HAL
open science

Estimation de la fiabilité des sources des bases de données évidentielles

Mouna Chebbah, Arnaud Martin, Boutheina Ben Yaghlane

To cite this version:

Mouna Chebbah, Arnaud Martin, Boutheina Ben Yaghlane. Estimation de la fiabilité des sources des bases de données évidentielles. Revue des Nouvelles Technologies de l'Information, 2011, E (21), pp.191-208. hal-00657353

HAL Id: hal-00657353

https://hal.science/hal-00657353

Submitted on 7 Jan 2012

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

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

Estimation de la fiabilité des sources des bases de données évidentielles

Mouna Chebbah*, Arnaud Martin**
Boutheina Ben Yaghlane***

*LARODEC, Université de Tunis, ISG Tunis
41 Rue de la Liberté, Cité Bouchoucha 2000 Le Bardo, Tunisie
Mouna.Chebbah@gnet.tn
**UMR 6074 IRISA, Université de Rennes1 / IUT de Lannion
Rue Edouard Branly BP 30219, 22302 Lannion cedex
Arnaud.Martin@univ-rennes1.fr
***LARODEC, Université de 7 Novembre à Carthage, IHEC Carthage
Carthage Présidence 2016, Tunisie
boutheina.yaghlane@ihec.rnu.tn

Résumé. Dans cet article, nous proposons une méthode permettant l'estimation des fiabilités [1] des sources à partir de toutes leurs fonctions de croyance stockées dans des bases de données évidentielles. Nous proposons également d'assurer le même niveau de fiabilité pour toutes ces fonctions de croyance. Les degrés de fiabilité des sources sont utilisés pour affaiblir leurs fonctions de croyance stockées dans des bases de données évidentielles. Cette méthode a été évaluée sur des données radar réelles et a montré une amélioration remarquable des fiabilités après affaiblissement.

1 Introduction

Les bases de données permettent de stocker une grande quantité d'informations qui sont, la plupart du temps, incertaines ou imprécises. Les données imparfaites sont négligées dans les bases de données classiques ce qui constitue une perte d'information puisque toute donnée peut être utile bien qu'elle soit incertaine, imprécise ou même incomplète.

Pour aborder ce problème, des bases de données évidentielles ont été proposées par Hewawasam et al. (2005) et Bach Tobji et al. (2008). Ce type de base de données stocke les données certaines ainsi qu'incertaines, ce qui garantit une bonne modélisation des informations mais engendre l'augmentation de données à stocker ainsi que le nombre des bases de données évidentielles pouvant contenir des informations redondantes.

La fusion d'informations permet, d'une part, de réduire la quantité d'informations disponibles dans les bases de données évidentielles, et d'autre part, d'aider les utilisateurs qu'ils soient humains ou logiciels à la prise de décision en résumant les différentes bases de données

1. Il ne s'agit pas de fiabilité précise mais plutôt de fiabilité estimée à partir des données disponibles sans avoir accès aux données réelles

en une seule facilement interprétable surtout dans les domaines où le nombre de sources est assez important comme par exemple la médecine, la détection de cibles. . . .

Un degré de confiance associé à chaque information incertaine doit alors être défini en vu d'être stocké dans une base de données évidentielle renseignant sur le degré de pertinence de cette information.

Fusionner revient à combiner différentes informations, pouvant être imparfaites, en provenance de diverses sources. Pour ce faire la *théorie des fonctions de croyance*, introduite par Dempster (1967) et Shafer (1976), offre plusieurs avantages. En effet, cette dernière est utilisée pour sa robustesse en terme de représentation et de combinaison de données incertaines et imprécises. La prise en considération de plusieurs sources hétérogènes lors de la combinaison peut induire l'apparition d'un conflit dû à une contradiction au niveau des informations fournies.

L'existence du conflit a favorisé l'apparition de plusieurs méthodes visant à le résoudre dans le cadre de la théorie des fonctions de croyance. Certaines méthodes proposent la résolution du conflit lors de la combinaison à travers l'application des règles adéquates comme celles proposées par Dubois et Prade (1988), Murphy (2000), Smets et Kennes (1994), Yager (1987) et Martin et Osswald (2007b). Ces règles permettent à la fois de combiner et de redistribuer le conflit de manières différentes sur les informations disponibles.

Une autre façon de gérer le conflit est de le réduire avant de combiner en affaiblissant les informations fournies par une source avec son degré de fiabilité. Cette méthode permet de différencier les informations fournies par une source fiable de celles fournies par une source de moindre fiabilité. Cette méthode nécessite une connaissance préalable des degrés de fiabilité des sources. Il est donc nécessaire de pouvoir estimer la fiabilité d'une source. Martin et al. (2008a) proposent une approche sans *a priori* fondée sur les informations fournies par les sources.

Dans cet article, nous proposons une méthode d'estimation de la fiabilité d'une source à partir de toutes les fonctions de masse qu'elle fournit et non pas à partir d'une seule fonction en se référant aux autres sources, avec l'hypothèse qu'au moins la moitié des sources est fiable. Notre méthode est particulièrement applicable sur les bases de données évidentielles du moment où ces dernières permettent de stocker toutes les fonctions de masse fournies par une source. Nous proposons également d'améliorer le niveau de fiabilité des différentes sources ainsi que les fonctions de masse qu'elles fournissent.

Le reste de cet article est organisé comme suit : dans la deuxième section nous présentons brièvement les notions de base de la théorie des fonctions de croyance, ensuite nous définissons les bases de données évidentielles dans la troisième section. La quatrième section présente notre méthode d'estimation du conflit et de la fiabilité d'une source et enfin nous présentons les résultats expérimentaux sur des données radar dans la cinquième et dernière section.

2 La théorie des fonctions de croyance

2.1 Formalisme

La théorie des fonctions de croyance est appelée aussi *théorie de l'évidence* ou encore *théorie de Dempster-Shafer* puisqu'elle est initiée par Dempster (1967) et Shafer (1976). Cette

théorie est un outil robuste pour la représentation des données imparfaites (imprécises et/ou incertaines). Nous présentons ici quelques concepts de base de cette théorie.

Soit $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ un ensemble fini de toutes les hypothèses possibles ω_i pour un problème donné. L'ensemble Ω représente *l'ensemble de discernement* ou *l'univers de discours* du problème en question.

Une *fonction de masse* est définie sur l'ensemble de tous les sous-ensembles possibles de Ω, noté 2^Ω et affecte à chaque sous-ensemble une valeur entre 0 et 1 représentant sa *masse de croyance élémentaire*. Cette dernière permet de représenter les connaissances incertaines et imprécises fournies par un expert (une source, un classifieur, ...). Formellement, une fonction de masse m^Ω est définie comme suit :

$$m^\Omega : 2^\Omega \mapsto [0,1] \tag{1}$$

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1 \tag{2}$$

Un sous-ensemble ayant une masse de croyance non-nulle est *un élément focal*. La masse affectée à un élément focal A représente le degré de croyance élémentaire d'une source à ce que la valeur réelle de l'attribut en question soit incluse ou égale à A.

On impose aussi en général $m^\Omega(\emptyset) = 0$ qui permet de rester en monde fermé. Si $m^\Omega(\emptyset)$ est non nulle, cette masse est interprétée comme le degré de croyance à ce que la valeur recherchée soit une hypothèse non énumérée dans Ω. Dans un monde fermé, Ω est supposé être exhaustif ce qui signifie que toutes les hypothèses possibles sont énumérées dans Ω exigeant ainsi une masse nulle affectée à l'ensemble vide.

Dans un monde ouvert proposé par Smets et Kennes (1994), Ω est supposé être non exhaustif donc une masse non nulle peut être affectée à l'ensemble vide.

La fonction de croyance (ou de crédibilité) bel^Ω représente le degré de croyance minimal affecté à un sous-ensemble de 2^Ω justifié par les informations disponibles. $bel^\Omega(A)$ mesure le degré auquel les informations données par une source ($B \subseteq A$) soutiennent A.

$$\begin{aligned} bel^\Omega : 2^\Omega &\to [0,1] \\ A &\mapsto \sum_{B \subseteq A, B \neq \emptyset} m^\Omega(B) \end{aligned} \tag{3}$$

La fonction de plausibilité pl^Ω représente le degré de croyance alloué aux propositions non contradictoires avec A. C'est le degré de croyance maximal en A.

$$\begin{aligned} pl^\Omega : 2^\Omega &\to [0,1] \\ A &\mapsto \sum_{A \cap B \neq \emptyset} m^\Omega(B) \end{aligned} \tag{4}$$

La fonction de plausibilité est la fonction duale de la fonction de crédibilité puisque :

$$pl^\Omega(A) = 1 - bel^\Omega(\bar{A}) \tag{5}$$

2.2 Les règles de combinaison

La théorie des fonctions de croyance représente un outil robuste pour la combinaison de différentes informations, représentées par des fonctions de masse définies sur le même ensemble de discernement, et fournies par différentes sources. La combinaison permet de confronter ces différentes informations (avis d'experts, ...) afin d'en obtenir une seule en vue d'une prise de décision.

Il existe un grand nombre de règles de combinaison mais dans cet article nous ne présentons que celles utilisées pour l'expérimentation de notre méthode.

La règle conjonctive de combinaison (notée CRC), présentée dans le modèle de croyances transférables de Smets et Kennes (1994), permet de combiner deux fonctions de masse distinctes m_1^Ω et m_2^Ω. Elle est définie comme suit :

$$m_{1\textcircled{\cap}2}^\Omega(A) = (m_1^\Omega \textcircled{\cap} m_2^\Omega)(A) = \sum_{B\cap C=A} m_1^\Omega(B) \times m_2^\Omega(C) \tag{6}$$

Cette règle est non normalisée du moment qu'elle autorise l'affectation d'une masse nulle à l'ensemble vide après la combinaison. Elle est applicable sous l'hypothèse du monde ouvert dans lequel la masse attribuée à l'ensemble vide représente le degré de croyance en une hypothèse inconnue et non énumérée dans Ω. Une normalisation de cette règle est présentée par Dempster (1967).

La fonction de masse $m_{1\oplus 2}^\Omega$ résultat de la combinaison de m_1^Ω et m_2^Ω par la règle de combinaison de Dempster est obtenue comme suit :

$$m_{1\oplus 2}^\Omega(A) = (m_1^\Omega \oplus m_2^\Omega)(A) = \begin{cases} \frac{m_{1\textcircled{\cap}2}^\Omega(A)}{1-m_{1\textcircled{\cap}2}^\Omega(\emptyset)} & \forall A \subseteq \Omega, \text{ si } A \neq \emptyset \\ 0 & \text{si } A = \emptyset \end{cases} \tag{7}$$

Cette règle est normalisée et elle est applicable avec l'hypothèse du monde fermé dans lequel on suppose que toutes les valeurs possibles que peut prendre un élément focal sont énumérées dans l'ensemble de discernement.

Yager (1987) a interprété $m^\Omega(\emptyset)$ comme un degré d'ignorance totale donc il l'a affecté à Ω. La règle de combinaison de Yager est définie pour deux fonctions de masse m_1^Ω et m_2^Ω comme suit :

$$\begin{cases} m_Y^\Omega(A) = m_{1\textcircled{\cap}2}^\Omega(A) & \forall A \in 2^\Omega,\ A \neq \Omega \text{ et } A \neq \emptyset \\ m_Y^\Omega(\Omega) = m_{1\textcircled{\cap}2}^\Omega(\Omega) + m_{1\textcircled{\cap}2}^\Omega(\emptyset) \\ m_Y^\Omega(\emptyset) = 0 \end{cases} \tag{8}$$

La solution de Dubois et Prade (1988) attribue la masse résultante de la combinaison de deux éléments focaux conflictuels (dont l'intersection est vide) à l'union de ces éléments. Cette règle est également définie pour deux fonctions de masse m_1^Ω et m_2^Ω comme suit :

$$\begin{cases} m_{DP}^\Omega(A) = m_{1\textcircled{\cap}2}^\Omega(A) + \sum_{B\cap C=\emptyset,\ B\cup C=A} m_1^\Omega(C) m_2^\Omega(B) & \forall A \subseteq \Omega,\ A \neq \emptyset \\ m_{DP}^\Omega(\emptyset) = 0 \end{cases} \tag{9}$$

Finalement, la règle de combinaison de Murphy (2000) consiste à prendre la moyenne des éléments focaux des deux fonctions de masse à combiner. Pour chaque élément focal A des N

fonctions de masse à combiner, la fonction de masse combinée est définie comme suit :

$$m^{\Omega}_{Murphy}(A) = \frac{1}{N}\sum_{k=1}^{N} m^{\Omega}_{k}(A) \tag{10}$$

Pour résumer, le tableau 1 présente ces règles, leurs hypothèses ainsi que leurs méthodes de distribution de $m^{\Omega}(\emptyset)$.

Règles de combinaison	**Hypothèse**	**Méthodes de redistribution de $\mathbf{m}(\emptyset)$**
Règle conjonctive de combinaison de Smets	Hypothèse du monde ouvert	$m^{\Omega}(\emptyset)$ n'est pas redistribuée
Règle de Dempster	Hypothèse du monde fermé	$m^{\Omega}(\emptyset)$ est redistribuée proportionnellement sur les éléments focaux
Règle de Yager	Hypothèse du monde fermé	$m^{\Omega}(\emptyset)$ est affectée à Ω
Règle de Dubois et Prade	Hypothèse du monde fermé	La masse résultante de la combinaison des éléments focaux conflictuels est attribuée à l'union de ces éléments
Règle de Murphy	Hypothèse du monde fermé et hypothèse du monde ouvert	Si au moins l'une des fonctions de masse est non normalisée alors $m^{\Omega}(\emptyset) \neq 0$ et cette masse n'est pas redistribuée

TAB. 1 – *Les méthodes de redistribution de* $m^{\Omega}(\emptyset)$

2.3 L'affaiblissement

La majorité des règles de combinaison ne font pas la distinction entre le conflit existant entre les sources et l'auto-conflit dû à la non-idempotence de la règle de combinaison utilisée, la notion d'auto-conflit a été introduite par Martin et al. (2008a). Une des origines du conflit est la non fiabilité d'au moins une des sources. La non fiabilité d'une source peut être réglée par l'affaiblissement des fonctions de masse avant la combinaison utilisant l'opérateur d'affaiblissement proposé par Shafer (1976). Du moment qu'on arrive à quantifier la fiabilité α de chaque source, on peut affaiblir les fonctions de masse associées comme suit :

$$\begin{cases} m^{\Omega,\alpha}(A) = \alpha\, m^{\Omega}(A) & \forall A \subset \Omega \\ m^{\Omega,\alpha}(\Omega) = (1-\alpha) + \alpha\, m^{\Omega}(\Omega) & \end{cases} \tag{11}$$

avec α le degré de fiabilité de la source et $1-\alpha$ le facteur d'affaiblissement.

Cet opérateur permet d'intégrer les degrés de fiabilité des sources au sein des fonctions de masse qu'elles fournissent. Cet opérateur atténue les masses sans toucher à l'ensemble des éléments focaux ce qui n'affecte pas la prise de décision puisque l'hypothèse la plus probable reste ainsi même après affaiblissement.

Il existe d'autres méthodes d'affaiblissement telle que la méthode d'affaiblissement séquentiel proposée par Schubert (2008) où un seuil de conflit maximal est initialement fixé. À chaque étape les conflits sont calculés et utilisés pour l'affaiblissement des fonctions de masse jusqu'à garantir un degré de conflit inférieur ou égal à ce seuil.

Cette méthode garantit un degré de conflit tolérable mais la complexité temporelle est assez élevée du moment que le nombre d'itérations ne peut pas être fixé dès le départ.

Zeng et Wu (2007) ont proposé un opérateur d'affaiblissement des fonctions de plausibilité permettant également l'affaiblissement des fonctions de masse par l'intermédiaire de leurs fonctions de plausibilité ainsi que la modification de l'ensemble des éléments focaux donc affecte indirectement la prise de décision.

Enfin, Mercier et al. (2005) ont considéré que la fiabilité d'une source dépend de la valeur réelle de la variable ou l'attribut en question, donc ils ont proposé d'utiliser un vecteur de fiabilité représentant la fiabilité de la source en fonction de chaque hypothèse de l'ensemble de discernement ainsi qu'un opérateur d'affaiblissement utilisant ce vecteur.

Une fois que la fiabilité d'une source est connue, elle est utilisée pour affaiblir toutes les fonctions de masse fournies par cette source avant leur combinaison avec d'autres fonctions de masse. Toutes ces fonctions de masse sont généralement stockées dans des bases de données évidentielles.

3 Les bases de données évidentielles

Une base de données évidentielle est une base de données qui contient des données parfaites, imparfaites ou même des données manquantes. L'imperfection (incertitude et/ou imprécision) dans les bases de données évidentielles est représentée grâce à la théorie des fonctions de croyance précédemment décrite.

Formellement, une base de données évidentielle est une base de données ayant X attributs (colonnes) et Y enregistrements (lignes), chaque attribut j ($1 \leq j \leq X$) possède un domaine D_j représentant toutes les valeurs de cet attribut : *C'est son ensemble de discernement* tel que défini par Bach Tobji et al. (2008). Une base de données évidentielle doit contenir au moins un attribut évidentiel qui prendra des valeurs évidentielles décrites par une fonction de masse au lieu d'une valeur certaine et précise. Une valeur évidentielle V_{ij} de l'enregistrement i ($1 \leq i \leq Y$) pour l'attribut j ($1 \leq j \leq X$) est une fonction de masse $m_{ij}^{D_j}$ telle que :

$$\begin{aligned} & m_{ij}^{D_j} : 2^{D_j} \rightarrow [0,1] \text{ avec :} \\ & m_{ij}^{D_j}(\emptyset) = 0 \text{ et } \sum_{x \subseteq D_j} m_{ij}^{D_j}(x) = 1 \end{aligned} \tag{12}$$

Les bases de données évidentielles sont utilisées dans différents domaines notamment pour le stockage des fonctions de masse de différents classifieurs présenté par Hewawasam et al. (2005).

Le tableau 2 est un exemple d'une table d'une base de données évidentielle contenant les cibles détectées par un capteur. L'attribut cible est un attribut évidentiel dont l'ensemble de discernement $\Omega = \{Avion\ (A),\ Helicopter\ (H),\ Missile\ (M)\}$.

Les informations stockées dans les bases de données évidentielles peuvent être :

- *des informations parfaites :* Une information parfaite est caractérisée par un seul élément focal qui est un singleton.
- *des informations probabilistes :* Lorsque tous les éléments focaux sont des singletons, la fonction de masse associée correspond à une distribution de probabilité. La valeur de l'attribut cible pour le deuxième enregistrement du tableau 2 est une information probabiliste.

id	Capteur (source)	Vue	Cible
1	S_1	V_1	$\{A\}(0.2)$ $\{A \cup H\}(0.6)$ $\{A \cup H \cup M\}(0.2)$
2	S_2	V_1	$\{A\}(0.5)$ $\{H\}(0.5)$
3	S_3	V_1	$\{A\}(0.5)$ $\{M\}(0.2)$ $\{A \cup M\}(0.3)$

TAB. 2 – *Exemple d'une base de données évidentielle*

- *des informations possibilistes :* La première ligne du tableau 2 représente une information possibiliste puisque les éléments focaux sont emboîtés.
- *des informations manquantes :* Une fonction de masse vide peut être interprétée comme un état d'ignorance totale ou d'information inexistante.
- *des informations évidentielles :* Des masses sont affectées à un ou plusieurs éléments focaux non emboîtés et non singletons. Le troisième enregistrement du tableau 2 représente une information évidentielle.

4 Estimation du conflit et de la fiabilité d'une source

Une base de données évidentielle stocke différentes fonctions de masse fournies par une source. Avec la présence de plusieurs sources, experts ou classifieurs il y aura autant de bases de données évidentielles que de sources donc une quantité énorme de données à traiter. L'intégration des différentes bases de données en une seule permet de réduire la quantité d'informations disponibles afin de faciliter les requêtes sur la base de données et la prise de décision éventuelle. Lors de l'intégration de différentes bases de données évidentielles, le principal problème rencontré réside dans l'intégration des valeurs évidentielles surtout quand elles sont conflictuelles.

Dans cet article, nous nous concentrons sur la fusion de plusieurs fonctions de masse stockées dans différentes bases de données. Nous proposons aussi d'enrichir ces bases de données par des informations sur le niveau de fiabilité des sources et de fiabilité des combinaisons. Ces fiabilités seront des indicateurs importants sur le degré de confiance des résultats des requêtes effectuées sur la base de données. L'utilisation d'une règle de combinaison paraît une solution simple permettant à la fois de combiner plusieurs fonctions de masse et de résoudre le conflit. Ainsi, la résolution du conflit se fait par la redistribution de la masse affectée à l'ensemble vide sur les éléments focaux de différentes manières, selon la règle utilisée, sans prendre en considération les fiabilités des sources et le degré de véracité des fonctions de masse fournies.

Le degré de fiabilité de la source doit être pris en considération avant la combinaison pour prévenir le conflit au maximum. La difficulté réside dans la quantification de la fiabilité d'une source afin d'en tenir compte avant la combinaison en affaiblissant ses fonctions de masse. Dans une base de données évidentielle, plusieurs fonctions de masse relatives à une source y

sont stockées, d'où la nécessité d'attribuer un degré de fiabilité global à la source qui prend en considération toutes les fonctions de masse qu'elle fournit.

Il existe des méthodes qui visent à estimer la fiabilité d'une source ou le facteur d'affaiblissement. Elouedi et al. (2004) ont proposé de trouver le facteur d'affaiblissement minimisant la distance entre les probabilités pignistiques calculées à partir des fonctions de masse affaiblies et les valeurs actuelles des données. Cette méthode nécessite une connaissance préalable des valeurs actuelles des données ce qui n'est pas toujours le cas, par exemple dans le domaine de détection de cible, les classes réelles ne peuvent pas être connues à l'avance et dans ce cas là notre méthode pourra être adéquate. La fiabilité d'une source est estimée ou calculée en se référant aux données réelles. Dans cet article, les données réelles ne sont pas disponibles donc nous utilisons une fiabilité approchée qui est estimée en se référant aux données fournies par d'autres sources.

4.1 Mesure du conflit

Martin et al. (2008b) considèrent que plus deux fonctions de masse sont éloignées plus les sources sont en conflit. Ainsi, une mesure de distance entre différentes fonctions de masse permet de quantifier le conflit entre leurs sources.

La distance de Jousselme et al. (2001) est utilisée dans cet article parce qu'elle permet de tenir compte des spécificités des fonctions de croyance puisque cette distance utilise le coefficient de Jaccard $\frac{|A\cap B|}{|A\cup B|}$ ce qui permet de tenir compte les cardinalités des éléments focaux. Une matrice D est aussi définie sur l'ensemble 2^Ω ce qui rend cette distance spécifique aux fonctions de croyance. La distance de Jousselme est donnée par :

$$d(m_1^\Omega, m_2^\Omega) = \sqrt{\frac{1}{2}(m_1^\Omega - m_2^\Omega)^t D(m_1^\Omega - m_2^\Omega)} \tag{13}$$

avec :

$$D(A,B) = \begin{cases} 1 & \text{si } A = B = \emptyset \\ \frac{|A\cap B|}{|A\cup B|} & \forall A, B \in 2^\Omega \end{cases} \tag{14}$$

D'autres distances, résumées par Florea et Bossé (2009), peuvent être également utilisées.

Le conflit entre deux sources S_1 et S_2 n'est autre que la distance séparant leurs fonctions de masse m_1^Ω et m_2^Ω :

$$Conf(S_1, S_2) = d(m_1^\Omega, m_2^\Omega) \tag{15}$$

Avec la présence de plusieurs sources, la mesure de conflit correspond à la distance entre la fonction de masse fournie par une source donnée et les autres fonctions de masse. Cette mesure de conflit peut être calculée de deux manières différentes. La première méthode consiste à calculer la moyenne des distances entre la fonction de masse de la source S_j en question avec les $s-1$ autres fonctions de masse ϵ :

$$Conf(j, \epsilon) = \frac{1}{s-1} \sum_{j=1, j'\neq j}^{s} Conf(j, j') \tag{16}$$

La seconde correspond au calcul de la distance entre la fonction de masse fournie par la source S_j et la fonction de masse m_{s-1}^Ω résultat de la combinaison des fonctions de masse des

$s-1$ différentes sources autre que la source en question :

$$Conf(j,s) = d(m_j^{\Omega}, m_{s-1}^{\Omega}) \tag{17}$$

Ces mesures de conflit ont les mêmes propriétés que la distance de Jousselme. La distance entre deux fonctions de masse reflète le conflit entre leurs sources. Un conflit égal à 0 implique un accord total entre les sources justifié par deux fonctions de masse identiques. Un conflit égal à 1 reflète un conflit total montrant que les deux sources ont fourni deux fonctions de masse complètement différentes.

Schubert (1996) utilise le conflit pour regrouper les sources en sous ensembles s'exprimant sur les mêmes éléments focaux. La combinaison ne se fait qu'aux sources d'un même sous ensemble sans quantifier leurs fiabilités.

4.2 Estimation de la fiabilité relative d'une source

L'affaiblissement peut être utilisé afin de prendre en considération la non-fiabilité relative d'une source avant la combinaison pour éliminer ou réduire le conflit qui pourra apparaître après. Nous faisons ici l'hypothèse que le conflit est issu de la non-fiabilité des sources ainsi que la majorité (plus que la moitié) des sources sont fiables.

Pour pouvoir affaiblir une fonction de masse, l'estimation de la fiabilité est nécessaire. La méthode d'estimation de fiabilité proposée par Martin et al. (2008b) est fondée sur la mesure du conflit. La *fiabilité relative* α_j d'une source S_j est une fonction décroissante de son conflit avec les autres sources telle que :

$$\alpha_j = (1 - Conf(j,s)^{\lambda})^{\frac{1}{\lambda}} \tag{18}$$

avec λ un réel quelconque.

4.3 Estimation de la fiabilité absolue d'une source

Pour s sources, l'estimation de la fiabilité relative revient à calculer pour chacune le conflit de sa fonction de masse par rapport aux autres qui servira à estimer sa fiabilité, ce degré de fiabilité est utilisé pour affaiblir la fonction de masse correspondante avant la combinaison.

Dans une base de données évidentielle, différentes fonctions de masse fournies par la même source sont stockées. Il faut alors tenir compte de toutes les fonctions de masse pour estimer la fiabilité absolue de la source que nous introduisons ici. La fiabilité relative ne prend en compte qu'une seule fonction de masse alors que la *fiabilité absolue* reflète le niveau général de fiabilité de toutes les fonctions de masse fournies par une source.

Ces informations concernant les fiabilités relatives et la fiabilité absolue d'une source pourront servir à l'enrichissement de la base de données évidentielle en indiquant le niveau de fiabilité de chaque source.

À partir de s bases de données évidentielles concernant s sources, chacune ayant X attributs évidentiels et Y enregistrements, chaque base de données évidentielle stocke $X \times Y$ différentes fonctions de masse par source. À partir de chaque fonction de masse, l'estimation des fiabilités relatives pourra être faite ; il y aura donc $X \times Y$ degrés de fiabilité par source.

Dans cet article nous proposons de prendre la *moyenne* des différentes fiabilités relatives comme fiabilité absolue de la source. Le choix de la moyenne se justifie par le fait que la

fiabilité d'une source est fixe, bien qu'elle puisse se tromper parfois donc sa fiabilité peut augmenter ou diminuer légèrement mais en moyenne elle garde le même niveau.

Dans cet article, nous ne traitons pas les cas où la fiabilité d'une source varie selon les hypothèses de 2^{Ω}.

En effet, affaiblir avec une fiabilité relative minimale réduit la fonction de masse au maximum ce qui peut mener à l'ignorance totale. De plus, l'affaiblissement avec une fiabilité relative maximale ne permettra pas dans certains cas de réduire le conflit du moment où une source peut être au moins une fois complètement fiable alors qu'elle ne l'est pas en général. Affaiblir avec la fiabilité moyenne permet de réduire le conflit tout en gardant l'intégrité de la fonction de masse. Le choix de la fiabilité moyenne permettra ainsi d'éviter les valeurs aberrantes.

La *fiabilité absolue* α_j^a d'une source S_j est donc la moyenne de ses Y fiabilités relatives α_{yj} :

$$\alpha_j^a = \frac{1}{Y} \sum_{y=1}^{Y} (\alpha_{yj}) \tag{19}$$

Régis et al. (2007) proposent une méthode pour déterminer si une source est pertinente ou pas sans quantifier son degré de pertinence.

4.4 Estimation de la fiabilité de la combinaison

Un degré de fiabilité relative α_j est attribué à chaque fonction de masse, qui pourra servir à estimer la fiabilité de la combinaison.

Si on a s sources fournissant chacune une fonction de masse différente, l'estimation de la fiabilité de chaque source se fera en calculant la distance entre sa fonction de masse et le reste des fonctions ce qui représente son conflit relatif. La fiabilité relative est calculée à partir de ce degré de conflit relatif. La fiabilité de la combinaison est la moyenne des s fiabilités relatives propres aux différentes sources concernant les fonctions de masse à combiner pour une observation donnée. La fiabilité d'une combinaison représente le degré de confiance moyen attribué à la fonction de masse résultante.

$$\alpha_c = \frac{1}{s} \times \sum_{j=1}^{s} (\alpha_j) \tag{20}$$

Ces degrés de fiabilité associés aux combinaisons pourront enrichir la base de données évidentielle en indiquant à l'utilisateur à quel point il pourra faire confiance à la fonction de masse fournie. Ces fiabilités ne sont utiles que pour l'utilisateur lors de la prise de décision.

Il est à noter que les équations (19) et (20) sont différentes : la première équation permet d'obtenir la fiabilité moyenne d'une source à partir de toutes ses fiabilités relatives correspondant à ses fonctions de masse, tandis que la seconde équation permet d'obtenir la fiabilité de la combinaison à partir des fiabilités relatives de toutes les fonctions de masse des différentes sources combinées.

5 Expérimentation

Afin de pouvoir tester la méthode précédemment décrite, nous avons considéré une base de données radar. Ces données ont été recueillies dans la chambre anéchoïque de l'ENSIETA

en plaçant une cible (maquette d'avion) et un capteur radar pouvant détecter la cible sous différents points angulaires. Le système d'acquisition est présenté par Martin et Radoi (2004). Une base de données a été proposée pour l'acquisition et le stockage des signaux par Toumi (2007). Nous considérons ainsi cinq cibles radar différentes (Mirage, F14, Rafale, Tornado, Harrier). Chaque table contient 250 représentations fréquentielles obtenues dans un domaine angulaire d'environ $60°$ et utilisant une bande de fréquence d'environ 6 GHZ. Pour caractériser les cibles, et donc renseigner la base de données, nous avons utilisé trois classifieurs différents : le k-plus proche voisin flou, le k-plus proche voisin crédibiliste et un réseau de neurones. Ces trois classifieurs sont considérés comme des sources, sur lesquelles on déduit des fonctions de masse tel que présenté par Martin et Radoi (2004). Ils ont donc fourni 250 fonctions de masse stockées dans trois tables différentes et permettant de classifier les cinq cibles radar différentes.

Notre but est d'intégrer ces trois tables en combinant les 250 fonctions de masse fournies par chaque source (classifieur) pour obtenir une seule table facilitant les requêtes sur la base de données et aider ainsi à la prise de décision.

La première étape consiste à estimer les conflits relatifs à chaque source pour chaque fonction de masse, donc chaque source aura 250 conflits relatifs. Le conflit absolu d'une source n'est autre que la moyenne de ses 250 conflits relatifs. Afin de calculer les conflits relatifs, nous avons utilisé deux types de méthode de calcul de distance :

- **Distance type1 :** correspondant au calcul du conflit donné par l'équation (16), *i.e.* à la moyenne des distances séparant une fonction de masse fournie par une source et les autres fonctions de masse sans utiliser une règle de combinaison.
- **Distance type2 :** correspondant au calcul du conflit donné par l'équation (17), *i.e.* à la distance séparant la fonction de masse fournie par une source et la fonction de masse combinée relative aux autres sources. Il existe plusieurs règles de combinaison pouvant être utilisées pour la combinaison des fonctions de masse telles que rappelées par Smets (2007) et Martin et Osswald (2007a), mais dans cet article nous avons utilisé uniquement celles décrites dans la Section 2.2.

Le conflit absolu initial propre à chaque source est donné dans le tableau 3, cette valeur de conflit est la moyenne des conflits relatifs. Chebbah et al. (2010b) présentent une modélisation détaillée du conflit dans les bases de données évidentielles.

Nous avons utilisé les conflits absolus initiaux pour estimer les fiabilités absolues initiales afin de s'en servir pour affaiblir les fonctions de masse des différentes sources avec différentes valeurs de λ (le paramètre de calcul de la fiabilité à partir du conflit de l'équation (18)). Après affaiblissement, nous avons recalculé les fiabilités absolues pour chaque valeur de λ. Disposant des valeurs des fiabilités absolues initiales et après affaiblissement en fonction de λ, les taux d'amélioration des fiabilités absolues ainsi que les taux de diminution des variances des fiabilités relatives sont présentés dans les figures 1, 2 et 3. Les figures 1, 2 et 3 représentent les taux d'amélioration des fiabilités absolues ainsi que les taux de diminution des variances des fiabilités relatives du k-plus proche voisin crédibiliste, k-plus proche voisin flou et réseau de neurones.

Nous remarquons que dans certains cas, le taux d'amélioration est négatif signifiant que la fiabilité absolue après affaiblissement est inférieure à la fiabilité absolue initiale ce qui n'est pas acceptable puisque l'affaiblissement améliore la fiabilité d'une source.

Le but de l'affaiblissement n'est pas uniquement d'améliorer la fiabilité des sources, mais de diminuer aussi la contribution d'une source lorsqu'elle n'est pas très fiable. En diminuant

Source	Type de distance	Règle de combinaison	Conflit Initial
K-plus proche voisin flou	Type 1	-	0.196
	Type 2	Règle conjonctive	0.326
		Règle de Dempster	0.125
		Moyenne	0.148
		Règle de Dubois et Prade	0.201
		Règle de Yager	0.262
K-plus proche voisin crédibiliste	Type 1	-	0.223
	Type 2	Règle conjonctive	0.371
		Règle de Dempster	0.086
		Moyenne	0.214
		Règle de Dubois et Prade	0.238
		Règle de Yager	0.331
Réseau de neurones	Type 1	-	0.301
	Type 2	Règle conjonctive	0.331
		Règle de Dempster	0.334
		Moyenne	0.292
		Règle de Dubois et Prade	0.319
		Règle de Yager	0.316

TAB. 3 – *Conflits absolus initiaux des sources*

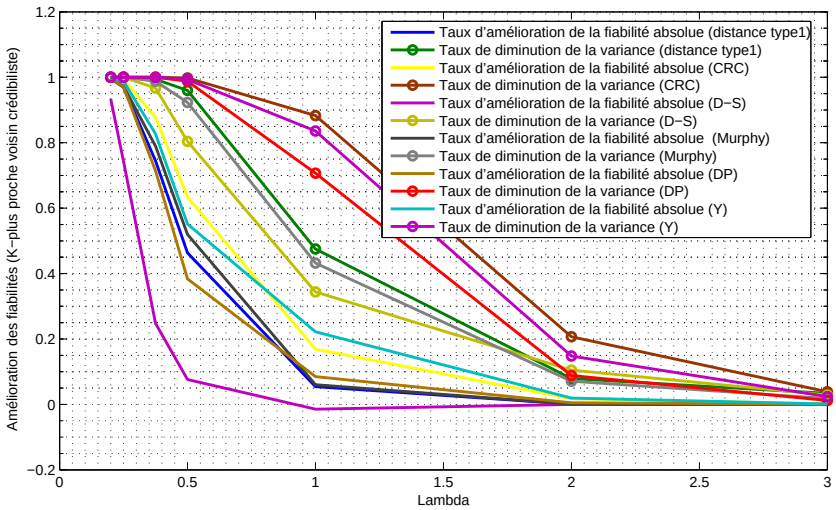

FIG. 1 – *Taux d'amélioration des fiabilités absolues et taux de diminution des variances des fiabilités relatives en fonction de λ pour le k-plus proche voisin crédibiliste*

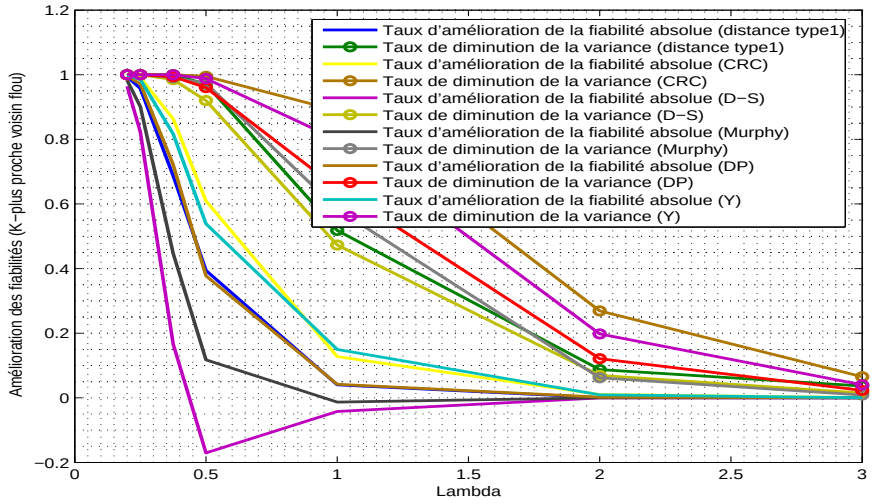

FIG. 2 – *Taux d'amélioration des fiabilités absolues et taux de diminution des variances des fiabilités relatives en fonction de λ pour le k-plus proche voisin flou*

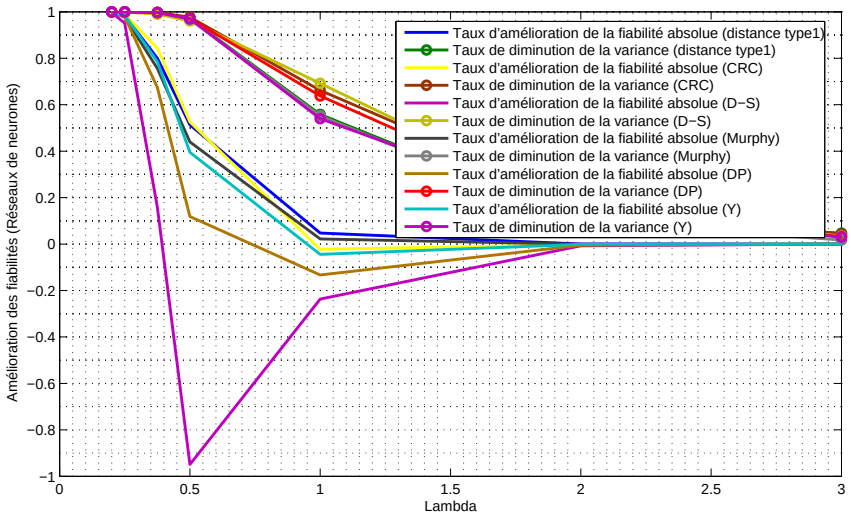

FIG. 3 – *Taux d'amélioration des fiabilités absolues et taux de diminution des variances des fiabilités relatives en fonction de λ pour un réseau de neurones*

la variance des fiabilités relatives, les cas où la source se trompe en étant plus ou moins fiable que d'habitude sont corrigés. L'utilisateur n'aura plus à se soucier des erreurs pouvant être commises par la source, il pourra plutôt utiliser l'information telle qu'elle est tout en tenant compte de la fiabilité absolue de la source. La diminution de la variance des fiabilités relatives indique une amélioration des fiabilités relatives.

Le choix de λ est conditionné par un taux d'amélioration positif puisque l'affaiblissement améliore la fiabilité. D'après les figures 2 et 3, pour $\lambda \leq 0.375$ les taux d'amélioration des fiabilités absolues sont positifs pour le réseau de neurones et le k-plus proche voisin flou quelque soit le type de distance et la règle de combinaison utilisée. D'après la figure 1, $\lambda \leq 0.5$ garantit un taux d'amélioration positif des fiabilités absolues pour le k-plus proche voisin crédibiliste quelque soit le type de distance utilisé.

La fiabilité est une fonction croissante de λ, plus λ est petite plus l'affaiblissement est important, d'où la valeur de λ doit être la plus grande possible pour affaiblir les fonctions de masse au minimum. La valeur $\lambda = 0.375$ garantit un taux de diminution positif pour le réseau de neurones et le k-plus proche voisin flou indépendamment du type de distance et de la règle de combinaison tout en affaiblissant le moins possible. La valeur $\lambda = 0.5$ est la valeur optimale pour le k-plus proche voisin crédibiliste pour améliorer la fiabilité absolue tout en affaiblissant le moins possible.

Les fiabilités absolues propres à chaque source avant et après affaiblissement ainsi que les pourcentages d'amélioration de la variance des fiabilités relatives associées pour les valeurs retenues de λ sont donnés dans le tableau 4.

Source	λ	**Type de distance**	**Règle de combinaison**	**Fiabilité initiale**	**Fiabilité après affaiblissement**	**Pourcentage de diminution de la variance**
K-plus proche voisin flou	0.375	Type 1	-	0.124	0.397	100%
		Type 2	Règle conjonctive	0.058	0.422	100%
			Règle de Dempster	0.195	0.234	98.5%
			Moyenne	0.168	0.303	99.1%
			Règle de D. et P.	0.121	0.433	99.4%
			Règle de Yager	0.084	0.456	100%
K-plus proche voisin crédibiliste	0.5	Type 1	-	0.278	0.518	96%
		Type 2	Règle conjonctive	0.153	0.417	99.8%
			Règle de Dempster	0.5	0.541	80.4%
			Moyenne	0.289	0.602	92.3%
			Règle de D. et P.	0.262	0.425	98.8%
			Règle de Yager	0.18	0.401	99.3%
Réseau de neurones	0.375	Type 1	-	0.067	0.333	99.7%
		Type 2	Règle conjonctive	0.056	0.357	99.7%
			Règle de Dempster	0.055	0.065	99%
			Moyenne	0.07	0.288	99.7%
			Règle de D. et P.	0.06	0.184	99.7%
			Règle de Yager	0.061	0.285	99.7%

TAB. 4 – *Fiabilités absolues et variances des fiabilités relatives après affaiblissement*

Nous remarquons bien l'amélioration au niveau des fiabilités après affaiblissement ainsi qu'au niveau des variances des fiabilités relatives. Prenons l'exemple de l'utilisation de la distance type 2 avec la moyenne comme règle de combinaison, la fiabilité initiale du réseau de neurones était de 0.07 avec une variance des fiabilités relatives de 0.292. Après affaiblissement, la fiabilité absolue est passée à 0.288 et la variance des fiabilités relatives est réduite à un taux de 0.0001.

D'après le tableau 4, il est clair que pour les valeurs retenues de λ, les fiabilités après affaiblissement sont supérieures aux fiabilités initiales ce qui signifie une amélioration considérable des fiabilités absolues. La plus grande amélioration réside dans la réduction des variances des fiabilités relatives du moment où les pourcentages d'affaiblissement sont presque égaux à 100%. Après affaiblissement, toutes les fonctions de masse ont le même niveau de fiabilités relatives qui correspond à la fiabilité absolue de leur source.

D'après les figures 1, 2 et 3, nous remarquons également que les taux d'amélioration de la fiabilité sont des fonctions croissantes de λ qui tendent vers 1, ce qui signifie qu'en choisissant des valeurs grandes de λ, les valeurs de fiabilité après affaiblissement tendent vers les valeurs de fiabilité initiale. En augmentant la valeur de λ, les fiabilités initiales et les fiabilités après affaiblissement se rejoignent pour atteindre un taux d'amélioration égal à 0.

Notons que les valeurs de fiabilité inférieures à 0.5 ne sont pas contradictoires avec l'hypothèse de la fiabilité d'au moins la moitié des sources du moment que ces valeurs de fiabilité sont calculées à partir d'une distance et les sources sont fiables en termes de taux de bonne reconnaissance.

Chebbah et al. (2010a) ont utilisé cette méthode d'estimation de la fiabilité d'une source avec un affaiblissement des *fonctions de plausibilité*, ce qui a donné de bons résultats en terme d'amélioration des fiabilités absolues et relatives.

6 Conclusion

Dans cet article, nous avons proposé une méthode permettant d'estimer la fiabilité d'une source à partir de toutes les fonctions de masse qu'elle fournit. Cette méthode a pour objectif d'utiliser les fiabilités estimées pour affaiblir les fonctions de masse stockées dans une base de données évidentielle afin de fusionner ces fonctions avec d'autres stockées dans différentes bases de données évidentielles. Cette fusion permettra d'aider l'utilisateur à la prise de décision en réduisant la quantité d'informations à traiter et en lui indiquant les degrés de fiabilité des sources et des informations combinées.

Comme perspective à ce travail, la proposition d'une méthode de transfert de masse autre que l'affaiblissement des masses permettrait de modifier l'ensemble des éléments focaux ce qui pourrait améliorer la qualité de classification des fonctions de masse, résultats de la combinaison.

Remerciements

Ce travail a été partiellement financé par l'Association Internationale Francophone d'Extraction et de Gestion des Connaissances (EGC), les laboratoires de recherche E3I2 (Brest, France) et LARODEC-ISG (Tunisie).

Références

Bach Tobji, M.-A., B. Ben Yaghlane, et K. Mellouli (2008). A new algorithm for mining frequent itemsets from evidential databases. In *Information Processing and Management of Uncertainty (IPMU'2008)*, Malaga, Spain, pp. 1535–1542.

Chebbah, M., B. Ben Yaghlane, et A. Martin (2010a). Reliability estimation based on conflict for evidential database enrichment. In *Workshop on the theory of belief functions*, Brest, France.

Chebbah, M., A. Martin, et B. Ben Yaghlane (2010b). Modélisation du conflit dans les bases de données évidentielles. In *Atelier EGC'2010 " Fouille de données complexes : compléxité liée aux données multiples"*, Hammamet, Tunisia, pp. 13–19.

Dempster, A. P. (1967). Upper and Lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics 38*, 325–339.

Dubois, D. et H. Prade (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence 4*, 244–264.

Elouedi, Z., K. Mellouli, et P. Smets (2004). Assessing Sensor Reliability for Multisensor Data Fusion Within The Transferable Belief Model. *IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics 34*(1), 782–787.

Florea, M. C. et E. Bossé (2009). Crisis management using Dempster Shafer theory : Using dissimilarity measures to characterize sources' reliability. In *C3I for Crisis, Emergency and Consequence Management*, Bucharest, Roumania.

Hewawasam, K., K. Premaratne, S. Subasingha, et M.-L. Shyu (2005). Rule mining and classification in imperfect databases. In *International Conference on Information Fusion*, Philadelphia, USA, pp. 661–668.

Jousselme, A.-L., D. Grenier, et E. Bossé (2001). A new distance between two bodies of evidence. *Information Fusion 2*, 91–101.

Martin, A., A.-L. Jousselme, et C. Osswald (2008a). Conflict measure for the discounting operation on belief functions. In *International Conference on Information Fusion*, Cologne, Germany, pp. 1003–1010.

Martin, A. et C. Osswald (2007a). Toward a combination rule to deal with partial conflict and specificity in belief functions theory. In *International Conference on Information Fusion*, Québec, Canada.

Martin, A. et C. Osswald (2007b). Une nouvelle règle de combinaison répartissant le conflit - applications en imagerie sonar et classification de cibles radar. *Traitement du Signal 24*(2), 71–82.

Martin, A., C. Osswald, J. Dezert, et F. Smarandache (2008b). General combination rules for qualitative and quantitative beliefs. *Journal of Advances in Information Fusion 3*(2), 67–82.

Martin, A. et E. Radoi (2004). Effective ATR Algorithms Using Information Fusion Models. In *International Conference on Information Fusion*, Stockholm, Sweden, pp. 161–166.

Mercier, D., B. Quost, et T. Denœux (2005). Contextual discounting of belief functions. In *Proceedings of ECSQARU'2005*, Barcelona, Spain, pp. 552–562.

Murphy, C. (2000). Combining belief functions when evidence conflicts. *Decision Support Systems 29*, 1–9.

Régis, S., A. Doncescu, et J. Desachy (2007). Théorie des fonctions de croyance pour la fusion et l'évaluation de la pertinence des sources d'informations : application à un bioprocédé fermentaire. *Traitement du Signal 24*(2).

Schubert, J. (1996). Specifiying nonspecific evidence. *International Journal of Intelligent Systems 11*, 525–563.

Schubert, J. (2008). Conflict management in Dempster-Shafer theory by sequential discounting using the degree of falsity. In *Information Processing and Management of Uncertainty (IPMU'2008)*, Málaga, Spain, pp. 298–305.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.

Smets, P. (2007). Analyzing the combination of conflicting belief functions. *Information Fusion 8*, 387–412.

Smets, P. et R. Kennes (1994). The Transferable Belief Model. *Artificial Intelligent 66*, 191–234.

Toumi, A. (2007). *Intégration des bases de connaissances dans les systèmes d'aide à la décision : Application à l'aide à la reconnaissance de cibles radar non-coopératives*. Ph. D. thesis, Université de Bretagne Occidentale, ENSIETA, Brest.

Yager, R. R. (1987). On the Dempster-Shafer Framework and New Combination Rules. *Information Sciences 41*, 93–137.

Zeng, C. et P. Wu (2007). A reliability discounting strategy based on plausibility function of evidence. In *International Conference on Information Fusion*, Québec, Canada.

Summary

The conflict appearing while combining several uncertain information reflects the degree of conflict between their sources. This conflict can be managed before the combination step by discounting belief functions using sources' reliability. In this paper, we suggest a generalized method for sources' reliability estimation based on distance measure and taking into account all its belief functions stored in an evidential database. This method is evaluated on real radar data and supplied good results in terms of sources' reliability improvement. Our method insures also the same level of relative reliabilities for all belief functions supplied by the same source and this level reflects the source's global reliability.