

Organiser la fusion d'informations par l'analyse formelle de concepts

Zainab Assaghir*, Mehdi Kaytoue*, Amedeo Napoli*, Henri Prade**

*LORIA, Campus Scientifique, B.P. 239 – 54500 Vandœuvre-lès-Nancy – France
{assaghiz,kaytouem,napoli}@loria.fr

**IRIT, 118 Route de Narbonne – 31062 Toulouse – France, prade@irit.fr

Résumé. Le problème étudié dans cet article est la fusion d'informations provenant de différentes sources : il s'agit de constituer une nouvelle information en fusionnant celles qui sont délivrées par les sources. Le problème de la fusion devient délicat surtout quand les sources fournissent des informations contradictoires. Parfois, les résultats de la fusion globale, appliquée à l'ensemble de toutes les sources, ne peuvent pas être directement utilisés pour une décision. Dans ce papier, nous utiliserons l'analyse formelle de concepts pour organiser les résultats des méthodes de fusion d'informations numériques. Cette approche permet d'associer un sous-ensemble de sources avec son résultat de fusion. Quand le résultat global de la fusion est imprécis, la méthode permet à l'utilisateur d'identifier un sous-ensemble maximal de sources qui a un résultat de fusion plus précis et utile. Enfin, une expérience en agronomie nous sert de première validation pour l'aide à la décision de pratiques agricoles.

1 Introduction

Le principal problème étudié dans cet article est celui de la fusion des informations numériques provenant de plusieurs sources et représentées dans le cadre de la théorie des possibilités. Nous utilisons l'Analyse Formelle de Concepts (AFC) pour organiser les résultats de fusion.

La fusion d'informations consiste à combiner des informations hétérogènes issues de plusieurs sources en une information globale permettant de mieux décider et agir (Bloch et al., 2001). Une source d'informations peut être un expert, un capteur, une base de données. L'information délivrée peut être numérique (nombre, intervalle de nombres) ou symbolique (terme linguistique...) et imparfaite (incertaine et/ou imprécise). Le problème de fusion des informations imparfaites issues de plusieurs sources est rencontré dans divers champs d'application, par exemple, en robotique (Abidi et Gonzalez, 1992), dans les systèmes d'informations (Cholvy, 1998), ou pour synthétiser les opinions des groupes d'experts (Cooke, 2001).

Plusieurs cadres ont été proposés pour modéliser et combiner les données imparfaites (Dubois et Prade, 1994; Dubois et al., 1992; Cholvy, 1998; Gebhardt et Kruse, 1998; Cooke, 2001; Cholvy et Moral, 2001; Chaudron et al., 2003; Phan-Luong, 2008; Destercke et al., 2009). Mais il n'existe pas de méthode unique même si le cadre de représentation de l'imperfection est choisi (Dubois et Prade, 1994).

Dubois et Prade soulignent, dans (Dubois et Prade, 1994), l'intérêt de la théorie des sous-ensembles flous pour le problème de la fusion et ont proposé plusieurs opérateurs de fusion d'informations où l'information est représentée par une distribution de possibilité. Plus récemment, un opérateur de fusion possibiliste fondé sur la notion de Sous-ensembles Maximaux Cohérents (SMC) a été proposé afin d'obtenir une vue globale lorsqu'aucune information concernant la fiabilité des sources n'est disponible (Dubois et al., 2000; Destercke et al., 2009). Tous ces travaux appliquent l'opérateur de fusion sur l'ensemble de toutes les sources et le résultat obtenu est utilisé tel quel.

D'autres approches (Chaudron et al., 2003; Phan-Luong, 2008) utilisent un treillis de concepts avec ses opérateurs pour combiner les informations symboliques. (Chaudron et al., 2003) ont proposé "*le modèle des cubes*" pour représenter et manipuler les informations. Chaque information est représentée par un cube (une conjonction finie de littéraux) et leur méthode de fusion satisfait deux critères. Le premier est le critère d'*endogénéité*, qui signifie que l'information résultante représente une des informations issues des sources. Le second critère est l'*intégrité* qui veut dire que l'information résultante contient au moins toutes les informations fournies par les sources. Des relations d'*ordre partiel*, de *réduction des cubes*, d'*anti-unification* sont définies sur l'espace des cubes pour construire un treillis des cubes qui représente les informations symboliques et leurs résultats de fusion. (Phan-Luong, 2008) utilise aussi la structure de treillis pour organiser les informations et considère des contraintes sur les résultats de la fusion comme dans l'approche de (Chaudron et al., 2003). Deux opérateurs d'*agrégation* et de *consensus* (Phan-Luong, 2008) sont définis pour fusionner les informations issues d'une même ontologie et les structurer dans un treillis.

Dans cet article, nous ne cherchons pas à définir un nouvel opérateur de fusion, mais à structurer de manière synthétique différents résultats de fusion en fonction de différents sous-ensembles de sources. Pour cela, nous utilisons l'Analyse Formelle de Concepts pour étudier les sous-ensembles de sources et leurs résultats de fusion lorsque les informations fournies sont représentées par des distributions de possibilité. L'AFC permet d'associer au sein de concepts formels des sous-ensembles maximaux de sources avec le résultat de la fusion des informations associées à ces sources. Les concepts sont partiellement ordonnés et forment un treillis complet, appelé *treillis de concepts*. Nous montrons que ce treillis contient les résultats partiels aussi bien que le résultat complet de la fusion à la différence des approches classiques qui considèrent seulement le résultat de la fusion appliquée à l'ensemble de toutes les sources, e.g. (Dubois et Prade, 1994; Dubois et al., 2000; Destercke et al., 2009). De plus, le treillis permet d'organiser les résultats de la fusion de différents sous-ensembles de sources et il fournit ainsi plusieurs bases de choix à l'utilisateur. En outre, le treillis permet aussi de préserver l'origine de l'information comme c'est aussi le cas dans la fusion d'informations symboliques en logique possibiliste multi-sources (Dubois et al., 1992). Une telle approche peut s'avérer utile dans toute application où il est nécessaire de synthétiser d'une manière partielle ou globale des valeurs qui proviennent de plusieurs sources. Ici, nous utilisons une expérimentation en agronomie pour aider aux choix de pratiques agricoles. Différents aspects de ce travail ont fait l'objet de communications à des conférences (Assaghir et al., 2010a,b), où nous avons abordé le problème de fusion d'informations représentées par des intervalles. Dans ce papier, nous présentons ces travaux et nous présentons également une extension dans le cadre de la fusion d'informations représentées par des distributions des possibilités triangulaires et trapézoïdales.

Le papier est organisé comme suit. La Section 2 rappelle quelques notions de bases de

la théorie des possibilités. Puis, la Section 3 présente les opérateurs de fusion possibiliste et leurs propriétés. Ensuite, la Section 4 rappelle les notions de base de l'Analyse Formelle de Concepts et les structures de patrons. Dans la Section 5, nous présentons comment organiser les résultats de la fusion dans un treillis pour des informations imprécises représentées par des intervalles, pour ensuite le généraliser dans la Section 6 à des informations représentées par des distributions de possibilité triangulaires et/ou trapézoïdales. Enfin, la Section 7 présente une application en agronomie pour l'aide à la décision de pratiques agricoles.

2 Préliminaires

2.1 Théorie des possibilités

La théorie des possibilités, introduite dans (Zadeh, 1978) et développée par (Dubois et Prade, 1985), permet de traiter l'incertitude et l'imprécision inhérentes à certaines données. Elle est fondée sur la notion de distribution de possibilité.

Soient une variable m dont la valeur est inconnue et V un domaine des valeurs de m . La distribution de possibilité de m , notée π_m , est une application de V dans $[0, 1]$ telle que pour tout $v \in V$ on associe $\pi_m(v)$ le "degré de possibilité" pour que $m = v$. $\pi_m(v) = 1$ indique que absolument rien n'empêche m de valoir v . En revanche, $\pi_m(v) = 0$ indique l'impossibilité que m vaille v .

La hauteur h d'une distribution est la plus grande valeur que la distribution peut atteindre et on écrit $h = \max_{v \in V} \pi_m(v)$. Une distribution de possibilité est dite normalisée si sa hauteur vaut 1. Ce qui signifie qu'un des états au moins est totalement possible.

Etant donné $\alpha \in]0, 1]$, les α -coupes strictes et régulières de la distribution sont définies respectivement par les ensembles $I^{\bar{\alpha}} = \{v \in V | \pi_m(v) > \alpha\}$ et $I^{\alpha} = \{v \in V | \pi_m(v) \geq \alpha\}$. Le support d'une distribution représente l' α -coupe stricte pour $\alpha = 0$ et le noyau représente l' α -coupe régulière pour $\alpha = 1$. Nous notons qu'une distribution peut s'interpréter comme un ensemble d'intervalles emboîtés $A_1 \subseteq \dots \subseteq A_n \subseteq V$ ayant des degrés de confiance $\alpha_1 \leq \dots \leq \alpha_n$. Dans la suite, la distribution de possibilité d'une variable m est notée π_m .

Etant donné un ensemble de distributions de possibilité, on dit qu'elles sont conflictuelles (non cohérentes) si l'intersection de leurs supports est vide. Elles sont à la fois partiellement conflictuelles et partiellement cohérentes si l'intersection de leurs supports n'est pas vide et l'intersection de leurs noyaux est vide. Enfin, elles sont dites cohérentes si leurs noyaux s'intersectent.

2.2 Illustration et exposition du problème

Dans cet article, nous considérons un ensemble de n sources d'informations $G = \{g_1, \dots, g_n\}$. Chaque source $g \in G$ fournit un élément d'informations concernant une variable $m \in M$. Cette information est représentée par une distribution de possibilité $\pi_{m(g)}$ pour la variable $m \in M$. Pour illustrer notre propos, nous considérons les deux cas suivants :

- Les distributions de possibilité appartenant à $\{0, 1\}$, c-à-d que les sources fournissent un intervalle pour la variable m comme dans la Table 1. Par exemple " m_1 appartient à $[1, 5]$ " selon la source g_1 " qui se traduit par $\pi_{m_1(g_1)} = 1$ si $m_1 \in [1, 5]$ et $\pi_{m_1(g_1)} = 0$ sinon.
- Les distributions de possibilité appartenant à l'intervalle $[0, 1]$, c-à-d que l'information est composée d'intervalles emboîtés. Par exemple, dans la Table 2, l'information de g_1

	m_1	m_2
g_1	[1, 5]	[1, 7]
g_2	[2, 3]	[1, 3]
g_3	[4, 7]	[6, 7]
g_4	[6, 10]	[8, 9]

	m_1		m_2	
	Support	Noyau	Support	Noyau
g_1	[1, 5]	[2, 3]	[1, 7]	[2, 5]
g_2	[2, 3]	[3, 3]	[1, 3]	[2, 2]
g_3	[4, 7]	[5, 6]	[6, 7]	[6, 6]
g_4	[6, 10]	[8, 9]	[8, 9]	[8, 8]

TAB. 1 – Informations fournies par TAB. 2 – Informations fournies par les sources sous les sources sous forme d'intervalles forme d'intervalles emboîtés

est donnée comme suit "la valeur de m_1 est dans [1, 5], mais les valeurs au sein de [2, 3] semblent être les plus vraisemblables". Cette information peut être vue comme une distribution de possibilité de support [1, 5] et de noyau [2, 3]. Nous utilisons ainsi la notation $\pi_{m_1(g_1)} = \langle [1, 5]; [2, 3] \rangle$.

Dans la suite, nous utiliserons les notations suivantes : n le nombre de sources, I_i l'intervalle donné par la source i et $\mathbb{I}_m = \{I_1, \dots, I_n\}$ est l'ensemble des valeurs données pour la variable m . f_m représente l'opérateur de fusion pour les intervalles.

3 Opérateurs de fusion possibiliste

Dans cette section, nous rappelons les méthodes de fusion les plus couramment utilisées en théorie des possibilités. Les méthodes de fusion peuvent être caractérisées par l'un des comportements suivants (Bloch et al., 2001; Dubois et Prade, 1994) : conjonctif, disjonctif et de compromis.

Le comportement conjonctif est le pendant d'une intersection d'ensembles. Le résultat d'un opérateur conjonctif est toujours compris dans chacun des éléments d'informations fournies par les différentes sources. Donc, l'opérateur conjonctif réduit l'incertitude globale et fournit un résultat plus précis que l'information provenant des sources prises séparément. Cela suppose que toutes les sources sont fiables. Le résultat obtenu pourra cependant être peu utile s'il reste insuffisamment précis. Il pourra aussi être vide en cas d'incohérence des informations fournies par les sources. Le résultat d'un opérateur conjonctif s'écrit, pour une variable m :

$$\hat{\pi}_m = \bigwedge_{i=1, \dots, n} \pi_{m(g_i)}$$

Ce qui correspond dans le cas des intervalles à leur intersection : $\hat{I}(m) = \bigcap_{i=1, \dots, n} I_i$ où I_i est l'intervalle fourni par la source g_i pour la variable m . Par exemple, dans la Table 1, $\hat{I}(m_1) = \emptyset$ représente l'intersection des intervalles de m_1 .

Le comportement disjonctif est le pendant d'une union d'ensembles. Le résultat d'un opérateur disjonctif contient toujours toutes les informations données par les sources. Un opérateur disjonctif augmente donc l'incertitude globale, et fournit un résultat moins précis que l'information de chacune des sources prise séparément. Il fait la supposition qu'au moins une des sources est fiable. Le résultat d'une telle opération est généralement fiable, mais peut être

très imprécis, ce qui réduit son utilité. L'opérateur disjonctif, pour une variable m , s'écrit :

$$\hat{\pi}_m = \bigvee_{i=1, \dots, n} \pi_m(g_i)$$

Ce qui correspond dans le cas des intervalles à leur union : $\hat{I}(m) = \bigcup_{i=1, \dots, n} I_i$. Par exemple, dans la Table 1, $\hat{I}(m_1) = [1, 10]$ qui représente le résultat de l'union des intervalles de m_1 .

Le comportement de compromis : le résultat d'un comportement de compromis se situe entre les résultats de la conjonction et la disjonction. De tels comportements sont généralement utilisés quand les informations fournies par les sources sont partiellement incohérentes. L'objectif d'un tel comportement est d'obtenir un résultat qui ait un bon équilibre entre informativité et fiabilité. Nous distinguons deux types de compromis :

- adaptatifs : un opérateur de compromis est appelé adaptatif si le résultat dépend du contexte (Bloch et al., 2001). Le but est de passer d'un comportement conjonctif à un comportement disjonctif au fur et à mesure que l'incohérence entre les informations augmente.

On retrouve alors la disjonction (resp. conjonction) en cas d'incohérence totale (resp. cohérence totale) entre les sources. Entre ces deux situations, le comportement est de compromis. Les méthodes utilisant les Sous-ensembles Maximaux Cohérents (SMC) en sont des bons représentants.

Méthode de fusion fondée sur les Sous-ensembles Maximaux Cohérents. Cette méthode consiste à utiliser un opérateur conjonctif sur les sous-ensembles non-conflictuels des sources, et appliquer ensuite un opérateur disjonctif sur les résultats de ces derniers (Dubois et Prade, 1994; Dubois et al., 2000; Destercke et al., 2009). Soit $\mathbb{E} = \{E_1, \dots, E_n\}$ l'ensemble des informations fournies par les sources pour une variable m . Un sous-ensemble $\mathbb{K} \subseteq \mathbb{E}$ est dit maximal cohérent ssi $\bigcap_{i=1, \dots, |\mathbb{K}|} K_i \neq \emptyset$ avec $K_i \in \mathbb{K}$ et il n'existe aucun sous-ensemble $\mathbb{K}' \subseteq \mathbb{E}$ tel que $\mathbb{K} \subseteq \mathbb{K}'$ où $\bigcap_{i=1, \dots, |\mathbb{K}'|} K'_i \neq \emptyset$ et $K'_i \in \mathbb{K}'$. Ainsi, on obtient p sous-ensembles maximaux cohérents et le résultat de la méthode s'écrit :

$$\hat{\pi}_m = \bigvee_{j=1, \dots, p} \bigwedge_{i=1, \dots, |\mathbb{K}_j|} K_i$$

On retrouve la conjonction en cas d'absence de conflit et la disjonction si les sources sont conflictuelles deux à deux. Nous détaillerons dans la suite un exemple de calcul du résultat de la fusion utilisant les SMC pour les intervalles et pour les distributions de possibilité.

- non-adaptatifs : un comportement de compromis est non-adaptatif quand il se comporte toujours de la même manière, quelque soit le contexte. Les moyennes arithmétiques pondérées constituent un exemple typique de tels opérateurs, et sont de loin les opérateurs de fusion les plus utilisés en pratique.

Propriétés des opérateurs de fusion. Généralement, tous les opérateurs de fusion sont commutatifs et idempotents. Les opérateurs conjonctifs et disjonctifs sont associatifs mais les opérateurs compromis ne le sont pas. En plus, si le résultat de la fusion n'est pas convexe, il est toujours possible de prendre son enveloppe convexe (perdre un peu d'informations dans le processus, mais gagner de l'efficacité de calcul). Les opérateurs conjonctifs fournissent un

résultat convexe mais ce n'est pas le cas pour les autres opérateurs en général (pour plus de détails voir (Destercke et al., 2009)).

En conclusion, pour combiner les informations numériques, un opérateur doit être choisi selon le contexte. Les opérateurs de fusion dépendent des informations disponibles sur les sources (fiabilité, conflit, etc.) et du contexte de l'étude. Parfois, le résultat global de la fusion n'est pas directement utile pour prendre une décision comme dans (Assaghir et al., 2009), où la fusion utilisant les SMC fournit un résultat imprécis. Dans cet article, nous proposons de considérer des sous-ensembles globaux ou partiels des sources avec leurs résultats de fusion respectifs pour résoudre ce problème.

4 Analyse formelle de concepts

4.1 Notions de base

Soient G un ensemble d'objets, M un ensemble de propriétés et R une relation binaire définie sur le produit cartésien $G \times M$, i.e. $R \subseteq G \times M$. Le triplet (G, M, R) est appelé *contexte formel* : pour un objet $g \in G$ et une propriété $m \in M$, $(g, m) \in R$ ou gRm signifie que l'objet g "possède" la propriété m . Un exemple d'un contexte formel est donné dans la Table 3 et g_1 possède la propriété m_1 .

Les deux opérateurs de dérivation $(\cdot)'$: $A' = \{m \in M \mid \forall g \in A \subseteq G : gRm\}$ et $B' = \{g \in G \mid \forall m \in B \subseteq M : gRm\}$ définissent une *connexion de Galois* entre l'ensemble de parties de G et celles de M . Le premier opérateur $(\cdot)'$ associe à un ensemble d'objets A l'ensemble des propriétés que les objets de A possèdent en commun. De manière duale, le second opérateur associe à un ensemble de propriétés B , l'ensemble des objets qui possèdent au moins toutes les propriétés de B .

Un couple (A, B) avec $A' = B$ et $B' = A$ est un *concept formel*. A et B sont respectivement appelés l'*extension* et l'*intension* du concept. L'ensemble de tous les concepts du contexte (G, M, R) est ordonné par la relation : $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 (\Leftrightarrow B_2 \subseteq B_1)$, e.g. le concept $(\{g_3\}, \{m_1, m_2\})$ est un sous-concept de $(\{g_1, g_2, g_3\}, \{m_1\})$. L'ensemble de tous les concepts ordonnés forme un treillis appelé *treillis de concepts* du contexte (G, M, R) donné par la Figure 1¹. Cette représentation à *étiquetage réduit* ne présente chaque objet et chaque propriété qu'une seule fois. Un concept est représenté par un nœud. Son extension (resp. intension) est composée de tous les objets (resp. propriétés) qui sont associés à lui et à ses sous-concepts (resp. sur-concepts). Le concept le plus haut dans le treillis est noté (\top) et le plus bas (\perp) .

Le treillis de concepts fournit une classification des objets d'un domaine. Il comporte les deux notions de maximalité et de généralisation/spécialisation :

	m_1	m_2
g_1	×	
g_2	×	
g_3	×	×
g_4		×

TAB. 3 – Contexte formel

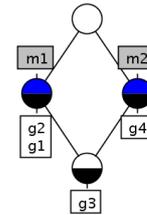


FIG. 1 – Treillis de concepts

1. Le treillis est construit en utilisant ConExp, <http://conexp.sourceforge.net/>

- Maximalité : un concept correspond à un sous-ensemble maximal des objets (extension) qui ont en commun un sous-ensemble des propriétés (intension) et vice versa.
- La généralisation/spécialisation est donnée par l'ordre partiel des concepts. Un concept est plus général que ses sous-concepts, quand son extension est plus grande et son intension est plus petite. De manière duale, un concept est plus spécifique que ses sur-concepts, si son extension est plus petite et son intension est plus large.

Cependant, les données du monde réel, par exemple en agronomie ou en chimie, sont complexes. Elles sont composées de nombres, d'intervalles de nombres ou de graphes. L'AFC ne peut être utilisée que si les données sont transformées sous forme binaire. Cette transformation est appelé *échelonnage conceptuel* ou *discrétisation*, et peut parfois poser des problèmes de perte d'informations. Par exemple, un objet possède la propriété m_1 (resp. m_2) dans la Table 3 ssi sa valeur pour cette propriété est inférieure à 7 (resp. supérieure à 5) dans la Table 1. Un autre choix de seuils fournit un treillis dont l'interprétation est différente. Par conséquent, la manipulation des données numériques par l'AFC n'est pas directe.

4.2 Structures de patrons

Les structures dites de "patrons" sont introduites par (Ganter et Kuznetsov, 2001) comme une extension de l'AFC et permettent la construction d'un treillis à partir de données complexes sans échelonnage. Le triplet $(G, (D, \sqcap), \delta)$ est appelé structure de patrons où G est un ensemble d'objets, (D, \sqcap) est un inf.-demi-treillis² d'éléments appelés patrons et δ est une fonction qui associe à tout objet $g \in G$ sa description $\delta(g)$ dans D . Les patrons de D sont ordonnés, pour tout c et d dans D :

$$c \sqsubseteq d \iff c \sqcap d = c$$

Les deux opérateurs de dérivation suivants forment une connexion de Galois entre les éléments de $(2^G, \subseteq)$ et ceux de (D, \sqsubseteq) , avec $A \in G$ et $d \in (D, \sqcap)$,

$$A^\square = \bigcap_{g \in A} \delta(g) \quad \text{et} \quad d^\square = \{g \in G \mid d \sqsubseteq \delta(g)\}$$

En effet, le premier opérateur $(.)^\square$ de la connexion de Galois prend un ensemble d'objets et retourne l'*infimum* de leur description. De manière duale, le second opérateur $(.)^\square$ donne pour une description le plus grand ensemble d'objets dont les descriptions sont représentées par cette description. Les concepts obtenus sont des couples (A, d) , tels que $A^\square = d$ et $A = d^\square$ et sont ordonnés par $(A_1, d_1) \leq (A_2, d_2) \iff A_1 \subseteq A_2 \iff d_2 \sqsubseteq d_1$ pour former un treillis de concepts.

En conclusion, les structures de patrons permettent de considérer directement les données complexes. Il suffit de définir l'opérateur *infimum* sur les descriptions des objets induisant une relation d'ordre. Ci-dessous, nous montrons qu'un opérateur de fusion peut être un infimum.

5 Organisation des résultats de la fusion d'informations par les structures de patrons

Nous présentons dans cette section une méthode fondée sur l'AFC pour organiser les résultats de la fusion de sources. Les sources vont jouer le rôle des objets, l'opérateur de fusion

2. Un inf.-demi-treillis est un ensemble ordonné où toute paire $\{x, y\}$ d'éléments admet une borne inférieure.

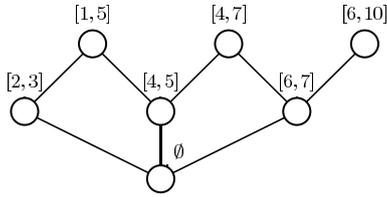


FIG. 2 – Un inf.-demi-treillis d'intervalles

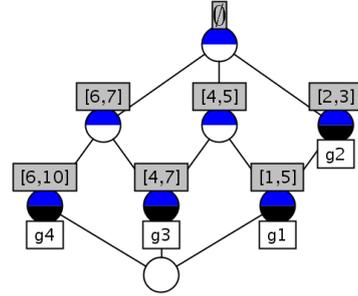


FIG. 3 – Le treillis des résultats conjonctifs pour m_1

représente l'infimum et l'inf.-demi-treillis est construit à partir des éléments d'informations fournies par les sources. Dans cette section, nous considérons le cas où les sources fournissent des intervalles, i.e. le cas des distributions de possibilités dans $\{0, 1\}$, et nous illustrons les exemples à partir de la Table 1. Nous utilisons aussi la notation f_m pour un opérateur de fusion qui retourne le résultat de fusion des n sources pour la variable $m \in M$.

Définition 1 (Espace de fusion) *Un espace de fusion D_m , est composé des informations fournies par les sources pour la variable m , et tous leurs résultats de fusion possibles suivant un opérateur de fusion f_m .*

Par exemple, pour la variable m_1 de la Table 1 et f_m représente l'intersection des intervalles, $D_{m_1} = \{[1, 5], [4, 7], [6, 10], [2, 3], [4, 5], [6, 7], \emptyset\}$.

5.1 Opérateurs de fusion conjonctifs et disjonctifs comme infimum

Considérons la variable $m \in M$ et son espace de fusion D_m qui correspond à un opérateur de fusion f_m . Si l'opérateur f_m est commutatif, idempotent et associatif, (D_m, f_m) est un inf.-demi-treillis. Les opérateurs conjonctifs et disjonctifs vérifient ces propriétés et on a $c \sqcap d = f_m(c, d), \forall c, d \in D_m$, ce qui signifie que l'infimum de deux éléments de D_m correspond à leur résultat de fusion.

Par exemple, si on considère la variable m_1 de la Table 1, et l'opérateur conjonctif qui correspond à l'intersection d'intervalles ($f_{m_1} = \cap$). La Figure 2 représente l'inf.-demi-treillis (D_{m_1}, f_{m_1}) . L'intervalle marquant chaque nœud est l'intersection de tous les intervalles marquant ses ascendants, i.e. le résultat de la fusion conjonctive des sources marquant les ascendants. Dans l'exemple, $f_{m_1}([4, 7], [6, 10]) = [6, 7]$ est le résultat de la fusion des informations provenant des objets g_3 et g_4 pour m_1 , et $f_{m_1}([2, 3], [1, 5]) = [2, 3]$ est celui de g_1 et g_2 . De plus, les éléments de D_{m_1} sont ordonnés tels que, pour tout c et d dans D_{m_1} :

$$c \sqcap d = c \Leftrightarrow c \subseteq d$$

Cet ordre est une instance de la relation de subsumption définie pour les structures de patrons. Dans cet exemple, tout intervalle est subsumé par un intervalle plus large, par exemple $[2, 3] \sqsubseteq [1, 5]$ puisque $[2, 3] \subseteq [1, 5]$. Donc, nous avons $[2, 3] \sqcap [1, 5] = [2, 3] \Leftrightarrow [2, 3] \sqsubseteq [1, 5]$ dans le demi-treillis qui correspond à $[2, 3] \cap [1, 5] = [2, 3] \Leftrightarrow [2, 3] \subseteq [1, 5]$ en terme d'inclusion d'intervalles.

5.2 Construction du treillis de concepts

Etant donné un ensemble de sources G , $m \in M$ est une variable, (D_m, f_m) est le inf.-demi-treillis des résultats de fusion et δ une fonction qui associe à tout objet son information pour la variable m , alors $(G, (D_m, f_m), \delta)$ est une structure de patrons.

Dans notre exemple de la Table 1, nous avons $(G, (D_{m_1}, f_{m_1}), \delta)$ une structure de patrons. (D_{m_1}, f_{m_1}) est l'espace de fusion de m_1 muni de l'intersection. Donc, à l'aide des opérateurs de dérivation définis dans la Section 4.2, nous pouvons calculer les concepts et les ordonner avec $\sqcap = \cap = f_{m_1}$ et $\sqsubseteq = \subseteq$. Revenons à l'exemple de la Table 1, les descriptions des sources g_1 et g_2 sont respectivement $\delta(g_1) = [1, 5]$ et $\delta(g_2) = [2, 3]$, et on a :

$$\begin{aligned} \{g_1, g_2\}^{\square} &= f_{m_1}([1, 5], [2, 3]) = [2, 3] \cap [1, 5] = [2, 3] \\ [2, 3]^{\square} &= \{g \in G \mid [2, 3] \subseteq \delta(g)\} = \{g_1, g_2\} \end{aligned}$$

Puisque $\{g_1, g_2\}^{\square} = [2, 3]$ et $[2, 3]^{\square} = \{g_1, g_2\}$, la paire $(\{g_1, g_2\}, [2, 3])$ est un concept. La Figure 3 représente le treillis des résultats de la fusion conjonctive obtenu à partir de l'exemple de la Table 1.

D'une manière équivalente, l'opérateur disjonctif est idempotent, commutatif et associatif. Donc, le triplet $(G, (D_m, f_m), \delta)$ est une structure de patrons, où D_m est l'espace de fusion suivant l'opérateur $f_m = \cup$ pour toute variable m . De plus, la relation d'ordre définie dans la structure de patrons se traduit dans pour la fusion disjonctive par la relation $c \cup d = c \Leftrightarrow c \supseteq d$. Dans l'exemple de la variable m_1 de la Table 1, $[1, 5] \sqsubseteq [2, 3]$ puisque $[1, 5] \supseteq [2, 3]$.

De nombreux algorithmes, discutés dans (Kuznetsov et Obiedkov, 2002), existent pour extraire et ordonner les concepts d'un contexte formel. Ainsi, ces algorithmes sont adaptés pour les structures de patrons, voir (Ganter et Kuznetsov, 2001; Kaytoue et al., 2010b). En particulier, Kaytoue et al. (Kaytoue et al., 2010b) montrent comment adapter l'algorithme *Close by one* pour les structures de patrons intervalles, et soulignent son passage à l'échelle.

5.3 Interprétation du treillis de concepts

Le treillis de concepts résultant fournit une classification des informations fournies par les sources et leurs résultats de fusion. Sur la Figure 3 (treillis construit à partir de l'opérateur conjonctif), l'extension d'un concept est lue avec l'étiquetage réduit. Cependant, les intensions sont données ici pour chaque concept (non réduit). Par exemple, le noeud marqué avec $[6, 7]$ représente le concept $(\{g_3, g_4\}, [6, 7])$. Puisque les concepts sont ordonnés, un concept fournit le résultat de la fusion des sous-ensembles des extensions de ses sous-concepts.

Un concept (A, d) de $(G, (D, \sqcap), \delta)$, est intéressant de plusieurs points de vue, par exemple le concept $(\{g_1, g_2\}, [2, 3])$ de la Figure 3.

- Son intension d fournit le résultat de fusion des informations des objets de son extension A , e.g. $[2, 3]$ est la fusion conjonctive (l'intersection) résultante des sources g_1 et g_2 .
- Le sous-ensemble A est maximal : l'ajout d'un objet à A conduit à un changement de d , e.g. $\{g_1, g_2\}$ est le sous-ensemble maximal des sources qui ont pour intersection $[2, 3]$.
- L'extension A garde l'origine de l'information, e.g. l'information $[2, 3]$ provient des sources g_1 et g_2 .

Ces propriétés s'appliquent pour tout concept d'un treillis de résultats conjonctifs ou disjonctifs. L'information représentée dans un concept en haut de treillis est plus riche que les informations de ces sous-concepts dans le treillis. La relation d'ordre induit ainsi une notion

de précision dans les deux cas de fusion conjonctive et disjonctive. Si on a deux concepts (A_1, d_1) et (A_2, d_2) tels que $(A_1, d_1) \leq (A_2, d_2)$ alors l'information contenue dans A_2 est plus riche que l'information contenue dans A_1 mais la précision est différente suivant l'opérateur de fusion. En effet, plus haut est le concept dans le treillis des résultats conjonctifs, plus son extension est large et plus petit est l'intervalle de son intention. En terme de fusion d'informations, plus on monte dans le treillis plus l'information contenue dans le concept est fiable et précise. Au contraire, avec le treillis des résultats disjonctifs, plus haut est un concept, plus l'ensemble des valeurs de son intension est large et plus sont nombreux les objets de son extension et l'information contenue dans son intension est imprécise.

Donc, la navigation dans le treillis permet plus de flexibilité pour aider à la décision. Par exemple, dans les travaux cités, la fusion de l'ensemble de toutes les sources correspond au concept le plus général (\top) dans le treillis. Ce résultat ne permet pas souvent de décider, e.g. l'intersection vide dans notre exemple. Par conséquent, la navigation dans le treillis permet d'identifier des sous-ensembles maximaux d'objets et leur fusion.

5.4 Treillis fondé sur les sous-ensembles maximaux cohérents

Avant d'introduire la méthode de construction du treillis des SMC, nous rappelons comment calculer les SMC de n intervalles.

Extraction des SMC des intervalles. Etant donnés n intervalles $\mathbb{I} = \{I_1, I_2, \dots, I_n\}$, un sous-ensemble $\mathbb{K} \subseteq \mathbb{I}$ est dit maximal cohérent si $\bigcap_{i=1}^{|\mathbb{K}|} K_i \neq \emptyset$ avec $K_i \in \mathbb{K}$ et il n'existe aucun sous-ensemble non vide $\mathbb{K}' \subseteq \mathbb{I}$ tel que $\mathbb{K} \subseteq \mathbb{K}'$ avec $\bigcap_{i=1}^{|\mathbb{K}'|} K'_i \neq \emptyset$ et $K'_i \in \mathbb{K}'$. Par exemple, dans la Table 1, le sous-ensemble $\mathbb{K} = \{I_1, I_2\} = \{[1, 5], [2, 3]\}$ est maximal cohérent pour la variable m_1 puisque $I_1 \cap I_2 \neq \emptyset$ et tout ajout d'un intervalle à \mathbb{K} conduit à une intersection vide.

En général, rechercher les SMC est un problème de complexité exponentielle e.g. (Malouf, 2007). (Dubois et al., 2000) ont introduit un algorithme linéaire pour trouver les SMC de n intervalles (voir Algorithme 1).

L'algorithme s'appuie sur l'ordre croissant des points extrêmes a_i et b_i des intervalles dans une nouvelle séquence $(c_q, q = 1, \dots, 2n)$. L'intersection maximale cohérente d'intervalles est atteinte seulement quand un élément c_q de type a (c-à-d une borne inférieure d'un intervalle) est suivi par un élément c_{q+1} de type b (c-à-d une borne supérieure d'un intervalle). La méthode de fusion consiste donc à appliquer l'union aux sous-ensembles maximaux cohérents obtenus. Par exemple, dans la Table 1, les sous-ensembles maximaux cohérents des intervalles $\{I_1, I_2, I_3, I_4\}$ de la variable m_1 , en utilisant l'Algorithme 1, sont $\{I_1, I_2\}$, $\{I_1, I_3\}$ et $\{I_3, I_4\}$ puisqu'ils vérifient la définition d'un sous-ensemble maximal cohérent et qui correspondent respectivement aux valeurs $[2, 3]$, $[4, 5]$ et $[6, 7]$ (Figure 4). Donc le résultat de la fusion des SMC est donné par : $[2, 3] \cup [4, 5] \cup [6, 7]$.

Treillis de résultats de fusion fondée sur les SMC. L'opérateur de fusion f_m utilisant la notion de SMC est idempotent et commutatif, mais il n'est pas associatif. Par exemple dans Table 1, $f_{m_1}(f_{m_1}([1, 5], [2, 3]), [4, 7]) = [2, 3] \cup [4, 7]$ et $f_{m_1}(f_{m_1}([1, 5], [4, 7]), [2, 3]) = [2, 3] \cup [4, 5]$. Par conséquent, l'opérateur de fusion des SMC ne possède pas les propriétés d'un infimum dans un demi-treillis. Cependant, l'opérateur de fusion retourne l'union des SMC. Nous proposons donc de chercher les SMC des sources et ensuite d'appliquer un opérateur

Algorithme 1 : Sous-ensembles maximaux cohérents sur des intervalles

Entrées : n intervalles $[a_i, b_i]$
Sorties : Liste de p sous-ensembles maximaux cohérents \mathbb{K}_j
 Liste = \emptyset ; $j = 1$; $\mathbb{K} = \emptyset$
 Mettre en ordre croissant $\{a_i, i = 1, \dots, n\} \cup \{b_i, i = 1, \dots, n\}$
 Les renommer $\{c_q, q = 1, \dots, 2n\}$
 avec $type(c_q) = a$ si $c_q = a_k$ et $type(c_q) = b$ si $c_q = b_k$ et $k \in \{1, \dots, n\}$
pour chaque $q = 1$ to $2n - 1$ **faire**
 si $type(c_q) = a$ **alors**
 Ajouter la source g_k à \mathbb{K} t.q. $c_q = a_k$
 si $type(c_{q+1}) = b$ **alors**
 Ajouter \mathbb{K} à Liste ($\mathbb{K}_j = \mathbb{K}$)
 $j = j + 1$
 sinon
 Enlever la source g_k de \mathbb{K} t.q. $c_q = b_k$

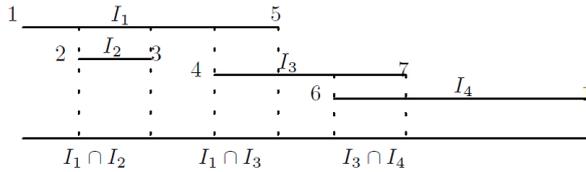


FIG. 4 – SMC de la variable m_1 dans la Table 1

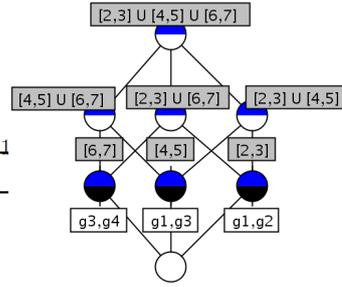


FIG. 5 – Treillis de concepts de SMC

disjonctif (union) au sein de ces derniers pour définir un inf.-demi-treillis puisque l'opérateur de fusion fondé sur les SMC représente l'union des SMC obtenus des n intervalles.

Formellement, on considère le triplet $(\mathcal{O}, (\mathcal{K}, \cup), \delta)$ comme une structure de patrons où les objets dans $\mathcal{O} \subseteq \mathcal{G}$ sont les origines des SMC des intervalles \mathbb{I}_m pour une variable m , i.e. l'ensemble de sources donnant le SMC. (\mathcal{K}, \cup) est un inf.-demi-treillis où \mathcal{K} est l'espace de fusion des informations fournies par les objets dans \mathcal{O} en utilisant un opérateur disjonctif. Considérons l'exemple de la Table 1, les SMC des intervalles de m_1 sont $[2, 3]$, $[4, 5]$ et $[6, 7]$ provenant respectivement de sous-ensembles des sources $\{g_1, g_2\}$, $\{g_1, g_3\}$ et $\{g_3, g_4\}$. Par la suite, \mathcal{O} représente l'ensemble $\{(g_1, g_2), (g_1, g_3), (g_3, g_4)\}$ avec $\delta((g_1, g_2)) = [2, 3]$ (qui signifie que l'intervalle $[2, 3]$ est attaché aux sources g_1 et g_2), $\delta((g_1, g_3)) = [4, 5]$ et $\delta((g_3, g_4)) = [6, 7]$. Ensuite, nous utilisons l'union d'intervalles comme un infimum pour construire le treillis.

Le treillis résultant, de la Table 1 pour la variable m_1 , est donné sur la Figure 5. Les extensions des concepts se lisent en utilisant l'étiquetage réduit. Les intensions sont données ici pour chaque concept. Par exemple sur la Figure 5, le concept situé à droite dans la seconde ligne est $\{(g_1, g_2), (g_1, g_3)\}, [2, 3] \cup [4, 5]$ présentant les valeurs de m_1 suivant les objets (g_1, g_2) et (g_1, g_3) . Ce qui signifie que l'information $[2, 3] \cup [4, 5]$ est fournie par les sources

" g_1 et g_2 " ou " g_1 et g_3 " (c-à-d " g_1 et " g_2 ou g_3 "). Nous pouvons ainsi remarquer que les concepts du treillis des SMC situés juste en dessus du concept \perp (le concept le plus bas dans le treillis) peuvent contenir des objets g_i (de la Table 1) en commun puisqu'un objet peut être cohérent avec plusieurs objets du contexte. Au contraire, dans les treillis des résultats conjonctifs et disjonctifs, un objet du contexte initial ne peut appartenir qu'à un seul concept. Par exemple l'objet g_1 appartient aux deux concepts $((g_1, g_2), [2, 3])$ et $((g_1, g_3), [4, 5])$ sur la Figure 5.

De plus, le treillis permet d'obtenir les sous-ensembles maximaux des sources qui ont fourni l'information contenue dans les intentions des concepts à partir des objets des extensions. Par exemple, les valeurs $[2, 3] \cup [4, 5]$ représentent le résultat de la fusion du sous-ensemble d'objets $\{g_1, g_2, g_3\}$ en utilisant l'opérateur fondé sur les SMC. Par conséquent, le concept \top , qui correspond à l'union de tous les SMC, est le résultat global de la fusion fondée sur les SMC appliquée à toutes les sources.

Donc, pour les concepts situés juste au dessus du concept (\perp) (la première ligne des concepts dans le treillis), les sous-ensembles maximaux des sources sont les sources données dans leurs extensions. Pour les autres concepts, nous considérons l'union des sous-ensembles d'objets des extensions de leurs sous-concepts si les extensions ont au moins une objet (du contexte initial) en commun. Revenons à la Figure 5 et considérons le concept $((g_1, g_2), (g_1, g_3), [2, 3] \cup [4, 5])$: les valeurs $[2, 3] \cup [4, 5]$ proviennent de $\{g_1, g_2\} \cup \{g_1, g_3\} = \{g_1, g_2, g_3\}$. Mais, dans le cas où les objets des extensions n'ont aucune intersection, le sous-ensemble maximal de sources est l'union des objets en excluant tout objet cohérent avec au moins un objet de chaque concept. Par exemple sur la Figure 5, considérons le concept $((g_1, g_2), (g_3, g_4), [2, 3] \cup [6, 7])$, le sous-ensemble maximal de sources fournissant les valeurs $[2, 3] \cup [6, 7]$ est $\{g_2, g_3, g_4\}$ puisque l'objet g_1 est cohérent avec g_2 d'une part et avec l'objet g_3 d'autre part.

Enfin, la méthode utilisée dans cette section pour obtenir le treillis des SMC ne considère pas les résultats de la fusion des SMC à la différence des opérateurs conjonctifs et disjonctifs. Ceci est dû à la non-associativité de l'opérateur fondé sur les SMC. Ainsi, le treillis ne contient pas tous les sous-ensembles de G avec la fusion de leur SMC puisque l'union est utilisée au sein des SMC des intervalles du contexte initial et non directement sur ces derniers intervalles eux mêmes. Néanmoins, le treillis nous permet de garder l'origine de l'information à la différence de plusieurs travaux cités. De plus, le treillis donne plus de flexibilité pour les utilisateurs dans le choix d'un sous-ensemble maximal des sources.

5.5 Traitement de plusieurs variables dans le treillis de concepts

Dans cette section, nous considérons que les sources fournissent des valeurs pour différentes variables. Comme par exemple, la Table 1 présente les objets décrits par des *vecteurs d'intervalles*. Un *vecteur d'intervalles* est un vecteur à p dimensions, composé de p intervalles où chaque intervalle correspond à une variable unique, par exemple la description de l'objet g_1 est $\delta(g_1) = \langle [1, 5], [1, 9] \rangle$. Ceci peut être intéressant pour calculer le résultat de la fusion pour toutes les variables simultanément.

Pour formaliser une structure de patrons dans ce cas, il suffit de définir un infimum, i.e. un opérateur de fusion dans notre cas, pour chaque dimension (variable). L'opérateur de fusion s'applique sur deux ou plusieurs vecteurs. Les vecteurs doivent être de la même taille et les variables sont indépendantes les unes des autres. On suppose un ordre canonique sur les dimensions des vecteurs et la fusion des vecteurs d'intervalles est définie comme étant la fusion

entre les intervalles de chaque dimension. Cela induit un ordre partiel des descriptions des objets (Kaytoute et al., 2010b). Le triplet $(G, (D, \sqcap), \delta)$ détermine une structure de patrons, où G est un ensemble des sources, (D, \sqcap) est un inf.-demi-treillis de patrons et δ est la fonction qui associe à tout $g \in G$ sa description dans D . Notons que \sqcap représente l'opérateur de fusion qui peut être choisi pour chaque dimension.

Revenons à la Table 1 et considérons les deux objets g_1 et g_2 . Les résultats de la fusion conjonctive pour les variables m_1 et m_2 sont respectivement $[2, 3]$ et $[1, 3]$ puisque $\langle [1, 5], [1, 7] \rangle \sqcap \langle [2, 3], [1, 3] \rangle = \langle [1, 5] \sqcap [2, 3], [1, 7] \sqcap [1, 3] \rangle = \langle [2, 3], [1, 3] \rangle$. L'opérateur disjonctif se traite de la même manière, par exemple, $\langle [1, 5], [1, 7] \rangle \sqcup \langle [2, 3], [1, 3] \rangle = \langle [1, 5] \sqcup [2, 3], [1, 7] \sqcup [1, 3] \rangle = \langle [1, 5], [1, 7] \rangle$ ce qui signifie que $[1, 5]$ et $[1, 7]$ sont respectivement les résultats de la fusion disjonctive pour les variables m_1 et m_2 .

Quant à l'opérateur fondé sur les SMC, on procède de la même manière que dans le cas d'une seule variable, c'est-à-dire trouver les SMC de chacune des variables et vérifier si les SMC de la variable m_i sont aussi des SMC pour la variable m_j . Par exemple, les SMC des intervalles de la variable m_1 sont $\{g_1, g_2\}$, $\{g_1, g_3\}$ et $\{g_3, g_4\}$ et ceux de la variable m_2 sont $\{g_1, g_2\}$ et $\{g_1, g_3\}$. Par conséquent on considère $\mathcal{O} = \{(g_1, g_2), (g_1, g_3), (g_3, g_4)\}$ et le contexte est donné

	m_1	m_2
(g_1, g_2)	$[2, 3]$	$[1, 3]$
(g_1, g_3)	$[4, 5]$	$[6, 7]$
(g_3, g_4)	$[6, 7]$	\emptyset

TAB. 4 – Le tableau résultant de pré-traitement pour deux variables

dans la Table 4 où la description de chaque objet dans \mathcal{O} est donnée par son sous-ensemble maximal de valeurs pour la variable, par exemple, le sous-ensemble $\{g_3, g_4\}$ décrit la variable m_1 par l'intervalle $[6, 7]$ et \emptyset pour la variable m_2 puisque $\{g_3, g_4\}$ n'est pas un SMC pour la variable m_2 . Notons finalement que l'algorithme de pré-traitement est de complexité quadratique $O(|SMC|)$, où SMC est l'ensemble de tous les SMC pour toutes les variables.

6 Traitement des distributions de possibilité

Dans cette section, nous présentons comment généraliser le cas des intervalles classiques aux intervalles flous, i.e. nous considérons que les objets décrivent la variable à l'aide d'une distribution de possibilité dans l'intervalle $[0, 1]$. Nous illustrons nos exemples à partir de la Table 2. Comme dans le cas des intervalles, les opérateurs de fusion conjonctifs et disjonctifs possèdent les propriétés d'un infimum (commutativité, associativité et idempotence). Par conséquent, pour une variable $m \in M$, (D_m, f_m) est un inf.-demi-treillis avec D_m est l'espace de fusion de distributions de la variable m pour un opérateur de fusion f_m .

L'espace de fusion de distributions (D_m) est composé des distributions données par les sources et leurs résultats de fusion possibles suivant un opérateur de fusion f_m . Une distribution de possibilité représente ici un intervalle flou qui peut être vu comme une collection (infinie) d'intervalles emboîtés définis par $\{v | \pi_{m(g)}(v) \geq \alpha\}$. Selon le niveau α considéré, l'intersection des coupes de niveau α des deux distributions peut être non vide, puis devenir vide pour des valeurs plus grandes de α . Ce qui conduit à passer d'un opérateur conjonctif à un opérateur disjonctif selon le niveau. Il en résulte que l'ensemble des résultats de la fusion aux différents niveaux ne constitue plus une distribution puisque les α -coupes ne sont plus emboîtées (celles des niveaux supérieurs étant trop grandes à cause de l'application de la disjonction, alors que dans les niveaux inférieurs la conjonction a pu être appliquée, conduisant à des coupes de niveaux plus petites). Les valeurs de transition entre l'application de la conjonction et de

la disjonction correspondent aux hauteurs de l'intersection de distributions " $\pi_{m(g_i)} \wedge \pi_{m(g_j)}$ ". A partir de cette valeur, les distributions ne sont plus cohérentes et le résultat de la fusion disjonctive est une union d'intervalles disjoints. En effet, il existe un ensemble fini des valeurs β_q telles que $0 < \beta_1 \leq \dots \leq \beta_{q+1} = 1$ et les intervalles sont emboîtés entre $] \beta_q, \beta_{q+1}]$ (Destercke et al., 2009). Notons $([I; J], [\theta, \gamma]) = \{[x, y] \subseteq \mathbb{R} \mid I \supseteq [x, y] \supseteq J\}$ l'ensemble des intervalles emboîtés entre deux niveaux θ et γ . Par exemple, $([6, 7], [\frac{20}{3}, \frac{20}{3}],]0, \frac{1}{3}])$ est un élément de l'espace de distributions des résultats conjonctifs dans la Table 2. De plus, le résultat de la fusion disjonctive de g_3 et g_4 est l'union des intervalles emboîtés entre $] \beta_q, \beta_{q+1}]$ pour l'ensemble des valeurs $0 < \beta_1 = \beta_2 = \frac{1}{3} < \beta_3 = 1$ (voir Figure 6 pour les valeurs de β_q).

On définit ainsi la relation d'ordre sur les distributions de possibilité comme étant la relation d'inclusion des sous-ensembles flous.

Treillis des résultats conjonctifs. Pour les opérateurs conjonctifs, le triplet $(G, (D_m, \wedge), \delta)$ est une structure de patrons et les opérateurs de dérivation sont définis pour tout $A \in G$ et $d \in D_m$ par :

$$A^\square = \bigwedge_{g \in A} \pi_{m(g)} \quad d^\square = \{g \in G \mid d \sqsubseteq \pi_{m(g)}\}$$

A^\square représente l'intersection des distributions des sources de A et d^\square représente les sources qui ont une distribution contenant d . Le résultat de A^\square est une distribution convexe.

$$\begin{aligned} \{g_3, g_4\}^\square &= \pi_{m(g_3)} \wedge \pi_{m(g_4)} = ([6, 7], [\frac{20}{3}, \frac{20}{3}],]0, \frac{1}{3}]) \\ ([6, 7], [\frac{20}{3}, \frac{20}{3}],]0, \frac{1}{3}])^\square &= \{g_3, g_4\} \end{aligned}$$

Une remarque relative à la fusion conjonctive d'un sous-ensemble de sources $A \subseteq G$ est que son résultat n'est pas normalisé, c'est-à-dire $\forall v \in V, \bigwedge_{g \in A} \pi_{m(g)} < 1$. Cela traduit un conflit entre les sources, puisqu'aucune valeur n'est trouvée simultanément complètement possible par les sources.

Moins le résultat de la conjonction est normalisé, plus le conflit entre les sources est important, et moins il est légitime de supposer que toutes les sources du sous-ensemble A sont fiables. Si malgré tout on continue de supposer que les sources sont fiables, on peut renormaliser le résultat de la fusion conjonctive, en divisant par la hauteur du résultat de la fusion qui représente le complément à 1 du degré de conflit entre les sources. Il faudra cependant passer à une attitude disjonctive en cas de conflit total (intersection vide des distributions).

Treillis des résultats disjonctifs. Pour la fusion disjonctive, $(G, (D_m, \vee), \delta)$ est la structure de patrons. Les opérateurs de dérivation se définissent comme dans l'opérateur conjonctif et la relation d'ordre est l'inclusion des sous-ensembles flous.

Le treillis des résultats conjonctifs (resp. disjonctifs) des distributions s'interprète comme celui des intervalles mais en ajoutant une notion de confiance à chacun des résultats représenté par l'intervalle $] \beta_q, \beta_{q+1}]$. Pour l'opérateur de fusion fondé sur les SMC, nous calculons d'abord les SMC des distributions et nous appliquons l'opérateur disjonctif sur ces derniers. Nous allons détailler maintenant une méthode de pré-traitement des informations fournies à partir de laquelle nous obtenons une structure de patrons pour construire le treillis des SMC.

Extraction des SMC des distributions. Partant de la définition d'une distribution de possibilité comme un ensemble d'intervalles emboîtés (ses α -coupes), nous pouvons calculer le résultat de la fusion de n sources. En effet, à chaque niveau α , les α -coupes des distributions forment un ensemble d'intervalles dont nous pourrions calculer les résultats de leur fusion.

Par la suite, pour tout $\mathbb{I}^\alpha = \{I_1^\alpha, \dots, I_n^\alpha\}$ qui correspond à l'ensemble de n α -coupes I_i^α ($i = 1, \dots, n$), on obtient un résultat de fusion des SMC des intervalles comme détaillé dans la section 5.4.

Formellement, soit \mathbb{K}^α les sous-ensembles d'intervalles de \mathbb{I}^α tels que $\bigcap_{i=1, \dots, |\mathbb{K}^\alpha|} K_i^\alpha \neq \emptyset$, avec $K_i^\alpha \subseteq \mathbb{I}^\alpha$, la méthode de fusion s'écrit alors : $\mathcal{K}^\alpha = \bigcup_{j=1, \dots, p} \mathbb{K}_j^\alpha$, où p est le nombre des sous-ensembles maximaux cohérents pour un niveau α donné. En général, \mathcal{K}^α est une union d'intervalles disjoints, et nous n'avons pas $\mathcal{K}^\alpha \supset \mathcal{K}^\beta$, $\forall \beta > \alpha$, par conséquent le résultat de la fusion globale de n sources n'est pas une distribution de possibilité, puisque les α -coupes ne sont pas emboîtées. Par conséquent, le calcul du résultat des SMC se fait suivant le même principe du calcul des résultats conjonctifs, c-à-d que les résultats sont des unions d'ensembles d'intervalles emboîtés disjoints entre $]\beta_q, \beta_{q+1}]$.

Considérons l'exemple de la Table 2, où il y a quatre distributions. Les SMC des supports des sources proviennent de $\{g_1, g_2\}$, $\{g_1, g_3\}$ et $\{g_3, g_4\}$ dont les hauteurs respectives sont $1, \frac{1}{3}$ et $\frac{1}{3}$. Donc, les valeurs de β_q correspondantes à notre exemple sont $0 < \beta_1 = \beta_2 = \frac{1}{3} < \beta_3 = 1$. Comme mentionné plus haut, à partir de niveau $\alpha = \frac{1}{3}$, les distributions de g_2 et celle de g_3 (respectivement g_3 et g_4) ne sont plus cohérentes (elles n'ont plus d'intersection). La Figure 6 montre le résultat global de la fusion pour $\alpha \in [0, 1]$, qui n'est pas une distribution. La Figure 6 montre ainsi les coupes de niveaux $\frac{1}{3}$ et 1.

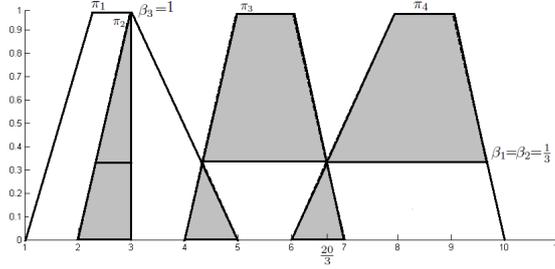


FIG. 6 – Distributions de la variable m_1 de la Table 2 (–) avec le résultat global de fusion des SMC (en gris).

La Figure 6 montre le résultat global de la fusion pour $\alpha \in [0, 1]$, qui n'est pas une distribution. La Figure 6 montre ainsi les coupes de niveaux $\frac{1}{3}$ et 1.

Construction du treillis. Nous utilisons la même méthode de traitement de données introduite dans la section 5.4, i.e. nous calculons les SMC avec leurs valeurs pour la variable m et nous construisons le treillis en utilisant l'opérateur disjonctif. Comme mentionné ci-dessus, les résultats de la fusion des SMC ne sont pas des distributions, mais ce sont des sous-ensembles des valeurs munis de niveaux de confiance. Par conséquent, pour construire le treillis des SMC, il faut considérer plusieurs cas suivant les valeurs de α .

Formellement, nous considérons $(\mathcal{O}, (K, \cup), \delta)$ comme une structure de patrons où les objets dans $\mathcal{O} \subseteq G$ sont les origines des SMC des distributions pour la variable m en respectant les valeurs de α . (K, \cup) est l'inf.-demi-treillis des descriptions munies de l'union et δ est une fonction qui associe à chaque objet dans \mathcal{O} sa description dans K . Chaque objet est décrit par sa valeur et un intervalle de confiance, c'est-à-dire un ensemble d'intervalles emboîtés entre deux niveaux. Par exemple, dans la Table 2, le SMC des sources g_3 et g_4 est

$$\begin{cases} \left[\left[\left[6, 7 \right]; \left[\frac{20}{3}, \frac{20}{3} \right] \right] & \text{si } \alpha \in]0, \frac{1}{3}] \\ \left[\left[\left[\frac{13}{3}, \frac{20}{3} \right]; [5, 6] \right] \cup \left[\left[\frac{20}{3}, \frac{29}{3} \right]; [8, 9] \right] \right] & \text{si } \alpha \in]\frac{1}{3}, 1] \end{cases}$$

Les éléments de K sont des couples $(R, [\gamma, \theta])$ où R est l'ensemble des intervalles emboîtés entre les deux niveaux γ et θ . Ces éléments sont ordonnés, pour tout $d_1 = (R_1; [\alpha_1, \beta_1])$ et $d_2 = (R_2; [\alpha_2, \beta_2])$, par : $d_1 \leq d_2 \Leftrightarrow R_1 \subseteq R_2$ et $[\gamma_1, \theta_1] \subseteq [\gamma_2, \theta_2]$. L'ensemble de concepts ordonnés forme le treillis de concepts dont le \top est le résultat de la fusion globale

des SMC pour les valeurs de $\alpha \in]0, 1]$. Ainsi, le treillis permet d'identifier des sous-ensembles maximaux avec leur résultat de fusion et un intervalle de confiance tout en gardant l'origine de l'information. Dans le cas où les sources sont en conflit total (intersection vide des distributions), le treillis des résultats par SMC est le même que le treillis des résultats disjonctifs puisque les SMC des sources seront les informations fournies par les sources.

7 Une application en agronomie

Données et problème. Les agronomes calculent des indicateurs pour évaluer l'impact des pesticides sur les pratiques agricoles et l'environnement en utilisant les caractéristiques d'un pesticide et les propriétés des champs. Par exemple, l'indicateur I_{gro} renseigne du risque qu'un produit atteigne les eaux souterraines (Van der Werf et Zimmer, 1998). Les agronomes établissent des diagnostics suivant la valeur calculée de I_{gro} . Une valeur inférieure à 7 indique que l'agriculteur doit changer ses pratiques (pesticide, sol, date, etc...). Au contraire, une valeur supérieure à 7 indique que les pratiques utilisées par l'agriculteur ne posent pas de problèmes pour l'environnement (Van der Werf et Zimmer, 1998). Les caractéristiques du pesticide proviennent des sources d'information (expert, livre, base de données) et dont les origines sont différentes (France, Angleterre, Etats-Unis), tant que les caractéristiques du champ sont des mesures de terrain (Van der Werf et Zimmer, 1998). Par conséquent les valeurs de chaque caractéristique varient suivant les sources.

Dans cet article, nous nous intéressons à l'utilisation du pesticide *sulcotrione* et son influence sur les eaux souterraines. Le *sulcotrione* est un herbicide vendu depuis 1993 et il est utilisé pour lutter contre les mauvaises herbes dans les cultures de maïs. Trois caractéristiques de *sulcotrione* sont utilisées dans le calcul de I_{gro} , appelées *DT50*, *koc*, et *ADI* (voir (Van der Werf et Zimmer, 1998) pour plus de détails). La Table 5 (données simplifiées) montre les valeurs des caractéristiques *DT50* et *koc* suivant 8 différentes sources d'informations, par exemple, la valeur de *DT50* selon le groupe d'experts européens *com98* est [2, 6]. Les sources sont hétérogènes, donnant des valeurs conflictuelles des variables par exemple *DT50* vaut [2, 6] selon *com98* et elle vaut [15, 74] selon *AGXl*. Ceci est dû aux conditions de mesure de la valeur de la variable. Par exemple, les champs en Angleterre ont des caractéristiques différentes que les champs en France ainsi que des conditions comme la température. Des sources d'informations sont consultées mais elles n'avaient pas d'informations pour la variable. Ces données sont prises en compte par la théorie des possibilités comme étant des cas d'ignorance totale, c'est-à-dire que toutes les valeurs de la variable sont possibles et la distribution représentant cette information vaut 1 partout sur le domaine des valeurs. Elles se traduisent aussi par le fait que la source est en accord avec les autres sources. Donc, les informations manquantes sont remplacées par l'intervalle $[\min, \max]^3$ qui est en accord avec toutes les autres sources. Par exemple, dans la Table 5, les valeurs de *DT50* selon la base de données *PM12* est [2, 74].

	DT50 day	koc L/kg
<i>BUS</i>	[2, 74]	[1.08, 940]
<i>PM11</i>	[15, 72]	[1.08, 940]
<i>PM12</i>	[2, 74]	[44, 940]
<i>PM13</i>	[2, 74]	[44, 940]
<i>INRA</i>	[2, 74]	[1.08, 8.98]
<i>Com98</i>	[2, 6]	[17, 160]
<i>AGXf</i>	[2, 6]	[1.08, 160]
<i>AGXl</i>	[15, 74]	[1.08, 160]

TAB. 5 – Caractéristiques de *Sulcotrione*

3. \min (resp. \max) représente le minimum (resp. maximum) de toutes les valeurs proposées par les sources.

La valeur de ADI pour le *sulcotrione* vaut 0.00005 pour toutes les sources. Les agronomes cherchent une valeur parmi les valeurs proposées pour la considérer dans le calcul de l'indicateur I_{gro} . Par conséquent, c'est un problème de fusion d'informations où une valeur doit être trouvée pour chacune des caractéristiques.

Pour combiner les différents éléments d'informations, un opérateur doit être choisi. Dans cette application, (1) les sources sont hétérogènes, (2) aucune information sur la fiabilité des sources n'est disponible. Donc, l'opérateur approprié dans cette situation est l'opérateur utilisant les SMC. Les SMC de la variable $DT50$ sont $[2, 6]$ et $[15, 72]$ respectivement $[1.08, 8.98]$ et $[44, 160]$ pour koc .

Premiers résultats. Les résultats de la fusion globale (proposé dans Destercke et al. (2009)) des SMC des deux variables $DT50$ et koc sont $[2, 6] \cup [15, 72]$ et $[1.08, 8.98] \cup [44, 160]$. Le calcul des bornes inférieure et supérieure de l'indicateur et ses conséquences pour les pratiques agricoles et la pollution de l'environnement sont détaillés dans (Assaghir et al., 2009). Nous considérons l'enveloppe convexe du résultat de fusion pour le calcul de l'indicateur. En utilisant ces valeurs, on obtient l'intervalle $[4, 10]$ pour I_{gro} . Cet intervalle n'est pas utile puisque ces valeurs ne sont pas toutes plus grandes que 7 ni plus petites que 7, donc l'analyste ne peut pas prendre une décision concernant les pratiques agricoles. L'intervalle obtenu de I_{gro} est également imprécis et ne permet pas les experts de prendre une décision. Par conséquent, la fusion globale des sources ne fournissant pas de résultats suffisamment précis, et il est intéressant de considérer des sous-ensembles de sources avec leurs résultats de fusion.

Construction du treillis et interprétation. Comme mentionné ci-dessus, la fusion de toutes les sources ne conduit pas à un résultat utile. Par conséquent, le choix d'un sous-ensemble des sources avec leur résultat de fusion permet de réduire l'imprécision contenue dans le résultat de la fusion globale. Nous utilisons donc la méthode détaillée dans la Section 5.4 qui permet de construire un treillis des sous-ensembles maximaux de sources et leurs résultats. Les SMC pour la variable $DT50$ sont K_1 et K_2 , resp. K_3 et K_4 pour koc (voir Table 6).

$K_1 = \{BUS, PM12, PM13, INRA, Com98, AGXf\}$
$K_2 = \{BUS, PM11, PM12, PM13, INRA, AGXl\}$
$K_3 = \{BUS, PM11, INRA, AGXf, AGXl\}$
$K_4 = \{BUS, PM11, PM12, PM13, Com98, AGXf, AGXl\}$

TAB. 6 – Origine des SMC

	DT50 (days)	koc (L/kg)
K_1	$[2, 6]$	\emptyset
K_2	$[15, 72]$	\emptyset
K_3	\emptyset	$[1.08, 8.98]$
K_4	\emptyset	$[44, 160]$

TAB. 7 – Table résultant du pré-traitement de la Table 5

La Table 7 résulte du pré-traitement de la Table 5, détaillé dans la Section 5.4. Le treillis résultant est donné sur la Figure 7 avec 15 concepts. Une extension est présentée avec étiquetage réduit. L'intension des concepts est obtenue à partir des intentions des sous-concepts, par exemple, l'intension du concept C_1 est $\{(DT50, [15, 72]), (koc, [44, 160])\}$. Mais, si les intensions de deux sous-concepts donnent des différentes valeurs pour le même attribut, alors l'union des valeurs est considérée. Par exemple, l'intension du concept C_2 est $\{(DT50, [2, 6] \cup [15, 72]), (koc, [44, 160])\}$ et les intensions de ses sous-concepts sont $\{(DT50, [2, 6])\}$, $\{(DT50, [15, 72])\}$ et $\{(koc, [44, 160])\}$. De plus, chaque intension d'un concept du treillis représente la fusion des SMC d'objets dans son extension.

Le concept le plus général dans le treillis correspond au résultat de la fusion de SMC de toutes les sources pour toutes les caractéristiques. Par exemple, le concept du bas à droite dans la Figure 7 est $(\{K_1\}, \{(DT50, [2, 6])\})$ où $[2, 6]$ est le résultat de fusion des SMC de

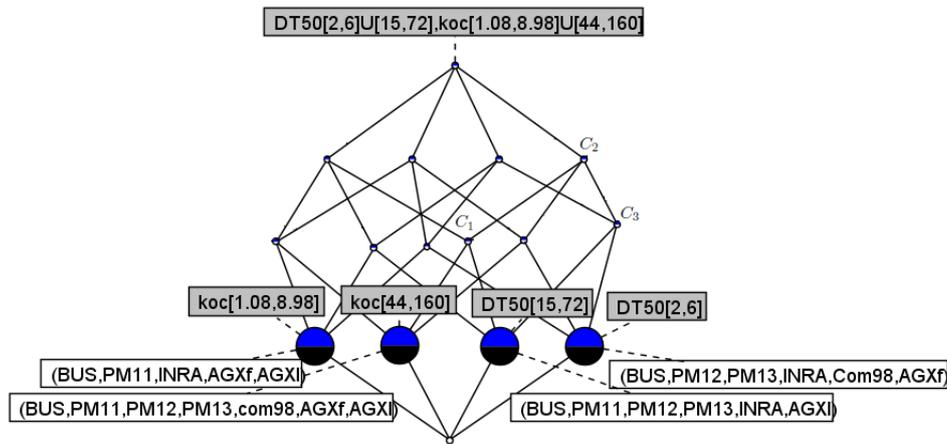


FIG. 7 – Treillis de concepts de la Table 7

l'ensemble $K_1 = \{BUS, PM12, PM13, INRA, Com98, AGXf\}$ et son sur-concept C_3 situé à droite est $(\{K_1, K_2\}, \{(DT50, [2, 6] \cup [15, 72])\})$ où $[2, 6] \cup [15, 72]$ est le résultat de la fusion de $K_1 \cup K_2 = \{BUS, PM11, PM12, PM13, INRA, AGXl, Com98, AGXf\}$.

Résultats et discussion. Le treillis de concepts permet aux utilisateurs de I_{gro} et aux experts de donner plusieurs diagnostics pour l'agriculteur. Le concept \top (qui correspond à la fusion globale calculée plus haut) ne permet pas de décider. Alors l'indicateur I_{gro} peut être calculé en choisissant d'autres valeurs dans le treillis. Par exemple, si on considère les valeurs de $DT50$ dans l'intervalle $[2, 6]$ et koc dans $[44, 160]$, on obtient l'intervalle $[9.97, 10]$ pour I_{gro} et les pratiques de l'agriculteur ne posent pas de problème pour l'environnement, puisque la valeur de I_{gro} est nécessairement plus grande que 7 dans l'intervalle calculé. Cependant, si $DT50 = [15, 72]$ et $koc = [1.08, 8.98]$, l'intervalle résultant pour I_{gro} est $[4.32, 4.32]$ indique que l'agriculteur doit changer ses pratiques puisque la valeur calculée ne dépasse pas le seuil 7. Pour le choix entre plusieurs résultats, l'utilisateur de la fusion devra faire appel à des critères extérieures (comme un classement des sources (Cholvy, 1995)).

De cette façon, nous obtenons, avec ces concepts, des résultats plus précis de I_{gro} , ce qui n'est pas le cas si on choisit le résultat global de la fusion représenté par le concept le plus général. Le treillis de concept permet donc d'identifier les sous-ensembles maximaux de sources associés à des évaluations plus précises.

8 Conclusion

Dans ce papier, nous avons présenté une méthode utilisant l'AFC, en particulier les "structures de patrons", qui permet d'organiser les résultats de la fusion d'informations numériques imparfaites représentées dans le cadre de la théorie des possibilités. Nous avons considéré deux représentations : i) par intervalles et ii) par distributions de possibilité. Nous avons utilisé les opérateurs conjonctifs et disjonctifs pour construire directement un treillis, et nous avons proposé un pré-traitement des données dans le cas des opérateurs non-associatifs.

Cette approche permet d'associer un sous-ensemble de sources avec son résultat de fusion. Quand le résultat global de la fusion appliquée à l'ensemble de toutes les sources est imprécis ou inutile, la méthode permet à l'utilisateur d'identifier un sous-ensemble de sources qui a un résultat de fusion plus précis et utile, et permettant éventuellement de laisser de côté des sources divergentes dans la mesure où elles ne sont pas complètement fiables.

La poursuite de ce travail consistera à considérer d'autres méthodes de fusion existantes dans la construction du treillis et à évaluer la méthode en utilisant des données réelles. De plus, lorsqu'on dispose de l'information concernant la fiabilité des sources, il est important d'en tenir compte pendant la construction du treillis, ainsi que de considérer des relations entre les données fournies par les sources comme par exemple la distance entre les valeurs pour permettre de fusionner et de rendre convexe le résultat obtenu (Kaytoue et al., 2010a).

Références

- Abidi, A. et R. Gonzalez (1992). *Data fusion in robotics and machine intelligence*. San Diego, CA, USA : Academic Press Professional.
- Assaghir, Z., P. Girardin, et A. Napoli (2009). Fuzzy logic approach to represent and propagate imprecision in agri-environmental indicator assessment. In *International Fuzzy Associations Systems, European Society for Fuzzy Logic and Technology (IFSA/EUSFLAT)*, pp. 707–712.
- Assaghir, Z., M. Kaytoue, A. Napoli, et H. Prade (2010a). Managing Information Fusion with Formal Concept Analysis. In *Modeling Decisions for Artificial Intelligence*, pp. 104–115.
- Assaghir, Z., M. Kaytoue, A. Napoli, et H. Prade (2010b). Organisation de la fusion d'information avec l'analyse formelle de concepts - Journées Nationales de l'Intelligence Artificielle Fondamentale (IAF).
- Bloch, I., A. Hunter, A. Ayoun, S. Benferhat, P. Besnard, L. Cholvy, R. Cooke, D. Dubois, et H. Fargier (2001). Fusion : general concepts and characteristics. *International Journal of Intelligent Systems* 16, 1107–1134.
- Chaudron, L., N. Maille, et M. Boyerc (2003). The cube lattice model and its applications. *Applied Artificial Intelligence* 17(3), 207 – 242.
- Cholvy, L. (1995). Automated reasoning with merged contradictory information whose reliability depends on topics. In *European Conferences on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pp. 125–132.
- Cholvy, L. (1998). Reasoning about merged information. In *Belief Change, Volume 3 of Handbook of defeasible Reasoning and Uncertainty management Systems*, pp. 233–263.
- Cholvy, L. et S. Moral (2001). Merging databases : Problems and examples. *Int. J. Intell. Syst.* 16(10), 1193–1221.
- Cooke, R. M. (2001). *Experts in uncertainty*. Oxford University Press.
- Destercke, S., D. Dubois, et E. Chojnacki (2009). Possibilistic information fusion using maximal coherent subsets. *IEEE Transactions on Fuzzy Systems* 17, 79–92.
- Dubois, D., H. Fargier, et H. Prade (2000). Multiple source information fusion : a practical inconsistency tolerant approach. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 1047–1054.

- Dubois, D., J. Lang, et H. Prade (1992). Dealing with multi-source information in possibilistic logic. In *European Conference on Artificial Intelligence*, pp. 38–42.
- Dubois, D. et H. Prade (1985). *Théorie des possibilités*. Paris : Masson.
- Dubois, D. et H. Prade (1994). Possibility theory and data fusion in poorly informed environments. *Control Eng. Practice* 2, 811–823.
- Ganter, B. et S. O. Kuznetsov (2001). Pattern Structures and Their Projections. In *Int. Conf. on Conceptual Structures*, pp. 129–142.
- Gebhardt, J. et R. Kruse (1998). Parallel combination of information sources. In *Belief Change, Volume 3 of Handbook of defeasible Reasoning and Uncertainty managemen Systems*, pp. 393–439.
- Kaytoue, M., Z. Assaghir, N. Messai, et A. Napoli (2010a). Two complementary classification methods for designing a concept lattice from interval data. In *Foundations of Information and Knowledge Systems*, pp. 345–362.
- Kaytoue, M., S. O. Kuznetsov, A. Napoli, et S. Duplessis (2010b). Mining gene expression data with pattern structures in formal concept analysis. *Information Sciences In Press*.
- Kuznetsov, S. O. et S. A. Obiedkov (2002). Comparing Performance of Algorithms for Generating Concept Lattices. *J. Exp. Theor. Artif. Intell.*, 14, 189–216.
- Malouf, R. (2007). Maximal consistent subsets. *Comput. Linguist.* 33(2), 153–160.
- Phan-Luong, V. (2008). A framework for integrating information sources under lattice structure. *Information Fusion* 9(2), 278 – 292.
- Van der Werf, H. et C. Zimmer (1998). An indicator of pesticide environmental impact based on a fuzzy expert system. *Chemosphere* 36(10), 2225–2249.
- Zadeh, L. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets & Systems* 1, 3–28.

Summary

The main problem addressed in this paper is the merging of numerical information provided by several sources (databases, experts...). Fusion results may not be in suitable form for being used in decision analysis. This is generally due to the fact that information sources are heterogeneous. In this paper, we propose the use of Formal Concept Analysis and more specifically pattern structures for organizing the results of fusion methods. This allows us to associate any subset of sources with its information fusion result. Then once chosen a fusion operator, a concept lattice is built. With examples throughout this paper, we show that concept lattice gives an interesting classification of fusion results. When the fusion global result is too imprecise, the method enables the users to identify what maximal subset of sources that would support a more precise and useful result.