de (Chavira, Darwiche, 2007) sur les réseaux bayésiens. Tous ces travaux ont montré que

l'utilisation de la théorie des possibilités dans son interprétation qualitative permet un gain considérable au niveau des traitements nécessaires à la propagation. Récemment, un autre travail a été proposé pour la propagation dans les structures naïves traitant des évidences souples (soft evidence) (Benferhat, Tabia, 2012) sous forme de distributions de possibilités sur les variables observées.

2.2.3. Intérêt

Les réseaux bayésiens et les réseaux possibilistes constituent des outils de représentation et d'analyse importants en présence d'informations incertaines. Néanmoins, les réseaux bayésiens nécessitent une connaissance parfaite relative à la composante numérique afin d'être opérationnels ce qui n'est pas le cas avec les réseaux possibilistes où l'ignorance totale est naturellement supportée par la théorie des possibilités (voir section 2.1.1). La modélisation ordinale est également un autre atout des réseaux possibilistes puisqu'elle offre aux experts une alternative très pratique pour la représentation de leurs connaissances incertaines. Ceci veut dire que les deux modèles sont des outils complémentaires qui doivent être utilisés en fonction du problème et des données disponibles. A titre d'exemple, nous pouvons citer le travail de (Caglioni *et al.*, 2014) où les auteurs explorent l'élicitation de connaissances incertaines décrivant le domaine de la spécialisation sociale dans les espaces métropolisés. Pour ce faire, ils ont besoin d'un cadre où les facteurs influant le domaine étudié peuvent être captés d'une manière qualitative sous une forme hiérarchique sans pour autant les quantifier. En outre, l'évaluation de l'impact de certaines variables sur les autres n'est pas précise et elle est modélisée par des fonctions d'appartenance floues qui génèrent des distributions de possibilités. Par conséquent, l'utilisation des réseaux bayésiens dans ce cas n'est pas très appropriée. Dans ce qui suit, nous présentons un autre exemple dans le domaine de recherche d'information où la notion de l'ignorance totale est inévitable et mieux exprimée avec des réseaux possibilistes. Dans (Boughanem *et al.*, 2009), les auteurs cherchent à sélectionner dans une collection de documents ceux susceptibles d'être pertinents. Leur modèle distingue, par exemple, le rejet d'un document comme non pertinents et sa sélection par l'utilisation des mesures de possibilité et de nécessité. La possibilité de la pertinence est utilisée pour éliminer les documents non pertinents alors que la nécessité de la pertinence permet de sélectionner les plus intéressants. De plus, la pondération d'un mot par une unique probabilité ne permet pas de capturer la double notion de spécialité et non spécialité, plus facilement modélisable grâce à la théorie des possibilités.

2.2.4. Solutions logicielles

Contrairement aux réseaux bayésiens, peu de logiciels manipulent des réseaux possibilistes. Aucune de ces solutions ne permet à la fois la définition, l'apprentissage et le raisonnement. Pour la propagation dans les réseaux possibilistes basés sur le produit, nous pouvons citer Pulcinella (Saffiotti, Umkehrer, 1991), POSSINFER (Gebhardt, Kruse, 1996) et PNT (*Possibilistic Networks Toolbox*) (Ben Amor *et al.*, 2003). PNT traite aussi les réseaux basés sur le min. Il n'existe aucune solution pour l'apprentissage des paramètres. Pour l'apprentissage de la structure des réseaux possibilistes

basés sur le produit, pour autant que nous le sachions, il existe une unique solution logicielle à savoir INeS (*Induction of Networks Structure*) (Borgelt *et al.*, 2009) implémentant certaines méthodes détaillées dans la section 4.2.

3. Apprentissage des paramètres des réseaux possibilistes quantitatifs

Dans la suite de ce papier, nous nous intéressons aux réseaux possibilistes basés sur le produit donc le symbole $|$ signifiera $|_{\times}$.

3.1. Cadre théorique

L'estimation des paramètres des réseaux bayésiens est effectuée en satisfaisant le principe de l'*entropie maximale* (Herskovits, Cooper, 1991). L'analogue de cette dernière dans le cadre possibiliste correspond à la spécificité minimale. Par conséquent, par analogie aux réseaux bayésiens, apprendre les paramètres d'un réseau possibiliste devrait revenir à estimer ses distributions de possibilités selon le principe de la *spécificité minimale* (Dubois, Hüllermeier, 2007) c'est-à-dire estimer les distributions de possibilités les moins spécifiques sous contraintes d'informations fournies par les données.

A notre connaissance, il n'existe pas de travaux dédiés à l'apprentissage des paramètres des réseaux possibilistes. Nous proposons ici deux solutions en nous inspirant des méthodes d'estimation de distributions de possibilités décrites dans les sections 2.1.2 et 2.1.3. La première solution la plus naïve consiste à utiliser une des méthodes de transformation de distributions de probabilités vers des distributions de possibilités. En premier lieu, nous apprenons les distributions de probabilités de chaque nœud $P(X_i|\text{Pa}(X_i))$ à partir de données précises. Ensuite, nous transformons ces distributions en des distributions de possibilités. La deuxième méthode consiste à utiliser l'estimation des distributions de possibilités en utilisant des histogrammes possibilistes décrits dans la section 2.1.3 et traitant des données imprécises. Nous allons démontrer dans la suite que ces deux solutions ne donnent pas de bons résultats.

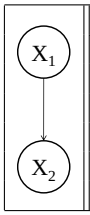
3.2. Apprentissage des paramètres basé sur les transformations

La première solution naïve consiste à utiliser les transformations des distributions de probabilités apprises à partir de données vers des distributions de possibilités (voir section 2.1.2). Notons que ces méthodes de transformation ne nous permettent d'apprendre que les paramètres des réseaux possibilistes basés sur le produit à partir de données précises (ou contenant des données manquantes en utilisant le principe de l'algorithme EM (Lauritzen, 1995)). Cette démarche se base sur le lien entre la *vraisemblance maximale* proposée dans le cadre probabiliste et une distribution de possibilité énoncé dans (Dubois, 2006) : *lorsqu'il manque des probabilités a priori, les fonctions de vraisemblance peuvent être interprétées comme des distributions de possibilités*. Rappelons que les méthodes de transformation manipulent les distributions

de probabilités soit directement soit à travers des intervalles de confiance. Dans ce qui suit, nous présentons un exemple d'application d'une des transformations de chaque catégorie, c'est-à-dire la *transformation optimale* et la *normalisation par le maximum*.

EXEMPLE 6. — Soit le réseau bayésien décrit dans le tableau 5 avec deux variables X_1 ternaire définie sur $D_1 = \{x_{11}, x_{12}, x_{13}\}$ et X_2 binaire définie sur $D_2 = \{x_{21}, x_{22}\}$. Les tableaux 6 et 7 sont les résultats de la transformation du réseau bayésien. $\pi(X_1, X_2)$ représente la distribution de possibilité jointe, le résultat de la transformation de la distribution de probabilités jointe du réseau bayésien. $\pi_t(X_1, X_2)$ est la distribution de possibilité jointe calculée à partir des distributions de possibilités conditionnelles (équation 11). L'exemple (a) utilise *la transformation optimale* et l'exemple (b) utilise l'adaptation de la *normalisation par le maximum* avec les intervalles de confiance. Nous remarquons que les valeurs de $\pi(X_1, X_2)$ et $\pi_t(X_1, X_2)$ sont différentes. □

Tableau 5. Exemple d'un réseau bayésien

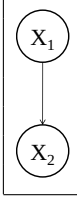


X_1	$P(X_1)$
x_{11}	$0,2$
x_{12}	$0,3$
x_{13}	$0,5$

X_1	X_2	$P(X_1\|X_2)$
x_{11}	x_{21}	$0,4$
x_{11}	x_{22}	$0,6$
x_{12}	x_{21}	$0,9$
x_{12}	x_{22}	$0,1$
x_{13}	x_{21}	$0,3$
x_{13}	x_{22}	$0,7$

X_1	X_2	$P(X_1, X_2)$
x_{11}	x_{21}	$0,08$
x_{11}	x_{22}	$0,12$
x_{12}	x_{21}	$0,27$
x_{12}	x_{22}	$0,03$
x_{13}	x_{21}	$0,15$
x_{13}	x_{22}	$0,35$

Tableau 6. Exemple (a) de transformation du réseau bayésien vers un réseau possibiliste par la transformation optimale

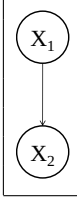


X_1	$\pi(X_1)$
x_{11}	$0,2$
x_{12}	$0,5$
x_{13}	1

X_1	X_2	$\pi(X_2\|X_1)$
x_{11}	x_{21}	$0,4$
x_{11}	x_{22}	1
x_{12}	x_{21}	1
x_{12}	x_{22}	$0,1$
x_{13}	x_{21}	$0,3$
x_{13}	x_{22}	1

X_1	X_2	$\pi_t(X_1, X_2)$	$\pi(X_1, X_2)$
x_{11}	x_{21}	$0,08$	$0,11$
x_{11}	x_{22}	$0,2$	$0,23$
x_{12}	x_{21}	$0,5$	$0,65$
x_{12}	x_{22}	$0,05$	$0,03$
x_{13}	x_{21}	$0,3$	$0,38$
x_{13}	x_{22}	1	1

Tableau 7. Exemple (b) de transformation du réseau bayésien vers un réseau possibiliste par l'adaptation de la normalisation par le maximum avec les intervalles de confiance



X_1	$\pi(X_1)$
x_{11}	$0,78$
x_{12}	1
x_{13}	1

X_1	X_2	$\pi(X_2\|X_1)$
x_{11}	x_{21}	$0,68$
x_{11}	x_{22}	1
x_{12}	x_{21}	1
x_{12}	x_{22}	$0,22$
x_{13}	x_{21}	$0,51$
x_{13}	x_{22}	1

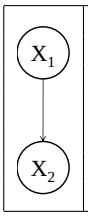
X_1	X_2	$\pi_t(X_1, X_2)$	$\pi(X_1, X_2)$
x_{11}	x_{21}	$0,54$	$0,28$
x_{11}	x_{22}	$0,78$	$0,41$
x_{12}	x_{21}	1	$0,89$
x_{12}	x_{22}	$0,22$	$0,11$
x_{13}	x_{21}	$0,51$	$0,51$
x_{13}	x_{22}	1	1

3.3. Apprentissage direct des paramètres à partir de données

La méthode de Joslyn (Joslyn, 1991) permet d'inférer directement des distributions de possibilités à partir de données. Elle permet de traiter une forme plus générique de données imparfaites (voir section 2.1.3). Nous essayons dans ce qui suit d'appliquer cette méthode pour apprendre les paramètres d'un réseau possibiliste à partir d'une base de données imprécises.

EXEMPLE 7. — Reprenons l'exemple de la base de données imprécises décrite dans le tableau 2. En appliquant la méthode de Joslyn, nous obtenons les distributions de possibilités du tableau 8. Pour les distributions de possibilités conditionnelles, nous calculons par exemple $\hat{\pi}^n(x_{21}|x_{11})$ en divisant $\hat{\pi}^{nn}(x_{11}, x_{21}) = 0,6$ par $\hat{\pi}^{nn}(x_{11}) = 0,6$ calculés avec l'équation 9 et nous normalisons par la suite les degrés de possibilités obtenus. Nous remarquons que les valeurs de $\hat{\pi}^n$=$\hat{\pi}^n(X_1, X_2)$, la distribution de possibilité jointe calculée à partir de données, et $\hat{\pi}^n_l$ = $\hat{\pi}^n_l(X_1, X_2)$, la distribution de possibilité jointe calculée à partir des distributions de possibilités conditionnelles (équation 11), sont différentes.

Tableau 8. Exemple d'apprentissage des paramètres d'un réseau possibiliste



X_1	$\hat{\pi}^n(X_1)$	X_1	X_2	$\hat{\pi}^n(X_1\|X_2)$	X_1	X_2	$\hat{\pi}^n_l$	$\hat{\pi}^n$
x_{11}	$0,75$	x_{11}	x_{21}	1	x_{11}	x_{21}	$0,75$	1
x_{12}	1	x_{11}	x_{22}	$0,33$	x_{11}	x_{22}	$0,25$	$0,33$
x_{13}	$0,5$	x_{12}	x_{21}	1	x_{12}	x_{21}	1	1
		x_{12}	x_{22}	1	x_{12}	x_{22}	1	1
		x_{13}	x_{21}	1	x_{13}	x_{21}	$0,5$	$0,66$
		x_{13}	x_{22}	$0,5$	x_{13}	x_{22}	$0,25$	$0,33$

□

3.4. Discussion

Dans les deux sections précédentes, nous avons appliqué les deux méthodes de la littérature utilisables pour apprendre les paramètres d'un réseau possibiliste : les transformations d'une distribution de possibilité vers une distribution de probabilités et l'estimation directe des distributions de possibilités. Nous avons montré sur des exemples simples qu'aucune de ces méthodes ne semble satisfaisante. Les exemples 6 et 7 montrent que ces méthodes ne sont valables que si elles sont appliquées à la distribution jointe globalement et pas à chaque distribution locale séparément. Les méthodes de transformation dans (Klir, Parviz, 1992; Dubois *et al.*, 1993, 2004; Mouchaweh *et al.*, 2006; Bouguelid, 2007; Dubois *et al.*, 2008) ont déjà été testées pour la transformation des réseaux bayésiens vers des réseaux possibilistes dans (Ben Slimen *et al.*, 2013). Cette étude confirme bien notre remarque. Les autres méthodes de transformation présentent le même inconvénient. Le problème vient du fait que la notion de marginalisation des mesures de probabilités impliquant l'opérateur addition est très différente de l'idée de marginalisation des mesures de possibilités impliquant le maximum. La méthode directe d'apprentissage des paramètres à partir de données im-

précises présente le même inconvénient. Par conséquent, il faudrait adapter ces transformations et la combinaison des distributions de possibilité et particulièrement les distributions conditionnelles afin d'obtenir une distribution de possibilité jointe qui domine la distribution de probabilités jointe initiale. L'application de ces méthodes doit donc se faire avec précaution : cherche-t-on à estimer en priorité des paramètres locaux (intérêt habituel des modèles graphiques probabilistes ou possibilistes), ou à estimer directement une loi jointe en perdant l'intérêt de la décomposition graphique? Il est aussi possible de s'interroger sur l'intérêt d'apprendre un réseau possibiliste avec des données précises. L'imprécision pourrait venir dans ce cas d'un nombre de données limité. Dans ce contexte, les réseaux crédaux (Cozman, 2000) sont une alternative plus intéressante.

4. Apprentissage de la structure des réseaux possibilistes quantitatifs

Plusieurs tentatives ont été faites pour apprendre la structure des réseaux possibilistes à partir de données mais tous ces travaux ont eu lieu avant les avancées sur les réseaux possibilistes en tant que modèles d'indépendance (Ben Amor, Benferhat, 2005) et sans se préoccuper des problèmes liés à l'apprentissage des paramètres. Dans ce qui suit, nous essayons de donner un aperçu sur les méthodes de l'état de l'art tout en nous inspirant de ce qui a été déjà développé dans le contexte de l'apprentissage des réseaux bayésiens où les méthodes d'apprentissage sont généralement organisées en trois familles : méthodes basées sur la recherche d'indépendances conditionnelles, méthodes à base de scores et méthodes hybrides.

4.1. Méthodes basées sur la recherche d'indépendances conditionnelles

Les approches basées sur la recherche d'indépendances conditionnelles sont toutes basées sur un principe identique décrit par les 3 points suivants :

1. Construire un graphe non dirigé contenant les relations entre les variables, à partir de tests d'indépendance conditionnelle.

2. Détecter les V-structures (sous la forme de $X_j \rightarrow X_i \leftarrow X_k$) en utilisant encore les résultats des tests d'indépendance conditionnelle précédents.

3. Inférer l'orientation des arcs à partir des contraintes imposées par le DAG.

Mesurer l'indépendance conditionnelle est une phase indispensable dans cette famille d'approche. Dans le cadre possibiliste, pour autant que nous le sachions, une seule tentative a été faite pour chercher des relations d'indépendance conditionnelle afin d'apprendre des réseaux possibilistes (Sangüesa *et al.*, 1998a) à partir de données précises. Dans ce travail, les auteurs proposent une mesure d'indépendance, notée $\text{Dep}(X_i, X_j, \alpha)$ et exprimée comme suit :

$$\text{Dep}(X_i, X_j, \alpha) = 1 - \sum_{x_{jl} \in D_j} \pi(x_{jl}) \sum_{x_{ik} \in \alpha - set} |\pi(x_{ik}) - \pi(x_{ik}|x_{jl})| \qquad (14)$$

où α-set $= \{x_{ik} \in D_i$ t.q. $|\pi(x_{ik}) - \pi'(x_{ik})| \geq \alpha\}$.

4.2. Méthodes à base de scores

4.2.1. Généralités

Les approches à base de scores parcourent de manière heuristique l'espace des DAG en essayant d'optimiser une fonction objectif caractérisant l'adéquation entre le modèle et les données. Pour que ces approches à base de scores soient réalisables en pratique, le score doit être décomposable localement comme la somme de scores locaux au niveau de chaque nœud. Comme dans le cas de l'apprentissage des paramètres, il semble logique de relier la notion de score d'un réseau possibiliste à la notion de spécificité. Nous disons alors qu'un réseau possibiliste appris est bon s'il a une spécificité minimale. La non-spécificité d'un réseau possibiliste, noté nsp(G), a été proposée par Sangüesa et al. (Sangüesa *et al.*, 1998a) et peut être exprimée par :

$$\text{nsp(G)} = \sum_{X_i \in V} \text{nsp}(\pi(X_i|\text{Pa}(X_i))) \tag{15}$$

où $\text{nsp}(\pi(X_i|\text{Pa}(X_i))) = \text{nsp}(\pi(X_i, \text{Pa}(X_i))) - \text{nsp}(\pi(\text{Pa}(X_i)))$
et $\text{nsp}(\pi(X_i|\text{Pa}(X_i))) = \text{nsp}(\pi(X_i))$ si $\text{Pa}(X_i) = \emptyset$.

Notons que la spécificité n'est pas directement définie par (Sangüesa *et al.*, 1998a) sur les distributions de possibilités conditionnelles, mais sur les familles $\{X_i, \text{Pa}(X_i)\}$, ce qui, à notre avis la rendrait décomposable uniquement sur les cliques du graphe.

Dans ce qui suit, nous classifions les méthodes selon la nature du score utilisé : global ou local. Les premiers travaux proposés dans ce sens ont essayé de proposer des scores globaux. Ensuite, d'autres travaux se sont basés sur ces scores en proposant des scores locaux.

4.2.2. Méthodes à base de scores globaux

(Borgelt, Gebhardt, 1997) ont proposé une fonction de score utilisant la somme pondérée des degrés de possibilités. Etant donnée une base de données imprécise $\mathcal{D}$ et un DAG G appris à partir de $\mathcal{D}$, la somme pondérée des degrés de possibilités, notée $Q(G, \mathcal{D})$, est exprimée comme suit :

$$Q(G, \mathcal{D}) = \sum_{\{x_{1k}, \ldots, x_{nl}\} \in \mathcal{D}} N(x_{1k}, ..., x_{nl}) \hat{\pi}^{nn}(x_{1k}, ..., x_{nl}) \tag{16}$$

$\hat{\pi}^{nn}(x_{1k}, ..., x_{nl})$ est calculé localement à partir de $\hat{\pi}^{nn}(X_i|\text{Pa}(X_i))$ par l'équation 9.

4.2.3. Méthodes à base de scores locaux

Borgelt et Kruse ont proposé deux méthodes qui sont à base de scores et qui traitent aussi des données imprécises (Borgelt, Kruse, 2003). En effet, les auteurs proposent une version possibiliste des deux méthodes initialement proposées pour l'apprentissage des réseaux bayésiens : K2 (Cooper, Herskovits, 1992), arbre couvrant de

poids maximal (Heckerman *et al.*, 1995). Les scores locaux utilisés par ces méthodes sont :

- Le gain de spécificité, noté S_{gain}, est exprimé par (Borgelt, Kruse, 2003) :

$$S_{gain}(X_i, X_j) = \text{nsp}(\pi(X_i)) + \text{nsp}(\pi(X_j)) - \text{nsp}(\pi(X_i, X_j)) \tag{17}$$

où les distributions de possibilités sont calculées par l'équation 9. Cette mesure est l'ancêtre de plusieurs variantes (Borgelt, Kruse, 2003) dont nous pouvons citer le ratio de gain de spécificité, les ratios de gain de spécificité symétriques ou le gain de spécificité conditionnelle.

- L'information mutuelle possibiliste, notée d_{mi}, propose l'adaptation directe de l'entropie probabiliste et elle est définie comme suit (Borgelt, Kruse, 2003) :

$$d_{mi}(X_i, X_j) = - \sum_{\substack{x_{ik} \in D_i \\ x_{jl} \in D_j}} \frac{N_{ik,jl}}{N} \log_2 \frac{N_{ik,jl}}{\min(N_{ik}, N_{jl})} \tag{18}$$

- La mesure χ^2 possibiliste, notée d_{χ^2}, représente l'analogue du χ^2 probabiliste et elle est définie comme suit (Borgelt, Kruse, 2003) :

$$d_{\chi^2}(X_i, X_j) = \sum_{\substack{x_{ik} \in D_i \\ x_{jl} \in D_j}} \frac{(\min(N_{ik}, N_{jl}) - N_{ik,jl})^2}{\min(N_{ik}, N_{jl})} \tag{19}$$

Une généralisation de ces scores à plus de deux variables se fait comme suit : pour une variable quelconque X_i, tous ses parents peuvent être combinés en une pseudo-variable. Ainsi, nous calculons le score entre X_i et une variable artificielle qui représente la combinaison de tous ses parents c'est-à-dire le produit cartésien de leurs domaines de définition.

4.3. Méthodes hybrides

Cette dernière famille d'approches combine les avantages des deux précédentes. En effet, les méthodes hybrides emploient les informations issues d'une première phase de recherche d'indépendances conditionnelles pour guider la phase suivante qui est une recherche dans l'espace des DAG optimisant un certain score. Sangüesa et al. (Sangüesa *et al.*, 1998a) ont proposé deux méthodes d'apprentissage hybrides à partir de données précises : la première permet d'apprendre des arbres par contre la deuxième induit des DAG. Les deux méthodes se basent essentiellement sur la mesure d'indépendance conditionnelle exprimée par l'équation 14.

Cette méthode apprend au premier lieu des sous-graphes non dirigés contenant chacun un nœud, ses parents et ses enfants. Ensuite, il faut combiner ces sous-graphes et les orienter en se basant sur la spécificité du DAG exprimée dans l'équation 15. Cette méthode utilise au choix trois transformations de distributions de probabilités vers des distributions de possibilités pour apprendre les paramètres des réseaux possibilistes.

4.4. Discussion

Dans cette section, nous essayons de présenter une vue globale sur les méthodes d'apprentissage de structure proposées dans la littérature et de détailler les limites de ces méthodes. Le tableau 9 récapitule les propriétés et les objectifs de chaque mesure de score ou de dépendance passée précédemment en revue. Pour les méthodes à base de scores globaux présentées dans la section 4.2.2, notons qu'aucune méthode n'a appliqué directement la notion de spécificité décrite par l'équation 15. La structure des réseaux possibilistes appris par les versions possibilistes des algorithmes de recuit simulé et arbre couvrant de poids maximal (voir section 4.2.3) est évaluée par Q la somme pondérée des degrés de possibilités exprimée dans l'équation 16. Pour ces méthodes, nous notons une discordance entre les définitions du score global et des scores locaux (S_{gain}, d_{mi} et d_{χ^2}). En effet, le score global c'est-à-dire la somme pondérée des degrés de possibilités n'est décomposable en aucun des scores locaux. Ce score ressemble à la log-vraisemblance utilisée dans l'apprentissage des réseaux bayésiens, qui est est égale à $\sum_{\{x_{1k},...,x_{nl}\}\in\mathcal{D}} N(x_{1k},...,x_{nl}) \log\ p(x_{1k},...,x_{nl})$. L'adaptation possibiliste de (Borgelt, Gebhardt, 1997) voit ainsi disparaître le log en proposant une somme pondérée des possibilités.

Pour la méthode hybride présentée dans la section 4.3, le problème majeur résidant dans sa mesure d'indépendance conditionnelle Dep est qu'elle se base sur une mesure de similarité pour laquelle nous trouvons plusieurs formulations contradictoires dans plusieurs travaux des mêmes auteurs (Sangüesa *et al.*, 1998a; Sangüesa *et al.*, 1998b). L'autre souci concerne l'absence de proposition automatique d'un seuil permettant de décider entre les deux hypothèses, seuil calculable à partir d'un taux d'erreur de type I désiré (comme cela est fait classiquement par les tests statistiques). En outre cette méthode peut ne pas trouver une solution parce qu'elle ne prend pas en considération la propriété d'acyclicité d'un DAG au cours de l'apprentissage mais une fois la structure du réseau possibiliste apprise, la méthode vérifie si c'est bien un DAG ou non. Par ailleurs, pour estimer les distributions de possibilités, cette méthode fait recours à trois transformations (Klir, Parviz, 1992; Dubois *et al.*, 1993, 2004) qui ne sont pas applicables pour les réseaux possibilistes (voir section 3.2).

Comme dans le cas de l'apprentissage de paramètres, il semblerait plus logique d'apprendre un réseau possibiliste à partir de données imprécises. Les travaux proposés dans ce contexte restent à améliorer autant au niveau théorique qu'expérimental.

5. Conclusion et perspectives

Dans cet article, un large aperçu de la littérature concernant l'apprentissage des réseaux possibilistes quantitatifs a été présenté. Pour l'apprentissage de leurs paramètres, nous avons testé les méthodes d'estimation des distributions de possibilités existantes et nous avons montré qu'elles ne donnent pas de bons résultats, l'ensemble des paramètres locaux ne coïncidant pas avec le modèle global qui aurait été appris

Tableau 9. Tableau récapitulatif des propriétés des mesures

Mesure	Equation	Nature des données	Grandeur à optimiser
Dep	14	précises	distance entre la distribution jointe et les distributions indépendantes
nsp	15	précises	spécificité
Q	16	imprécises	divergence par rapport à un réseau initial (inconnu)
S_{gain}	17	imprécises	divergence entre la spécificité de la distribution jointe et celle des distributions indépendantes
d_{mi}	18	imprécises	différence entre la distribution jointe et les distributions indépendantes
d^2_{χ}	19	imprécises	différence entre la distribution jointe et les distributions indépendantes

avec les mêmes données. Cela rend le problème d'apprentissage des paramètres une piste de recherche encore ouverte.

Certains des travaux présentés dans cet état de l'art traitent de l'apprentissage des réseaux possibilistes à partir des données précises. Ces réseaux représentent une approximation qui cause une grande perte d'information. En outre, Les travaux récents sur les réseaux crédaux (Cozman, 2000) semblent être une alternative plus intéressante lorsque l'imprécision vient dans ce cas d'un nombre de données limité.

Concernant l'apprentissage de la structure des réseaux possibilistes basés sur le produit, les travaux existants restent limités et manquent de bases théoriques solides. Ainsi, il pourrait être intéressant de vérifier la cohérence théorique des scores proposés dans la littérature par rapport à la spécificité et de vérifier les mesures d'indépendance en se basant sur les avancées sur les réseaux possibilistes en tant que modèles d'indépendance.

Nous avons pu montrer que l'apprentissage des réseaux possibilistes quantitatifs est un problème qui reste encore à étudier sur des bases théoriques solides. La simple adaptation de méthodes proposées pour les réseaux bayésiens en remplaçant les distributions de probabilités par des distributions de possibilités n'est pas satisfaisante. En revanche, les réseaux possibilistes qualitatifs sont plus adaptés à la représentation de l'incertitude subjective sans exiger de valeurs numériques difficiles à fournir par des experts.

Une autre piste consiste à mettre en place des méthodes d'évaluation concernant l'apprentissage des réseaux possibilistes, en se basant sur les procédures existantes

dans le cadre des réseaux bayésiens comme l'échantillonnage des réseaux possibilistes et la proposition de mesures d'évaluation permettant de comparer la distribution jointe apprise et la distribution théorique.

De par le fait que les différents algorithmes évoqués dans cet article (propagation, apprentissage, évaluation) ont été traités séparément et avec divers niveaux de maturité, il n'existe aucune boîte à outils logicielle satisfaisante pour manipuler les réseaux possibilistes. L'implémentation d'un tel logiciel est indispensable pour aider à une meilleure utilisation de ces modèles.

Bibliographie

Adina C. (2006). Possibilistic networks for uncertainty knowledge processing in student diagnosis. *Annals of Dunarea de Jos*, p. 69–73.

Augustin T., Coolen F. P., Cooman G. de, Troffaes M. C. (2014). *Introduction to imprecise probabilities*. John Wiley & Sons.

Ayachi R., Ben Amor N., Benferhat S., Haenni R. (2010). Compiling possibilistic networks: Alternative approaches to possibilistic inference. In *Proceedings of the twenty-sixth conference annual conference on uncertainty in artificial intelligence (UAI-10)*, p. 40–47. Corvallis, Oregon, AUAI Press.

Ben Amor N., Benferhat S. (2005). Graphoid properties of qualitative possibilistic independence relations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 13, n° 01, p. 59–96.

Ben Amor N., Benferhat S., Mellouli K. (2003). Anytime propagation algorithm for min-based possibilistic graphs. *Soft Computing*, vol. 8, n° 2, p. 150–161.

Benferhat S., Smaoui S. (2007). Hybrid possibilistic networks. *International journal of approximate reasoning*, vol. 44, n° 3, p. 224–243.

Benferhat S., Tabia K. (2012). Inference in possibilistic network classifiers under uncertain observations. *Annals of Mathematics and Artificial Intelligence*, vol. 64, n° 2-3, p. 269–309.

Ben Slimen Y., Ayachi R., Ben Amor N. (2013). Probability-possibility transformation. In *Fuzzy logic and applications*, vol. 8256, p. 122-130. Springer.

Borgelt C., Gebhardt J. (1997). Learning possibilistic networks with a global evaluation method. In *Proc. 5th european congress on intelligent techniques and soft computing (eufit97, aachen, germany)*, vol. 2, p. 1034–1038.

Borgelt C., Kruse R. (2003). Operations and evaluation measures for learning possibilistic graphical models. *Artificial Intelligence*, vol. 148, n° 1, p. 385–418.

Borgelt C., Steinbrecher M., Kruse R. (2009). *Graphical models: representations for learning, reasoning and data mining* (vol. 704). Wiley.

Boughanem M., Brini A., Dubois D. (2009). Possibilistic networks for information retrieval. *International journal of approximate reasoning*, vol. 50, n° 7, p. 957–968.

Bouguelid M. S. (2007). Contribution à l'application de la reconnaissance des formes et la théorie des possibilités au diagnostic adaptatif et prédictif des systèmes dynamiques. Thèse de doctorat. *Université de Reims Champagne-Ardenne, France*.

Caglioni M., Moreno D., Dubois D., Prade H., Tettamanzi A. G., et al. (2014). Mise en oeuvre pratique de réseaux possibilistes pour modéliser la spécialisation sociale dans les espaces métropolisés. *23 ème rencontres francophones sur la logique floue et ses applications*, p. 267–274.

Chavira M., Darwiche A. (2007). Compiling bayesian networks using variable elimination. In *International joint conference on artificial intelligence*, p. 2443–2449.

Cooper G. F. (1990). The computational complexity of probabilistic inference using bayesian belief networks. *Artificial intelligence*, vol. 42, n° 2, p. 393–405.

Cooper G. F., Herskovits E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, vol. 9, n° 4, p. 309–347.

Cozman F. G. (2000). Credal networks. *Artificial intelligence*, vol. 120, n° 2, p. 199–233.

Darwiche A., Goldszmidt M. (1994). On the relation between kappa calculus and probabilistic reasoning. In *Proceedings of the tenth international conference on uncertainty in artificial intelligence*, p. 145–153.

De Campos L. M., Huete J. F. (1999a). Independence concepts in possibility theory: Part i. *Fuzzy sets and systems*, vol. 103, n° 1, p. 127–152.

De Campos L. M., Huete J. F. (1999b). Independence concepts in possibility theory: Part ii. *Fuzzy sets and systems*, vol. 103, n° 3, p. 487–505.

De Campos L. M., Huete J. F. (2001). Measurement of possibility distributions. *International Journal of General System*, vol. 30, n° 3, p. 309–346.

Dempster A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, p. 325–339.

Dubois D. (2006). Possibility theory and statistical reasoning. *Computational statistics & data analysis*, vol. 51, n° 1, p. 47–69.

Dubois D., Cerro L. F. del, Herzig A., Prade H. (1994). An ordinal view of independence with application to plausible reasoning. In *Proceedings of the tenth international conference on uncertainty in artificial intelligence*, p. 195–203.

Dubois D., Foulloy L., Mauris G., Prade H. (2004). Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable computing*, vol. 10, n° 4, p. 273–297.

Dubois D., Hüllermeier E. (2007). Comparing probability measures using possibility theory: A notion of relative peakedness. *International Journal of Approximate Reasoning*, vol. 45, n° 2, p. 364–385.

Dubois D., Prade H. (1998). Possibility theory: qualitative and quantitative aspects. In *Quantified representation of uncertainty and imprecision*, vol. 1, p. 169–226. Springer.

Dubois D., Prade H. (2000). Possibility theory in information fusion. In *Information fusion, 2000*, vol. 1, p. PS6–P19.

Dubois D., Prade H. (2006). Représentations formelles de l'incertain et de l'imprécis. In *Concepts et méthodes pour l'aide à la décision - outils de modélisation*, vol. 1, p. 99–137.

Dubois D., Prade H., Harding E. F. (1988). *Possibility theory: an approach to computerized processing of uncertainty* (vol. 2). Plenum press New York.

Dubois D., Prade H., Sandri S. (1993). On possibility/probability transformations. In *Fuzzy logic*, vol. 12, p. 103–112. Springer.

Dubois D., Prade H., Smets P. (2008). A definition of subjective possibility. *International journal of approximate reasoning*, vol. 48, n° 2, p. 352–364.

Fonck P. (1992). Propagating uncertainty in a directed acyclic graph. In *Proceedings of the fourth information processing and management of uncertainty conference*, vol. 92, p. 17–20.

Fonck P. (1994a). Conditional independence in possibility theory. In *Proceedings of the tenth international conference on uncertainty in artificial intelligence*, p. 221–226.

Fonck P. (1994b). Réseaux d'inférence pour le raisonnement possibiliste. Thèse de doctorat. *Université de Liege, Faculté des Science*.

Gebhardt J., Kruse R. (1996). On a tool for possibilistic reasoning in relational structures. In *Fuzzy systems*, vol. 2, p. 1471-1475.

Gebhardt J., Kruse R. (1997). Background and perspectives of possibilistic graphical models. In *Qualitative and quantitative practical reasoning*, vol. 1244, p. 108–121. Springer.

Giang P. H., Shenoy P. P. (1999). On transformations between probability and Spohnian disbelief functions. In *Proceedings of the fifteenth conference on uncertainty in artificial intelligence*, p. 236–244.

Guyonnet D., Bourgine B., Dubois D., Fargier H., Côme B., Chilès J.-P. (2003). Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, vol. 129, n° 1, p. 68–78.

Heckerman D., Geiger D., Chickering D. M. (1995). Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, vol. 20, n° 3, p. 197–243.

Henrion M., Provan G., Del Favero B., Sanders G. (1994). An experimental comparison of numerical and qualitative probabilistic reasoning. In *Proceedings of the tenth international conference on uncertainty in artificial intelligence*, p. 319–326.

Herskovits E., Cooper G. F. (1991). An entropy-driven system for construction of probabilistic expert systems from databases. In *Proceedings of the sixth annual conference on uncertainty in artificial intelligence*, p. 117–128. New York, NY, USA, Elsevier Science Inc.

Hou Y., Yang B. (2010). Probability-possibility transformation for small sample size data. In *Fuzzy systems and knowledge discovery*, vol. 4, p. 1720–1724.

Jensen F. V. (1996). *An introduction to Bayesian networks* (vol. 74). UCL press London.

Joslyn C. (1991). Towards an empirical semantics of possibility through maximum uncertainty. In *Fourth world congress of the international fuzzy systems association: Artificial intelligence*, p. 86-89.

Klir G. J., Mariano M. (1987). On the uniqueness of possibilistic measure of uncertainty and information. *Fuzzy sets and systems*, vol. 24, n° 2, p. 197–219.

Klir G. J., Parviz B. (1992). Probability-possibility transformations: a comparison. *International Journal of General System*, vol. 21, n° 3, p. 291–310.

Kruse R., Borgelt C. (1995). Learning probabilistic and possibilistic networks : Theory and applications. In *Eighth european conference on machine learning*, p. 3–16.

Lauritzen S. L. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, vol. 19, n° 2, p. 191–201.

Masson M.-H., Denœux T. (2006). Inferring a possibility distribution from empirical data. *Fuzzy sets and systems*, vol. 157, n° 3, p. 319–340.

Mouchaweh M., Bouguelid M., Billaudel P., Riera B. (2006). Variable probability-possibility transformation. *Twenty fifth European Annual Conference on Human Decision Making and Manual Control*, p. 417-428.

Pearl J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Pourret O., Naïm P., Marcot B. (2008). *Bayesian networks: a practical guide to applications* (vol. 73). John Wiley & Sons.

Sabbadin R. (2001). Towards possibilistic reinforcement learning algorithms. In *Fuzzy systems, 2001. the 10th ieee international conference on*, vol. 1, p. 404–407.

Saffiotti A., Umkehrer E. (1991). Pulcinella: A general tool for propagating uncertainty in valuation networks. In *Proceedings of the seventh conference on uncertainty in artificial intelligence*, p. 323–331.

Sangüesa R., Cabós J., Cortes U. (1998a). Possibilistic conditional independence: A similarity-based measure and its application to causal network learning. *International Journal of Approximate Reasoning*, vol. 18, n° 1, p. 145–167.

Sangüesa R., Cortés U., Gisolfi A. (1998b). A parallel algorithm for building possibilistic causal networks. *International Journal of Approximate Reasoning*, vol. 18, n° 3, p. 251–270.

Shafer G. (1976). *A mathematical theory of evidence* (vol. 1). Princeton university press Princeton.

Smets P., Kennes R. (1994). The transferable belief model. *Artificial intelligence*, vol. 66, n° 2, p. 191–234.

Yager R. R. (1982). Level sets for membership evaluation of fuzzy subsets. *Recent advances in fuzzy subset and possibility theory, Pergamon press*, p. 90–97.

Yager R. R. (1992). On the specificity of a possibility distribution. *Fuzzy Sets and Systems*, vol. 50, n° 3, p. 279–292.

Zadeh L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Journal of Information Science*, vol. 8, p. 199-249.

Zadeh L. A. (1999). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, vol. 100, p. 9–34.