
Approximation de l'inférence *MAP* via les transformations probabilistes-possibilistes

Salem Benferhat Amélie Levray Karim Tabia

CRIL, Univ. Artois - CNRS UMR 8188,
62300 Lens, France

{benferhat,levray,tabia}@cril.univ-artois.fr

Résumé

Dans ce papier, nous utilisons les modèles graphiques et les transformations probabilistes-possibilistes afin de donner une approximation de l'inférence *MAP* dans les réseaux crédaux. Dans un premier temps, nous présentons les transformations du cadre probabiliste imprécis vers le cadre possibiliste, nous étudions exactement deux transformations de la littérature que nous comparons expérimentalement. Ensuite, nous présentons trois critères de choix que nous utilisons dans l'expérience pour restreindre le domaine de réponse retourné. L'idée est de comparer notre approche, qui consiste à transformer un réseau crédal en un réseau possibiliste pour effectuer l'inférence *MAP*, avec les approches (exactes et approchées) existantes dans les réseaux crédaux. L'expérience conduite est la suite de précédents travaux mais de manière approfondie avec l'introduction des nouveaux critères de choix. La dernière partie du papier discute les résultats de cette expérimentation.

Abstract

This paper uses belief graphical models and probability-possibility transformations to give an approximation of *MAP* inference in credal networks. We first present transformations from credal to possibilistic network, we study exactly two different transformations from the literature to compare them during the experimentation. Then we present three criteria that we use in the process of our experiment. The idea is to compare our approach, which consists in transforming a credal network into a possibilistic one to perform *MAP* inference, with both exact and approximate *MAP* inference in credal networks. This experiment is the continuation of our previous work but in-depth and with the use of new criteria of choice. The last part of this paper discusses the results of the experiment.

1 Introduction

L'inférence *MAP* (Maximum A Posteriori) dans les réseaux Bayésiens est un problème de grand intérêt et est

étudié depuis des années [11, 12, 16, 17, 18]. De ce fait, il existe une variété de méthodes et d'algorithmes pour calculer et trouver la configuration ayant le degré de probabilité le plus élevé étant donné un ensemble de variables observées. Malgré tout, les réseaux Bayésiens, aussi connus qu'ils soient, peuvent parfois sembler inadaptés pour représenter des informations, telles que la connaissance d'un groupe d'experts, ou alors pour encoder des informations incomplètes ou mal connues. C'est pourquoi, nous avons besoin de cadre qui nous donne plus de flexibilité sur les paramètres du modèle. C'est dans cet esprit que les réseaux crédaux [4] ont été conçus. Ils généralisent les réseaux Bayésiens et offrent plus d'expressivité puisqu'ils représentent l'information à l'aide de probabilités imprécises. Le problème que l'on rencontre à vouloir raisonner avec un cadre aussi expressif est que cela implique une complexité de calcul plus élevée. Néanmoins, des méthodes et des algorithmes pour calculer des requêtes *MAP* existent et donnent des résultats plutôt concluants. Ces méthodes, cependant, ne sont pas très efficaces en termes de complexité de calcul et lorsqu'on fait face à des problèmes à taille réelle (avec beaucoup de variables). Le but de ce papier est de fournir une nouvelle méthode pour inférer dans les réseaux crédaux, plus précisément effectuer l'inférence *MAP*. Cette méthode est basée sur les transformations probabilistes-possibilistes. En cause : ces méthodes existent déjà et leurs propriétés ont déjà fait l'objet de travaux [2, 3, 6, 8, 10, 19, 20]. De plus, l'inférence *MAP* dans les réseaux possibilistes est aussi déjà connue, ce qui nous donne une bonne base pour ce travail. Ainsi, dans ce papier nous étudions deux transformations des probabilités imprécises vers les possibilités (qui ont été adaptées aux modèles graphiques dans [3]) et effectuons une vaste expérimentation sur l'inférence *MAP*. Ce travail suit la ligne directe de nos précédents travaux [3], où nous avons analysé plusieurs problèmes concernant la commutativité des

transformations (des probabilités imprécises vers les possibilités), tels la marginalisation et le conditionnement, deux notions indispensables quand on s'intéresse aux requêtes *MAP*. Ce papier propose aussi un début d'expérimentation pour observer l'efficacité de l'inférence *MAP*. Ce papier¹ est organisé comme suit :

- La première partie rappelle les notions de base des modèles graphiques utilisés dans notre expérimentation.
- La seconde partie présente le problème de l'inférence *MAP* et ses difficultés.
- La troisième partie introduit notre approche, les transformations, et les critères.
- Pour enfin, terminer sur une analyse des résultats.

2 Bref rappel sur les réseaux crédaux et les réseaux possibilistes

Dans cette section, nous présentons les principaux modèles graphiques utilisés, les réseaux crédaux et les réseaux possibilistes. Mais d'abord, on introduit les réseaux Bayésiens.

2.1 Réseaux Bayésiens

Les réseaux Bayésiens (*BN*) sont des modèles graphiques probabilistes [5] très connus, basés sur la théorie des probabilités, et spécifiés par deux composants :

- un *composant graphique* : un graphe acyclique dirigé (DAG) avec des nœuds représentant chaque variable A_i et les flèches encodant les relations de dépendance entre les variables.
- un *composant numérique* : où on associe à chaque variable A_i une table de probabilités locale $p(A_i|par(A_i))$ pour A_i dans le contexte de ses parents $par(A_i)$.

Cette représentation aux moyens de modèles graphiques nous permet d'encoder de manière compacte une distribution de probabilités sur un ensemble de variables. La distribution jointe encodée par le graphe *BN* est calculée grâce à la règle de chaînage suivante :

$$P(A_1, \dots, A_n) = \prod_{i=1}^n P(A_i|par(A_i)). \quad (1)$$

2.2 Réseaux crédaux

Les réseaux crédaux sont aussi des modèles graphiques probabilistes, et de la même manière que les réseaux Bayésiens sont basés sur la théorie des probabilités, les réseaux crédaux sont eux basés sur la théorie des probabilités imprécises [21, 13]. Une notion clé de cette théorie est celle

d'ensemble crédal qui est souvent interprété comme un ensemble de croyances imprécises, dans le sens où la mesure de probabilités réelle se trouve dans l'ensemble mais il n'existe aucun moyen de la déterminer exactement due au manque de connaissance. Dans le but de caractériser un ensemble crédal, on peut utiliser un ensemble de points extrêmes, des intervalles de probabilités ou des contraintes linéaires. Dans notre cas, nous utilisons des distributions de probabilités à intervalles qui sont un moyen très naturel et commun pour définir l'imprécision et les informations malconnues. Dans une distribution de probabilités à intervalles *IP*, chaque interprétation $\omega_i \in \Omega$ est associée à un intervalle de probabilités $IP(\omega_i)=[l_i, u_i]$ où l_i (resp. u_i) dénote la borne inférieure (resp. borne supérieure) de la probabilité de ω_i . Notons qu'alors qu'une distribution de probabilités standard p induit un ordre complet sur l'ensemble des mondes possibles Ω , une distribution de probabilités à intervalles *IP* peut induire un ordre partiel puisque certaines interprétations peuvent être incomparables en cas d'intervalles qui se chevauchent. Définissons maintenant les modèles graphiques probabilistes basés sur les ensembles crédaux, autrement dit les réseaux crédaux [4, 15].

Définition 1 (Réseaux crédaux) *Un réseau crédal* $CN = \langle G, K \rangle$ est un modèle graphique probabiliste où

- $G = \langle V, E \rangle$ est un graphe acyclique dirigé, comme l'est un *BN*, avec $V = \{A_1, \dots, A_n\}$ un ensemble de variables et E est un ensemble d'arcs.
- $K = \{IP_1, IP_2, \dots, IP_n\}$ est une collection de distributions de probabilités à intervalles, chaque IP_i est associée à une variable A_i dans le contexte de ses parents $par(A_i)$.

Notons qu'en pratique, dans les tables locales, on peut aussi spécifier un ensemble de points extrêmes à la place d'une distribution de probabilités à intervalles comme on le trouve dans le logiciel JavaBayes².

2.3 Réseaux possibilistes

Un réseau possibiliste $\mathcal{PN} = \langle G, \Theta \rangle$ est défini par :

- un *composant graphique* G qui est un graphe acyclique dirigé.
- un *composant numérique* Θ , un ensemble de tables de possibilités locales $\pi(A_i|par(A_i))$.

La sémantique associée à un réseau possibiliste est la théorie des possibilités où un degré représente l'incertitude et ce degré peut être interprété de plusieurs façons, par exemple, comme étant un degré de probabilités supérieures [9]. La théorie des possibilités est régie par plusieurs axiomes comme pour la théorie des probabilités. Ainsi, toutes les tables locales doivent satisfaire la condition de normalisation définie par : $\forall i = 1..n$, pour chaque contexte de parent $par(a_i)$, $\max_{a_i \in D_i} (\pi(a_i | par(a_i))) = 1$.

1. Ce papier est la version traduite d'un papier en cours de soumission.

2. <http://www.cs.cmu.edu/~javabayes/Home/>

Dans le cadre possibiliste, la distribution de possibilités jointe est factorisée en utilisant la contrepartie possibiliste de la règle de chaînage suivante :

$$\pi(a_1, a_2, \dots, a_n) = \otimes_{i=1}^n (\pi(a_i | \text{par}(a_i))) \quad (2)$$

où \otimes dénote l'opérateur du produit ou l'opérateur du min en fonction de l'interprétation que l'on fait de l'échelle possibiliste, soit quantitative (avec le produit) soit qualitative (avec le min). Dans ce travail, on ne s'intéresse qu'à la partie produit.

3 L'inférence MAP : aspects pratiques et calculatoires

Lorsqu'on s'intéresse au raisonnement dans les modèles graphiques, on est principalement intéressé par deux tâches d'inférence : trouver la configuration la plus probable (inférence MAP) et calculer des distributions marginales. Ces problèmes sont généralement calculatoirement très difficiles. Dans cette section, nous développons un peu plus la notion d'inférence dans les modèles graphiques.

3.1 Type de requêtes d'inférence

L'inférence dans les modèles graphiques probabilistes consiste, de manière générale, au calcul de la probabilité d'un événement. Dans les réseaux crédaux, cela revient à calculer la borne inférieure et la borne supérieure de la probabilité d'un événement. Soit $V = \{A_1..A_n\}$ l'ensemble des variables du modèle. Soit $O \subseteq V$ l'ensemble des variables observées et soit une instance $o \in O$. Soit $Q \subseteq V$ l'ensemble des variables de requêtes et une instance $q \in Q$. Il existe trois principaux types de requêtes :

- Calcul de la probabilité d'un événement d'intérêt q (Pr) étant donnée une évidence o . Le problème de décision associé est si $P(q|o) \geq s \in [0, 1]$.
- Calcul de l'explication la plus plausible (MPE). Étant donnée une observation o sur plusieurs variables, l'objectif est de calculer l'instance la plus probable q de toutes les variables restantes (non observées) Q . Notons qu'ici, $O \cup Q = \mathcal{A}$ et $Q \cap O = \emptyset$. Le problème de décision correspondant est si $P(q, o) \geq s \in [0, 1]$.
- Calcul du maximum a posteriori (MAP). Étant donnée une observation o sur plusieurs variables O , l'objectif est de trouver l'instance la plus probable q de l'ensemble de variables de requêtes Q . Dans les requêtes MAP, $Q \cap O = \emptyset$. Notons que lorsque Q et O couvre l'ensemble des variables V , le problème MAP se réduit à trouver l'explication la plus plausible (MPE).

3.2 Complexité calculatoire

La table suivante (Table 1) liste les résultats de complexité des requêtes décrites ci dessus dans les réseaux

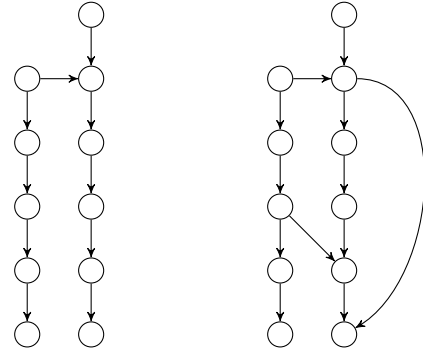


FIGURE 1 – A gauche : polytree, à droite : multiply-connected

Bayésiens et dans les réseaux crédaux [15]. On sépare ces réseaux en différentes catégories selon leur forme, ici nous nous intéressons à deux types de réseaux en particulier, les réseaux dits *polytree* et les réseaux dits *multiply-connected*. Un réseau *polytree* est une structure qui admet un seul chemin (indirect) entre deux nœuds du réseau et aussi dans lequel un nœud peut avoir plusieurs parents (dans le cas où un nœud n'admet qu'un parent au plus, on travaille alors avec des structures dites *tree*). Alors qu'une structure *multiply-connected* est une structure dans laquelle au moins une paire de nœuds admet plusieurs chemins (le graphe indirect associé comporte un cycle). Reportez-vous à la Figure 1 pour plus de détails.

En pratique, la taille des réseaux est souvent grande. Ce qui motive l'apparition de méthodes d'inférence approchée et dans ce papier, on vous propose une nouvelle méthode pour l'inférence MAP dans les réseaux crédaux. Et ce en transformant le réseau crédal CN en réseau possibiliste PN. Notons que la complexité de l'inférence dans les réseaux possibilistes est similaire à la complexité dans les réseaux Bayésiens.

Exemple 1 Voyons un exemple de requête MAP et un exemple de requête MPE dans un réseau possibiliste. En considérant le réseau possibiliste, sur l'ensemble des variables $V = \{A, B, C, D\}$, représenté par la Figure 2. Dans cet exemple, nous souhaitons calculer la requête MAP suivante : on cherche la configuration de D la plus plausible sachant que $A = F$. Pour répondre à ce genre de questions, il existe de nombreux algorithmes tels que l'élimination de variables ou l'arbre de jonction. Par ce biais, le calcul nous donne : $\pi(D = T|A = F) = 1$ et $\pi(D = F|A = F) = .6$, ainsi le résultat du maximum à posteriori de D sachant A est $D = T$. En comparaison, calculer la requête MPE pour l'observation $A = F$ revient à calculer le degré de possibilité le plus élevé pour l'ensemble des variables non-observées B, C, D . Alors, on calcule pour chaque configuration de A, B, C, D où $A = F$ le degré de possibilité, et la réponse à la requête, est l'instance avec le degré le plus

	Requêtes	Polytree	Bounded treewidth	Multiply-connected
Réseaux Bayésien	<i>Pr</i>	Polynomial	Polynomial	PP-Complet
	<i>MPE</i>	Polynomial	Polynomial	NP-Complet
	<i>MAP</i>	NP-Complet	NP-Complet	NP ^{PP} -Complet
Réseaux crédaux	<i>Pr</i>	NP-Complet	NP-Complet	NP ^{PP} -Complet
	<i>MPE</i>	Polynomial	Polynomial	NP-Complet
	<i>MAP</i>	Σ_2^P -Complet	Σ_2^P -Complet	NP ^{PP} -Difficile

TABLE 1 – Résumé de la complexité des requêtes dans les réseaux Bayésiens et les réseaux crédaux

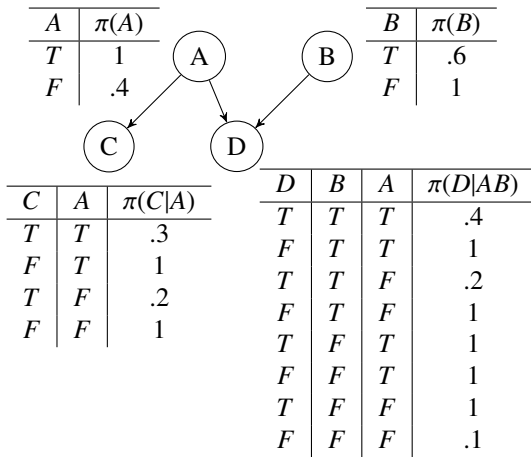


FIGURE 2 – Exemple d'un réseau possibiliste avec 4 variables booléennes

élevé, ici, $\pi(A = F, B = F, C = F, D = T) = .4$.

4 Une approche basée sur les transformations probabiliste-possibiliste

Plusieurs transformations probabiliste-possibiliste existent comme on le rappelle dans la sous-section 4.1. Des méthodes les généralisant aux probabilités imprécises ont été proposées par Masson et Denoeux [14] et plus encore dans [6]. Dans cette section, nous présentons deux de ces transformations.

4.1 Transformation probabilités \rightarrow possibilités

La théorie des probabilités et la théorie des possibilités sont deux cadres qui ont été étudiés de manière approfondie et plusieurs passerelles ont été construites entre les deux [10, 22]. Nous avons maintenant des transformations qui permettent de passer des probabilités vers les possibilités et inversement. Dubois et Prade ont, par exemple, proposé la transformation appelée **Optimal transformation (OT)**

[8] qui est définie par :

$$\pi(\omega_i) = \sum_{j/p(\omega_j) \leq p(\omega_i)} p(\omega_j) \quad (3)$$

Les transformations, de manière générale, sont tenues de satisfaire des principes de base afin de préserver le plus possible l'information. Dans ce sens, la transformation **OT** est considérée comme celle qui préserve le mieux l'information tout en satisfaisant ces principes³. Il existe d'autres travaux sur les transformations que l'on peut trouver dans [2, 3, 6, 8, 10, 19, 20].

4.2 Des probabilités imprécises vers les possibilités

La première transformation que nous étudions est celle de Masson et Denoeux dans [14], où les auteurs apprennent une distribution de possibilités à partir de données empiriques en transformant les intervalles de confiance en distributions de possibilités. Le premier point consiste à considérer une distribution de probabilités à intervalles comme moyen d'encoder l'ordre partiel \mathcal{M} sur Ω . Ainsi, soit \mathcal{M} représentant l'ordre partiel encodé par la distribution de probabilités à intervalles IP et soit \mathcal{C} l'ensemble des extensions linéaires (ordre total) qui sont compatibles avec l'ordre partiel \mathcal{M} , la transformation MD se déroule comme suit. Pour chaque extension linéaire $C_l \in \mathcal{C}$ et pour chaque interprétation $\omega_i \in \Omega$, on cherche la distribution de probabilités compatible qui donne le degré de possibilités le plus élevé, quand la transformation **OT** est utilisée pour obtenir la distribution de possibilités⁴.

$$\pi^{C_l}(\omega_i) = \max_{p^1 \dots p^n} \sum_{p_j \leq p_i} p_j \quad (4)$$

Ensuite, la distribution de possibilités finale est construite en prenant la valeur maximale, pour chaque interprétation, des distributions de possibilités π^{C_l} obtenues pour chaque extension linéaire. Cette transformation tente d'une part de préserver l'ordre des interprétations induit par IP et d'autre

3. Parmi ces principes se trouvent : la condition de consistance, la préservation de la préférence et le principe de maximum de spécificité.

4. En effet, la transformation MD se réduit à **OT** lorsqu'on considère des singletons à la place des intervalles de probabilités.

part de préserver le principe de dominance qui requiert que $\forall \phi \subseteq \Omega, P(\phi) \leq \Pi(\phi)$ (La probabilité est bornée par le degré de possibilités).

Il y a deux principaux inconvénients à cette transformation, le premier concerne la complexité de calcul. A savoir, le nombre d'extensions linéaires que l'on doit considérer est dans le pire des cas $N!$ (où N est le nombre de mondes possibles). En second, la distribution obtenue n'est pas nécessairement optimale, et pas, non plus, la plus spécifique dominant la distribution de probabilités à intervalles considérée. Du coup, nous avons besoin de considérer une autre approche pour transformer une distribution imprécise en une distribution de possibilités. Cette transformation se rapporte à la notion de distribution cumulative inférieures et supérieures. Dans notre précédent papier [3], nous avons déjà parlé de la transformation MD , nous avons aussi évoqué la distribution cumulative supérieure \bar{F} construite à partir d'une seule extension linéaire $C_l \in C$. Cette fois-ci, dans le but d'être le plus complet possible, l'on ne considère plus une seule distribution de possibilités mais deux, en ajoutant la distribution cumulative inférieure. Étant donné un ensemble d'intervalles de probabilités et une relation d'ordre \leq_{C_l} construite à partir d'une extension linéaire C_l entre les éléments ω_i , on peut facilement construire une p-box $[F, \bar{F}]$. Étant donné les ensembles consécutifs $A_i = \{\omega_i, \forall \omega_i \in \Omega \text{ et t.q. } \omega_i \leq_{C_l} \omega_j \text{ ssi } i < j\}$, les distributions cumulatives inférieure et supérieure correspondants à Ω sont, respectivement :

$$\underline{F}(\omega_i) = \underline{P}(A_i) = \max\left(\sum_{\omega_j \in A_i} l_j, 1 - \sum_{\omega_j \notin A_i} u_j\right) \quad (5)$$

$$\bar{F}(\omega_i) = \bar{P}(A_i) = \min\left(\sum_{\omega_j \in A_i} u_j, 1 - \sum_{\omega_j \notin A_i} l_j\right) \quad (6)$$

Ces deux distributions obtenues sont deux distributions de possibilités qui encadrent (dans le sens dominant) les bornes de la distribution de probabilités imprécises. En dehors des considérations algorithmiques, l'utilisation des transformations par le biais de p-box [7] est intéressante malgré le fait qu'il n'existe pas de relation spécifique entre les deux (dans le sens où aucune ne généralise l'autre), puisqu'il existe des résultats entre les deux qui ont été démontrés [1]. Maintenant, regardons comment appliquer ces transformations dans le cadre des modèles graphiques.

4.3 Réseau crédal \rightarrow réseau possibiliste

Une méthode simple et directe pour transformer un réseau crédal en réseau possibiliste et qui nous donne l'avantage de préserver les relations d'indépendance entre les variables, est défini ainsi :

Définition 2 (Transformation graphique crédal-possibiliste) Soit CN un réseau crédal, \mathcal{PN}_{CN} est un réseau possibiliste obtenu depuis CN et défini par :

- une composante graphique G qui est le même graphe que pour le réseau crédal ainsi \mathcal{PN}_{CN} encode les mêmes relations d'indépendance que CN .
- un ensemble de tables de possibilités locales π_i qui sont le résultat de la transformation des ensembles crédaux locaux IP_i par TR , une transformation de probabilités imprécises vers les possibilités.

Exemple 2 Soit CN un réseau crédal représenté par la Figure 3 sur les deux variables binaires A et B . En utilisant la transformation MD , le réseau crédal CN de la Figure 3 sera transformé en un réseau possibiliste \mathcal{PN} que l'on peut voir sur la Figure 4.

Concernant les principes respectés par ces transformations, nous avons étudiés plusieurs questions dans [3]. En effet, nous avons répondu à ces deux questions : i) Est-ce que la distribution $\pi_{\mathcal{PN}}$ domine IP_{CN} (la distribution jointe encodé par CN) ? et ii) Est-ce que l'ordre partiel des interprétations induit par IP_{CN} est préservé par la transformation TR ?

En réponse à la première question, pour les mondes élémentaires $\omega_i \in \Omega$, nous garantissons que la distribution de possibilités calculée domine bien les degrés de probabilités correspondant, dans le cas où le réseau crédal est un réseau Bayésien (à savoir, tous les intervalles sont des singletons). Concernant les événements arbitraires $\phi \subseteq \Omega$, la question reste ouverte. Si nous utilisons la transformation OT , la distribution de possibilités jointe obtenue ne garantit pas de dominer la distribution jointe de probabilités. En réponse à la seconde question, il n'existe aucune garantie que l'ordre des interprétations encodé par la distribution jointe sera le même après la transformation. Pour plus de détails, reportez-vous à [3]. Ces problèmes sont la raison pour laquelle nous ne pouvons proposer de méthode exacte d'inférence dans les réseaux crédaux. Mais notons que c'est l'absence de garantie sur la préservation de l'ordre qui rend notre méthode approchée.

5 Étude expérimentale

Dans cette section, nous donnons les résultats d'une expérience approfondie où l'on utilise de nouveaux critères pour le choix des résultats des requêtes MAP dans les réseaux crédaux.

5.1 Configuration de l'expérience et critères d'évaluation

Dans la précédente expérimentation effectuée dans [3], nous avons utilisé pour critère de choisir toutes interprétations qui n'étaient pas interval-dominées (voir Définition 3). Dans cette expérience, nous nous sommes rendus compte que les résultats n'étaient pas très convaincants.

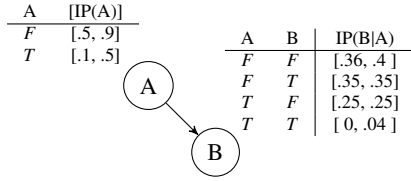


FIGURE 3 – Exemple d'un réseau crédal CN .

Une raison qui l'explique est le nombre d'interprétations qui est retourné et qui se trouve être trop grand, ce qui est dû en partie à la taille des réseaux. Dans cette nouvelle expérimentation que nous faisons, nous réduisons le nombre de résultats retournés en utilisant trois nouveaux critères. Rappelons, d'abord, la définition d'*Interval-dominance* qui se réfère aux instances non-dominées des variables de requêtes.

Définition 3 (Interval-dominance) Une instance q_i des variables de requêtes Q domine une autre instance q_j si et seulement si $\underline{IP}(q_i|o) > \overline{IP}(q_j|o)$ où o est une instance de variables observées O .

Nous considérons maintenant trois nouveaux critères pour réduire le nombre de résultats. Il existe plusieurs raisons pour lesquelles on veut réduire ce nombre. Dans un premier cas, dans certaines applications où l'on utilise les requêtes *MAP*, on s'attend à une réponse limitée. Si le résultat contient toutes les interprétations disponibles, soit l'on se trouve dans un cas de totale ignorance, soit l'algorithme est juste mauvais. C'est pour cela qu'on introduit ces trois critères très connus *Maximax*, *Maximin*, et *Hurwicz*. Ces trois critères sont principalement utilisés dans la prise de décision sous incertitude depuis le début des années 1950. Ils sont, ci après, détaillés.

Le critère *Maximax* peut-être vu comme une approche optimiste. Il examine les gains maximaux des alternatives et choisit l'alternative qui a le meilleur gain. La définition 4, donne une définition formelle du critère *Maximax* dans le cadre des probabilités imprécises.

Définition 4 (Critère Maximax) Une instance q_i des variables de requêtes Q est un résultat de l'inférence *MAP* si et seulement si $\overline{IP}(q_i|o) \geq \max\{1 - \sum_{q_j \neq q_k} \underline{IP}(q_j|o), \forall q_k\}$, où o est un instance des variables observées O .

D'un autre côté, le critère *Maximin*, aussi connu sous le nom de *Wald's Maximin criterion*, est une approche pessimiste. Il suggère que le décideur examine seulement les gains minimums de chaque alternative et choisit l'alternative qui a le gain le moins mauvais. Dans notre cas, la définition de ce critère appliqué au contexte des probabilités imprécises est donnée par la suivante :

Définition 5 (Critère Maximin) Une instance q_i des variables de requêtes Q est un résultat de l'inférence *MAP* si et seulement si $\underline{IP}(q_i|o) \geq \max\{1 - \sum_{q_j \neq q_k} \overline{IP}(q_j|o), \forall q_k\}$, où o est un instance de variables observées O .

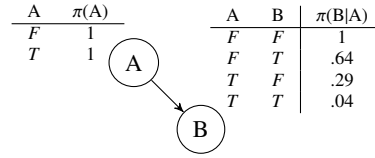


FIGURE 4 – Réseau possibiliste \mathcal{PN}_{CN} obtenu par transformation de CN (Fig. 3).

Réseaux	Topologie	#Nœuds	$ D_{A_i} $
<i>Alarm</i>	Multiply-connected	37	4
<i>Insurance</i>	Multiply-connected	27	5
<i>Poly</i>	Polytree	10	4
<i>Multi</i>	Multiply-connected	6	4

TABLE 2 – Description des réseaux crédaux utilisés.

Le dernier critère que nous passons en revue est le célèbre critère d'*Hurwicz*, suggéré par Leonid Hurwicz en 1951, qui sélectionne le gain minimum et le gain maximum pour chaque alternative donnée. Le critère d'*Hurwicz* tente de trouver un terrain d'entente entre les deux extrêmes, que représentent le critère optimiste et le pessimiste, en assignant une certaine proportion, α au critère optimiste et son contre poids $1 - \alpha$ au critère pessimiste. Cet index reflète l'attitude du décideur face au risque. Un décideur prudent assignera $\alpha = 1$ ce qui ramène le critère *Hurwicz* au critère *Maximin*. Un décideur aventureux mettra $\alpha = 0$ ce qui réduit le critère *Hurwicz* au critère *Maximax*.

La définition 6 définit formellement le critère d'*Hurwicz* avec les probabilités à intervalles en utilisant le coefficient α .

Définition 6 (Critère Hurwicz) Soit une instance q_i des variables de requêtes Q pour o une instance de variables observées de O , $a = \{\alpha * (1 - \sum_{q_i \neq q_k} \overline{IP}(q_i|o)) + (1 - \alpha) * (1 - \sum_{q_i \neq q_k} \underline{IP}(q_i|o)), \forall q_k\}$. Alors q_i est un résultat de l'inférence *MAP* si et seulement si $a = \max_{q_j} \{\alpha * (1 - \sum_{q_j \neq q_k} \overline{IP}(q_j|o)) + (1 - \alpha) * (1 - \sum_{q_j \neq q_k} \underline{IP}(q_j|o)), \forall q_k\}$.

Dans le cadre de notre expérimentation, nous avons choisi de poser $\alpha = 0.5$, ne donnant ni l'avantage à l'approche pessimiste ou optimiste.

Avant de présenter les résultats, nous donnons la description des benchmark utilisés dans l'expérimentation dans la table 2.

Présentons maintenant avec de plus amples détails l'étude expérimentale effectuée. En gardant en tête ce qui a été fait dans la précédente étude, nous poursuivons notre analyse plus loin, en incluant trois critères afin d'élaguer le nombre de résultats retourné par le réseau crédal. Nous utilisons aussi deux distributions de possibilités comme résultats à la transformation *CD*. En vérité, c'est une amélioration de la transformation *CD* utilisée dans [3], que l'on pouvait interpréter comme la distribution de possibilités cumulative supérieure, en ajoutant sa contre-partie, la dis-

tribution de possibilités cumulative inférieure. Ainsi, dans cet étude nous fournissons :

- Les taux de précision comparant l’algorithme exact développé dans JavaBayes avec :
 - l’algorithme d’inférence approchée utilisé dans le programme *GL2U*.
 - la précision fournie par l’inférence *MAP* dans les réseaux possibilistes obtenus par la transformations *MD*.
 - la précision fournie dans les réseaux possibilistes obtenues en utilisant les distributions de possibilités cumulatives inférieure et supérieure.
- Les taux d’inclusion de chaque méthode : on calcule le nombre d’instances de nos méthodes incluses dans les instances retournées par la méthode exacte.
- Les taux de précision et d’inclusion de chaque méthode en utilisant les trois critères sur les résultats de la méthode exacte (JavaBayes).
- Les taux de résultats : on compare le nombre de résultats retourné par l’algorithme avec le nombre de résultats possibles. L’idée est de mettre en lumière le niveau d’incertitude ou confusion retournée pour chaque algorithme.
- Tous ces résultats en variant le nombre de variables de requêtes de 1 à 5. Dans chacun des cas, on effectue l’expérience sur 200 fichiers.

5.2 Résultats

Les résultats sont donnés dans le graphique suivant (Figure 5). Afin de rester concis, nous n’y montrons que la comparaison entre *MD*, *CD* et *GL2U* sur un seul type de réseaux : les polyarbres et la comparaison des critères *Interval-dominance*, *Maximin* et *Hurwicz*. On a choisi d’oublier le critère *Maximin* puisque les résultats sont similaires avec ceux des critères *Maximax* et *Hurwicz*. Une information que l’on présente sur ce graphe, plus précisément sur le premier diagramme, est le nombre de fichiers auquel l’algorithme *GL2U* a pu répondre. En effet, nous déroulons tous les algorithmes sur approximativement 200 fichiers pour chaque nombre de variables de requêtes, mais dans certains cas, l’algorithme *GL2U* ne peut pas fournir de réponse. Du coup, le point sur le diagramme représente le taux de fichiers répondus. Par exemple, pour une seule variable de requête, la méthode *GL2U* n’a répondu qu’à 70% des 200 fichiers.

Dans un premier temps, notons que pour les autres types de fichiers, nous observons les mêmes résultats avec plus ou moins d’écart dans les taux de précision. Les résultats diffèrent de manière significative en fonction du nombre de variables de requêtes et du domaines de ces variables. En effet, si nous considérons trois variables de requêtes avec un domaine de cinq valeurs pour chaque variable, le nombre de configuration à prendre en compte peut aller jus-

qu’à 125, ce qui veut dire que lorsqu’on transforme des probabilités à intervalles vers les possibilités, nous devons considérer au plus 125 extensions linéaires. Ce qui résulte en une distribution de possibilités qui est bien plus imprécise (où il y a partout des 1). Et du coup, on observe moins de précision. Cependant, nous pouvons argumenter qu’en terme des réponses incluses, nous avons de bons résultats. De plus, dans la table 3 qui suit, nous montrons les résultats sur le nombre d’interprétations retourné pour l’inférence *MAP*. En fait, pour les critères *Interval-dominance* et *Maximax*, on fournit le taux d’interprétations répondues sur le nombre d’interprétations possible. Et ce, pour les algorithmes exact (de JavaBayes), et la même chose pour *MD*, *CD*, *GL2U*. Nous fournissons aussi, le ratio de réponses incluses, le nombre de réponses (des méthodes approchées) qui sont incluses dans l’ensemble des réponses retournées par l’algorithme exacte.

Dans la table 3, nous pouvons réellement voir trois type de résultats :

- i) Dans le cas du critère *Interval-dominance*, le nombre d’instances retournées en résultat de l’inférence *MAP* est élevé, ce qui révèle trop d’ignorance. Au contraire, lorsqu’on utilise le critère *Maximax*, on observe, avec un taux de 0.36, plus de précision et notons que dans ce cas, *CD* donne un nombre d’instances assez similaire.
- ii) Lorsqu’on compare le taux de réponses retournées pour la transformation *MD*, nous remarquons qu’en fait cette transformation nous donne le réseau possibiliste le plus précis en respect avec le réseau crédal. De plus, il montre de meilleurs résultats en terme de réponses incluses comparé aux deux autres approches.
- iii) La dernière information qui ressort de ces résultats repose sur l’attitude de l’algorithme *GL2U*, en effet il retourne, de manière générale, un plus grand nombre d’interprétations, même comparé à l’approche exacte. Et lorsqu’on considère plusieurs variables de requêtes, le nombre de réponses retournées grimpe rapidement pour atteindre tout l’ensemble des mondes possibles.

En se concentrant sur les résultats donnés par le graphique, nous observons, que même si l’algorithme approchée *GL2U* est plus précis, il ne tient pas la route lorsque le nombre de variables de requêtes atteint ou dépasse 3, ce qui est une mauvaise nouvelle sachant que dans ce genre d’applications, on s’attend à pouvoir travailler avec un grand nombre de variables. Cette information est représentée dans le graphique par les points (·) qui évalue le taux de fichiers où *GL2U* a pu répondre, et on peut s’apercevoir qu’au delà de trois variables de requêtes, le nombre de fichiers dégringole jusqu’à atteindre 0. Ainsi, la méthode que l’on propose, par transformation avec *MD* est une alternative plutôt intéressante à *GL2U* puisqu’elle nous donne les

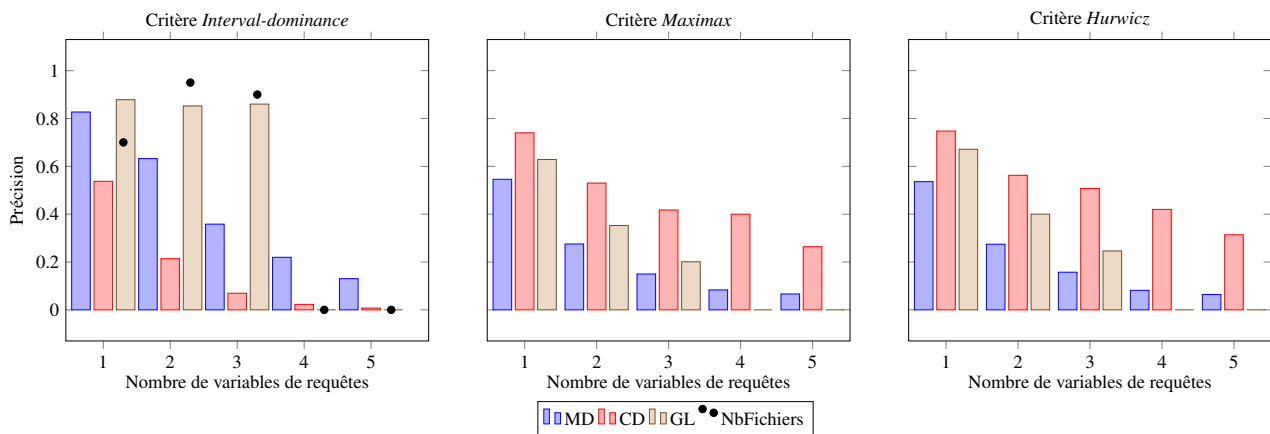


FIGURE 5 – Résultats sur les réseaux de type polyarbres

Critères	<i>MD</i>	<i>CD</i>	<i>GL2U</i>	
.794	.685	.36	.88	% $ rponses / D_Q $
<i>Interval-dominance</i>	.967	1	.891	% Inclusion
.36	.685	.36	.88	% $ rponses / D_Q $
<i>Maximax</i>	.546	.74	.629	% Inclusion

TABLE 3 – Ratio de réponses retournées sur l'ensemble possible des mondes disponibles VS Ratio de réponse incluses

second meilleurs résultats dans le cas du critère *Interval-dominance*. En ce qui concerne l'algorithme *CD*, nous reconnaissons la moindre efficacité qu'il présente en utilisant le critère *Interval-dominance*, et ce résultat s'explique par le nombre d'interprétations retournées par JavaBayes qui est trop important. Au contraire, nous observons l'effet inverse avec les trois autres critères, où *CD* nous donne les meilleurs résultats autant en termes de précision que d'inclusion des réponses.

Globalement, les trois algorithmes que l'on analyse ici, tendent à montrer le même comportement envers le nombre de variables de requêtes, ce qui insinue que les taux de précision baissent tous quand le nombre de variables augmente. Pourtant, on peut tout de même conclure que *GL2U* n'est définitivement pas à considérer puisque il atteint un timeout trop vite. Et que *MD* et *CD* nous offrent la possibilité de travailler l'inférence *MAP* sur plus de réseaux mais plus encore de travailler avec des réseaux plus grands et avec plus de variables de requêtes. Nous pouvons, en dernier lieu, inférer que l'utilisation du critère *Interval-dominance* n'est pas une bonne idée puisqu'il retourne un nombre d'interprétations trop important qui ne nous offrent ainsi pas les informations suffisantes pour prendre des décisions. En effet, avec un aussi grand nombre de réponses retournées, on ne peut rien faire dans des cas de classifications ou de recommandations. Enfin, en général, une approche optimiste (en utilisant le critère *Maximax*) est la meilleure variante des trois proposées.

6 Remarques finales

Dans ce travail, nous avons effectué une étude expérimentale sur les réseaux crédaux. Le but était d'améliorer le processus d'inférence *MAP*. La méthode que nous proposons consiste à transformer le réseau crédal en réseau possibiliste. Les avantages revendiqués sont l'amélioration de la complexité de calcul, l'utilisation de critères de choix qui nous permettent de restreindre le nombre de résultat retournés. Nous avons montré que, premièrement, utiliser l'algorithme approché (*GL2U*) sur les réseaux crédaux n'était pas calculatoirement intéressant étant donné la limite qu'il montre lors de plusieurs variables de requêtes. Ensuite, l'utilisation du critère *Maximax* améliore la qualité de l'algorithme *CD* de manière significative, et nous permet d'interroger de plus grands réseaux et en nombre plus important. En conclusion, l'algorithme *CD* représente une bonne alternative pour approximer l'inférence *MAP* dans les réseaux crédaux. Cette expérience, toutefois, n'est conduite que dans le cadre possibiliste quantitatif, nous allons par la suite expérimenter la même chose dans le cadre possibiliste qualitatif, et donner une comparaison des deux cadres (qualitatif vs. quantitatif).

Références

- [1] Baudrit, C et D Dubois: *Practical representations of incomplete probabilistic knowledge*. Compu-

- tational Statistics & Data Analysis, 51(1) :86–108, 2006. <https://doi.org/10.1016/j.csda.2006.02.009>.
- [2] Benferhat, S., A. Levray et K. Tabia: *On the Analysis of Probability-Possibility Transformations : Changing Operations and Graphical Models*. Dans *ECS-QARU 2015, Compiègne, France, July 15-17, 2015*.
- [3] Benferhat, S., A. Levray et K. Tabia: *Probability-Possibility Transformations : Application to Credal Networks*. Dans *SUM 2015, Québec City, QC, Canada, Sept 16-18, 2015. Proc.*, pages 203–219, 2015. http://dx.doi.org/10.1007/978-3-319-23540-0_14.
- [4] Cozman, F. G.: *Credal networks*. Artificial Intelligence, 120(2) :199 – 233, 2000. <http://www.sciencedirect.com/science/article/pii/S0004370200000291>.
- [5] Darwiche, A.: *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009, ISBN 978-0-521-88438-9.
- [6] Destercke, S., D. Dubois et E. Chojnacki: *Transforming Probability Intervals into Other Uncertainty Models*. Dans *EUSFLAT 2007 proc.*, tome 2, pages 367–373, Ostrava, Czech Republic, 2007.
- [7] Destercke, S, D Dubois et E Chojnacki: *Unifying practical uncertainty representations - I : Generalized p-boxes*. Int. J. Approx. Reasoning, 49(3) :649–663, 2008. <https://doi.org/10.1016/j.ijar.2008.07.003>.
- [8] Dubois, D., L. Foulloy, G. Mauris et H. Prade: *Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities*. Reliable Computing, pages 273–297, 2004.
- [9] Dubois, D et H Prade: *When upper probabilities are possibility measures*. Fuzzy Sets and Systems, 49(1) :65 – 74, 1992, ISSN 0165-0114. <http://www.sciencedirect.com/science/article/pii/016501149290110P>.
- [10] Klir, G. J. et J. F. Geer: *Information-Preserving Probability- Possibility Transformations : Recent Developments*. Dans *Fuzzy Logic*, pages 417–428. 1993.
- [11] Kwisthout, J.: *Treewidth and the Computational Complexity of MAP Approximations*. Dans *7th European Workshop, PGM 2014, Utrecht, The Netherlands, Sept. 17-19, 2014. Proc.*, pages 271–285, 2014. http://dx.doi.org/10.1007/978-3-319-11433-0_18.
- [12] Kwisthout, J., H.L. Bodlaender et L.C. van der Gaag: *The Necessity of Bounded Treewidth for Efficient Inference in Bayesian Networks*. Dans *ECAI 2010 - 19th, Lisbon, Portugal, August 16-20, 2010, Proc.*, pages 237–242, 2010. <http://www.booksonline.iospress.nl/Content/View.aspx?piid=17749>.
- [13] Levi, I.: *The enterprise of knowledge : an essay on knowledge, credal probability, and chance / Isaac Levi*. MIT Press Cambridge, Mass, 1980, ISBN 0262120828.
- [14] Masson, M. H. et T. Denoeux: *Inferring a possibility distribution from empirical data*. Fuzzy Sets and Systems, pages 319–340, 2006.
- [15] Mauá, D., C. P. de Campos, A. Benavoli et A. Antonucci: *Probabilistic Inference in Credal Networks : New Complexity Results*. J. Artif. Intell. Res. (JAIR), 50 :603–637, 2014.
- [16] Mauá, D.D, C.P. de Campos et F.G. Cozman: *The Complexity of MAP Inference in Bayesian Networks Specified Through Logical Languages*. Dans *Proc. of the 24th IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 889–895, 2015. <http://ijcai.org/Abstract/15/130>.
- [17] Pearl, J.: *Probabilistic reasoning in intelligent systems - networks of plausible inference*. Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann, 1989.
- [18] Shimony, S.E. et C. Domshlak: *Complexity of probabilistic reasoning in directed-path singly-connected Bayes networks*. Artif. Intell., 151(1-2) :213–225, 2003. [http://dx.doi.org/10.1016/S0004-3702\(03\)00110-3](http://dx.doi.org/10.1016/S0004-3702(03)00110-3).
- [19] Slimen, Y. Ben, R. Ayachi et N. Ben Amor: *Probability-Possibility Transformation : - Application to Bayesian and Possibilistic Networks*. Dans *WILF13 2013, Genoa, Italy, Nov. 19-22, 2013.*, pages 122–130.
- [20] Sudkamp, T.: *On probability-possibility transformations*. Fuzzy Sets and Systems, pages 73–81, 1992.
- [21] Walley, P.: *Towards a unified theory of imprecise probability*. International Journal of Approximate Reasoning, 24(2–3) :125 – 148, 2000.
- [22] Zadeh, L. A.: *Fuzzy sets as a basis for a theory of possibility*. Fuzzy Sets and Systems, 100 :9–34, 1999.