Source Independence in the Theory of Belief Functions

Mouna Chebbah

ANNÉE 2014

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne
(en cotutelle)

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Informatique

Ecole doctorale MATISSE

présentée par

Mouna Chebbah

préparée à l'unité de recherche UMR 6074 IRISA
Institut de Recherche en Informatique et Systèmes Aléatoires
ENSSAT

Source Independence in the Theory of Belief Functions

Thèse soutenue à l'université de Rennes 1
le 25/06/2014

devant le jury composé de :

Eric LEFÈVRE
Professeur des universités, Université d'Artois, France / rapporteur

Weiru LIU
Professeur des universités, Queen's University, Belfast, UK / rapporteur

Zied ELOUEDI
Professeur des universités, Université de Tunis, Tunisie / examinateur

Ludovic LIETARD
Maître de conférences HDR, Université de Rennes 1, France / examinateur

Arnaud MARTIN
Professeur des universités, Université de Rennes 1, France / directeur de thèse

Boutheina BEN YAGHLANE
Professeur des universités, Université de Carthage, Tunisie / co-directrice de thèse

Abstract

The theory of belief functions manages uncertainty and proposes a set of combination rules to aggregate beliefs of several sources. Some combination rules mix evidential information where sources are independent; other rules are suited to combine evidential information held by dependent sources. Information on sources' independence is required to justify the choice of the adequate type of combination rules. In this thesis, we suggest a method to quantify sources' degrees of independence that may guide the choice of the appropriate type of combination rules. In fact, we propose a statistical approach to learn sources' degrees of independence from all provided evidential information. There are three main uses of estimating sources' degrees of independence: First, we use sources' degree of independence to guide the choice of combination rules to use when aggregating beliefs of several sources. Second, we propose to integrate sources' degrees of independence into sources' beliefs leading to an operator similar to the discounting. Finally, we define a new combination rule weighted with sources' degree of independence.

Résumé

La fusion d'informations issues de plusieurs sources cherche à améliorer la prise de décision. Pour réaliser cette fusion, la théorie des fonctions de croyance utilise des règles de combinaison faisant bien souvent l'hypothèse de l'indépendance des sources. Cette forte hypothèse n'est, cependant, ni formalisée ni vérifiée. Elle est supposée pour justifier le choix du type de règles à utiliser sans avoir, pour autant, un moyen de la vérifier. Nous proposons dans ce rapport de thèse un apprentissage de l'indépendance cognitive de sources d'information. Nous détaillons également une approche d'apprentissage de la dépendance positive et négative des sources. Les degrés d'indépendance, de dépendance positive et négative des sources ont principalement trois utilités. Premièrement, ces degrés serviront à choisir le type de règles de combinaison à utiliser lors de la combinaion. Deuxièmement, ces degrés exprimés par une fonction de masse sont intégrés par une approche d'affaiblissement avant de réaliser la combinaison d'information. Une troisième utilisation de cette mesure d'indépendance consiste à l'intégrer dans une nouvelle règle de combinaison. La règle que nous proposons est une moyenne pondérée avec ce degré d'indépendance.

Acknowledgements

This section is the most difficult to write. Many people supported me to finish my PhD, it is not easy to express my gratitude to these people.

First, I would like to express my deepest thanks to my advisors Pr. Arnaud Martin (*University of Rennes 1, France*) and Pr. Boutheina Ben Yaghlane (*University of Carthage, Tunisia*) for their continuous support, guidance, encouragement and presence. I would like to express my sincere gratitude to them for introducing me to an interesting area of research and also for all valuable discussion we have had during these three years.

Second, I would like to thank members of my thesis committee: Pr. Éric Lefèvre (*University of Artois, France*), Pr. Weiru Liu (*Queen's University, United Kingdom*), Pr. Zied Elouedi (*University of Tunis, Tunisia*) and Dr. Ludovic Lietard (*University of Rennes 1, France*) for accepting to review my dissertation.

I would like to address my thanks to all colleagues in my laboratory LARODEC and my team DRUID for their help and kindness. A special thank for all colleges in IUT of Lannion for their encouragement, kindness and support.

Then, I gratefully acknowledge all my family, friends and closest persons. All my thanks and love to my lovely parents, comprehensive husband and gentle brothers for their continuous encouragement and motivation these last years. I also thank all my friends in Lannion for their presence and help in all difficult periods, I really appreciated their advices and jokes

Finally, I owe so much to my director Arnaud Martin for his patience and his understanding.

L'indépendance des sources dans la théorie des fonctions de croyance

1 Introduction

La théorie des fonctions de croyance issue des travaux de (Dempster, 1967) et (Shafer, 1976) permet une bonne modélisation des données imprécises et/ou incertaines et offre un outil puissant pour fusionner des informations issues de plusieurs sources. Pour ce faire, les données incertaines et imprécises des différentes sources sont modélisées par des fonctions de masse et combinées afin de mettre en évidence les croyances communes et assurer une prise de décision plus fiable.

Le choix de la règle de combinaison à appliquer repose sur des hypothèses d'indépendance de sources. En effet, certaines règles de combinaison comme celles de (Dempster, 1967; Smets, 1990; Yager, 1987; Dubois and Prade, 1988) combinent des fonctions de croyance dont les sources sont supposées indépendantes par contre les règles prudente et hardie proposées par (Denœux, 2006a) n'exigent pas d'hypothèse d'indépendance. L'indépendance cognitive est une hypothèse fondamentale pour le choix des règles de combinaison à appliquer.

Les indépendances évidentielle, cognitive et doxastique ont été définies dans la cadre de la théorie des fonctions de croyance. D'une part, les travaux de (Ben Yaghlane, 2002) étudient principalement l'indépendance doxastique des variables. D'autre part, les travaux de (Shafer, 1976) ont défini l'indépendance cognitive des variables. Dans cette thèse nous nous sommes focalisés sur l'indépendance cognitive des sources, nous proposons une approche statistique pour l'estimation de l'indépendance cognitive de deux sources. Deux sources sont cognitivement indépendantes si elles ne communiquent pas entre elles et si elles n'ont pas le même corpus de croyance[1]. La méthode proposée permet d'étudier le comportement général de deux sources et de les comparer pour déceler toute dépendance pouvant exister entre elles. Dans le cas de sources dépendantes, nous proposons d'étudier le type de cette dépendance, c'est-à-dire analyser les données pour mettre en évidence des sources plutôt *positivement* ou *négativement* dépendantes. Nous avons, également, proposé une généralisation de cette approche statistique pour plusieurs sources.

[1] Le corpus de croyance est l'ensemble de connaissances ou d'informations acquises par une source.

Cette approche statistique a pour but de guider la combinaison. En effet, nous proposons trois solutions; la première solution consiste à justifier le choix du type de règle de combinaison par le degré de dépendance ou d'indépendance des sources. La deuxième solution consiste à intégrer les degrés d'indépendance, dépendance positive et négative dans les fonctions de masse afin de justifier l'hypothèse d'indépendance des sources. Enfin, comme troisième solution, nous proposons une nouvelle règle de combinaison qui est une moyenne pondérée de la combinaison conjonctive et la combinaison prudente. L'approche proposée a été illustrée sur des données générées.

2 Théorie des fonctions de croyance

La théorie des fonctions de croyance initialement introduite par (Dempster, 1967), formalisée ensuite par (Shafer, 1976) est employée dans des applications de fusion d'informations. Nous présentons ci-dessous un résumé de quelques principes de base de cette théorie. Soit un *cadre de discernement* $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ l'ensemble de toutes les hypothèses exclusives et exhaustives. Le cadre de discernement est aussi *l'univers de discours* d'un problème donné. L'ensemble $2^\Omega = \{A|A \subseteq \Omega\} = \{\emptyset, \omega_1, \omega_2, \ldots, \omega_n, \omega_1 \cup \omega_2, \ldots, \Omega\}$, est l'ensemble de toutes les hypothèses de Ω ainsi que leurs disjonctions.

Une *fonction de masse* est une fonction de 2^Ω vers l'intervalle $[0, 1]$ qui affecte à chaque sous-ensemble une *masse de croyance élémentaire.* Cette fonction de masse fournie par une source d'information[2] est une représentation des connaissances incertaines et imprécises. Formellement, une fonction de masse, notée m^Ω, est définie comme suit :

$$m^\Omega : 2^\Omega \rightarrow [0, 1] \quad (1)$$

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

Dans le cadre de la théorie des fonctions de croyance, plusieurs règles de combinaison sont proposées pour la fusion d'informations. Les fonctions de masse sont issues de différentes sources et sont définies sur le même ensemble de discernement. La combinaison permet de synthétiser ces différentes informations en vue d'une prise de décision plus fiable. Le choix des règles de combinaison dépend de certaines hypothèses initiales, les opérateurs de type conjonctif tels que (Dempster, 1967; Smets, 1990; Yager, 1987; Dubois and Prade, 1988) peuvent être employés lorsque les sources sont *cognitivement indépendantes* par contre les règles prudente et hardie proposées par (Denœux, 2006a) ne suppose pas une telle hypothèse.

Après l'acquisition d'une fonction de masse, une information certaine peut ap-

[2]La source peut être un expert humain, un classificateur, un capteur, ...

paraître confirmant que l'hypothèse vraie est (ou n'est pas) dans l'un des sous-ensembles de 2^Ω. Dans ce cas, la fonction de masse doit être mise à jour afin de prendre en considération cette nouvelle information certaine. Cette mise à jour est réalisée par l'opérateur de *conditionnement* proposé par (Smets and Kruse, 1997).
Le *déconditionnement* est l'opération inverse permettant de retrouver une fonction de masse la moins informative à partir d'une fonction de masse conditionnée.

En combinant des fonctions de masse, un degré de conflit peut surgir reflétant un certain désaccord entre les sources. La non fiabilité d'une source peut être réglée par l'affaiblissement des fonctions de masse avant la combinaison en utilisant l'opérateur d'affaiblissement proposé par (Shafer, 1976). Une fonction de masse, m^Ω, est affaiblie par la fiabilité α de sa source comme suit:

$$\begin{aligned} {}^{\alpha}m^{\Omega}(A) &= \alpha \times m^{\Omega}(A) \quad \forall A \subset \Omega && (3) \\ {}^{\alpha}m^{\Omega}(\Omega) &= 1 - \alpha \times (1 - m^{\Omega}(\Omega)) && (4) \end{aligned}$$

Smets (Smets, 1993) a justifié cette procédure.

La prise de décision dans la théorie des fonctions de croyance peut être fondée sur des probabilités pignistiques notées $BetP$ issues de la transformation pignistique proposée par (Smets, 2005). Cette transformation calcule une probabilité pignistique à partir des fonctions de masse en vue de prendre une décision.

3 Classification non-supervisée

Dans les travaux de cette thèse, nous proposons d'utiliser un algorithme de classification non-supervisée de type C-moyenne, utilisant une distance sur les fonctions de masse définie par (Jousselme et al., 2001) comme proposé par (Ben Hariz et al., 2006; Chebbah et al., 2012a; Chebbah et al., 2012b). Soit un ensemble T contenant N objets $o_i : 1 \leq i \leq N$ à classifier dans C clusters. Les valeurs des o_i sont des fonctions de masse m_i^Ω définies sur un cadre de discernement Ω. Une mesure de dissimilarité $D(o_i, Cl_k)$ permet de mesurer la dissimilarité entre un objet o_i et un cluster Cl_k comme suit :

$$D(o_i, Cl_k) = \frac{1}{n_k} \sum_{j=1}^{n_k} d(m_i^\Omega, m_j^\Omega) \tag{5}$$

$$d(m_1^\Omega, m_2^\Omega) = \sqrt{\frac{1}{2}(m_1^\Omega - m_2^\Omega)^t \underline{\underline{D}}(m_1^\Omega - m_2^\Omega)}, \tag{6}$$

$$\underline{\underline{D}}(A, B) = \begin{cases} 1 & \text{si } A = B = \emptyset \\ \dfrac{|A \cap B|}{|A \cup B|} & \forall A, B \in 2^\Omega \end{cases} \tag{7}$$

La disimilarité d'un objet o_i et un cluster Cl_k est définie par la moyenne des distances entre la fonction de masse m_i^Ω valeur de cet objet et toutes les n_k fonctions de masse valeurs des $o_j : 1 \leq j \leq n_k$ objets contenus dans le cluster Cl_k. Chaque objet est affecté au cluster qui lui est le plus similaire (ayant une valeur de disimilarité minimale) de manière itérative jusqu'à ce qu'une répartition stable soit obtenue.

À la fin de la classification non-supervisée, C clusters contenant chacun un certain nombre d'objets sont obtenus. Nous supposons que le nombre de clusters C est égal à la cardinalité du cadre de discernement ($C = |\Omega|$).

4 Indépendance

L'indépendance a été introduite en premier dans le cadre de la théorie des probabilités pour modéliser l'indépendance statistique des évènements. Les fonctions de masse peuvent être perçues comme des probabilités subjectives fournies par des sources s'exprimant sur un problème étant donné un ensemble de connaissances ou d'informations appelé *corpus de croyance*. Dans le cas d'une hypothèse *d'indépendance cognitive* des sources, les corpus de croyance doivent être distincts et aucune communication entre les sources n'est tolérée.

Nous proposons une démarche statistique comme détaillée dans (Chebbah et al., 2012a; Chebbah et al., 2012b; Chebbah et al., 2013; Chebbah et al., 2014) afin d'étudier l'indépendance cognitive de deux sources.
Nous introduisons d'abord la mesure d'indépendance de deux sources s_1 et s_2, notée $I_d(s_1, s_2)$, comme étant l'indépendance de s_1 de s_2. Cette mesure vérifie les axiomes suivants :

1. Non-négative : L'indépendance d'une source s_1 de s_2, $I_d(s_1, s_2)$ est une valeur qui est, soit nulle si s_1 est complètement dépendante de s_2, soit strictement positive.

2. Normalisée : $I_d(s_1, s_2) \in [0, 1]$, si I_d est nulle alors s_1 est complètement dépendante de s_2. Si $I_d = 1$, alors s_1 est complètement indépendante de s_2 autrement c'est un degré de $]0, 1[$.

3. Non-symétrique : Si s_1 est indépendante de s_2, cela n'implique pas forcement que s_2 soit indépendante de s_1. Les sources s_1 et s_2 peuvent être simultanément indépendantes avec des degrés d'indépendance égaux ou différents.

4. Identité : $I_d(s_1, s_1) = 0$.

L'approche proposée est une approche statistique pour mesurer le degré d'indépendance cognitive. Nous proposons ainsi de classifier toutes les fonctions de masse des deux sources et de comparer les clusters obtenus. La classification non supervisée détaillée dans la Section 3 regroupe les objets ayant pour valeurs des fonctions de masse similaires. Si les clusters des deux sources sont similaires, alors il est fort probable qu'elles soient dépendantes. Nous proposons d'apparier les clusters des sources et de quantifier leurs similarités.

L'algorithme de classification non-supervisée est appliqué aux fonctions de masse fournies par différentes sources séparément et puis ces clusters sont comparés dans le but de voir s'il y a un lien entre eux. Plus les liens entre ces clusters sont forts plus les sources ont tendance à être dépendantes. Soient deux sources s_1 et s_2, fournissant chacune N fonctions de masse pour les mêmes objets. Après avoir classifié les fonctions de masse de s_1 et s_2, la matrice de correspondance des clusters M est obtenue par :

$$M_1 = \begin{pmatrix} \beta^1_{1,1} & \beta^1_{1,2} & \dots & \beta^1_{1,C} \\ \dots & \dots & \dots & \dots \\ \beta^1_{k,1} & \beta^1_{k,2} & \dots & \beta^1_{k,C} \\ \dots & \dots & \dots & \dots \\ \beta^1_{C,1} & \beta^1_{C,2} & \dots & \beta^1_{C,C} \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} \beta^2_{1,1} & \beta^2_{1,2} & \dots & \beta^2_{1,C} \\ \dots & \dots & \dots & \dots \\ \beta^2_{k,1} & \beta^2_{k,2} & \dots & \beta^2_{k,C} \\ \dots & \dots & \dots & \dots \\ \beta^2_{C,1} & \beta^2_{C,2} & \dots & \beta^2_{C,C} \end{pmatrix} \tag{8}$$

avec

$$\beta^i_{k_i,k_j} = \frac{|Cl^i_{k_i} \cap Cl^j_{k_j}|}{|Cl^i_{k_i}|} \tag{9}$$

Notons que $\beta^i_{k_i,k_j}$ est la similarité des clusters $Cl^i_{k_i}$ de s_i et $Cl^j_{k_j}$ de s_j par rapport à s_i avec $\{i,j\} \in \{1,2\}$ et $i \neq j$.

Une fois les deux matrices de correspondances M_1 et M_2 calculées, une correspondance entre les clusters est établie. Chaque cluster est lié au cluster qui lui est le plus similaire, ayant le β maximal, en vérifiant que deux clusters de la même source ne peuvent pas être liés au même cluster de l'autre source.

La recherche de correspondances des clusters est faite pour les deux sources. Deux correspondances différentes peuvent être obtenues pour les deux sources.

Une fois la correspondance des clusters établie, une fonction de masse définissant l'indépendance de chaque couple de clusters est déduite. Ceci revient à avoir un agent ayant les correspondances des clusters (k_i, k_j) avec les similarités correspondantes $\beta^i_{k_i,k_j}$ comme corpus de croyance pour s'exprimer sur l'indépendance de ces clusters. Après appariement de clusters, les clusters de s_1 sont liés aux clusters de s_2 qui leur sont similaires et ceux de s_2 sont également liés aux clusters de s_1 les plus similaires. Différents appariements sont obtenus pour s_1 et s_2. Nous définissons l'indépendance de chaque couple de clusters liés (k_1, k_2) comme une fonction de masse définie sur le cadre de discernement $\mathcal{I} = \{\bar{I}, I\}$, où $\bar{I}$ représente la dépendance et I l'indépendance:

$$\left\{ \begin{array}{l} m^{\mathcal{I}}_{k_i,k_j}(I) = \alpha^i_{k_i,k_j}\,(1 - \beta^i_{k_i,k_j}) \\ m^{\mathcal{I}}_{k_i,k_j}(\bar{I}) = \alpha^i_{k_i,k_j}\,\beta^i_{k_i,k_j} \\ m^{\mathcal{I}}_{k_i,k_j}(\mathcal{I}) = 1 - \alpha^i_{k_i,k_j} \end{array} \right. \tag{10}$$

Le coefficient $\alpha^i_{k_i,k_j}$ est un degré de fiabilité utilisé pour tenir compte du nombre d'objets contenus dans les clusters de la source référente. Une fonction de masse est définie pour chaque couple de clusters appariés pour chacune des sources. Pour avoir une fonction de masse sur l'indépendance globale de chaque source, toutes ces fonctions de masse sont combinées avec la moyenne. La combinaison de ces C fonctions de masse est une fonction de masse $m_i^{\mathcal{I}}$ décrivant l'indépendance globale de la source s_i par rapport à s_j:

$$\begin{cases} m_i^{\mathcal{I}}(I) = \dfrac{1}{C}\displaystyle\sum_{k_i=1}^{C} m_{k_i,k_j}^{\mathcal{I}}(I) \\ m_i^{\mathcal{I}}(\bar{I}) = \dfrac{1}{C}\displaystyle\sum_{k_i=1}^{C} m_{k_i,k_j}^{\mathcal{I}}(\bar{I}) \\ m_i^{\mathcal{I}}(\bar{I}\cup I) = \dfrac{1}{C}\displaystyle\sum_{k_i=1}^{C} m_{k_i,k_j}^{\mathcal{I}}(\bar{I}\cup I) \end{cases} \tag{11}$$

Les probabilités pignistiques calculées à partir de la fonction de masse combinée permettent la prise de décision sur l'indépendance des sources. L'indépendance de la source s_1 de la source s_2, $I_d(s_1, s_2)$ n'est autre que la probabilité pignistique de I, $I_d(s_1, s_2) = BetP(I)$ et $\bar{I}_d(s_1, s_2) = BetP(\bar{I})$ ce qui revient à écrire I_d comme suit :

$$\begin{cases} I_d(s_i, s_j) = \dfrac{1}{C}\displaystyle\sum_{k_i=1}^{C} [\alpha_{k_i,k_j}^i\, \beta_{k_i,k_j}^i + \dfrac{1}{2}(1-\alpha_{k_i,k_j}^i)] \\ \bar{I}_d(s_i, s_j) = \dfrac{1}{C}\displaystyle\sum_{k_i=1}^{C} [\alpha_{k_i,k_j}^i\, (1-\beta_{k_i,k_j}^i) + \dfrac{1}{2}(1-\alpha_{k_i,k_j}^i)] \end{cases} \tag{12}$$

Si $I_d(s_i, s_j) < \bar{I}_d(s_i, s_j)$, alors s_i est dépendante de s_j, dans le cas contraire s_i est indépendante de s_j. Notons que cette approche d'estimation de l'indépendance, dépendance positive et négative a été généralisée pour plusieurs sources.

5 Dépendance positive ou négative

Dans le cas de sources dépendantes I_d n'est pas suffisante pour indiquer le type de la dépendance. Deux sources dépendantes peuvent être positivement ou négativement dépendantes. Nous définissons une mesure de conflit entre les clusters de s_i et s_j quantifiant cette dépendance que nous qualifions de positive ou négative. Si les clusters liés ne sont pas conflictuels alors s_i est positivement dépendante de s_j sinon elle est négativement dépendante. Nous définissons alors le conflit entre les deux clusters dépendants $Cl_{k_i}^i$ et $Cl_{k_j}^j$ ($\{i, j\} \in \{1, 2\}$ et $i \neq j$) à partir de la moyenne des distances entre les fonctions de masse des objets en commun :

$$\begin{cases} Conf(Cl^i_{k_i}, Cl^j_{k_j}) = \dfrac{1}{|Cl^i_{k_i} \cap Cl^j_{k_j}|} \displaystyle\sum_{l \in E(Cl^i_{k_i}, Cl^j_{k_j})} d(m^{\Omega,i}_l, m^{\Omega,j}_l) & \text{si } |Cl^i_{k_i} \cap Cl^j_{k_j}| \neq 0 \\ 1 & \text{sinon} \end{cases} \tag{13}$$

avec

$$E(Cl^i_{k_i}, Cl^j_{k_j}) = \{l \in [1, n], n = |Cl^i_{k_i} \cap Cl^j_{k_j}|, m^{\Omega,i}_l \in Cl^i_{k_i} et\ m^{\Omega,j}_l \in Cl^j_{k_j}\} \tag{14}$$

Cette mesure de conflit est la moyenne des conflits (Chebbah et al., 2010b; Chebbah et al., 2010a; Chebbah et al., 2011) entre les objets contenus dans les clusters $Cl^i_{k_i}$ et $Cl^j_{k_j}$. Le conflit est calculé pour chaque couple de clusters liés. Une fonction de masse définie sur le cadre de discernement $\mathcal{P} = \{I, P, \bar{P}\}$ (où P représente la dépendance positive et $\bar{P}$ la dépendance négative) décrivant la dépendance est obtenue pour chaque couple de clusters :

$$\begin{cases} m^{\mathcal{P}}_{k_i,k_j}[\bar{I}](P) = 1 - Conf(Cl^i_{k_i}, Cl^j_{k_j}) \\ m^{\mathcal{P}}_{k_i,k_j}[\bar{I}](\bar{P}) = Conf(Cl^i_{k_i}, Cl^j_{k_j}) \end{cases} \tag{15}$$

Notons que le conflit entre les clusters reflète la contradiction entre ces clusters. Plus le conflit est important, plus les sources sont dépendantes négativement mais par contre moins il est important plus les sources sont dépendantes positivement. Ces fonctions de masse sont conditionnelles puisque la dépendance positive ou négative des clusters n'est mesurée qu'avec une forte hypothèse de dépendance des clusters liés. L'hypothèse de dépendance ou encore de non indépendance des clusters explique le fait que les fonctions de masse de l'équation (4.27) soient conditionnées sur $\bar{I}$ ou encore sur $\{P \cup \bar{P}\}$. Afin de pouvoir combiner les fonctions de masse (10) et (4.27) pour tenir compte du degré de dépendance des clusters dans la fonction de masse de la dépendance positive ou négative, il faut déconditionner les fonctions de masse conditionnelles et redéfinir les deux fonctions de masse sur un cadre le discernement commun $\mathcal{P}$. Les fonctions de masse obtenues après déconditionnement sont alors :

$$\begin{cases} m^{\mathcal{P}}_{k_i,k_j}(P \cup I) = 1 - Conf(Cl^i_{k_i}, Cl^j_{k_j}) \\ m^{\mathcal{P}}_{k_i,k_j}(\bar{P} \cup I) = Conf(Cl^i_{k_i}, Cl^j_{k_j}) \end{cases} \tag{16}$$

D'autre part, le cadre de discernement $\mathcal{I} = \{\bar{I}, I\}$ peut être raffiné en raffinant l'hypothèse $\bar{I} = \{P \cup \bar{P}\}$, ceci mènera au cadre de discernement raffiné $\mathcal{P}$. Les fonctions de masse marginales de la dépendance des clusters liés de l'équation (10) deviennent après raffinement :

$$\begin{cases} m^{\mathcal{P}}_{k_i,k_j}(I) = \alpha^i_{k_i,k_j}\,(1-\beta^i_{k_i,k_j}) \\ m^{\mathcal{P}}_{k_i,k_j}(P\cup\bar{P}) = \alpha^i_{k_i,k_j}\,\beta^i_{k_i,k_j} \\ m^{\mathcal{P}}_{k_i,k_j}(I\cup P\cup\bar{P}) = 1-\alpha^i_{k_i,k_j} \end{cases} \tag{17}$$

Nous définissons ainsi la fonction de masse de l'indépendance, dépendance positive et dépendance négative de chaque couple de clusters liés de s_i et s_j après combinaison conjonctive des fonctions de masse des équations (4.28) et (4.29) définies sur le cadre de discernement $\mathcal{P}$:

$$\begin{cases} m^{\mathcal{P}}_{k_i,k_j}(I) = \alpha^i_{k_i,k_j}\,(1-\beta^i_{k_i,k_j}) \\ m^{\mathcal{P}}_{k_i,k_j}(P) = \alpha^i_{k_i,k_j}\,\beta^i_{k_i,k_j}(1-Conf(Cl^i_{k_i},Cl^j_{k_j})) \\ m^{\mathcal{P}}_{k_i,k_j}(\bar{P}) = \alpha^i_{k_i,k_j}\,\beta^i_{k_i,k_j}Conf(Cl^i_{k_i},Cl^j_{k_j}) \\ m^{\mathcal{P}}_{k_i,k_j}(I\cup P) = (1-\alpha^i_{k_i,k_j})\,(1-Conf(Cl^i_{k_i},Cl^j_{k_j})) \\ m^{\mathcal{P}}_{k_i,k_j}(I\cup\bar{P}) = (1-\alpha^i_{k_i,k_j})\,Conf(Cl^i_{k_i},Cl^j_{k_j}) \end{cases} \tag{18}$$

La fonction de masse générale sur la dépendance de la source s_i par rapport à s_j est donnée par :

$$m^{\mathcal{P}}(A) = \frac{1}{C}\sum_{k_i=1}^{C} m^{\mathcal{P}}_{k_i,k_j}(A) \tag{19}$$

avec $\{i,j\} \in \{1,2\}$ et $i \neq j$, où k_i est le cluster de la source s_i associé au cluster k_j de la source s_j. Cette fonction de masse représente ainsi l'ensemble des croyances élémentaires sur l'indépendance, la dépendance positive et négative de la source s_i face la source s_j. Le degré d'indépendance est la probabilité pignistique de l'hypothèse I, $BetP(I)$, ceux des dépendances positive et négative sont respectivement $BetP(P)$ et $BetP(\bar{P})$. Le manuscrit de thèse détaille les résultats expérimentaux de l'approche.

6 Utilisation de l'indépendance, dépendance positive ou négative

La mesure $I_d(s_i,s_j)$ informe sur l'indépendance ou *a contrario* la dépendance de la source s_i par rapport à la source s_j permettant par exemple de choisir la règle de combinaison à utiliser ou encore intégrer cette information dans ses fonctions de masse. Quant au moins l'une des sources s_i ou s_j est dépendante de l'autre ($I_d(s_i,s_j) < \bar{I}_d(s_i,s_j)$ ou $I_d(s_j,s_i) < \bar{I}_d(s_j,s_i)$), il est alors préférable d'utiliser les règles de (Denœux, 2006a) sinon les règles de (Dubois and Prade, 1988; Martin and Osswald, 2007b; Smets and Kennes, 1994; Yager, 1987) permettent par exemple de redistribuer la masse de l'ensemble vide.

D'autres utlisations de la mesure d'indépendance, dépendance positive ou négative consiste à les intégrer soit dans les fonctions de masse afin de supposer l'indépendance des sources ou encore de les intégrer dans une nouvelle règle de combinaison.

6.1 Intégration de l'indépendance dans une fonction de masse

L'indépendance est généralement une information supplémentaire nécessaire à la fusion d'informations, mais non prise en compte dans le formalisme choisi. Nous proposons d'appuyer sur le principe de l'affaiblissement afin de tenir compte de l'indépendance dans les fonctions de masse en vue de la combinaison.

En effet, lors de la combinaison conjonctive par exemple, l'hypothèse d'indépendance cognitive des sources d'informations est nécessaire. Si les sources sont dépendantes on peut penser qu'elles ne devraient pas être combinées par ce biais. Cependant, comme le montre la Section 4 les sources peuvent avoir des degrés de dépendance et d'indépendance. L'information fournie sur l'indépendance n'est pas catégorique.

Dans ce cas, il suffit d'appliquer la procédure d'affaiblissement détaillée dans (Smets, 1993) sur la fonction de masse m^{Ω} de la source s_i en considérant l'indépendance donnée par la fonction de masse de l'équation (19).

À présent, nous distinguons la dépendance positive de la dépendance négative. Si une source est dépendante positivement d'une autre source, il ne faut pas en tenir compte et donc tendre vers un résultat de combinaison qui prendrait cette première source comme un élément neutre. Enfin si une source est dépendante négativement d'une autre source, il peut être intéressant de marquer cette dépendance conflictuelle en augmentant la masse sur l'ensemble vide.

Pour réaliser ce schéma, nous proposons d'affaiblir les fonctions de masse d'une source s_i en fonction de sa mesure d'indépendance à une autre source s_j, donnée par la fonction de masse $m_i^{\mathcal{I}}$ de l'équation (19).

Nous considérons ici une fonction de masse d'une source m^{Ω} en fonction de son indépendance ou dépendance à une autre source comme détaillé dans (Chebbah et al., 2014). Ainsi la fonction de masse affaiblie par les degrées d'indépendance, dépendance positive et négative de la source est définie comme suit :

$$\left\{\begin{array}{ll} m^{\Omega}[I](X) = m^{\Omega}(X) & \\ m^{\Omega}[\bar{P}](X) = m^{\Omega}(X) & m^{\Omega}(X) = 1 \text{ si } X = \emptyset, \text{ 0 sinon} \\ m^{\Omega}[P](X) = m^{\Omega}(X) & m^{\Omega}(X) = 1 \text{ si } X = \Omega, \text{ 0 sinon} \end{array}\right. \tag{20}$$

Cette procédure réalisée pour la source s_i en rapport à la source s_j peut être réalisée pour la source s_j au regard de la source s_i. Ainsi les deux fonctions de masse obtenues peuvent être combinées par la règle de combinaison conjonctive qui suppose l'indépendance. Des illustrations sont détaillées dans la suite de ce manuscrit de thèse.

6.2 Intégration de l'indépendance dans une règle de combinaison mixte

Les degrés d'indépendance, dépendance positive et négative sont utiles soit pour guider le choix du type de règles de combinaison à utiliser, soit pour les intégrer dans les fonctions de masse afin de pouvoir supposer l'indépendance cognitive dans la combinaison.

Les degrés d'indépendance, dépendance positive ou négative sont des degrés dans $[0, 1]$. Si l'indépendance des sources est 1, les sources sont complètement indépendantes et si l'indépendance est 0, les sources sont complètement dépendantes. Le choix des règles de combinaison dans les deux cas extrême est assez simple. Quand le degré de dépendance est dans $]0, 1[$, nous proposons une nouvelle règle de combinaison mixte qui est une moyenne pondérée des combinaisons conjonctive et prudente. La combinaison mixte de deux fonctions de masse m_1^Ω et m_2^Ω fournies par deux sources s_1 et s_2 telle que leur degré d'indépendance est $\gamma = I_d(s_1, s_2)$ est défini comme suit :

$$m_{Mixte} = \gamma * m_{\textcircled{\cap}} + (1 - \gamma) * m_{\textcircled{\wedge}} \qquad (21)$$

La masse combinée d'un élément focal A, $m_{Mixte}(A)$, est la moyenne de sa masse combinée avec la règle disjonctive et sa masse combinée avec la règle prudente calibrée avec le degré d'indépendance des sources. La règle mixte tend vers la combinaison conjonctive quand les sources sont indépendantes, et vers la règle prudente quand les sources sont dépendantes. La règle mixte est:

- Commutative: La règle conjonctive et la règle prudente sont commutatives, le degré d'indépendance des sources est symétrique donc la règle mixte est commutative.
- Associative: La règle conjonctive et la règle prudente sont associative mais la règle mixte ne l'est pas.
- Idempotente: L'indépendance d'une source d'elle même est 1, dans ce cas la règle mixte et la règle prudente sont équivalentes. La règle prudente est idempotente donc la règle mixte l'est aussi.
- L'élément neutre et l'élément absorbant: La règle mixte ne possède ni un élément neutre ni un élément absorbant.

7 Conclusion

Lors des travaux de recherche de cette thèse, nous nous sommes focalisés sur l'estimation de l'indépendance, dépendance positive et négative des sources. En effet, nous avons proposé une approche statistique afin d'estimer ces degrés d'indépendance/dépendance. Ces degrés serviront à guider le choix du type de règles de combinaison à appliquer. En effet, les règles de combinaison du type conjonctive et/ou disjonctive telles que (Dubois and Prade, 1988; Martin and Osswald, 2007b; Smets and Kennes, 1994; Yager, 1987) combinent des fonctions de masse dont les sources sont indépendantes par contres les règles prudente et hardi combinent des fonctions de masse dont les sources sont dépendantes. L'information sur l'indépendance des sources peut aussi être intégrée dans les fonctions de masse afin de supposer l'indépendance des sources. Le degré d'indépendance est aussi utilisé dans une règle de combinaison mixte que nous avons proposée. La règle mixte est une moyenne pondérée des combinaisons conjonctive et prudente.

Contents

List of Tables

List of Figures

Abbreviations and notations

In the following, a list as exhaustive as possible of abbreviations and notations used in this thesis:

Belief functions

- Θ, Ω: are two distinct frames of discernment; they can be compatible or not. The frame Ω can be indexed with the first letter of variables' names (for example: Ω_d is the frame of discernment of the variable disease);
- $\omega_1, \omega_2, \ldots, \omega_n$: hypothesis in Ω; they are singletons;
- $\theta_1, \theta_2, \ldots, \theta_n$: hypothesis in Θ; they are also singletons;
- n: number of hypotheses in a frame of discernment, for example $n = |\Omega|$;
- $\Omega \times \Theta$: is the cartesian product of Ω and Θ;
- $\uparrow$: vacuous extension;
- $\downarrow$: marginalization;
- $\Uparrow$: deconditioning;
- m, m^Ω, m_j^Ω, m_j: is a mass function, m is a mass function defined on any frame of discernment Ω and provided by a source j. Normality condition is not required in this mass function;
- m_Ω^Ω: a vacuous mass function;
- $m_\emptyset^\Omega$: a contradictory mass function;
- A^w: a simple support function focused on A with a degree of support w;
- bel, bel^Ω, bel_j^Ω: belief function;
- pl, pl^Ω, pl_j^Ω: plausibility function;
- q, b: communality and implicability functions;
- Bel, Pl: normalized belief and plausibility functions;

- $BetP_m^{\Omega}$, $BetP_m$, $BetP$: pignistic probability;
- A, B, C, D: focal elements of a mass function m^{Ω}; $A\ B\ C\ D \subseteq \Omega$;
- $\bar{A}$: complement of A in Ω; $\bar{A} = \Omega \setminus \{A\}$;
- $\mathcal{F}$: the set of all focal elements of a mass functions;
- $|F|$: number of focal elements in a mass function;
- F_S: set of same focal elements that have same masses;
- $|F_S|$: number of same focal elements that have same masses;
- F_{NC}: set of not conflicting focal elements;
- $|F_{NC}|$: number of not conflicting focal elements;
- F_{CO}: set of conflicting focal elements;
- $|F_{CO}|$: number of conflicting focal elements;
- φ: is the core of a mass function;
- N: is the number of mass functions;
- A^w: a simple support function focused on A with a degree of support w;
- $m_{A^\star}^{\Omega}$: a categorical mass function focused on $A^\star$, $A^\star \subseteq \Omega$. When $A^\star = \Omega$, the mass function is vacuous and when $A^\star = \emptyset$, it is contradictory;
- ${}^{\alpha}m$, ${}^{\alpha}m^{\Omega}$, ${}^{\alpha}m_j^{\Omega}$: a discounted mass function;
- $m[A]$, $m^{\Omega}[A]$: conditioned mass function;
- M: is the number of sources;
- i, j: indexes for sources and their mass functions;
- s_i, s_j: sources i and j. When $M = 2$, $\{i, j\} \in \{1, 2\}$ and $i \neq j$. Mass functions m_i and m_j are mass functions respectively provided by s_i and s_j;
- EC: evidential corpus;
- Ev: evidence.

Distances

- m': the transpose of m considered as a vector;
- d_{PS}: Perry and Stephanou's distance;
- d_{BP}: Blackman and Popoli's distance;
- d_R: Ristic and Smets distance generalizing Bhattacharrya's distance (Ristic and Smets, 2006);
- d_{RS}: Ristic and Smets distance using Dempster's conflict (Ristic and Smets, 2006);
- d_{Fl}: distance of (Florea et al., 2009b);
- $DifBetP$: distance on pignistic probabilities proposed by (Liu, 2006);
- d: Jousselme distance;
- p: an integer such that $p \geq 1$;
- $d^{(p)}$, L_p: Minkowski distance;
- $d_C^{(p)}$: Cuzzolin distance generalizing the Minkowski distance on the theory of belief functions;
- $d^{(1)}$, L_1: Manhatten distance;
- $d^{(2)}$, L_2: Euclidean distance;
- $d^{(\infty)}$, L_∞: Chebychev distance;
- U: upper triangular matrix of the Cholesky decomposition;
- Inc: inclusion matrix;
- Int: intersection matrix.

Conflict

- k, $m_{1 \cap\!\!\!\!\bigcirc 2}(\emptyset)$, $m_{\cap\!\!\!\!\bigcirc}(\emptyset)$: Dempster's conflict;
- K: re-normalizing constant;
- $\mathrm{Con}(m_1, m_2)$: weight of conflict between two mass functions;
- $\cos(m_1, m_2)$: a cosine-based measure;
- $||.||$: the norm of a vector;

- a_s: auto-conflict after s times sequential combinations of identical mass functions m;
- $\omega^\star$: hypothesis not mentioned on Ω;
- λ: a real not null;
- $\text{Conf}(s_i, s_j)$: conflict between two sources.

Evidential databases

- EDB: evidential database;
- c: number of attributes in an EDB;
- a_j: attribute j;
- Ω_{a_j}: domain of the attribute a_j;
- V_{ij}: evidential value of the object i for the j^{th} attribute;
- m_{ij}, $m_{ij}^{\Omega_{a_j}}$: a mass function value of attribute "j" for object "i". Mass functions m_{ij} can be certain, probabilistic, possibilistic, evidential and even missing.

Clustering

- ECM: Evidential C-means;
- BKM: Belief C-modes;
- C: number of clusters Cl_k $(1 \leq k \leq C)$;
- N: number of objects to be classified;
- n_k: number of objects classified into cluster Cl_k;
- T: a set of n objects $o_i : 1 \leq i \leq n$ to classify;
- Q: cluster's center;
- Ω_c: frame of discernment of possible classes $\Omega_c = \{Cl_1, Cl_2, \ldots, Cl_c\}$;
- $\bar{v}_i$: barycenter of clusters classes;
- dis: dissimilarity of an object and a class mode;
- D: distance between an object and cluster's mode;
- s: similarity between two objects;

- S: similarity between an object and a cluster;
- q: index of objects;
- p: index of attributes.

Independence

- H_1, H_2: hypotheses;
- X, Y: stochastic variables, $x \subseteq \Omega_X$ and $y \subseteq \Omega_Y$ are possible values of X and Y;
- Ω_X, Ω_Y: domains of X and Y respectively;
- $\propto$: proportionality;
- m_A, m_B: mass functions induced by two distinct evidences;
- W: weighting function;
- Υ: angle between two mass functions considered as vectors in a $2^{|\Omega|}$ dimensional space;
- $\mathrm{I_d}(s_1, s_2)$: independence degree of s_1 on s_2;
- $\mathrm{Ind}(s_1, s_2)$: overall independence of s_1 and s_2;
- k_i, k_j, k_1, k_2: indexes of clusters;
- l, q: indexes for objects to classify;
- D: a distance of an object to cluster, $D(o_i, Cl_k)$ is the distance between the object o_i and the cluster Cl_k;
- N_k: number of objects classified into a cluster Cl_k;
- $Cl^i_{k_i}$: the k_j^{th} cluster of the source i, such that $1 \leq k_i \leq C$;
- $\beta^i_{k_i,k_j}$, $\beta^i(k_i, k_j)$: similarity between clusters $Cl^i_{k_i}$ and $Cl^j_{k_j}$;
- M_i: similarity matrix of the source s_i;
- $\mathcal{I}$: a frame of discernment of the independence, $\mathcal{I} = \{I,\ \bar{I}\}$ such that $\bar{I}$ is for dependent and I for independent hypotheses;
- $\mathcal{P}$: a frame of discernment on the independence, positive and negative independence of sources, $\mathcal{P} = \{I,\ P,\ \bar{P}\}$ such that I is for independence, P for positive dependence and $\bar{P}$ for negative dependence hypotheses;

- $m^{\mathcal{I},i}_{k_i,k_j}$: a mass function for the independence of linked clusters $Cl^i_{k_i}$ and $Cl^j_{k_j}$ according to s_i;
- $m^{\mathcal{I},i}$: a mass function of the independence of a source s_i;
- $m^{\mathcal{P},i}_{k_i,k_j}$: a mass function of the independence, positive and negative independence of matched clusters $Ck^i_{k_i}$ and $Ck^j_{k_j}$ according to s_i;
- $m^{\mathcal{P}}_i$: a mass function on a given source's (s_i) independence, positive and negative dependence;
- $\alpha^i_{k_i}$: a reliability degree taking into account the number of mass functions in a cluster $Cl^i_{k_i}$;
- $\mathrm{Conf}(Cl^i_{k_i}, Cl^j_{k_j})$: the conflict between matched clusters, $Cl^i_{k_i}$ of s_i and $Cl^j_{k_j}$ of s_j.

Combination rules

- PCR: Proportional Conflict Redistribution;
- $CWAC$: Combination With Adapted Conflict;
- $CCAC$: Cautious Combination With Adapted Conflict;
- m_{12}, m^{Ω}_{12}: a mass function issued from the combination of two distinct mass functions m^{Ω}_1 and m^{Ω}_2 using any combination rule;
- $m_{\oplus}$, $m_{1\oplus 2}$: the orthogonal sum of m_1 and m_2;
- $m_{\bigcirc\!\!\!\!\cup}$, $m_{1\bigcirc\!\!\!\!\cup 2}$: combined mass function with the disjunctive rule of combination;
- m_Y: combined mass function with Yager's rule;
- m_{DP}: combined mass function with Dubois and Prade's rule;
- $m_{\bigcirc\!\!\!\!\cap}$: the conjunctive combination of mass functions;
- $m_{\bigcirc\!\!\!\!\leftrightarrow}$: combined mass function with $CWAC$;
- $m_{\bigcirc\!\!\!\!\wedge}$: combined mass function with the unnormalized cautious rule;
- $m_{\bigcirc\!\!\!\!\wedge^*}$: normalized cautious combination of mass functions;
- $m_{\bigcirc\!\!\!\!\vee}$: bold combination of mass functions;
- $m_{\odot}$: combined mass function with $CCAC$;
- m_{Mean}: mean combination of mass functions;

- m_{Flo}: combination of mass functions with the rule proposed in (Florea et al., 2006; Florea, 2007);
- m_{Mixed}: proposed mixed combination that is a weighted average of the conjunctive and cautious combinations;
- $m^{\Omega}[I]$, $m^{\Omega}[P]$, $m^{\Omega}[\bar{P}]$: mass functions defined on any frame of discernment and conditioned on their source's independence I, positive P, and negative dependence $\bar{P}$;
- W_1, W_2, W: weighting functions;
- W': similarity functions;
- γ: sources' degree of independence (overall independence degree).

1
Introduction

Real-world problems are stained by imperfect information. Unfortunately, modeling these problems cannot be done with certainty because information may be imprecise, uncertain or even not available. Smithson advances a detailed analysis of imperfect data (Smithson, 1989). In fact, many areas are stained with uncertainty like machine learning, medical diagnosis, risk analysis, target recognition, etc. Imperfect information is either *inconsistent*, *imprecise* or *uncertain* (Smets, 1997).

Inconsistent information is incoherent and the value of the variable or the attribute is impossible and not consistent. *Uncertain* information is provided by a source that is not sure about their truth. For example, in the case of lack of information in medical diagnosis, a doctor may hesitate between two diagnoses and supplies an uncertain diagnosis which may be insufficient for decision making. On the other hand, information can be certain but we don't discern which subset is the truth. Information is *imprecise* when it is certain without discerning exactly the truth. In this case, a doctor may be sure about his diagnosis about which disease can be, but the disease may be one of a set of possible diseases. Imprecise information is certain but not exact. By misuse of language, uncertainty is used instead of imperfection due to their closeness. Information may be also *objective* or *subjective*. Objective information evolves out of measurements or objective sources. However, subjective information is collected from sources who are expressing their own opinions or beliefs.

Decision making is more difficult with the use of imperfect information. It may be easier to decide on certain information; however available data are not always entirely certain. Even though imperfect information cannot be avoided, we may find appropriate tools to cope with it.

Many theories have been proposed to manage imperfect information, such as fuzzy sets theory (Zadeh, 1965), possibility theory (Dubois and Prade, 1988; Zadeh, 1999), imprecise probabilities (Walley, 1991), rough set theory, credal set theory (Abellán and Moral, 2000), random set theory (Kendall, 1974) and the theory of belief functions (Dempster, 1967; Shafer, 1976). Although the theory of probabilities has been widely used to manage uncertainty, it cannot handle the case of total ignorance. In fact, when sources ignore partially or totally values of some parameters, hypotheses are equiprobable. Equiprobability handles both equiprobable hypotheses and ignorant sources. Ignorance is not well modeled with the theory of probabilities.

The theory of belief functions is a powerful tool for representing imperfect information that can be imprecise, uncertain, missing or incomplete. Usual databases are used to store a high quantity of structured data which are perfect. When uncertainty is modeled with the theory of belief functions, such information namely *evidential information* may be stored in evidential databases (Hewawasam et al., 2005; Bach Tobji, 2012). Since available information tends to be imperfect, the use of evidential databases supporting both certain and uncertain data is of a great interest.

In many fields such as medical diagnosis and banking, evidential databases may store similar information provided by different sources. For example, in medical diagnosis, several doctors provide same or different (evidential) information when diagnosing the same patient. In such case, each diagnosis is stored in the evidential database of the doctor. Integrating evidential databases helps decision makers when handling all available information. Taking into account several evidential databases to make a decision is considered as a difficult task; the use of only one integrated evidential database makes this task easier and more pleasant for decision makers. When integrating several evidential databases, evidential information are combined. The combination of several evidential information helps users and decision makers to reduce the degree of uncertainty by confronting several opinions. The theory of belief functions presents a strong framework for combination (Dubois and Prade, 1988; Murphy, 2000; Smets and Kennes, 1994; Martin and Osswald, 2007a; Yager, 1987). A conflict may arise during the combination due to the discord between different sources of evidential information. This conflict may be redistributed through combination rules. The conflict appears because of the unreliability of at least one source, which can be avoided before combining by taking into account sources' reliabilities. Discounting operator can be used to balance evidential information with their sources' degrees of reliability.

First, we tackle the problem of estimating sources' degrees of reliability. We propose a method which aims to estimate source's reliability degree using information stored in its database. This method generalizes the approach detailed in (Martin et al., 2008) in order to estimate reliabilities of evidential databases' sources. Indeed, source's reliability is estimated in order to discount its evidential information before combining them with other evidential information supplied by different sources. Once source's reliability is estimated, it can be used to discount all evidential information. The conflict defined to estimate reliability degrees is also used to learn independence degrees of sources.

Then, we detail our method of estimating sources' independence. The Oxford dictionary defines the term *dependence* as “the state of relying on or being controlled by someone or something else”; however, two sources are dependent if they are confident and one of them is controlling the other. The Oxford dictionary defines also the term *independence* as “the fact or state of being independent”. Nevertheless, two sources are independent when no one is controlling between each other. For example, in social networks, two persons are dependent if one of them is controlling the other or if he relies on and adopts opinions of the other. However, two persons are independent if no one is controlling the other and no one is adopting beliefs of the opponent. Social networks are a great example of dependence where users are linked and then they are dependent. In the theory of belief functions, belief holders may be experts, persons, algorithms, etc. When combining several evidential information stored in evidential databases, beliefs provided by several belief holders are aggregated by stressing common points in their beliefs.

In the theory of belief functions, many combination rules are proposed to aggregate beliefs. Some combination rules like (Dubois and Prade, 1988; Martin and Osswald, 2007a; Murphy, 2000; Smets and Kennes, 1994; Yager, 1987; Lefèvre and Elouedi, 2013) are fitted to the aggregation of evidential information provided by *cognitively independent* sources. Otherwise the *cautious*, *bold* (Denœux, 2008) and *mean* combination rules can be applied when sources are *cognitively dependent* because they are idempotent and tolerate redundant information. A source is assumed to be cognitively independent on another one when the knowledge of beliefs of that source, does not affect beliefs of the other one. Information on the independence of sources guides the choice of the type of combination rules to be used. For example, when beliefs are completely dependent, only cautious or bold rules can be used. When evidential information is completely independent, a set of combination rules (Dubois and Prade, 1988; Martin and Osswald, 2007a; Murphy, 2000; Smets and Kennes, 1994; Yager, 1987; Lefèvre and Elouedi, 2013) can be applied.

Some researches are focused on cognitive and evidential independence (Shafer, 1976; Smets, 1993); others (Ben Yaghlane et al., 2002a; Ben Yaghlane et al., 2002b) tackle doxastic independence of variables. Cognitive dependence is defined as the change of beliefs on one variable if a new evidence bears on the other variable. In this thesis, we are focusing on measuring only the independence of sources, not the independence of variables.

In this thesis, we suggest a novel statistical approach for estimating sources’ independence from all their evidential information. Evidential information is stored in evidential databases. There are three possible uses of sources’ degree of independence: First, it guides the choice of the combination rule to aggregate evidential information.

In fact, sources' degree of independence justifies hypothesis made when choosing the type of combination rules. These degrees may motivate a choice of combination rules that supposes the independence of sources or contrary, it may justify the choice of an idempotent combination rule.

Second, sources' degree of independence may be integrated in evidential information that they provide leading to an operator similar to the discounting. That combination integrates information on sources' independence in evidential information provided by these sources. Afterward, hypothesis on the independence of sources may be made to choose the appropriate type of combination rules.

Finally, we propose a new mixed combination rule. The proposed rule weighs the conjunctive and cautious rules with sources' degree of independence. When sources' degree of independence is neither 0 nor 1 but a level over $[0, 1]$, we propose a new combination rule weighted with sources' degree of independence leading to the conjunctive rule of combination when sources are fully independent (Smets, 1990) and to the cautious rule when they are fully dependent (Denœux, 2008).

Our thesis is organized in the following four chapters:

- In Chapter 2, we recall some basic concepts of the theory of belief functions. It proposes strong tools to model uncertain and/or imprecise information. Evidential information is provided by sources that can be distinct and independent or not. Many combination rules are also proposed in the theory of belief functions, they aggregate several sources beliefs. In addition, operations on frames of discernment are proposed to convert mass functions from any frame of discernment to a compatible (or not) frame. Converting several mass functions defined on different frames of discernment allows the combination of that mass functions.

- In Chapter 3, we define evidential databases and a conflict measure between sources of that databases. We also propose a method to a estimate source's reliability in order to resolve the conflict during the combination. We propose also a clustering algorithm that classifies objects stored in evidential databases. Clustering technique gather similar objects into the same cluster in order to study the source's overall behavior. The proposed clustering algorithm minimizes the conflict between objects in the same clusters; the conflict between objects in different clusters are maximized. The use of a clustering algorithm is of interest in next chapters.

- Chapter 4 is about independence concept in the theory of belief functions. Many researches are focused on variables' independence. Evidential, cognitive and doxastic independence of variables are defined in the theory of belief functions. This thesis does not focus on variables' independence but on sources' independence. In this chapter, we propose a method for learning sources' independence from all

evidential information. Two sets of evidential information assessed by two different sources are classified into two sets of clusters. Clusters of both sources are matched and the independence of each couple of matched clusters is quantified in order to estimate sources' degrees of independence.
We propose also a statistical approach to ascertain if the dependence is positive or negative. In the case of positive dependent sources, they are communicating or their evidential corpus is almost the same. In the case of negative dependent sources, sources are also either communicating or their evidential corpora are almost identical but one of the sources provides the opposite of information it knows. In this chapter, we propound a refinement of sources' independence degrees. Thus if sources are dependent, we learn sources' degrees of positive and negative dependence.

- Chapter 5 guides the choice of the type of combination rules according to sources' degrees of independence. Therefore in a case of dependent sources, only idempotent combination rules that tolerate redundant information may be used. Evidential information provided by independent sources can be combined by any combination rules that are not necessarily idempotent. In cases of strongly dependent or independent sources, the choice of combination rules is quite easy. However, in a case of an independence degree over $]0,1[$, the choice is not enough easy. Therefore, we propose a new combination rule that takes into account sources' degrees of independence. The proposed combination rule is a weighted average using sources' degrees of independence.
Another way to take consideration of sources' degrees of independence is to integrate it into mass functions provided by that sources. Justification of such combination is detailed in this chapter.

- Finally, in Chapter 6 conclusions are drawn and some perspectives of this thesis are presented.

2
Basics of the theory of Belief functions

Contents

Summary

Among theories handling uncertainty, *the theory of belied functions* detailed in this chapter deals with imprecise and uncertain information. This thesis handles sources cognitive independence when uncertainty is modeled with the theory of belief functions. In fact, sources cognitive independence is learned from all evidential information that they provide. In this first chapter, we display some basic notions of the theory of belief functions because it provides flexible tools to cope with uncertainty.

2.1 Introduction

Uncertainty theories like the *theory of probabilities*, the *theory of fuzzy sets* (Zadeh, 1965), the *theory of possibilities* (Dubois and Prade, 1988) and the *theory of belief functions* model and manage uncertain data. The theory of belief functions also called *Dempster-Shafer theory* was first introduced by (Dempster, 1967) and (Shafer, 1976) for quantifying beliefs. Thereafter (Smets and Kennes, 1994) proposed an interpretation of this theory: the Transferable Belief Model (TBM).

In the TBM, the representation of beliefs is on two levels: a *credal level*[1] and a *pignistic level*[2]. The credal level is split into a *static* part and a *dynamic* part. In the static part, beliefs are quantified and represented; they are combined in the dynamic part. In the pignistic level, decisions are made with regard to the risk and the earnings associated to these decisions.

The theory of belief functions is a mathematical theory that extends probability theory by giving up the additivity constraint as well as the equal probability in the case of ignorance. Therefore, in probability theory equal probabilities do not distinguish equally probable events from the case of ignorance. In the theory of belief functions, cases of *uncertainty*, *incompleteness* and *ignorance* are modeled and distinguished. In this theory, *justified degrees of support* are assessed according to an *evidential corpus*. Evidential corpus is the set of all evidential pieces of evidence held by a source that justifies degrees of support awarded to some subsets.

In the framework of belief functions, uncertainty is modeled and several pieces of evidence provided by different bodies of evidence are combined in order to have synthetic information that takes into account all pieces of evidence. Thus, this theory deals with imprecise and/or uncertain data provided by several belief holders and also combines them. Combining evidential information aggregates the beliefs of various sources by emphasizing common points in their faiths.

This chapter is a synopsis of basic concepts of the theory of belief functions; in the first section, we will introduce fundamental notions of the theory of belief functions such as mass functions, some particular cases, equivalent belief functions and pignistic transformation used for decision making. In the second section, we will detail operations on frames of discernment to be used to express mass functions in a common frame when they are initially defined on different, compatible or incompatible frames of discernment. The third section is about combination, thus we summarize some combination rules that

[1] from Latin "credo" means "I believe"

[2] from Latin "pignus" means "a bet"

will be used in the sequel but more details will be given in the last chapter of this thesis. Finally, before concluding, methods for building belief functions are displayed.

2.2 Belief functions

Let $\Omega = \{\omega_1,\ \omega_2,\ \ldots,\ \omega_n\}$ be a set of n elementary, non empty and mutually exclusive hypotheses related to a given problem. The set Ω is called *frame of discernment*, *universe of discourse* or *domain of reference*. Among Ω only one hypothesis is true. The theory of belief functions can be used to assess degrees of belief to some hypotheses when the true hypothesis cannot be defined with certainty.

Smets defined a *closed world assumption* where all possible hypotheses are enumerated in Ω; an *open world assumption* which admits the existence of a set of unknown hypotheses which can include the truth (Smets, 1988; Smets, 1990). Under the closed world assumption, we suppose that Ω is exhaustive. This assumption was admitted in Shafer's model (Dempster, 1967) but it is, in some cases, a bit difficult to enumerate from the beginning all the hypotheses related to a given problem, thus the use of the open world assumption. Therefore, Ω is not necessarily exhaustive under the open world assumption.

Let 2^Ω be a set of all subsets of Ω. It is made of hypotheses and unions of hypotheses from Ω. This set 2^Ω is called *power set* and defined as follows:

$$2^\Omega = \{A : A \subseteq \Omega\} \tag{2.1}$$

Subsets of 2^Ω are called *propositions* or *events*.

Example 2.1 Let us consider a problem of medical diagnosis where a doctor examines a patient, he can identify some diseases with uncertainty from identified symptoms. For example, suppose that possible diseases are either flu, pharyngitis or bronchitis. Therefore the frame of discernment Ω_d is formed of flu F, pharyngitis P and bronchitis B: $\Omega_d = \{F,\ P,\ B\}$.
The corresponding power set is: $2^{\Omega_d} = \{\emptyset,\ F,\ P,\ F\cup P,\ B,\ F\cup B,\ P\cup B,\ F\cup P\cup B\}$[3].
Along this thesis, this example will be used to illustrate some notions.

A *basic belief assignment* (*bba*) is a mapping from 2^Ω to $[0, 1]$ that allocates a degree of justified support over $[0, 1]$ to some subsets A of 2^Ω. A basic belief assignment also called *mass function* is held by an agent, a source or a belief holder and defined as follows:

[3]Set notations $F\cup P$, $F\cup B$, $P\cup B$ and $F\cup P\cup B$ are equivalent to $\{F, P\}$, $\{F, B\}$, $\{P, B\}$ and $\{F, P, B\}$

$$\begin{aligned} m^{\Omega}: \quad & 2^{\Omega} \rightarrow [0,1] \\ & \sum_{A \subseteq \Omega} m^{\Omega}(A) = 1 \end{aligned} \tag{2.2}$$

The degree of support of A, $m^{\Omega}(A)$, also called *basic belief mass* (*bbm*) or *mass* for short, is the degree of support that is committed exactly to A (the degree of belief that the true hypothesis is in A) justified by available information. If further information arises, that amount can be committed to subsets of A. A subset A having a non null mass is called a *focal element*. For example if $m^{\Omega_d}(F \cup B) = 0.2$, $F \cup B$ is a focal element.

The couple $(\mathcal{F}, m^{\Omega})$ is called *body of evidence* such that $\mathcal{F}$ is the set of all focal elements of a given mass function m^{Ω}.

The *core* φ of a mass function is the union of all its focal elements and is defined as follows:

$$\varphi = \bigcup_{A: m^{\Omega}(A) > 0 \ A \subset \mathcal{F}} A \tag{2.3}$$

Example 2.2 Let us take the same example of medical diagnosis, the doctor assigned these degrees of support when diagnosing a new patient:

$$m^{\Omega_d}(F) = 0.6, \ \ m^{\Omega_d}(P \cup B) = 0.4$$

The belief holder, which is the doctor in this example, believes that the disease of the patient is flu with a degree of support 0.6; it is a bronchitis or a pharyngitis with a degree of 0.4.
Note that F and $P \cup B$ are focal elements, $(\mathcal{F}, m^{\Omega_d})$ is the body of evidence such that $\mathcal{F} = \{F, \ P \cup B\}$ and $\varphi = \{F \cup P \cup B\}$ is the core of m^{Ω_d}.

A mass, $m^{\Omega}(A)$, assigned to a proposition A from 2^{Ω}, $A \subseteq \Omega$, represents explicitly the doubt between hypotheses in A. For example if $A = F \cup B$, $m^{\Omega}(F \cup B)$ is the degree of support of $F \cup B$ without supporting any subset of A (it does not support F nor B but it supports $F \cup B$). The mass, $m^{\Omega}(F \cup B)$, is the degree of belief on $F \cup B$ which cannot be committed to its subsets F and B if that transfer is *not justified*.

The degree of support $m^{\Omega}(\Omega)$ is the part of belief assigned to the whole frame of discernment which cannot be committed to its subsets. This degree of support represents the amount of *ignorance*.

The basic belief mass $m^{\Omega}(\emptyset)$ represents the degree of belief that is not committed to any subset, (Smets, 1992b) interpreted that amount as a degree of conflict or contradiction between evidences. That amount can be also interpreted as a degree of belief on an hypothesis non enumerated on the frame of discernment Ω, interpretations of that uncommitted degree of support is detailed in (Smets, 1992b; Smets, 2007). (Shafer, 1976) assumed a normality condition such that:

$$m^{\Omega}(\emptyset) = 0 \tag{2.4}$$

In this case the closed world assumption is admitted and such mass function is called a *normalized basic belief assignment*.
A *non-normalized mass function* may be transformed into a normalized one using an operator of normalization defined as follows:

$$\begin{cases} m^{\star\Omega}(A) = \dfrac{m^{\Omega}(A)}{1 - m^{\Omega}(\emptyset)} & \forall A \subseteq \Omega \\ m^{\star\Omega}(\emptyset) = 0 \end{cases} \tag{2.5}$$

Example 2.3 For the same frame of discernment $\Omega_d = \{F,\ P,\ B\}$. Suppose that a doctor assessed these degrees of support:

$$m^{\Omega}(\emptyset) = 0.2,\ \ m^{\Omega}(F) = 0.6,\ \ m^{\Omega}(F \cup B) = 0.2$$

The normalization of this mass function gives:

$$m^{\star\Omega_d}(\emptyset) = 0,\ \ m^{\star\Omega_d}(F) = 0.75,\ \ m^{\star\Omega_d}(F \cup B) = 0.25$$

Sets F and $F \cup B$ are focal elements, therefore $\mathcal{F} = \{F,\ F \cup B\}$ is the set of focal elements and $(\mathcal{F}, m^{\Omega})$ is the body of evidence.
The core of m^{Ω_d} is: $\varphi = \{F\} \cup \{F \cup B\} = \{F \cup B\}$.

(Shafer, 1976) required that the mass of the empty set is null ($m^{\Omega}(\emptyset) = 0$), therefore the normality condition is required in Shafer's model. Under a closed world assumption, the frame of discernment is supposed to be exhaustive where all the possible hypotheses are enumerated on Ω and thus $m^{\Omega}(\emptyset) = 0$. Smets proposed the open world where a positive mass can be allocated to the empty set because of the non exhaustivity of the frame of discernment or the combination of mass functions induced by contradicting evidences (Smets, 1988; Smets, 1992b).

2.2.1 Particular belief functions

Mass function is the common representation of evidential knowledge. Basic belief masses are degrees of support justified by available evidences. Other functions model exactly the same evidences. This section, is a synopsis of some particular mass functions.

Categorical mass functions

A *categorical mass function* is a normalized mass function which has a unique focal element $A^\star$. This mass function is noted $m^\Omega_{A^\star}$ and defined as follows:

$$m^\Omega_{A^\star}(A) = \begin{cases} 1 & \text{if } A = A^\star \subset \Omega \\ 0 & \forall A \subseteq \Omega \text{ and } A \neq A^\star \end{cases} \tag{2.6}$$

Example 2.4 Assume that the doctor identified symptoms showing that the patient caught either a pharyngitis or a bronchitis ($A^\star = \{P \cup B\}$), the corresponding mass function is a categorical belief function: $m^\Omega_{A^\star}(\{P \cup B\}) = 1$.

When all sources supply categorical mass functions, the theory of belief functions corresponds to the classical propositional logic.

We distinguish two particular cases of categorical mass functions: the *vacuous mass functions* when $A^\star = \Omega$ and the *contradictory mass functions* if $A^\star = \emptyset$.

Vacuous mass functions

A *vacuous mass function* is a particular categorical mass function focused on Ω. It means that a vacuous mass function is normalized and has a unique focal element which is Ω. This type of mass functions is defined as follows:

$$m^\Omega_{A^\star}(A) = \begin{cases} 1 & \text{if } A = \Omega \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

with $A^\star = \Omega$. Vacuous mass function emphasizes the case of *total ignorance.*

Example 2.5 Suppose that the doctor was unable to identify the disease of the patient, thus the situation of total ignorance arises. The mass function supplied by the doctor is vacuous and is defined as follows:

$$m^{\Omega_d}(\Omega_d) = 1, \;\; m^{\Omega_d}(A) = 0 \;\; \forall A \subset 2^{\Omega_d}, \;\; \forall A \neq \Omega_d$$

In this case, the doctor does not support any subset of 2^{Ω_d} and the unit is attributed to the whole frame of discernment.

Contradictory mass functions

A *contradictory mass function* is a particular categorical mass function focused on the empty set. A contradictory mass function is unnormalized, having a unique focal element which is the empty set "$\emptyset$". A contradictory mass function is defined as follows:

$$m^{\Omega}_{A^\star}(A) = \begin{cases} 1 & \text{if } A = \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

with $A^\star = \emptyset$. Such mass function emphasizes the case where the true hypothesis is not enumerated in Ω, therefore Ω is not exhaustive.

Example 2.6 Suppose that the doctor observed symptoms never seen before. Hence, the patient does not suffer from neither a flu, nor a pharyngitis nor a bronchitis ($A^\star = \emptyset$). The mass function supplied by this doctor will be a contradictory mass function defined as follows:

$$m^{\Omega_d}_{A^\star}(\emptyset) = 1, \; m^{\Omega_d}_{A^\star}(A) = 0 \quad \forall A \subset 2^{\Omega_d} \text{ and } A \neq \emptyset$$

In this case, the total degree of support is attributed to the empty set. No hypothesis from the frame of discernment is true, thus the true hypothesis is not enumerated in Ω_d. There is no hypothesis more possible than others. This mass function is defined under the open world assumption.

Dogmatic mass functions

A *dogmatic mass function* is a mass function where Ω is not a focal element. A dogmatic mass function is defined as follows:

$$m^{\Omega}(\Omega) = 0 \tag{2.9}$$

Mass functions of examples 2.2 and 2.6 are dogmatic.

Bayesian mass functions

A *Bayesian mass function* is a mass function which all focal elements are elementary hypotheses; it is defined as follows:

$$\begin{cases} m^{\Omega}(A) \in]0,1] & \text{if } |A| = 1 \\ m^{\Omega}(A) = 0 & \text{otherwise} \end{cases} \tag{2.10}$$

As all focal elements are single points, this mass function is a *probability distribution*. Figure 2.1 illustrates focal elements of a Bayesian mass function where A, B, C and D are single elements.

Example 2.7 Suppose that a doctor assigned these masses when diagnosing a new patient:

$$m^{\Omega_d}(F) = 0.2, \;\; m^{\Omega_d}(P) = 0.5, \;\; m^{\Omega}(B) = 0.3$$

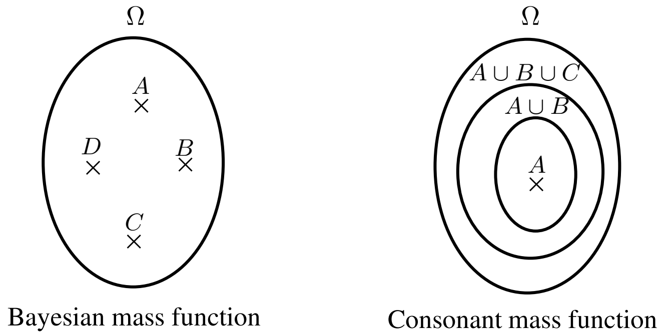

Figure 2.1: Focal elements of bayesian and consonant mass functions

This mass function is bayesian and the corresponding probability distribution is:

$$p(F) = 0.2,\ \ p(P) = 0.5 \text{ and } p(B) = 0.3$$

Consonant mass functions

A *consonant mass function* is a mass function which focal elements are nested ($A_1 \subset A_2 \subset \ldots \subset \Omega$). Figure 2.1 illustrates focal elements of consonant mass functions.

Example 2.8 This following mass function is consonant:

$$m^{\Omega_d}(F) = 0.2,\ \ m^{\Omega_d}(F \cup P) = 0.4,\ \ m^{\Omega_d}(F \cup P \cup B) = 0.4$$

This mass function has common characteristics with possibilities.

Certain mass functions

A *certain mass function* is a categorical mass function (a mass supporting a unique focal element) such that its focal element is an elementary hypothesis. This mass function emphasizes the case of total certainty as the source supports only one hypothesis with certainty. Certain mass function is defined as follows:

$$m^{\Omega}(A) = \begin{cases} 1 & \text{if } A = \omega \in \Omega \\ 0 & \forall A \subseteq \Omega \text{ and } A \neq \omega \end{cases} \tag{2.11}$$

Example 2.9 The doctor is sure that the patient has a flu therefore:

$$m^{\Omega_d}(F) = 1$$

Simple support functions

A *simple support function* is a mass function which has only one focal element other than the frame of discernment Ω. This unique focal element is called *the focus of the simple support function.* A simple support function is defined in (Shafer, 1976) and (Smets, 1995) as follows:

$$m^{\Omega}(B) = \begin{cases} w & \text{if } B = \Omega \\ 1 - w & \text{if } B = A \text{ for some } A \subset \Omega \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

where A is the *focus* of the simple support function and $w \in [0, 1]$.
A simple support function is also noted A^w where w is the degree of support of the frame of discernment Ω and the complement of w to 1 is the degree of support of the focus A. Figure 2.2 is an example of focal elements of a simple support function focused on A.

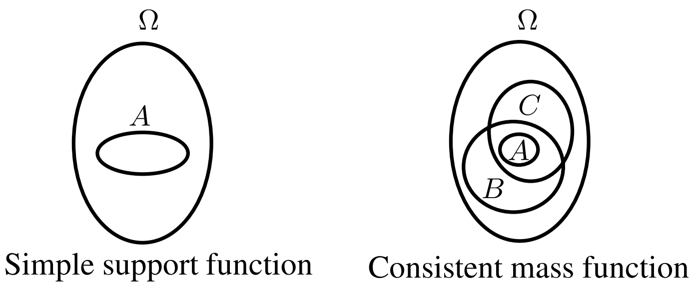

Figure 2.2: Focal elements of consistent and simple support mass functions

Example 2.10 Suppose the frame of discernment $\Omega_d = \{F, P, B\}$ and assume a mass function m^{Ω_d} defined on Ω_d:

$$m^{\Omega_d}(P \cup B) = 0.6, \;\; m^{\Omega_d}(\Omega_d) = 0.4$$

m^{Ω_d} is a simple support function focused on $\{P \cup B\}$, it can also be noted $\{P \cup B\}^{0.4}$.

Consistent mass functions

A *consistent mass function* is a function which all focal elements have a non empty intersection. For such mass functions, at least one focal element is common to all the focal ones. Figure 2.2 illustrates a case of consistent mass function where all focal elements $\{A,\ B,\ C\}$ intersect.

Example 2.11 The following mass function is an example of a consistent mass function defined on Ω_d:

$$m^{\Omega_d}(P) = 0.3,\ \ m^{\Omega_d}(P \cup B) = 0.5,\ \ m^{\Omega_d}(P \cup F) = 0.2$$

Note that $P \cap \{P \cup B\} \cap \{P \cup F\} = P$.

2.2.2 Transformations of belief functions

Other functions related to mass functions model differently the same pieces of evidence. These functions are used, amongst others, to simplify computations. They are also mappings from 2^Ω to $[0,1]$.

Belief (or credibility) function

A *belief function*, noted bel^Ω, is the minimal degree of belief justified by available information. Although mass functions measure the belief committed *exactly* to some subsets A from 2^Ω, the credibility of a subset, $bel^\Omega(A)$, is the *total* belief on A. To compute the total belief on A, the masses of proper subsets B of A, $m^\Omega(B)$, must be summed to $m^\Omega(A)$. Therefore, $bel^\Omega(A)$ is obtained by summing masses of subsets of A. The belief function is defined by:

$$\begin{aligned} \mathrm{bel}^\Omega :\ \ & 2^\Omega \to [0,1] \\ & \mathrm{bel}^\Omega(A) = \sum_{B \subseteq A, B \neq \emptyset} m^\Omega(B) \end{aligned} \tag{2.13}$$

Furthermore, the mass function that produces a given belief function is unique and therefore it can be recovered from the belief function as follows:

$$\begin{cases} m^\Omega(A) = \displaystyle\sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} \mathrm{bel}^\Omega(B) & \forall A \subset \Omega,\ A \neq \emptyset \\ m^\Omega(\emptyset) = 1 - \mathrm{bel}^\Omega(\Omega) \end{cases} \tag{2.14}$$

$\bar{A}$ is the complement of A in Ω. As the empty set is included in both of A and $\bar{A}$, it is discarded from the sum.

Properties

- *Sub-additivity:* $\mathrm{bel}^\Omega(A) + \mathrm{bel}^\Omega(\bar{A}) \leq 1$.
- *Monotonicity:* $A \subseteq B \Rightarrow \mathrm{bel}^\Omega(A) \leq \mathrm{bel}^\Omega(B)$.
- $\mathrm{bel}^\Omega(A) + \mathrm{bel}^\Omega(B) \leq \mathrm{bel}^\Omega(A \cup B)$.
- $\mathrm{bel}^\Omega(\Omega) = 1$ and $\mathrm{bel}^\Omega(\emptyset) = 0$ under the closed world assumption, only $\mathrm{bel}^\Omega(\Omega) \leq 1$ (or equivalently $\mathrm{bel}^\Omega(\emptyset) \neq 0$) is required under the open world assumption.
- $\mathrm{bel}^\Omega(A_1 \cup A_2 \cup \ldots \cup A_z) \geq \sum_i \mathrm{bel}^\Omega(A_i) - \sum_{i>j} \mathrm{bel}^\Omega(A_i \cap A_j) - \ldots - (-1)^z$ $\mathrm{bel}^\Omega(A_1 \cap A_2 \cap \ldots \cap A_z)$.

Plausibility function

The *plausibility function*, also noted pl^Ω, is the maximum amount of *potential support* that could be given to a subset. The plausibility of an event $A \subseteq \Omega$, $\mathrm{pl}^\Omega(A)$, is the *maximum* amount of belief that *could be* given to A. It is measured by summing masses of propositions compatible with A. plausibility function is defined as follows:

$$\begin{aligned} \mathrm{pl}^\Omega : \quad & 2^\Omega \rightarrow [0,1] \\ & \mathrm{pl}^\Omega(A) = \sum_{A \cap B \neq \emptyset,\ B \subseteq \Omega} m^\Omega(B) \end{aligned} \tag{2.15}$$

Also:

$$\mathrm{pl}(A) = 1 - \mathrm{bel}(\bar{A}) \tag{2.16}$$

The mass function that produces a given plausibility function is unique and therefore it can be recovered as follows:

$$\begin{cases} m^\Omega(A) = \sum_{A \subseteq B} (-1)^{|B|-|A|+1} \mathrm{pl}^\Omega(\bar{A}) \\ m^\Omega(\emptyset) = 1 - \mathrm{pl}^\Omega(\Omega) \end{cases} \tag{2.17}$$

Properties

- *Over-additivity:* $\mathrm{pl}^\Omega(A) + \mathrm{pl}^\Omega(\bar{A}) \geq 1$
- *Monotonicity:* $A \leq B \Rightarrow \mathrm{pl}^\Omega(A) \leq \mathrm{pl}^\Omega(B)$
- $\mathrm{pl}^\Omega(A \cup B) \leq \mathrm{pl}^\Omega(A) + \mathrm{pl}^\Omega(B)$
- $\mathrm{bel}^\Omega(A) \leq \mathrm{pl}^\Omega(A)$

Under the closed world assumption, $m(\emptyset) = 0$ and $\mathrm{bel}(\Omega) = \mathrm{pl}(\Omega) = 1$ but with the open world assumption the mass $m(\emptyset)$ can be viewed as a missing mass or a not committed mass equal to $1 - \mathrm{pl}(\Omega)$.

Implicability function

The *implicability function*, b^{Ω}, is also related to a mass function but it has no interpretation. It is used to simplify computations. The *implicability function* is defined as follows:

$$\begin{aligned} \mathrm{b}^{\Omega}: \quad & 2^{\Omega} \rightarrow [0,1] \\ & \mathrm{b}^{\Omega}(A) = \sum_{B \subseteq A} m^{\Omega}(B) \end{aligned} \tag{2.18}$$

and:

$$\mathrm{b}^{\Omega}(A) = \mathrm{bel}^{\Omega}(A) + m^{\Omega}(\emptyset) \tag{2.19}$$

The mass function that produces an implicability function is unique and can be recovered as follows:

$$m^{\Omega}(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} b^{\Omega}(B), \, \forall A \subseteq \Omega \tag{2.20}$$

Commonality function

The *commonality function*, q^{Ω}, has also computational uses but has no interpretation. It is defined as follows:

$$\begin{aligned} \mathrm{q}^{\Omega}: \quad & 2^{\Omega} \rightarrow [0,1] \\ & \mathrm{q}^{\Omega}(A) = \sum_{A \subseteq B} m^{\Omega}(B) \end{aligned} \tag{2.21}$$

The mass function that produces a commonality function is unique and can be recovered as follows:

$$m^{\Omega}(A) = \sum_{A \subseteq B} (-1)^{|B|-|A|} \mathrm{q}^{\Omega}(B), \, \forall A \subseteq \Omega \tag{2.22}$$

Example 2.12 Table 2.1, is an example of a mass function provided by a doctor with the corresponding belief, plausibility, implicability and commonality functions.

Table 2.1: Mass, belief, plausibility, implicability and commonality functions

	$\mathbf{m}^{\mathbf{\Omega_d}}$	$\mathrm{bel}^{\mathbf{\Omega_d}}$	$\mathrm{pl}^{\mathbf{\Omega_d}}$	$\mathbf{b}^{\mathbf{\Omega_d}}$	$\mathrm{q}^{\mathbf{\Omega_d}}$
$\emptyset$	0	0	0	0	1
F	0.05	0.05	0.54	0.05	0.54
P	0.3	0.3	0.88	0.3	0.88
$F \cup P$	0.2	0.55	0.94	0.55	0.48
B	0.06	0.06	0.45	0.06	0.45
$F \cup B$	0.01	0.12	0.7	0.12	0.29
$P \cup B$	0.1	0.46	0.95	0.46	0.38
Ω_d	0.28	1	1	1	0.28

2.2.3 Pignistic transformation

In the credal level, degrees of belief are assessed and mass functions can be combined. In the pignistic level, decisions are made according to a criteria. One criteria for decision-making consists on choosing the most probable hypothesis from Ω. Decision-making is done on the basis of pignistic probabilities (Smets, 2005) noted $BetP_m^{\Omega}$ calculated from m^{Ω} for each hypothesis ω_i from Ω.

The *pignistic transformation* consists on dividing the mass attributed to a proposition A on the hypotheses which train it. The pignistic transformation is a mapping from Ω to $[0, 1]$ defined as follows:

$$\begin{aligned} BetP_m^{\Omega}\colon\quad & \Omega \to [0,1] \\ & BetP_m^{\Omega}(\omega_i) = \omega_i \mapsto \frac{1}{(1 - m^{\Omega}(\emptyset))} \times \sum_{B \subseteq \Omega,\ \omega_i \in B,\ B \neq \emptyset} \frac{m^{\Omega}(B)}{|B|} \end{aligned} \tag{2.23}$$

Decision is made according to the maximum of pignistic probabilities.

Example 2.13 To finish with the same problem described in example 2.1, suppose that the doctor gave this mass function:

$$m^{\Omega_d}(F) = 0.2,\ \ m^{\Omega_d}(P) = 0.4,\ \ m^{\Omega_d}(B \cup P) = 0.3,\ \ m^{\Omega_d}(\Omega_d) = 0.1$$

The doctor can decide about a patient's disease after the pignistic transformation of the supplied mass function:

$$BetP\{F\} \simeq 0.23, BetP\{P\} \simeq 0.58, BetP\{B\} \simeq 0.19$$

The patient seems to have a pharyngitis.

2.3 Common space

Before assessing degrees of support, hypotheses have to be enumerated in order to have an exhaustive frame of discernment. As Shafer said ((Shafer, 1976), chapter 6), "*a single frame of discernment can embody only a small subset of the immense collection of concepts and distinctions that any thinker can call to his aid*", different but compatible frames of discernment embody different and compatible collections.

2.3.1 Compatible frames of discernment: coarsening and refinement

The idea is to obtain a frame of discernment Ω from another frame of discernment Θ by splitting or merging some or all subsets of Θ. *Coarsening* and *refinement* concepts are defined to establish relationships between different and compatible frames of discernment in order to express beliefs on any one of them.

Let Θ and Ω, two different and compatible frames of discernment. The set Ω is a refinement of Θ if it is obtained by splitting all or some hypotheses from Θ (Shafer, 1976). Conversely, Θ is a coarsening of Ω obtained by grouping together hypotheses of Ω. Let us define a refining σ which is a mapping from $2^{\Theta} \rightarrow 2^{\Omega}$ satisfying:

$$\begin{cases} \sigma(\theta) \neq \emptyset & \forall\, \theta \in \Theta \\ \sigma(\theta) \cap \sigma(\theta') = \emptyset & \text{if } \theta \neq \theta' \\ \bigcup\limits_{\theta \in \Theta} \sigma(\theta) = \Omega & \end{cases} \tag{2.24}$$

For each $\theta \in \Theta$, $\sigma(\theta)$ is obtained by splitting the elements of θ in Ω (Shafer, 1976). The set Ω is a *refinement* of Θ and Θ is a *coarsening* of Ω. Figure 2.3 emphasizes the frames of discernment Ω and Θ where Θ is a coarsening of Ω and Ω is a refinement of Θ.

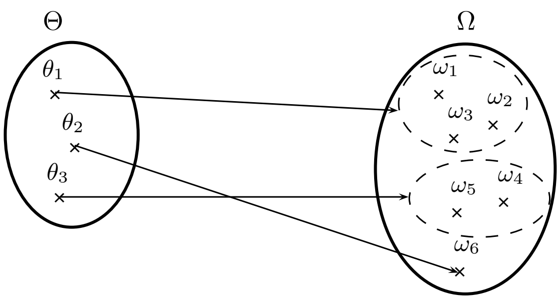

Figure 2.3: Coarsening Θ of Ω and refinement σ of Θ

Example 2.14 Let us illustrate with the same example 2.1, Θ = {Flu, Pharyngitis, Bronchitis}. A possible refinement of Θ is:
Ω = {Flu type A, Flu type B, Flu type C, *Pharyngitis*, *Bronchitis*}.
where:

σ(Flu) = {Flu type A, Flu type B, Flu type C}

σ(Pharyngitis) = {Pharyngitis}

σ(Bronchitis) = {Bronchitis}

Also, Θ is a *coarsening* of Ω.

2.3.2 Product space

In some applications, pieces of evidence may be defined on different frames of discernment. To assess flexibly justified degrees of support in different frames, some tools provide the redefinition of these pieces under a common space. Suppose that $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{n_1}\}$ and $\Theta = \{\theta_1, \theta_2, \ldots, \theta_{n_2}\}$, two different frames of discernment. The frame of discernment $\Omega \times \Theta$ is composed of the *Cartesian product* of Ω and Θ (Shafer, 1976), $\Omega \times \Theta$ is the product space given as follows:

$$\Omega \times \Theta = \{(\omega_1, \theta_1),\ (\omega_1, \theta_2),\ \ldots,\ (\omega_1, \Theta),\ \ldots,\ (\Omega, \Theta)\} \tag{2.25}$$

In this section, the transformations of mass functions in different spaces $\Omega \times \Theta$, Ω and Θ are detailed.

Vacuous extension

The *vacuous extension* (Smets, 1993) is a tool to extend a mass function defined on a frame of discernment Ω (or Θ) to the product frame $\Omega \times \Theta$. How to express a mass function m^Ω on the product space $\Omega \times \Theta$?

The vacuous extension, noted $\uparrow$, consists on a transfer of basic belief masses of each focal element B to its *cylindrical extension*[4] as follows:

$$m^{\Omega\uparrow\Omega\times\Theta}(A) = \begin{cases} m^\Omega(B) & \text{if } A = B \times \Theta,\ B \subseteq \Omega \\ 0 & \text{otherwise} \end{cases} \tag{2.26}$$

Figure 2.4 emphasizes the vacuous extension of m^Ω where the mass of a focal element B is transferred to its cylindrical extension.

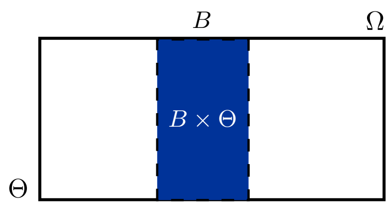

Figure 2.4: Vacuous extension of m^Ω in the finer frame $\Omega \times \Theta$

Example 2.15 Let $\Omega_d = \{flu\ F,\ pharyngitis\ P,\ bronchitis\ B\}$ and $\Omega_I = \{severe\ S,$ $moderate\ M,\ chronic\ C\}$ be two frames of discernment and the product frame

[4] $B \times \Theta$ is the cylindrical extension of B

$\Omega_d \times \Omega_I = \{(F,S),\ (F,M),\ (F,C),\ (P,S),\ (P,M),\ (P,C),\ (B,S),\ (B,M),\ (B,C)\}$ is schematically represented by figure 2.5. The doctor provides the following mass function:

$$m^{\Omega_d}(F) = 0.5,\ \ m^{\Omega_d}(\Omega_d) = 0.5$$

To extend m^{Ω_d} from Ω_d to $\Omega_d \times \Omega_I$, the mass of each focal element is transferred to its cylindrical extension. The joint mass function $m^{\Omega_d \uparrow \Omega_d \times \Omega_I}$ is given as follows:
$m^{\Omega_d \uparrow \Omega_d \times \Omega_I}(F, \Omega_I) = 0.5$, $m^{\Omega_d \uparrow \Omega_d \times \Omega_I}(\Omega_d, \Omega_I) = 0.5$.

Ω_d			
F	(F,S)	(F,M)	(F,C)
P	(P,S)	(P,M)	(P,C)
B	(B,S)	(B,M)	(B,C)
	S	M	C Ω_I

Figure 2.5: Product space $\Omega_I \times \Omega_d$

Marginalization

Marginalization is the inverse operation that expresses a mass function defined on the product space $\Omega \times \Theta$, $m^{\Omega\times\Theta}$, in the coarser frame Ω or Θ. The marginalization, noted $\downarrow$, transfers basic belief masses of each focal element $B \subseteq \Omega \times \Theta$ to its *projection*[5] on Ω or Θ as follows:

$$m^{\Omega\times\Theta\downarrow\Omega}(A) = \sum_{\{B\subseteq\Omega\times\Theta,\ B\downarrow\Omega = A\}} m^{\Omega\times\Theta}(B), \qquad \forall A \subseteq \Omega \qquad (2.27)$$

Figure 2.6 emphasizes the marginalization of $m^{\Omega\times\Theta}$ where the mass of a focal element B is transferred to its projection on Ω.

Example 2.16 Suppose the following mass function defined on the joint frame $\Omega_d \times \Omega_I$:

$$m^{\Omega_d\times\Omega_I}((F,M)\cup(P,C)) = 0.2,\ \ m^{\Omega_d\times\Omega_I}(B,C) = 0.3,\ \ m^{\Omega_d\times\Omega_I}(\Omega_d,\Omega_I) = 0.5$$

To have the marginal mass function $m^{\Omega_d\times\Omega_I\downarrow\Omega_d}$, the mass of each focal element is

[5] $A = B \downarrow \Omega$ is the projection of B on Ω

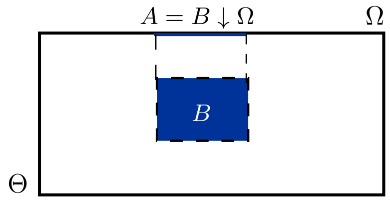

Figure 2.6: Marginalization of m^{Ω} on Ω

transferred to its projection on Ω_d as follows:

$$m^{\Omega_d \times \Omega_I \downarrow \Omega_d}(F \cup P) = 0.2, \;\; m^{\Omega_d \times \Omega_I \downarrow \Omega_d}(B) = 0.3, \;\; m^{\Omega_d \times \Omega_I \downarrow \Omega_d}(\Omega_d) = 0.5$$

2.4 Methods for merging belief functions

Combination of several belief functions aggregates beliefs of several bodies of evidence, that are induced by different evidences. Some combination rules work with a strong assumption of bodies of evidence independence. Other combination rules tolerate redundancy and combine beliefs induced by dependent bodies of evidence. In some particular cases, combined mass function can be certain which yields to the conditioning.

2.4.1 Some combination rules

There are a great number of combination rules proposed in the framework of belief functions (Dempster, 1967; Yager, 1987; Dubois and Prade, 1988; Smets, 1990; Denœux, 2006a; Martin and Osswald, 2007a). This section is a synopsis of combination rules that will be used, all combination rules will be detailed in Chapter 5. Combination rules merge a set of mass functions into only one mass function in order to summarize them and facilitate decision-making.

Let s_1 and s_2 be two *distinct* and *cognitively independent* sources providing two different mass functions m_1^{Ω} and m_2^{Ω} defined on the same frame of discernment Ω. Combining these mass functions induces a third one m_{12}^{Ω} defined on the same frame of discernment Ω.

The first combination rule (Dempster, 1967) was proposed by Dempster and defined for two mass functions m_1^Ω and m_2^Ω induced by two distinct bodies of evidence as follows:

$$m_{1\oplus 2}^\Omega(A) = (m_1^\Omega \oplus m_2^\Omega)(A) = \begin{cases} \dfrac{\sum\limits_{B\cap C=A} m_1^\Omega(B) \times m_2^\Omega(C)}{1 - \sum\limits_{B\cap C=\emptyset} m_1^\Omega(B) \times m_2^\Omega(C)} & \forall A \subseteq \Omega,\ A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases} \tag{2.28}$$

The basic belief mass of the empty set is null ($m^\Omega(\emptyset) = 0$), therefore this rule verifies the normality condition and works under the closed world assumption. Note that this combination rule is applied to combine the mass functions, cores of which intersect.

In order to solve the problem enlightened by Zadeh's counter example (Zadeh, 1984) where Dempster's rule of combination produced unsatisfactory results, many combination rules appeared. (Smets, 1990) proposed an open world where a positive mass can be allocated to the empty set interpreted as the non exhaustivity of the frame of discernment. Therefore the *conjunctive rule of combination* for two mass functions m_1^Ω and m_2^Ω also induced by two distinct bodies of evidence is defined in (Smets, 1990) as follows:

$$m_{1\bigcirc\!\!\!\!\cap\, 2}^\Omega(A) = (m_1^\Omega \bigcirc\!\!\!\!\cap\, m_2^\Omega)(A) = \sum_{B\cap C=A} m_1^\Omega(B) \times m_2^\Omega(C) \tag{2.29}$$

Even if Smets interpreted the basic belief mass, $m^\Omega(\emptyset)$, as the amount of conflicts between evidences that induced m_1^Ω and m_2^Ω (Smets, 2007), that amount is not really a conflict because it includes some degree of auto-conflict due to the non idempotence of the conjunctive combination (Martin et al., 2008). The conflict and auto-conflict will be detailed in the next chapter.

Commonality function is used to simplify the conjunctive combination especially when a great number of belief functions (N mass functions with $N > 2$) is involved in the combination. The combined mass function is obtained by computing the commonality function from each mass function using equation (2.21) then, multiplying all these commonalities as follows:

$$\mathrm{q}^\Omega(A) = \prod_{j=1}^{N} \mathrm{q}_j^\Omega(A) \qquad\qquad \forall A \subseteq \Omega \tag{2.30}$$

Finally, combined mass function is obtained by converting back the multiplied commonality function to a mass function using equation (2.22).

Example 2.17 In table 2.2, three mass functions are given by three different doctors when diagnosing the same patient (example 2.2), these three mass functions are combined using commonality functions.

Table 2.2: Conjunctive combination

	$\mathbf{m_1^{\Omega_d}}$	$\mathbf{m_2^{\Omega_d}}$	$\mathbf{m_3^{\Omega_d}}$	$\mathrm{q_1^{\Omega_d}}$	$\mathrm{q_2^{\Omega_d}}$	$\mathrm{q_3^{\Omega_d}}$	$\mathrm{q^{\Omega_d}}$	$\mathbf{m^{\Omega_d}}$
$\emptyset$	0	0	0	1	1	1	1	0.793
F	0.05	0.3	0.5	0.54	0.6	0.5	0.162	0.162
P	0.3	0.4	0	0.88	0.7	0	0	0
$F \cup P$	0.2	0.1	0	0.48	0.3	0	0	0
B	0.06	0	0.5	0.45	0.2	0.5	0.045	0.045
$F \cup B$	0.01	0	0	0.29	0.2	0	0	0
$P \cup B$	0.1	0	0	0.38	0.2	0	0	0
Ω_d	0.28	0.2	0	0.28	0.2	0	0	0

The conjunctive rule is used only when both sources are *reliable*. (Smets, 1990) proposed also to use the *disjunctive rule of combination* when only one source is unreliable[6]. The disjunctive rule of combination of two mass functions m_1^Ω and m_2^Ω is defined as follows:

$$m_{1 \textcircled{\cup} 2}^\Omega(D) = (m_1^\Omega \textcircled{\cup} m_2^\Omega)(D) = \sum_{B \cup C = D} m_1^\Omega(B) \times m_2^\Omega(C) \qquad (2.31)$$

Like the use of commonality functions for the conjunctive combination, implicability function is used to simplify the disjunctive combination. The combined mass function is obtained by computing the implicability function from each mass function using equation (2.18) then, multiplying all these implicabilities as follows:

$$b^\Omega(A) = \prod_{j=1}^{N} b_j^\Omega(A) \qquad \forall A \subseteq \Omega \qquad (2.32)$$

Finally, combined mass function is obtained using equation (2.20).

Example 2.18 In table 2.3, the same mass functions as in table 2.2 are combined with the disjunctive rule of combination.

Finally, the *mean combination rule*, m_{Mean}^Ω, of two mass functions m_1^Ω and m_2^Ω is the average of these ones. Therefore, for each focal element A of N mass functions, the combined one is defined as follows:

$$m_{Mean}^\Omega(A) = \frac{1}{N} \sum_{i=1}^{N} m_i^\Omega(A) \qquad (2.33)$$

Besides the idempotence of this combination rule, it verifies normality condition ($m^\Omega(\emptyset) = 0$) if combined mass functions are normalized ($\forall i \in N,\ m_i^\Omega(\emptyset) = 0$). We note

[6] No information about which source is the unreliable one.

Table 2.3: Disjunctive combination

	$\mathbf{m_1^{\Omega_d}}$	$\mathbf{m_2^{\Omega_d}}$	$\mathbf{m_3^{\Omega_d}}$	$\mathbf{b_1^{\Omega_d}}$	$\mathbf{b_2^{\Omega_d}}$	$\mathbf{b_3^{\Omega_d}}$	$\mathbf{b^{\Omega_d}}$	$\mathbf{m^{\Omega_d}}$
$\emptyset$	0	0	0	0	0	0	0	0
F	0.05	0.3	0.5	0.05	0.3	0.5	0.0075	0.0075
P	0.3	0.4	0	0.3	0.4	0	0	0
$F \cup P$	0.2	0.1	0	0.55	0.8	0.5	0.22	0.2125
B	0.06	0	0.5	0.06	0	0.5	0	0
$F \cup B$	0.01	0	0	0.12	0.3	1	0.036	0.0285
$P \cup B$	0.1	0	0	0.46	0.4	0.5	0.092	0.092
Ω_d	0.28	0.2	0	1	1	1	1	0.6595

also that this combination rule is commutative but not associative. However, it does not matter, because many mass functions can be combined at once.

All combination rules described above work under a strong assumption of cognitive independence since they are used to combine mass functions induced by two distinct bodies of evidence. This strong assumption is always assumed but never verified. (Denœux, 2008) proposed a family of conjunctive and disjunctive rules based on triangular norms and conorms. The *cautious* and *bold* rules of combination are members of that family and combine mass functions for which independence assumption is not verified.

2.4.2 Canonical decomposition

Shafer ((Shafer, 1976), chapter 4) distinguished four types of mass functions. *Simple support function* is a mass function supporting homogeneous evidences where a given subset on the frame of discernment is supported. *Separable support function* includes simple support functions as well as their combination with Dempster's rule of combination. *Support function* is obtained by coarsening the frame of discernment of separable support functions and finally, *belief function* includes simple support functions, separable support functions, support functions and non-dogmatic belief functions. Schematically, belief functions are modeled in (Shafer, 1976) as follows:

$$\left\{\begin{array}{l}\text{Simple}\\\text{support}\\\text{function}\end{array}\right\} \subset \left\{\begin{array}{l}\text{Separable}\\\text{support}\\\text{function}\end{array}\right\} \subset \left\{\begin{array}{l}\text{Support}\\\text{function}\end{array}\right\} \subset \left\{\begin{array}{l}\text{Belief}\\\text{function}\end{array}\right\}$$

Table 2.4: Canonical conjunctive decomposition of a non-dogmatic mass function

$\mathbf{2^{\Omega}}$	$\mathbf{m^{\Omega_d}}$	q^{Ω_d}	$\mathbf{w^{\Omega_d}}$
$\emptyset$	0	1	4/3
F	0.5	0.8	3/8
P	0.2	0.5	3/5
$F \cup P$	0.2	0.3	1/3
B	0	0.1	1
$F \cup B$	0	0.1	1
$P \cup B$	0	0.1	1
Ω_d	0.1	0.1	1

(Smets, 1995) distinguished *n-separable*[7] support functions from *u-separable*[8] support functions. A mass function $m^{\Omega\star}$ is a n-separable support function if:

$$m^{\Omega\star} = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w(A)} \tag{2.34}$$

A mass function m^{Ω} is u-separable support function if:

$$m^{\Omega} = \bigcirc\!\!\!\!\cap_{\emptyset \neq A \subset \Omega} A^{w(A)} \tag{2.35}$$

A simple support function is a *non-dogmatic mass function* which supports only one subset of the frame of discernment as defined in section 2.2.1. The mass function $A^{w(A)}$ is a simple support function focused on A. Shafer named this representation of separable support functions the *canonical decomposition.* Decompositions of equations (2.34) and (2.35) are unique as long as m^{Ω} is *non-dogmatic.* The canonical conjunctive decomposition (Smets, 1995) of a non-dogmatic mass function m^{Ω} is unique. The weights of evidence $w(A)$ are given from commonalities as follows:

$$\omega(A) = \prod_{B \subseteq A} \mathrm{q}(B)^{(-1)^{|B|-|A|+1}} \tag{2.36}$$

Example 2.19 Table 2.4 illustrates an example of calculation of the canonical conjunctive decomposition.

The *cautious combination* (Denœux, 2008) of two mass functions m_1^{Ω} and m_2^{Ω} issued from probably dependent sources is defined as follows:

$$m_1^{\Omega} \bigcirc\!\!\!\!\wedge\, m_2^{\Omega} = \bigcirc\!\!\!\!\cap_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)} \tag{2.37}$$

[7] n for normalized.
[8] u for unnormalized

Where $A^{w_1(A)}$ and $A^{w_2(A)}$ are simple support functions focused on A with weights w_1 and w_2 issued from the canonical decomposition (Smets, 1995) of m_1^Ω and m_2^Ω respectively, note also that $\wedge$ is the *min operator* of simple support functions weights. When, the min operator $\wedge$, is replaced by, the *max operator* $\vee$, the *bold combination rule* is obtained (Denœux, 2008). Both cautious and bold rules combine mass functions issued from dependent sources, but the cautious rule is more fitted to reliable sources, otherwise the bold rule fits to unreliable ones. Both bold and cautious combination rules are commutative, associative and idempotent.

Example 2.20 In table 2.5, we illustrate the combination of two mass functions with Dempster's, conjunctive, disjunctive, mean and cautious rules.

Table 2.5: Combining using different combination rules

	$\mathbf{m_1^\Omega}$	$\mathbf{m_2^\Omega}$	$\mathbf{m_{1\oplus 2}^\Omega}$	$\mathbf{m_{1\textcircled{\cap}2}^\Omega}$	$\mathbf{m_{1\textcircled{\cup}2}^\Omega}$	$\mathbf{m_{Mean}^\Omega}$	$\mathbf{m_{1\textcircled{\wedge}2}^\Omega}$
$\emptyset$	0	0	0	0.168	0	0	0.2763
P	0.12	0.3	0.2596	0.216	0.036	0.21	0.1737
H	0.3	0.4	0.5433	0.452	0.12	0.35	0.3401
$P \cup H$	0.2	0.1	0.1058	0.088	0.34	0.15	0.0724
M	0	0	0	0	0	0	0
$M \cup P$	0	0	0	0	0	0	0
$H \cup M$	0.1	0	0.024	0.02	0.04	0.05	0.0362
Ω	0.28	0.2	0.0673	0.056	0.464	0.24	0.1013

2.4.3 Conditioning

When handling a mass function, a new evidence can arise confirming that a proposition A is true (or false). Therefore, the mass function has to be revised in order to take consideration of this new information. Basic belief masses of some focal elements B have to be redistributed. This is achieved by the conditioning operator.

Conditioning a mass function m^Ω over a subset $A \subseteq \Omega$ consists on restricting the frame of possible propositions 2^Ω to the set of subsets compatible with A, subsets having a non empty intersection with A. Therefore, the mass allocated to each focal element $B \subseteq \Omega$ is transferred to $\{B \cap A\}$. The obtained mass function, being the result of the unnormalized Dempster's rule of conditioning (Dempster, 1967), is noted

$m^{\Omega}[A] : 2^{\Omega} \rightarrow [0,1]$ such that (Smets and Kruse, 1997):

$$m^{\Omega}[A](C) = \begin{cases} 0 & \text{for } C \cap A = \{\emptyset\} \\ \sum_{B \cap A = C} m^{\Omega}(B) & \text{otherwise} \end{cases} \quad (2.38)$$

Figure 2.7 illustrates the transfer of basic belief masses from B to C.

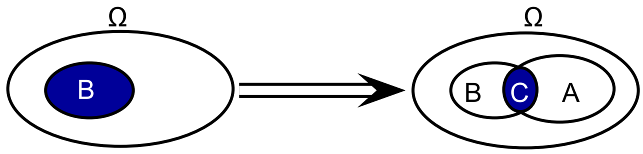

Figure 2.7: Transfer of beliefs with the conditioning

This unnormalized Dempster's rule of conditioning provides unnormalized conditional mass functions as a non null mass can be attributed to the empty set.
Dempster's rule of conditioning is defined in (Dempster, 1967) as follows:

$$m^{\Omega}[A](C) = \begin{cases} \dfrac{\sum_{X \subseteq \bar{A}} m^{\Omega}(C \cup X)}{1 - \sum_{X \subseteq \bar{A}} m^{\Omega}(X)} & \text{for } C \subseteq A \\ 0 & \text{for } C \not\subseteq A \text{ or } C = \emptyset \end{cases} \quad (2.39)$$

(Smets, 1992b) justifies the unnormalized Dempster's rule of conditioning. In fact, assume a belief function induced by an evidential corpus which supports propositions that have a non-empty intersection with $\bar{A}$, thus non-null masses are assigned to that propositions. If a new evidence appears and confirms that no positive masses should be allocated to subsets supporting $\bar{A}$, thus $\text{bel}(\bar{A})$ is the conflict between the initial evidential corpus and the new evidence. Two possible solutions arise; that amount of conflict is either kept on the "contradictory state" that is the empty set or redistributed among still possible hypotheses by a normalization process. The second solution does not verify the homomorphism requirement (Gärdenfors, 1988) and does not respect the condition of non-increase of plausibilities after conditioning. Therefore, only the first solution is held.
Note that Smets defines an evidential corpus as the set of all pieces of evidence held by a source.

Example 2.21 After a primary diagnosis, the doctor assessed the following mass function:

$$m^{\Omega_d}(F) = 0.2,\ \ m^{\Omega_d}(P \cup B) = 0.5,\ \ m^{\Omega_d}(\Omega_d) = 0.3$$

If, after a deep diagnosis, the doctor is sure that the patient suffers from a pharyngitis, the initial mass function m^{Ω_d} has to be conditioned to take consideration of this new certain information. Conditioning m^{Ω_d} with the unnormalized Dempster's rule of conditioning is given as follows:

$$m^{\Omega_d}[P](\emptyset) = 0.2,\ \ m^{\Omega_d}[P](P) = 0.8$$

Conditioning m^{Ω_d} with Dempster's rule of conditioning gives: $m[P]^{\Omega_d}(P) = 1$.

A mass function can also be conditioned on a subset from another frame of discernment. Thus a mass function, m^{Ω}, defined on Ω can be conditioned on a subset $\theta \subset \Theta$ as follows:

$$m^{\Omega}[\theta] = (m^{\Omega \uparrow \Omega \times \Theta} \textcircled{\scriptsize{$\cap$}} m^{\Theta}[\theta]^{\uparrow \Omega \times \Theta})^{\downarrow \Omega} \tag{2.40}$$

2.5 Building belief functions

Belief functions are induced by bodies of evidence according to distinct evidential corpora. In other words, sources provide belief functions to assert their uncertainty on the basis of evidential corpora. Unfortunately, sources do not model always their uncertainty with the theory of belief functions; therefore we use generated belief functions to illustrate proposed methods.

2.5.1 Least commitment principle

Before introducing the deconditioning, *least commitment principle* or *principle of minimal commitment* is defined as follows:

Definition 2.1 *When several belief functions are compatible with a set of constraints,* the least informative *according to some informational ordering (if it exists) should be selected.*

Definition 2.2 *The principle of minimal commitment indicates that, given two equally supported beliefs, only one of which can apply, the most appropriate is the least committed (Smets, 1993).*

(Smets, 1993) claims that "*the principle of minimal commitment formalizes this idea: one should never give more support than justified to any subset of Ω. It satisfies a form of skepticism, non-commitment, or conservatism in the allocation of belief*". The concept of commitment was introduced to create and ordering of mass functions defined on the same frame of discernment Ω (Yager, 1986; Dubois and Prade, 1986; Dubois and Yager, 1987; Hsia, 1991).

According to the least commitment principle, the least-committed mass function according to some constraints is chosen from a set of compatible mass functions. For example, let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ a frame of discernment, $\mathrm{bel}(\{\omega_1\}) = 0.2$ and $\mathrm{bel}(\{\omega_2, \omega_3\}) = 0.5$ the only known evidences over Ω. The main question is how to construct a mass function knowing these partial constraints?
To construct a mass function knowing only these constraints, all mass functions verifying these constraints are enumerated and the least committed one is chosen.

To define the principle, let pl_1^Ω and pl_2^Ω (or equivalently bel_1^Ω and bel_2^Ω), two plausibility functions (belief functions) on Ω such that:

$$\begin{cases} \mathrm{pl}_1^\Omega(A) \leq \mathrm{pl}_2^\Omega(A) & \forall A \subseteq 2^\Omega \\ \text{or} & \\ \mathrm{bel}_1^\Omega(A) \geq \mathrm{bel}_2^\Omega(A) & \forall A \subseteq 2^\Omega \end{cases} \tag{2.41}$$

The least commitment principle for unnormalized mass functions is also expressed with plausibilities but the inequality of belief functions becomes:

$$\mathrm{bel}_1^\Omega(A) + m_1^\Omega(\emptyset) \geq \mathrm{bel}_2^\Omega(A) + m_1^\Omega(\emptyset) \qquad \forall A \subseteq 2^\Omega \tag{2.42}$$

Thus pl_2^Ω is *less committed* than pl_1^Ω if there are at least *one strict inequality* else pl_2^Ω is *no more committed* than pl_1^Ω. The contradictory mass function ($m^\Omega(\emptyset) = 1$) is the *most committed* belief function and the vacuous mass function ($m^\Omega(\Omega) = 1$) is the *least committed one*.

2.5.2 Deconditioning

The *deconditioning* consists on retrieving a deconditioned mass function from a conditional one. If $m^\Omega[A]$ is a conditional mass function where the hypothesis A is assumed to be surely true, it is hard to retrieve m^Ω, the initial mass function before the conditioning on A. Whereas, it is possible to find the least committed mass function such that its conditioning on A is $m^\Omega[A]$. Deconditioning $m^\Omega[A]$ into m^Ω is given as follows:

$$m^\Omega(C \cup \bar{A}) = m^\Omega[A](C) \quad \forall C \subseteq 2^\Omega \tag{2.43}$$

Figure 2.8 illustrates the case where a conditional mass function $m^\Omega[A]$ is deconditioned and the basic belief mass of C is transferred to $C \cup \bar{A}$.

The deconditioning detailed above removes a strong assumption on the truth of

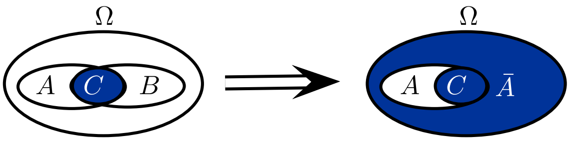

Figure 2.8: Deconditioning

$A \subseteq \Omega$ from the conditional mass function $m^{\Omega}[A]$. Sometimes, mass functions are conditional to a subset of another frame of discernment, that is the case of $m^{\Omega}[\theta]$, a mass function conditional to $\theta \subset \Theta$. Figure 2.9 illustrates the case of $m^{\Omega}[\theta]$ where the mass of B holds only when θ is assumed.

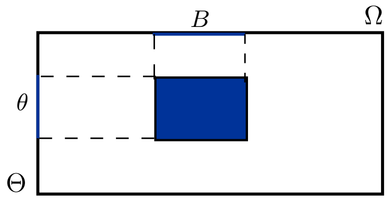

Figure 2.9: Conditional mass function

Suppose a situation where a mass function m^{Ω} defined on Ω when θ holds, the purpose is to find $m^{\Omega\times\Theta}$ such that:

$$(m^{\Omega\times\Theta} \circledcirc (m^{\Theta}[\theta])^{\uparrow\Omega\times\Theta})^{\downarrow\Omega} = m^{\Omega}[\theta] \tag{2.44}$$

According to the least commitment principle, $m^{\Omega\times\Theta}$ is obtained as follows:

$$m^{\Omega}[\theta]^{\Uparrow\Omega\times\Theta}((\theta \times B) \cup (\Theta \times \bar{B})) = m^{\Omega}[\theta](B), \qquad \forall B \subset \Omega \tag{2.45}$$

Note that $m^{\Omega\times\Theta} = m^{\Omega}[\theta]^{\Uparrow\Omega\times\Theta}$ and is called *ballooning extension*. Figure 2.10 illustrates the ballooning extension.

2.5.3 Discounting

Sometimes, it is possible to quantify the reliability of the body of evidence assessing degrees of support. The reliability of information sources reflects both its degrees of expertise and trust. When handling a mass function, we have to take into account the

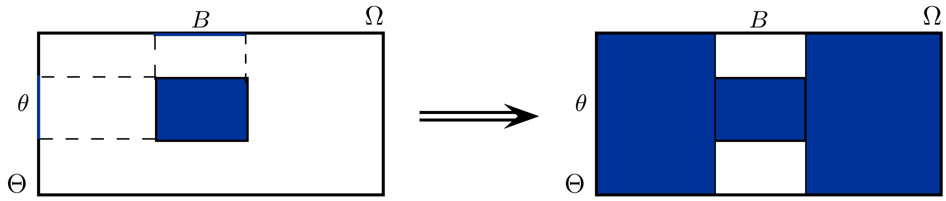

Figure 2.10: Ballooning extension

degree of reliability of its source.

The taking into account the reliability of sources is made with an operator named *discounting* as this operation calibrates beliefs with their sources reliability degrees. Thus, the degree of reliability of a source is taken into account by integrating it into all its mass functions. Using discounting operation in belief functions was first introduced in (Shafer, 1976).

Discounting a mass function m^{Ω} consists on weighting every mass $m^{\Omega}(B)$ by a coefficient $\alpha \in [0, 1]$ called *reliability* and $(1 - \alpha)$ is *the discount rate*. The discounted mass function is given by:

$$\begin{cases} {}^{\alpha}m^{\Omega}(A) = \alpha \times m^{\Omega}(A) & \forall A \subseteq 2^{\Omega} \setminus \Omega \\ {}^{\alpha}m^{\Omega}(\Omega) = 1 - \alpha(1 - m^{\Omega}(\Omega)) \end{cases} \tag{2.46}$$

Properties

- If $\alpha = 1$, the source is fully reliable; therefore, discounting does not change degrees of support: ${}^{\alpha}\mathbf{m}^{\mathbf{\Omega}} = \mathbf{m}^{\mathbf{\Omega}}$.

- If $\alpha = 0$, the source is fully unreliable; therefore, discounting cuts down the mass function to a vacuous mass function: ${}^{\alpha}\mathbf{m}^{\mathbf{\Omega}} = \mathbf{m}^{\mathbf{\Omega}}_{\mathbf{\Omega}}$ with m^{Ω}_{Ω} is a vacuous mass function.

The amount of belief due to the source unreliability is transferred to the frame of discernment as a degree of ignorance.

Example 2.22 Suppose that a doctor gave this mass function:

$$m^{\Omega_d}(F \cup P) = 0.5, \ \ m^{\Omega_d}(F \cup B) = 0.5$$

Suppose that the degree of reliability of that doctor is 0.3. The discounted mass function is:

$$
\begin{aligned}
{}^{\alpha}m^{\Omega_d}(F \cup P) = 0.5 \times 0.3 = 0.15, \\
{}^{\alpha}m^{\Omega_d}(F \cup B) = 0.5 \times 0.3 = 0.15, \\
{}^{\alpha}m^{\Omega_d}(\Omega_d) = 1 - 0.3 \times (1 - 0) = 0.7
\end{aligned}
$$

(Smets, 1993) justified the discounting operator by supposing that the mass function provided by the source is unchanged if the source is fully reliable. The case of total ignorance appears when the source is unreliable thus conditioned mass functions are obtained as follows:

$$
\begin{aligned}
m^{\Omega}[R](A) &= m^{\Omega}(A) && (2.47) \\
m^{\Omega}[\bar{R}](A) &= m^{\Omega}_{\Omega}(A) && (2.48)
\end{aligned}
$$

with:

$$
\begin{cases} m^{\Omega}(A) = 1 & \text{if } A = \Omega \\ m^{\Omega}(A) = 0 & \forall A \subseteq 2^{\Omega} \setminus \Omega \end{cases} \tag{2.49}
$$

With R the hypothesis confirming that the source is reliable and $\bar{R}$ the hypothesis that the source is unreliable, the frame of discernment $\mathcal{R} = \{R, \bar{R}\}$ describes the sources reliability and $m^{\mathcal{R}}$ is the mass function about the source's reliability described as follows:

$$
\begin{cases} m^{\mathcal{R}}(R) = \alpha \\ m^{\mathcal{R}}(\mathcal{R}) = 1 - \alpha. \end{cases} \tag{2.50}
$$

To combine mass functions, $m^{\Omega}[R]$, provided by a source and, $m^{\mathcal{R}}$, emphasizing beliefs about source's reliability, they have to be defined on the same product space $\mathcal{R} \times \Omega$. Therefore, $m^{\mathcal{R}}$ is transformed into $m^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}$ using the vacuous extension as follows:

$$
m^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}(Y) = \begin{cases} m^{\mathcal{R}}(X) & \text{if } Y = \Omega \times X, \quad X \subseteq \mathcal{R} \\ 0 & \text{otherwise} \end{cases} \tag{2.51}
$$

The conditional mass function $m^{\Omega}[R]$ has to be deconditioned as follows:

$$
m^{\Omega}[R]^{\Uparrow\Omega\times\mathcal{R}}((A \times R) \cup (\Omega \times \overline{R})) = m^{\Omega}[R](A), \quad A \subseteq \Omega \tag{2.52}
$$

Finally, $m^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}$ and $m^{\Omega}[R]^{\Uparrow\Omega\times\mathcal{R}}$ defined on the same product space $\Omega \times \mathcal{R}$ can be combined using the conjunctive rule of combination as follows:

$$
m^{\Omega\times\mathcal{R}}_{\textcircled{\cap}}(A) = m^{\mathcal{R}\uparrow\Omega\times\mathcal{R}} \textcircled{\cap} m^{\Omega}[R]^{\Uparrow\Omega\times\mathcal{R}}(A), \quad \forall A \subset \Omega \times \mathcal{R} \tag{2.53}
$$

The combined mass function $m_{\textcircled{\cap}}^{\Omega\times\mathcal{R}}$ is then marginalized on Ω as follows:

$$m^{\Omega\times\mathcal{R}\downarrow\Omega}(A) = \sum_{\{B\subseteq\Omega\times\mathcal{R}\,|\,Proj(B\downarrow\Omega)=A\}} m_{\textcircled{\cap}}^{\Omega\times\mathcal{R}}(B) \tag{2.54}$$

Thus:

$$^{\alpha}m^{\Omega}(A) = m^{\Omega\times\mathcal{R}\downarrow\Omega}(A) \tag{2.55}$$

Mercier proposed to use a mass function on source's reliability according to subsets of Ω and generalized the discounting operator to a contextual discounting (Mercier et al., 2005; Mercier, 2006; Mercier et al., 2008). (Zeng and Wu, 2007) proposed to discount plausibility functions. Researches like (Mercier et al., 2006; Mercier et al., 2008; Martin et al., 2008; Huynh, 2009; Chebbah et al., 2010a; Elouedi et al., 2010; Florea et al., 2010; Chebbah et al., 2011; Liu et al., 2011; Yang et al., 2013; Frikha, 2014) aim on estimating the discounting factor.

2.5.4 Random generation

A mass function can be obtained by the inverse pignistic transformation in the purpose of quantifying future realizations of a random variable X when its probability distribution is known (Denœux, 2006b; Aregui and Denœux, 2007). It can also be obtained by deconditioning or discounting a mass function to have a conditioned or discounted mass function from another one. Finally, it can even be obtained using the multivalued mapping (Liu et al., 1992) or by collecting experts opinions (Wong and Lingras, 1994; Bryson and Mobolurin, 1999; Ben Yaghlane et al., 2006b; Ben Yaghlane et al., 2006a).

Unfortunately, it is not always possible to obtain mass functions as cited above. In that case mass functions can be generated randomly, algorithm 1 generates a set of N random mass functions. Number of focal elements and their masses are independently and randomly chosen according to universal law.

(Burger and Destercke, 2012) proposed an efficient algorithm for random generation

Algorithm 1 Random mass functions generation

Require: $|\Omega|$, N: number of mass functions
  for $i = 1$ to N **do**
    Choose randomly $|\mathcal{F}|$, the number of focal elements on $[1, |2^{\Omega}|]$.
    Choose randomly $|\mathcal{F}|$ focal elements noted $\mathcal{F}$.
    Divvy the interval $[0, 1]$ into $|\mathcal{F}|$ continuous sub-intervals.
    Basic belief masses of focal elements are intervals sizes.
  end for
  return N mass functions

of mass functions.

2.6 Conclusion

In this chapter, we have reviewed basics of the theory of belief functions that handles uncertain information. We started this chapter by introducing belief functions and decision making. Then we detailed tools for representing belief functions defined on either compatible or incompatible frames in a common one. Next, we introduced some combination rules, and more details will be given in the last chapter. Finally, some methods for building belief functions are pointed out.

This chapter recalls some basics of the theory of belief functions. In the sequel of this report, uncertainty is modeled with the theory of belief functions. The next chapter deals with evidential clustering algorithms to classify belief functions which can be stored in evidential databases. We propose a clustering algorithm minimizing a conflict between objects of the same cluster. Clustering algorithm is then used for learning sources independence degree.

3

Conflict and clustering in the theory of belief functions

Contents

Summary

In the previous chapter, some basic concepts of the theory of belief functions are detailed. Indeed, when sources cannot provide certain information, they provide mass functions according to a given evidential corpus.
Mass functions are gathered in evidential databases, such type of databases stores both certain and evidential information. Thus, an evidential database stores all mass functions provided by the same source for some objects. Classifying objects from an evidential database groups together similar objects and provides an information about sources overall behavior. In this chapter, we propose a conflict measure between evidential databases and a clustering algorithm minimizing the conflict between objects into clusters. The use of clustering algorithm is to compare several sources overall behavior in order to estimate their independence or dependence.

3.1 Introduction

Information is plentiful in several fields, it is scattered and not exploited. Grouping information concerning the same fields gives a high quantity of data to be stored and information become bulky justifying the use of databases. Databases integrate a high quantity of information provided by different sources, these data may be conflicting because their sources may disagree. Most of researches are based on perfect and certain data but data stored in databases or data warehouses are almost imperfect (incomplete or uncertain) and may be conflicting; however, users need these data for decision making. Therefore, perfect and imperfect data have to be exploited and the conflict has to be solved.

Evidential databases (Hewawasam et al., 2005; Bach Tobji et al., 2008; Bach Tobji, 2012) are databases storing both certain and uncertain data where uncertainty is modeled with the theory of belief functions that is detailed in Chapter 2. A database stores some or all mass functions provided by a source according to some objects; thus, many evidential databases may be stored according to the number of sources. Note that sources are any possible sources of information that can be a human expert, a classifier, a sensor, etc.

In some fields like sensoring, one may manage a considerable number of evidential databases as many sensors may observe the same objects from different points. These evidential databases may be conflicting reflecting the conflict between evidences observed by sources. This conflict can be defined as a degree of discord between the beliefs of that sources. It may be noticed either by combining mass functions stored in evidential databases or by comparing mass functions with a similarity measure. The conflict appearing in the combination of conflicting evidential information incited the introduction of several methods intended to solve it. Some of these methods propose to solve the conflict when combining, like in (Yager, 1987; Dubois and Prade, 1988; Smets and Kennes, 1994; Murphy, 2000; Martin and Osswald, 2007b). These combination rules hide the conflict regardless of its causes. Therefore, the conflict does not appear in the combined information because combination rules redistribute it with different manners. Other methods consider that the main reason of conflict is the relative unreliability of at least one source. Thus, conflict solving can be insured by discounting the evidential information before combining with sources' degrees of reliability as detailed in Section 2.5.3; however, this method requires a preliminary knowledge of this degree of reliability.

In this chapter, we present an overview of conflict in the theory of belief functions, some interpretations and methods for solving it. We present also some distances in the theory of belief functions. The distance between mass functions informs also about their similarity; thus, distant mass functions are dissimilar but near ones are quite similar. Next, we detail a conflict measure between evidential databases as we proposed in (Chebbah et al., 2010b; Chebbah et al., 2010a). This conflict measure is used hereafter for classifying objects stored in evidential databases. The proposed clustering algorithm classifies objects of the same evidential database by minimizing the conflict between them in the same cluster.

This chapter is organized as follows: The second section reviews conflict interpretations and methods to solve it. The third section is an overview of distances between mass functions in the theory of belief functions; In the forth section, we define evidential databases as proposed in (Bach Tobji et al., 2008; Bach Tobji, 2012) and detail our method of estimating the conflict between evidential databases (Chebbah et al., 2010b; Chebbah et al., 2010a). Section 5 is an overview of clustering algorithms proposed in the theory of belief functions, we detail also our algorithm of evidential clustering. Finally before concluding, illustrations of the proposed algorithm are proposed in Section 6.

3.2 Conflict in the theory of belief functions

The main reason of conflict arising when combining mass functions provided by distinct and independent sources is their *unreliability*. Some degree of conflict can be tolerated but in some cases this degree can be alarming and must be eliminated. To eliminate the conflict, its issue must be detected. The conjunctive rule of combination (equation (2.29)) is the only combination rule that keeps $m(\emptyset) > 0$; the conflict is kept on the empty set and an expert system is used to solve the problem if possible (Smets, 2007). A positive mass on the empty set alarms the expert system who tries to look for origins of that conflict and finds the appropriate solution without revising belief functions. Several combination rules can be chosen according to conflict's issue.

3.2.1 Origins of the conflict

We detail below the main interpretations of the conflict issued from the combination of beliefs provided by distinct sources.

- **Conflict due to sources unreliability** (Huynh, 2009): When combining two mass functions provided by different sources, one of these sources may be totally or partially unreliable. If the unreliable source is not known, the disjunctive rule of combination performs well. Therefore, when combining two mass functions provided by two distinct and independent sources from which an unknown one is unreliable, the disjunctive rule of combination can be used to combine that mass functions.

Discounting operator detailed in Section 2.5.3 is used when at least one source is partially unreliable and its degree of reliability is known or can be learned or estimated. Schubert proposed to solve the conflict by discounting each piece of evidence with the degree of conflict that it contributes on, in an iterative way till reaching an *aprior* fixed level of acceptable conflict (Schubert, 2008; Schubert, 2011).

Table 3.1 summarizes methods used to solve the conflict appearing when at least one of the sources is unreliable.

Table 3.1: Methods for solving a conflict due to sources unreliability

Sources reliability	**Solutions**
An unknown source is unreliable	Disjunctive rule of combination (Smets, 1990)
A known source is unreliable, its degree of reliability can be learned or estimated	Discounting mass functions provided by the unreliable source with its degree of reliability as detailed in Section 2.5.3
Both sources are reliable	Conjunctive rule of combination and derived rules

- **Sources reporting about different objects**: In that case, sources must be grouped according to objects they report about. Only mass functions provided by two distinct and independent sources that have different evidential corpora can be combined. Therefore, in a case of a high conflict between sources, they may have different corpora about different objects or problems. Combining is performed on mass functions that reports about the same object at the same time.

Example 3.1 Suppose having mass functions of table 3.2 given by a doctor when examining some patients; the frame of discernment is $\Omega_d = \{F, P, B\}$. These mass functions emphasize a doctor's diagnostics of several patients p_1, p_2 and p_3. None from these mass functions can be combined because they are referred to different patients for several examinations. Only mass functions of the same patient for the same examinations done by independent and distinct doctors can be combined.

Table 3.2: Bbas given by different sources at different periods

Patient	**Examinations**	**Massfunction**
p_1	e_1	$F(1)$
p_3	e_1	$F \cup P(0.5),\ F(0.5)$
p_1	e_2	$\Omega(0.5)$
p_2	e_1	$F \cup B(0.5),\ \Omega_d(0.5)$
p_3	e_2	$H(1)$
p_2	e_2	$B(O.5),\ P(0.5)$

- **Conflict due to closed world assumption:** Suppose m_1 and m_2 two mass functions defined on 2^Ω to combine under the closed world assumption; in that case the frame of discernment Ω is assumed to be exhaustive. A conflict may appear when the hypothesis of the exhaustivity of Ω is not true. The closed world assumption is not maintained when some alternatives are forgotten when enumerating alternatives of the frame of discernment or when Ω evolved in time.

  One solution proposed by (Smets, 2007) is to add an extra alternative $\omega^\star$ to Ω that includes all those hypotheses not mentioned in Ω. Therefore $\Omega^\star = \Omega \cup \{\omega^\star\}$ is the extended frame; thus $\forall A \subseteq \Omega$ the mass of A, $m(A)$, must be transferred to $\{A \cup \omega^\star\}$ ($m(\{A \cup \omega^\star\}) = m(A)$). In this case, conflict disappears completely as $\{\omega^\star\}$ belongs to all prepositions of both mass functions; therefore, there is no empty intersection between focal elements.

  A problem can arise when making decision if $\omega^\star$ is the most probable solution. In this case we should condition that mass function on Ω before the pignistic transformation.

3.2.2 Conflict measures

The first measure of conflict was introduced in (Dempster, 1967) as the mass on the empty set $m_{1 \textcircled{\cap} 2}(\emptyset)$ issued from the conjunctive combination of two mass functions m_1 and m_2 defined as follows:

$$k = \sum_{A \cap B = \emptyset} m_1(A) m_2(B) \tag{3.1}$$

And the following re-normalizing constant K (equation (5.1)) redistributes the amount k to have normal belief functions:

$$K = \frac{1}{1-k} \tag{3.2}$$

Dempster (Dempster, 1967) and Yager (Yager, 1983) proposed also a weight of conflict between two belief functions noted $\text{Con}(Bel_1, Bel_2)$ and defined as follows:

$$\text{Con}(m_1, m_2) = log\ K = log \frac{1}{1-k} = -log(1-k) \tag{3.3}$$

The weight $\text{Con}(m_1, m_2)$ is in $[0, \infty[$ in contrast to K which is equal or greater than 1. When mass functions do not conflict $k = 0$ and $\text{Con}(m_1, m_2) = 0$ but when belief functions flatly contradict then $m_1 \oplus m_2$ does not exist, $k = 1$ and $\text{Con}(m_1, m_2) = \infty$. (Martin et al., 2008) do not consider $m(\emptyset)$ as an indicator about the conflict between sources because it contains an amount of auto-conflict (Osswald and Martin, 2006) owed on the characteristic of the combination rule which is not idempotent and the absorbing character of the empty set.

(Smets, 1992b) justifies the interpretation of k as a conflict and proposes the open world assumption as a solution. The amount k is issued from the combination of mass functions induced by an evidential corpus EC and a new contradicting evidence Ev. Suppose that a body of evidence[1] provided a belief function induced by EC supporting a preposition A such that positive degrees of support are allocated to some non-empty subsets of A. If further evidence Ev rejects the preposition A, no positive masses should be allocated to the non-empty subsets of A. Amounts of belief initially assessed to subsets of A have to be redistributed in two possible ways; the first solution is to allocate it to the contradictory subset denoted $\emptyset$ or to redistribute it among possible subsets with a normalizing process.

The second solution cannot be retained as that solution does not satisfy the homomorphism requirement. Thus k is not redistributed among till possible hypothesis; this amount is interpreted as a contradictory or conflict between the evidential corpus that induced an initial mass function and a new evidence.

We do not consider the amount k as a sufficient measure of conflict because the empty set is the absorbing element in the combination and also most of the combination rules are not idempotent. The non-idempotence of combination rules can imply that $m \textcircled{\cap} m(\emptyset) \neq 0$, thus k is not really a conflict.

Liu noticed in (Liu, 2006) that Dempster's conflict is not enough informative about the conflict between evidences; a distance on pignistic probabilities together with Dempster's conflict points out that amount of conflict. In addition to Dempster's conflict of equation (3.1), a distance on pignistic probabilities $BetP_{m_1}$ and $BetP_{m_2}$ transforma-

[1] source

tions of m_1 and m_2 with equation (2.23) is defined as follows:

$$DifBetP(m_1, m_2) = max_{A \subseteq \Omega}(|BetP_{m_1}(A) - BetP_{m_2}(A)|) \tag{3.4}$$

A high "*distance between betting commitment*" of two mass functions m_1 ans m_2 with a high Dempster's conflict $\sum\limits_{A \cap B = \emptyset} m_1(A)m_2(B)$ indicates a conflict between bodies of evidence that induced m_1 and m_2.

(Jousselme and Maupin, 2012) proposed a cosine-based measure of conflict for two mass functions m_1 and m_2 as follows:

$$cos(m_1, m_2) = \quad 1 - \frac{Pl_1' Pl_2}{||Pl_1||.||Pl_2||} \tag{3.5}$$

where $||.||$ denotes the norm of normalized plausibility functions Pl_1 and Pl_2. The cosine function computes the angle between two vectors. Orthogonality is interpreted as a conflict.

Martin et al. noticed that Dempster's conflict k is not a conflict measure because it includes some degree of auto-conflict due to the non-idempotence of the conjunctive rule of combination (Martin and Osswald, 2006b; Osswald and Martin, 2006; Martin et al., 2008). The combination of identical mass functions provided by distinct and independent sources gives some degree of conflict $k \geq 0$ although mass functions are identical and not conflicting. The *auto-conflict*, which is the intrinsic conflict of a mass function m, is given by the following equation:

$$a_s = \left(\bigcirc\!\!\!\!\cap_{i=1}^{s} m \right) (\emptyset) \tag{3.6}$$

The conjunctive combination $\bigcirc\!\!\!\!\cap$ is given in equation (2.29), the auto-conflict of order s is the s times sequential combination of identical mass function m. Note that $a_s \leq a_{s+1}$, meaning that combining identical mass functions m many times leads to high degree of Dempster's conflict tending to 1. Martin et al. proposed to use a distance between mass functions as a conflict measure (Martin and Osswald, 2006b; Osswald and Martin, 2006; Martin et al., 2008); thus, dissimilar mass functions in terms of distance are conflicting but near ones are not or a little bit conflicting.

(Florea and Bossé, 2009) defined an intrinsic conflict given as follows:

$$\sum_{A, B \subseteq \Omega} m(A)m(B)\frac{|A \cup B| - |A \cap B|}{|A \cup B|} \tag{3.7}$$

That conflict is non-null for consonant mass functions. Although, it is not the case of the "*internal conflict*" in (Daniel, 2010). Note that intrinsic conflict and auto-conflict are different.

(Destercke and Burger, 2012) proposed a conflict measure between mass functions as an extension of the conflict between sets. Authors of (Elouedi et al., 2004; Martin et al., 2008; Lefèvre et al., 2011; Martin, 2012; Boubaker et al., 2013) proposed a conflict measure based on a distance. This assumption is assumed along this chapter and will be extended to evidential databases in Section 3.4. Some distances are detailed in the following section.

3.3 Distances in the theory of belief functions

The mass of the empty set issued from the combination of several belief functions is considered as an indicator about a conflict between evidences held by sources (Dempster, 1967). That mass is not so informative about conflict between mass functions as it includes some degree of auto-conflict (Martin et al., 2008). Therefore, distances are used to quantify the conflict between mass functions. In a general case, a distance measures the closeness of two points; it computes the nearness of that points. In the framework of the theory of belief functions, a distance computes the closeness of beliefs; it computes how much two mass functions are similar.

(Florea et al., 2009b) propose an overview of some distances in the theory of belief functions; (Jousselme and Maupin, 2010; Jousselme and Maupin, 2012) propose an exhaustive and comprehensive survey of distances in belief functions framework. Before enumerating distances proposed to quantify similarity of mass functions, we summarize in table 3.3 some similarity measures of focal elements. These similarity coefficients quantify interactions and closeness of focal elements of mass functions.

Some distances on belief functions are surveyed in (Florea et al., 2009b; Jousselme and Maupin, 2010; Jousselme and Maupin, 2012), we enumerate hereafter some of them:

- **Perry and Stephanou's distance:** (Perry and Stephanou, 1991) proposed a distance based on (Stephanou and Lu, 1988) for a classification purpose. The proposed distance quantifies the gap of information between mass functions when they are considered separately and the combined mass function. That distance compares the amount of information available in each mass function m_1 and m_2

Table 3.3: Similarity functions

Similarity	Coefficient
Dice	$\dfrac{2\|A\cap B\|}{\|A\|+\|B\|}$
Sokal & Sneath 2	$\dfrac{\|A\cap B\|}{2\|A\cup B\|-\|A\cap B\|}$
Kulczynski 2	$\dfrac{\|A\cap B\|}{2\|A\|}+\dfrac{\|A\cap B\|}{2\|B\|}$
Ochiai	$\dfrac{\|A\cap B\|}{\sqrt{\|A\|\,\|B\|}}$
Jaccard	$\dfrac{\|A\cap B\|}{\|A\cup B\|}$

separately and their combined mass function m_{12} as follows:

$$d_{PS}(m_1, m_2) = |\,\mathcal{F}_1 \cup \mathcal{F}_2\,| \left(1 - \frac{|\,\mathcal{F}_1 \cap \mathcal{F}_2\,|}{|\,\mathcal{F}_1 \cup \mathcal{F}_2\,|}\right) + (m_{12} - m_1)'(m_{12} - m_2) \tag{3.8}$$

with $m_{12} = m_1 \oplus m_2$, and $\mathcal{F}_i$ are focal elements of mass functions m_i. This distance is on two components; the first component $|\,\mathcal{F}_1 \cup \mathcal{F}_2\,| \left(1 - \frac{|\,\mathcal{F}_1 \cap \mathcal{F}_2\,|}{|\,\mathcal{F}_1 \cup \mathcal{F}_2\,|}\right)$ measures dissimilarity between focal elements of m_1 and m_2. The second component $(m_{12} - m_1)'(m_{12} - m_2)$ measures the change of information relative to the orthogonal sum.

- **Blackman and Popoli's distance:** (Blackman and Popoli, 1999) proposed a distance for association algorithms based on Dempster's conflict $m_{12}(\emptyset)$ (detailed in Chapter 2) as follows:

$$d_{BP}(m_1, m_2) = -2log\left(\frac{1 - m_{1 \odot 2}(\emptyset)}{1 - max(m_{1\odot 1}(\emptyset), m_{2\odot 2}(\emptyset))}\right) + (m_1 + m_2)' g_A - m_1' G m_2 \tag{3.9}$$

with g_A, a vector whose elements are $\dfrac{|A|-1}{|X|-1}$ and $G = g_A g_A'$ with elements $\dfrac{(|A|-1)(|B|-1)}{(|X|-1)^2}$, $\forall A,\ B \subseteq X$. This distance is also composed of two components; the first one $-2log\left(\frac{1 - m_{1 \odot 2}(\emptyset)}{1 - max(m_{1\odot 1}(\emptyset), m_{2\odot 2}(\emptyset))}\right)$ is called "*attribute distance*" and the second is $(m_1 + m_2)' g_A - m_1' G m_2$ called "*ignorance distance*".

Unfortunately, Blackman and Popoli's distance is not positive and thus it is non-metric.

- **Minkowski distance (L_p):** The Minkowski distance, noted L_p, between two mass functions m_1 and m_2 is defined as follows:

$$d^{(p)}(m_1, m_2) = \left(\left[(Um_1 - Um_2)^{\frac{p}{2}} \right]' \left[(Um_1 - Um_2)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}} \tag{3.10}$$

such that U is the upper triangular matrix of the Cholesky decomposition and p an integer greater than 1. The case where $p = 1$ corresponds to Manhattan distance, $p = 2$ for the Euclidean distance and $p = \infty$ corresponds to Chebyshev distance.

(Cuzzolin, 2009) proposed the L_p distance on belief functions as follows:

$$d^{(p)}(m_1, m_2) = \left(\sum_{A \subseteq \Omega} |\mathrm{Bel}_1(A) - \mathrm{Bel}_2(A)|^p \right)^{\frac{1}{p}} \tag{3.11}$$

such that Bel_i are normalized belief functions of m_i. Note that equation (3.11) is equivalent to equation (3.10) when $U = \mathrm{Inc}'$ such that Inc is the inclusion matrix defined as follows:

$$\mathrm{Inc}(A, B) = \begin{cases} 1 & \text{if } A \subseteq B \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

- **Manhattan distance (L_1):** Manhatten distance (Klir and Wierman, 1999; Harmanec, 1999) is obtained from equation (3.10) such that $p = 1$ as follows:

$$d^{(1)}(m_1, m_2) = \left(\left[(Um_1 - Um_2)^{\frac{1}{2}} \right]' \left[(Um_1 - Um_2)^{\frac{1}{2}} \right] \right) \tag{3.13}$$

When $U = \mathrm{Inc}'$, $d^{(1)}$ is given by the following equation:

$$d^{(1)}(m_1, m_2) = \left(\sum_{A \subseteq \Omega} |Bel_1(A) - Bel_2(A)| \right) \tag{3.14}$$

(Denœux, 2001) defined $d^{(1)}$ as follows:

$$d^{(1)}(m_1, m_2) = \left(\sum_{A \subseteq \Omega} |Pl_1(A) - Pl_2(A)| \right) \tag{3.15}$$

Pl_i are normalized plausibility functions of m_i. Manhattan distance proposed in

(Denœux, 2001) is deduced from equation (3.10) when $p = 1$ and $U = \text{Int}$ an intersection matrix defined as follows:

$$\text{Int}(A, B) = \begin{cases} 1 & \text{if } A \cap B \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

- **Euclidean distance (L_2):** (Cuzzolin, 2008) proposed an extension of the Euclidean distance from the probability theory to the theory of belief functions by replacing probability values by masses. The Euclidean distance between two mass functions m_1 and m_2 is defined as follows:

$$d_C^{(2)}(m_1, m_2) = \sqrt{\sum_{A\subseteq\Omega} |m_1(A) - m_2(A)|^2} \tag{3.17}$$

  Ristic and Smets proposed also in (Ristic and Smets, 2006) an extension of the Euclidean distance from probability theory to the theory of belief functions. The Euclidean distance between two mass functions m_1 and m_2 is given by the following equation:

$$d^{(2)}(m_1, m_2) = \sum_{A\subseteq\Omega}\sum_{B\subseteq\Omega} m_1(A)m_2(B) \tag{3.18}$$

  This distance is always equal to 1 for any mass functions m_1 and m_2:

$$\begin{aligned} \sum_{A\subseteq\Omega}\sum_{B\subseteq\Omega} m_1(A)m_2(B) &= \sum_{A\subseteq\Omega} m_1(A) \sum_{B\subseteq\Omega} m_2(B) \\ &= \sum_{A\subseteq\Omega} m_1(A) \\ &= 1 \end{aligned}$$

- **Chebyshev distance (L_∞):** Chebyshev distance is the limit of Minkowski distance in equation (3.10) when p tends to $+\infty$. Chebyshev distance is given as follows:

$$d^{(\infty)} = \max_{A\subseteq X} \left\{ |(Um_1)'e_A - (Um_2)'e_A| \right\} \tag{3.19}$$

  (Tessem, 1993) proposed an error measure from Chebyshev distance as follows:

$$d_{Bet}^{(\infty)}(m_1, m_2) = \max_{A\subseteq X} \left\{ |(Bet\ m_1)'e_A - (Bet\ m_2)'e_A| \right\} \tag{3.20}$$

  and

$$d_{\text{Inc}}^{(\infty)}(m_1, m_2) = \max_{A\subseteq X} \left\{ |(\text{Inc}'\ m_1)'e_A - (\text{Inc}'\ m_2)'e_A| \right\} \tag{3.21}$$

- **Bhattacharyya's distance:** (Ristic and Smets, 2006) proposed an extension of the Bhattacharyya's distance from probability theory to the theory of belief functions as follows:

$$d_R(m_1, m_2) = \sqrt{1 - \sum_{A \subseteq \Omega} \sum_{B \subseteq \Omega} \sqrt{m_1(A)m_2(B)}} \tag{3.22}$$

We notice that this distance is always null because $\sum_{A \subseteq \Omega} \sum_{B \subseteq \Omega} m_1(A)m_2(B)$ is always equal to 1, thus it has no use.
Therefore, (Florea et al., 2009b) proposed a generalization of Bhattacharyya's distance extension to the theory of belief functions as follows:

$$d_{Fl}(m_1, m_2) = |1 - \sum_{A \subseteq \Omega} \sqrt{m_1(A)m_2(A)}|^p \tag{3.23}$$

with p any positive number.

- **Jousselme distance (Jousselme et al., 2001):** This distance is specific to the theory of belief functions because of the matrix D which is defined on 2^Ω.
This distance is balanced with Jaccard's coefficient $\frac{|A \cap B|}{|A \cup B|}$ as similarity measure between focal elements allowing the consideration of focal elements cardinality. Jousselme distance is defined as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^t \underline{\underline{D}}(m_1 - m_2)} \tag{3.24}$$

with :

$$\underline{\underline{D}}(A, B) = \begin{cases} 1 & \text{if A=B} \\ \dfrac{|A \cap B|}{|A \cup B|} & \forall A, B \in 2^\Omega \end{cases} \tag{3.25}$$

We notice that Jousselme distance is an extension of Euclidean distance when $\underline{\underline{D}}$ is the identity matrix; Euclidean distance can be written as follows:

$$d(m_1, m_2) = \sqrt{(m_1 - m_2)^t \underline{\underline{D}}(m_1 - m_2)} \tag{3.26}$$

with :

$$\underline{\underline{D}}(A, B) = \begin{cases} 1 & \text{if A=B} \\ 0 & \forall A, B \in 2^\Omega \end{cases} \tag{3.27}$$

The matrix $\underline{\underline{D}}$ is positive definite as proved in (Bouchard et al., 2013) and the distance of equation (3.24) is a full metric. Jousselme distance has been widely used for discounting rate estimation (Deng et al., 2008; Guan et al., 2008; Martin, 2009; Florea et al., 2009a; Chebbah et al., 2010b; Chebbah et al., 2010a; Chebbah et al., 2011).

(Diaz et al., 2006) proposed an extension of Jousselme distance by replacing the similarity matrix $\underline{\underline{D}}$ emphasizing interaction between focal elements thanks to Jaccard coefficient by any other similarity coefficient as those detailed in table 3.3. They proposed to use Dice index and the corresponding similarity matrix is given as follows:

$$\underline{\underline{D}}(A,B) = \begin{cases} 1 & \text{if A=B=}\emptyset \\ \dfrac{2|A \cap B|}{|A|+|B|} & \forall A, B \in 2^{\Omega} \setminus \emptyset \end{cases} \tag{3.28}$$

In that case, we cannot claim that the distance is a full metric or not without proving that similarity matrix $\underline{\underline{D}}$ using Dice index is either positive definite or semi-positive definite.

- **Ristic and Smets distance (Ristic and Smets, 2006):** Ristic and Smets defined a distance that they call "additive global dissimilarity measure" for two mass functions m_1 and m_2 from their amount of Dempster's conflict $m_{1 \textcircled{\scriptsize{$\cap$}} 2}(\emptyset)$ as follows:

$$d_{RS}(m_1, m_2) = -log(1 - m_{1 \textcircled{\scriptsize{$\cap$}} 2}(\emptyset)) \tag{3.29}$$

  Unfortunately, Ristic and Smets distance is a non-metric measure, it is not defined when $m_{1 \textcircled{\scriptsize{$\cap$}} 2}(\emptyset) = 1$. Also $d_{RS}(m, m) \neq 0$.

Many other distances are proposed in belief functions framework like (Zouhal and Denœux, 1998; Denœux, 2000; Denœux, 2001; Ristic and Smets, 2006; Liu, 2006). An exhaustive survey of distances and their properties as well as new distances are proposed by Jousselme and Maupin in (Jousselme and Maupin, 2012). In some researches, a distance may enlighten about the conflict; conflict measures are detailed in the previous section.

3.4 A new conflict measure between evidential databases

An *evidential database* (*EDB*), also called *D-S database*, is a database containing certain and/or uncertain data, uncertainty is modeled with the theory of belief functions as presented in (Hewawasam et al., 2005) and (Bach Tobji et al., 2008; Bach Tobji, 2012). In this section, we introduce evidential databases and then we detail our method of estimating conflict between such databases as proposed in (Chebbah et al., 2010b; Chebbah et al., 2010a; Chebbah et al., 2011).

3.4.1 Evidential databases

An evidential database is a database with N records (objects) and c attributes such that each attribute a_j $(1 \leq j \leq c)$ has a domain Ω_{a_j} enumerating all its possible values.

Ω_{a_j} is the *frame of discernment* of the j^{th} attribute (Hewawasam et al., 2005).

An EDB must have at least one *evidential attribute*, values of this attribute are uncertain expressed by mass functions as defined in (Bach Tobji et al., 2008). An *evidential value* V_{ij} for the i^{th} record ($1 \leq i \leq n$) and the j^{th} attribute is a mass function defined as follows:

$$\begin{array}{l} m_{ij}^{\Omega_{a_j}} : 2^{\Omega_{a_j}} \rightarrow [0,1] \text{ with :} \\ m_{ij}^{\Omega_{a_j}}(\emptyset)\text{=}0 \text{ and } \sum\limits_{A \subseteq \Omega_{a_j}} m_{ij}^{\Omega_{a_j}}(\text{A})\text{=}1 \end{array} \tag{3.30}$$

Example 3.2 Table 3.4 is an example of a table of an evidential database, it contains diseases of some patients that are examined by a doctor. The attribute *disease* is the only evidential attribute in this evidential table, its frame of discernment is $\Omega_d = \{\text{Flu } F,\ \text{Pharyngitis } P, \text{Bronchitis } B\}$ enumerating some possible diseases.

Table 3.4: Example of an EDB

id	**First name**	**Last name**	**Disease**
1	David	Johanson	$P(0.2)$ $\{P \cup F\}(0.6)$ $\{P \cup F \cup B\}(0.2)$
2	Andrew	Smith	$P(0.5)$ $B(0.5)$
3	Joshua	Clark	$P(0.5)$ $F(0.2)$ $\{P \cup F\}(0.3)$

This evidential database stores data of different levels of certainty. It stores:

- **Perfect information:** The focal element is a single point having the hole mass of belief. Value of the evidential attribute is a certain belief function.
- **Probabilistic information:** The value of the evidential attribute is a mass function with several focal elements which are singletons. Line 2 of table 3.4 is an example of probabilistic information.
- **Possibilistic information:** Line 1 of table 3.4 is a possibilistic information as focal elements are nested.

- **Missing information:** Missing information is represented by a mass function with only one focal element which is the frame of discernment Ω_{a_j}, the corresponding mass function is a vacuous belief function.

- **Evidential information:** When information is neither perfect, nor probabilistic, nor possibilistic nor missing, then it is evidential. The mass function defined for the attribute target in line 3 of table 3.4 is an evidential information.

Evidential databases can be used in several areas such as classification where they stock mass functions supplied by different classifiers such as in (Hewawasam et al., 2005). Sources of information stored in EDBs are numerous and the quantity of stored information is high, although they almost report about the same objects. In this thesis, we consider the fusion of these information. The fusion decreases their quantity and helps users for decision making, but the main problem is the type of combination rules to use. The choice of the type of combination rule depends on the independence of evidential databases' sources. Sources' independence estimation is detailed in Chapter 4. That method is based on the clustering algorithm detailed in Section 3.5.3. That clustering algorithm minimizes the conflict in clusters. An evidential database is used to store different mass functions supplied by a source, therefore the number of evidential databases is dependent on the number of sources. Having M sources implies the existence of M evidential databases such that every EDB belongs to a unique source.

Also, having a high number of EDBs implies that the quantity of data to be stored is high and these data may sometimes represent the same pieces of information.
Integrating M evidential databases reduces the quantity of data to be stored and also helps users in decision making, thus decision makers have to take into account only one integrated EDB which resumes M ones. When integrating evidential values from several EDBs, a conflict may appear due to the variety of EDBs sources which might be conflicting.

Example 3.3 Table 3.5 and table 3.6 are two tables of two different EDBs to integrate, these EDBs are supplied by different sources namely s_1 and s_2. Integrating these tables consists on combining stored mass functions (values of the evidential attribute disease for the same objects).

To integrate these two tables, values of the attribute disease have to be combined. Therefore, mass functions values of the attribute disease for the first object in both tables (line 1 in EDB of s_1 and line 1 in EDB of s_2) have to be combined; these two mass functions are certain belief functions which are not contradictory. Combining these two values does not raise any problem contrary to second values of this attribute for both EDBs because a non null mass is affected to the $\emptyset$ alarming about an eventual

Table 3.5: A table of an EDB of a source s_1

Examination	**Time**	**Disease**
e_1	t_1	$m^{\Omega}_{s_1 1} : P$
e_2	t_2	$m^{\Omega}_{s_1 2} : P(0.5)$ $B(0.5)$

Table 3.6: Another table of an EDB of a second source s_2

Examination	**Time**	**Disease**
e_1	t_1	$m^{\Omega}_{s_2 1} : P$
e_2	t_2	$m^{\Omega}_{s_2 2} : P(0.2)$ $B(0.6)$ $\{P \cup B\}(0.2)$

conflict (Dempster, 1967).

Table 3.7 is the integrated table of table 3.5 and table 3.6, the conflict is marked by a non null mass attributed to the $\emptyset$.

Table 3.7: Integration of tables of s_1 and s_2

id	**Time**	**Target**
1	t_1	P
2	t_2	$P(0,2)$ $B(0.4)$ $\emptyset(0.4)$

Different manners of conflict solving are presented in Section 3.2, one manner is to prevent the conflict before it happens by discounting masses using the source's degree of reliability. This method takes into account the degree of reliability of each source before combining their mass functions, thus the conflict is eliminated or reduced from the beginning and even if it appears after the combination it will be bearable.

In the following, we suggest a method which aims to solve the conflict appearing when combining several mass functions while integrating tables from several EDBs.

3.4.2 Conflict estimation

(Martin et al., 2008) consider that the more mass functions are distant the more they are conflicting. Thus, the distance between two mass functions reflects the degree of conflict between them and also the degree of conflict between their sources.

Jousselme distance, detailed in (Jousselme et al., 2001) and already formulated in equation (3.24), is used because it takes into account interactions between focal elements with Jaccard coefficient but distances detailed in Section 3.3 can also be used. The degree of conflict between two sources (s_1 and s_2) is the distance between their corresponding mass functions, respectively m_1 and m_2.

$$\mathrm{Conf}(s_1, s_2) = d(m_1, m_2) \tag{3.31}$$

Example 3.4 Let us take the example of mass functions in tables 3.5 and 3.6, note that $m^{\Omega}_{s_ij}$ is the j^{th} mass function value of the attribute disease for the j^{th} record and the i^{th} source (s_i).

If the distance $d(m_{s_11}, m_{s_21}) = 0$ then $\mathrm{Conf}(s_1, s_2) = 0$. For this first record, both sources agree on the value of the attribute target.
If $d(m_{s_12}, m_{s_22}) = 0.2236$ then $\mathrm{Conf}(s_1, s_2) = 0.2236$. For this second record, both sources disagree partially on the value of the attribute target and the degree of conflict is about 0.2236.

This distance measure is a binary one because it computes the distance between only two mass functions reflecting the conflict between their sources with a restriction on the number of sources which has to be equal to 2.
When the number of sources exceeds 2 ($M > 2$), the conflict of a given source s_j can be computed in two different ways, each way is a different type of distance, therefore we distinguish two types of distance.

- **Distance type 1:** is the mean of distances between a mass function m_j supplied by the source s_j and all other mass functions without using a combination rule. For M sources, the conflict of the source s_j is the mean of distances between m^{Ω}_j, the mass function provided by s_j and all the other $M-1$ mass functions one by one. For each mass function m_i from all $M-1$ mass functions (all the mass functions except m_j supplied by s_j, the source subject of the conflict estimation); distances between m_j and each m_i are computed. Therefore $M-1$ values of distance are obtained. The conflict of s_j is the mean of these $M-1$ distances.

$$\mathrm{Conf}(s_j, s_M) = \frac{1}{M-1} \times \sum_{i=1, i\neq j}^{M-1} d(m_j, m_i) \tag{3.32}$$

Example 3.5 Suppose that we have three sources s_1, s_2 and s_3 supplying respectively $m_{s_1}^{\Omega}$, $m_{s_2}^{\Omega}$, and $m_{s_3}^{\Omega}$ in table 3.8.

Table 3.8: Distance type 1

Source	**Mass functions**	**Distance type 1**	**Conflict**
s_1	$m_{s_1}:\ P(0.5)$ $\{P \cup B \cup F\}(0.5)$	$d(m_{s_1}, m_{s_2}) = 0.5323$ $d(m_{s_1}, m_{s_3}) = 0.6164$	0.57435
s_2	$m_{s_2}:\ P(0.2)$ $B(0.6)$ $\{P \cup B\}(0.2)$	$d(m_{s_2}, m_{s_1}) = 0.5323$ $d(m_{s_2}, m_{s_3}) = 0.7832$	0.65775
s_3	$m_{s_3}:\ F(0.6)$ $\{P \cup F\}(0.4)$	$d(m_{s_3}, m_{s_1}) = 0.6164$ $d(m_{s_3}, m_{s_2}) = 0.7832$	0.6998

- **Distance type 2:** is the distance between a mass function m_j supplied by the source s_j and the combined mass function of all others except m_j. This method needs the use of a combination rule to combine the $M-1$ mass functions. Combination rules detailed in Chapter 5 may be used in this context as well as those not quoted.

For M sources, the conflict of the source s_j with all the other sources corresponds to the distance between m_j^{Ω}, the mass function supplied by this source, and m_M^{Ω} representing the combined mass function of the $M-1$ mass functions provided by all the other sources.

Example 3.6 Let us continue with the same example 3.5 using Dempster's rule of combination. The conflict of each source using distance type 2 is given in table 3.9:

Table 3.9: Distance type 2

Source	**Mass functions**	**Distance type 2**	**Conflict**
s_1	$m_{s_1}:\ P(0.5)$ $\{P \cup B \cup F\}(0.5)$	$d(m_{s_1}, m_M) = 0.4082$	0.4082
s_2	$m_{s_2}:\ P(0.2)$ $B(0.6)$ $\{P \cup B\}(0.2)$	$d(m_{s_2}, m_M) = 0.6650$	0.6650
s_3	$m_{s_3}:\ F(0.6)$ $\{P \cup F\}(0.4)$	$d(m_{s_3}, m_M) = 0.7282$	0.7282

Remark: Temporal complexity of distance type 2 is lower than temporal complexity of distance type 1 especially for a great number of mass functions. This difference in temporal complexities is due to the number of uses of Jousselme distance which is great in distance type 1 (proportionally to the number of mass functions) and low in distance type 2 (it is used only once).

This method is a generalization of that proposed in (Martin et al., 2008) for estimating reliabilities of evidential databases sources (Chebbah et al., 2010b; Chebbah et al., 2010a; Chebbah et al., 2011). The proposed method, detailed in algorithm 2, is in three steps; in the first step we compute the conflict of a source against all other existing sources as detailed above. Then, the reliability of this source is estimated on the basis of its conflict values and finally, all data are discounted proportionally to their source's reliability degree.

Algorithm 2 Reliability estimation and discounting mass functions

Require: Evidential databases (EDB_1, EDB_2, $\dots$, EDB_M) for M sources (s_1, s_2, $\dots$, s_M)
  for $i = 1$ to M **do**
    Step 1: Compute the conflict, $\mathrm{Conf}(s_i, s_M)$, of the source s_i according to all other $M-1$ sources using either distance type 1 or distance type 2.
    Step 2: Estimate the reliability α_i of s_i from its conflict: $\alpha_i = (1 - \mathrm{Conf}(s_i, s_M)^\lambda)^{(\frac{1}{\lambda})}$, with λ a real not null.
    Step 3: Discount mass functions stored in EDB_i with α_i .
  end for
  return α_i reliability of s_1.

Conflict is used to discount mass functions stored in evidential databases in order to estimate its source's reliability. When source's reliability is estimated, all mass functions are discounted before the combination when integrating evidential databases. In the following, the conflict is used in a clustering approach. Objects will be classified in groups minimizing the conflict between objects of the same cluster.

3.5 Clustering in the theory of belief functions

In machine learning, clustering techniques classify objects which values are uncertain into clusters. In this section, we detailed evidential C-means and evidential C-modes algorithms which are unsupervised as class labels of objects in the training set are not known *a priori*. In the end of this chapter, we detail our algorithm of evidential clustering with illustrations on random data in the following section.

3.5.1 Evidential C-means

The C-means technique is widely used in classification of objects when their attributes values are numerical. For example, a database may store several objects with numerical attributes values. Classifying objects from that databases groups together homogeneous objects in the same category. Such techniques are useful in data mining. In C-means algorithm, no *prior* knowledge on objects' classes is required because that algorithm is unsupervised whereas number of clusters C has to be fixed. Two main parameters are emphasized in C-means algorithm:

- Centers: Suppose a database with N objects o_i $(1 \leq i \leq N)$ and c attributes a_j $(1 \leq j \leq c)$; each object o_i is a vector of c values V_{ij} according to attributes a_j. Cluster center is also a vector $Q = \{q_1, \dots, q_j, \dots, q_c\}$ where q_j is the mean of values of attribute a_j for all objects that are in the same cluster.

- Distance: Any distance like Euclidean distance can be used between any object and centers of clusters according to the attributes.

Algorithm 3 enumerates main steps for C-means algorithm. Each object is allocated to the nearest cluster (for that the distance between the object and its center is minimal) in an iterative way until reaching an unchanged partition.

Masson and Denoeux defined in (Denœux and Masson, 2003; Denœux and Masson, 2004; Masson and Denœux, 2004) a credal partition as the N-tuple, $\{m_1, \dots, m_N\}$, where m_i is the mass function about membership of an object o_i to one or more clusters from $\Omega_c = \{Cl_1, Cl_2, \dots, Cl_C\}$. Indeed, a set $\Omega_c = \{Cl_1, Cl_2, \dots, Cl_C\}$ is the set of classes; the membership of an object to any class of Ω_c is uncertain and formalized with a mass function. The set of all mass functions about memberships of all objects is a credal partition.

(Masson and Denœux, 2008) proposed an evidential C-means algorithm called ECM based on credal partitions. An object o_i membership is represented by a mass function m_i as partial knowledge on that object's class. Credal partitions show objects membership to one or more clusters. The main difference between ECM and C-means is computation of m_i. Thus, $\forall A \subseteq \Omega_c \setminus \emptyset$, $m_i(A)$ is the barycenter $\bar{v}_i$ of the centers of classes composing A. Formally:

$$m_i(A) = \bar{v}_i,\ \forall A \subseteq \Omega_c \setminus \emptyset \tag{3.33}$$

with:

$$\bar{v}_i = \frac{1}{|A|} \sum_{k=1}^{C} s_{ki} v_k \tag{3.34}$$

where:

$$s_{ci} = \begin{cases} 1 & \text{if } Cl_k \in A \\ 0 & \text{else} \end{cases} \tag{3.35}$$

Note that v_k are centers of classes ($k \in [1, C]$). A distance is then used for ECM (Masson and Denœux, 2008). ECM is an evidential clustering of certain object data; (Denœux and Masson, 2004) proposed also a clustering of relational data. A constrained evidential clustering is also proposed for object and relational data (Antoine et al., 2010; Antoine et al., 2011; Antoine et al., 2012).

3.5.2 Belief C-modes

The C-modes technique is proposed to extend C-means technique in order to deal with categorical attributes. Thus, C-modes algorithm classifies objects that attributes' values are categorical. The two main following parameters are used in C-modes algorithm:

- Modes: A cluster mode is most frequently encountered categories in a cluster. Suppose a cluster Cl_k containing n_k objects $\{o_1, o_2, \ldots, o_{n_k}\}$. Note that V_{ij} are values of attribute a_j ($1 \leq j \leq c$) for objects o_i. Meaning that each object $o_i = \{V_{i1}, \ldots, V_{ij}, \ldots, V_{ic}\}$ has c values for each attribute a_j. A cluster mode $Q = \{q_1, \ldots, q_j, \ldots, q_c\}$ where q_j is the most frequent category for attribute a_j encountered in Cl_k. The category q_j is the most frequent value of attribute a_j for all objects in a cluster Cl_k.

- Dissimilarity measure: C-modes algorithm uses a simple matching dissimilarity measure to quantify the dissimilarity between an object o_i and a cluster mode Q defined as follows:

$$dis(o_i, Q) = \sum_{j=1}^{c} \delta(x_{ij}, q_j) \tag{3.36}$$

  where:

$$\delta(x_{ij}, q_j) = \begin{cases} 0 & \text{if } x_{ij} = q_j \\ 1 & \text{if } x_{ij} \neq q_j \end{cases} \tag{3.37}$$

Algorithm 3 is C-modes algorithm that classifies objects in homogeneous clusters when their values are categorical and certain. The algorithm classifies iteratively objects to the most similar cluster until reaching an unchanged cluster partition where no change in the membership of any object. When a cluster partition is obtained, a new object to classify is attributed to the cluster that mode is the most similar to that object. Cluster partition is not unique because it depends on initial cluster partition.

(Ben Hariz et al., 2006) adapted C-modes algorithm to uncertain attributes values. The proposed algorithm of belief C-modes (BKM) classifies objects when their attribute values are uncertain when uncertainty is modeled with the theory of belief functions. Note that BKM classifies N objects o_i ($1 \leq i \leq N$) from an evidential

Algorithm 3 C-modes and C-means algorithm

Require: N: objects to classify, C: number of clusters
  Choose C initial centers (modes) randomly.
  repeat
    for $i = 1$ to N **do**
      1. Compute the dissimilarity between o_i and all clusters centers (modes) using a distance or a dissimilarity measure.
      2. Allocate the object to the cluster that center (mode) is the nearest (most similar) according to the distance (dissimilarity) measure.
    end for
  until Cluster partition is unchanged
  return Cluster partition

database EDB where V_{ij} is the value of attribute a_j $(1 \leq j \leq c)$ for that object o_i. Attributes a_j are categorical; their domains are Ω_{a_j} and their values are mass functions verifying (2.2). Therefore, the algorithm is almost the same although clusters' modes and dissimilarity measure are adapted to support attributes evidential values as follows:

1. Modes: Suppose a cluster Cl_k of n_k objects $\{o_1, o_2, \dots, o_{n_k}\}$ with $o_i = \{V_{i1}, \dots, V_{ij}, \dots, V_{ic}\}$. The mode of Cl_k is noted $Q = \{q_1, q_2, \dots, q_c\}$ where q_j is the mean combination (equation (2.33)) of attributes a_j values of all objects in Cl_k.

2. Dissimilarity measure: Clusters modes q_j are mass functions m_j issued from the mean combination of attributes a_j values of all objects in the same cluster. The dissimilarity between an object and a cluster mode is the sum of Jousselme distance (3.24). Thus, the distance between V_{ij}, the mass function value of the attribute j for an object i, and one value of a cluster mode q_i is given as follows:

$$d(V_{ij}, q_j) = d(m_{ij}, m_j) \tag{3.38}$$

   where $m_{ij} = V_{ij}$, $m_j = q_j$ and $d(m_{ij}, m_j)$ is given by equation (3.24). The dissimilarity between an object o_i and a cluster mode Q is given as follows:

$$D(o_i, Q) = \sum_{j=1}^{c} d(V_{ij}, q_j) \tag{3.39}$$

(Ben Hariz et al., 2007) proposed also a method to choose initial modes, an incremental BKM where number of clusters may be incremented by one after learning clusters' partition (Ben Hariz and Elouedi, 2010b) and a decremental number of clusters (Ben Hariz and Elouedi, 2010a).

3.5.3 A new evidential clustering technique minimizing the conflict

The previous clustering algorithms use a distance and a dissimilarity measure; however Shubert proposed in (Schubert, 2003; Schubert, 2004) a clustering technique based on the conflict. Another perception of the conflict will be used in our clustering algorithm detailed here after. In this section, we will detail a new clustering technique to classify objects; their attributes values are evidential and classes are unknown. Proposed clustering algorithm uses a distance on belief functions given in (Jousselme et al., 2001) such as proposed by Ben Hariz et al. in BKM. Jousselme distance is adapted in BKM to quantify dissimilarities between objects and clusters modes. These are sets of mass functions; each one is the combination of an attribute's values of all objects classified into that cluster. An object is attributed to the cluster having the minimum dissimilarity to its mode.

Temporal complexity of BKM is quite high as clusters modes and distances are computed in each iteration. The combination by the mean rule to compute modes values leads to mass functions with a high number of focal elements. Hence, the bigger the cluster is, the least significant is the distance.

Therefore, we propose a clustering technique to classify objects that attributes values are uncertain; however uncertainty is modeled with the theory of belief functions. In the proposed algorithm, we do not use any cluster mode to avoid the growth of focal elements' number in clusters' modes. Temporal complexity is also significantly reduced because all distances are computed only once. Temporal complexities will be compared in the next section.

To classify objects o_i into C clusters, we use a clustering algorithm with a distance on belief functions given by (Jousselme et al., 2001). The number of clusters C is assumed to be known.
Proposed clustering technique is based on a conflict measure which quantifies how much is in conflict an object o_i with a cluster Cl_k. At first, we define the dissimilarity between two objects o_i and o_l as follows:

$$s(o_i, o_l) = \frac{1}{c}\sum_{j=1}^{c} d(m_{ij}, m_{lj}) \tag{3.40}$$

with d is Jousselme distance of equation (3.24). The conflict between two objects o_i and o_l is then defined as follows:

$$\mathrm{Conf}(o_i, o_l) = s(o_i, o_l) \tag{3.41}$$

The dissimilarity between an object o_i and a cluster Cl_k is the mean of the dissimilarities between o_i and all objects o_q that are classified into cluster Cl_k as follows:

$$S(o_i, Cl_k) = \frac{1}{n_k} \sum_{q=1}^{n_k} s(o_i, o_q) \tag{3.42}$$

The conflict between o_i and Cl_k is defined as follows:

$$\mathrm{Conf}(o_i, Cl_k) = S(o_i, Cl_k) \tag{3.43}$$

Each object is allocated to the most similar cluster in an iterative way till reaching an unchanged cluster partition. It is obvious that clusters number C must be known.
The evidential clustering algorithm is detailed in algorithm 4. In the first step clusters are initialized by random objects; then each object is allocated to the most similar cluster or equivalently the cluster that minimizes the conflict until reaching an unchanged cluster partition. The proposed algorithm minimizes the conflict into clusters; thus it

Algorithm 4 Evidential clustering

Require: N: objects to classify, C: number of clusters
  Initialize clusters with C random objects.
  repeat
    for $i = 1$ to N **do**
      1. Compute the conflict between o_i and all clusters using equation (3.43).
      2. Allocate the object to the cluster that minimizes the conflict. That object is allocated to the cluster with which it has the minimal degree of conflict. If the object's new cluster is different from its cluster in the last iteration, then cluster partition is updated.
    end for
  until Cluster partition is unchanged
  return Cluster partition

maximizes the conflict between clusters. Note that cluster partition of the proposed algorithm is one of the possible solutions of ECM algorithm. Indeed, distances to clusters modes (if clusters modes are computed at the end of the algorithm) are optimized and minimized.

Temporal complexity of the proposed algorithm is significantly optimized as pairwise distances are computed once a time from the beginning. We do not use any cluster mode. Consequently, there will be no problem of increasing number of focal elements because attributes' values are not combined. Optimization of complexity and number of focal elements of centers will be emphasized in the following section.

3.6 Experiments

To illustrate the proposed algorithm of evidential clustering and especially to compare it with the Belief C-modes algorithm, we generated randomly mass functions as detailed in algorithm 1. The comparison concerns some criterion like classification results, distances and variances of mass functions into clusters; finally the optimization of temporal complexity. These points are detailed in the following:

- Results comparison: C-means and C-modes algorithms are known to be convergent but clusters partition is not unique. Unfortunately, clusters partition of that clustering techniques depends on initial centers that are generally randomly chosen. In fact, evidential C-means, belief C-modes and evidential clustering algorithms are also convergent algorithms but clusters partition are numerous and dependent on the choice of initial centers. To compare results of our algorithm of evidential clustering to possible clusters partition that can provide belief C-modes; we firstly generated randomly a set of mass functions and performed our evidential clustering algorithm on that mass functions. Then, computed the distance of each object to all clusters' centers in order to check to which class would be attributed that mass functions according to the belief C-modes criteria. However, to affect a mass function to a cluster in the evidential clustering, we compute the means of distance of that mass function with all mass functions into each cluster and then attribute that mass function to the most similar cluster (having the minimal mean of distances). On the other hand, according to the belief C-means, a mass function is attached to the cluster such that the distance between that mass function and the cluster's center is minimal.

  To check if the evidential clustering solution is one of the possible solutions of belief C-modes, we performed our algorithm on randomly generated mass functions; then, computed the distance between each mass function and clusters' centers (they do not initially exist in the evidential clustering technique; we computed center after obtaining clusters partition). Finally, we looked for the cluster of each mass function according to the belief C-modes criteria. Finally, we compared clusters of mass functions obtained with the evidential clustering and possible classes of mass functions according to the minimal distance to clusters' centers.

  At the end of evidential clustering and once cluster partitions are obtained, we noticed that the obtained partition is one possible solution of the belief C-modes algorithm. Thus, clusters partition of evidential clustering are optimal according to the minimal distance to centers of clusters. Table 3.10, is an example of generated mass functions for different sizes of frames of discernment, the number of clusters is the same as the size of the frame of discernment (that choice is only for illustration)[2]. All generated mass functions are classified with the evidential clustering algorithm in the cluster with the minimal distance to clusters centers. Hence, clusters partition of evidential clustering is one possible solution of belief C-means algorithm.

Table 3.10: Tests of results of evidential clustering

Number of generated mass functions	$\|\Omega\|$	number of clusters	Number of mass functions in the optimal cluster[3]
100	3	3	100
	4	4	
	5	5	
	6	6	
	7	7	
	8	8	
	9	9	
	10	10	
	11	11	

- Distances and variances into clusters: For the purpose of comparing the evidential clustering and belief C-modes algorithms, we performed both algorithms on a set of 100 mass functions and computed the mean of distances into clusters and their variances. Thus, once clusters' partitions are obtained with both algorithms, the mean of pairwise distances of objects into the same clusters are obtained. For each cluster, we computed pairwise distances of objects classified into the same cluster; then, we computed the mean of distances into each cluster as well as variances of th distances. Figures 3.1 and 3.2 shows that distances between objects classified into the same cluster are slightly improved. Pairwise distances of objects into the same cluster are minimized.

- Temporal complexity: The main asset of the proposed clustering algorithm is the gain of run-times. The proposed algorithm insure a gain in the run-time of the clustering. In figure 3.3, we generated randomly 100 mass functions in a frame of discernment $|\Omega| \in [2, 10]$, the size of the frame of discernment and the number of clusters are assumed to be the same. In plots of figure 3.3, we notice that the run-time of evidential clustering is optimized according to that of belief C-modes. In figure 3.4, we also generated randomly 100 mass functions

[2]We assumed only one evidential attribute for illustrations but similar results are obtained for several attributes

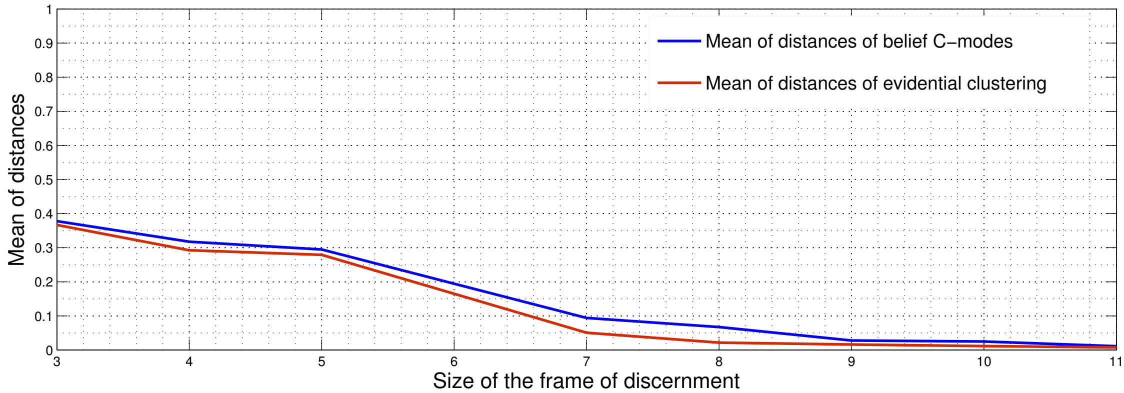

Figure 3.1: Mean of distances between objects classified into the same clusters

in a frame of discernment $|\Omega| \in [2, 10]$, there are 5 clusters. In plots of figure 3.3, only Ω varies in $[2, 10]$. We notice also a gain in the run-time of evidential clustering is optimized according to that of belief C-modes. In figure 3.5, we also generated randomly 200 mass functions in a frame of discernment $|\Omega| = 5$. In plots of figure 3.3, only the number of clusters varies in $[2, 10]$. We notice also a gain in the run-time of evidential clustering is optimized according to that of belief C-modes. Finally, figure 3.6 shows a big gain in the run-time of evidential clustering according to that of belief C-modes when the number of mass functions varies, $N \in [10, 1000]$. Temporal complexity of the evidential clustering algorithm is optimized and that optimization is especially noticed when the number of mass functions to classify is high and also when the frame of discernment contains many hypotheses.

Belief clustering technique provides a cluster partition that minimizes distances to centers. Distances between objects into the same clusters are also optimized. The main advantage of the belief clustering algorithm according to belief C-modes algorithm is the optimization of the temporal complexity. In fact, run-time of the belief clustering algorithm is better than the run-time of the belief C-modes. The optimization of run-time depends on the size of the frame of discernment $|\Omega|$, the number of clusters C and number of mass functions N.

3.7 Conclusion

In this chapter, we proposed a conflict measure for evidential databases. In fact, sources provide a set of imperfect information that is stored in evidential databases. When

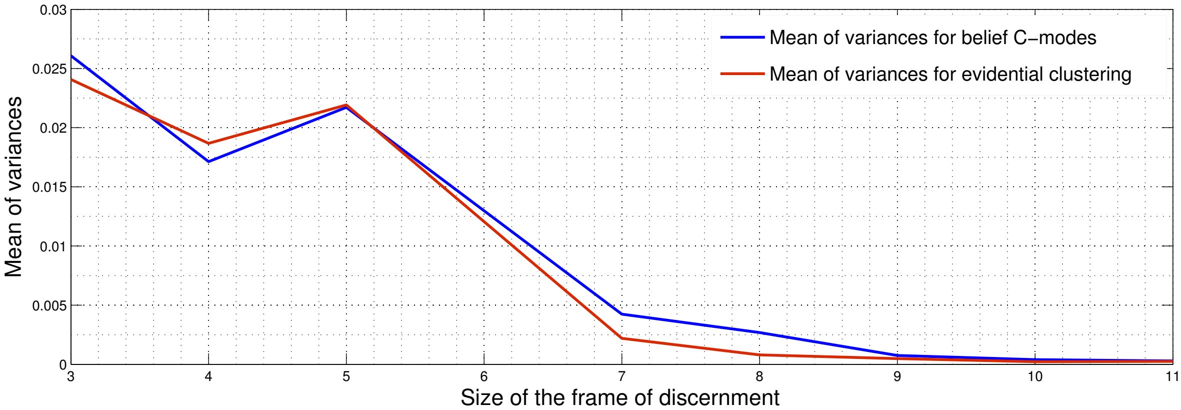

Figure 3.2: Mean of variances of distances between objects classified into the same clusters

using several evidential databases, a conflicting may appear reflecting the discord between beliefs of sources. The proposed conflict measure for evidential databases aims to estimate the disagreement between sources. We proposed an estimation of sources' reliabilities from that conflict measure in order to discount mass functions stored in evidential databases before the combination.

We proposed also an overview of evidential clustering that classifies mass functions. The belief clustering algorithm detailed in Section 3.5.3 is a clustering algorithm that classifies mass functions and minimizes the distance between objects and clusters' centers. Our proposed algorithm minimizes the conflict into clusters and also optimizes the run-time computation. The proposed algorithm will be used for estimating sources' independence in the next chapter.

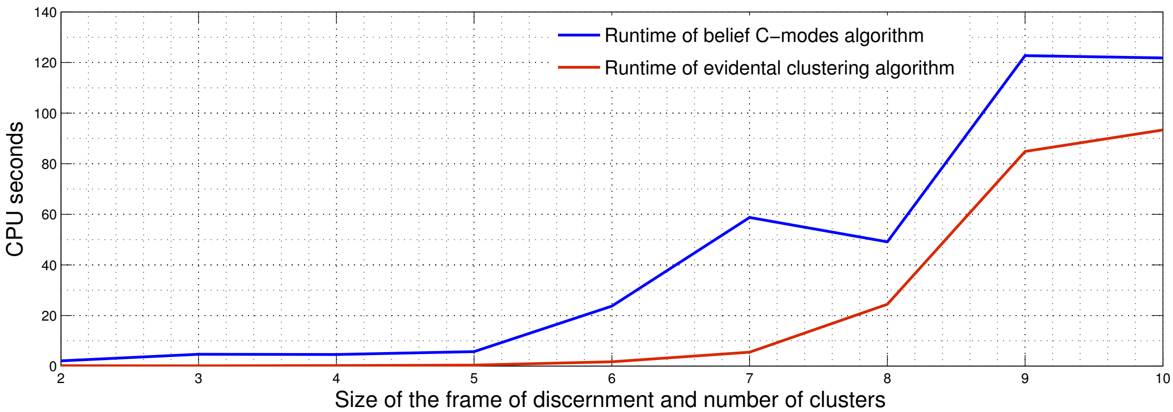

Figure 3.3: Comparison of run-times of belief C-modes and evidential clustering when $N = 100$, $C = |\Omega|$ and $|\Omega| \in [2, 10]$

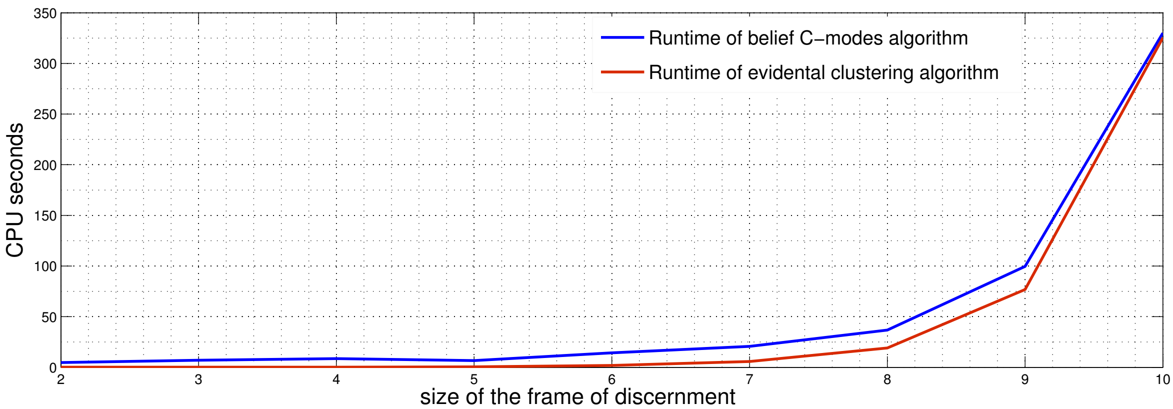

Figure 3.4: Comparison of run-times of belief C-modes and evidential clustering when $N = 100$, $C = 5$ and $|\Omega| \in [2, 10]$

Figure 3.5: Comparison of run-times of belief C-modes and evidential clustering when $N = 200$, $|\Omega| = 5$ and $C \in [2, 10]$

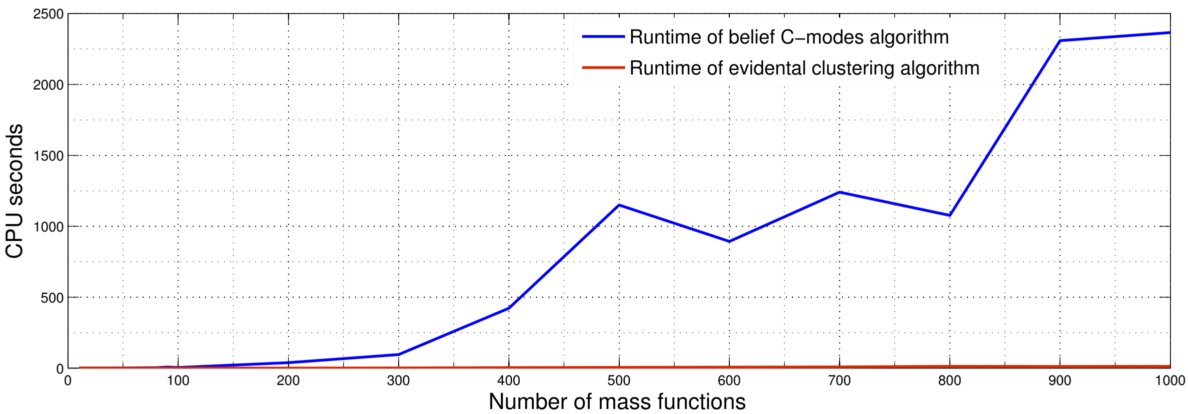

Figure 3.6: Comparison of run-times of belief C-modes and evidential clustering when $N \in [10, 1000]$, $\mid \Omega \mid = 5$ and $C = 5$

4
Sources independence estimation

Contents

Summary

An evidential database stores objects that values are mass functions. Classifying objects stored in an evidential database groups together similar objects in the same group. Similar objects are those which are not conflicting, thus conflict into clusters is minimized and that between clusters is maximized. Conflicting objects are those which values have conflicting focal elements.

In this chapter, we detail our statistical approach to estimate sources' independence. Our approach is based on the clustering algorithm detailed in the previous chapter in order to estimate sources' degrees of independence, positive and negative dependence.

4.1 Introduction

In Chapter 2, basics of the theory of belief functions are introduced. Uncertainty is modeled with the theory of belief functions and represented by mass functions which can be stored in evidential databases introduced in Chapter 3. In the previous chapter, evidential databases are emphasized as well as evidential clustering algorithms that can be used to classify objects stored in such databases. The clustering algorithm minimizing the conflict into clusters, introduced in Chapter 3, will be used to estimate sources independence degrees.

Some researches are focused on doxastic independence of variables such as (Ben Yaghlane et al., 2000; Ben Yaghlane et al., 2002a; Ben Yaghlane et al., 2002b; Ben Yaghlane, 2002); others (Shafer, 1976; Smets, 1993) tackled cognitive and evidential independence of variables. Variables independence can also be defined in terms of irrelevance and non-interactivity. This chapter is focused on measuring the independence of sources. We also present an overview of variables independence in the theory of belief functions framework although our research are focused on sources' independence.

We suggest a statistical approach to estimate the independence of sources on the basis of all evidential information that they provide. The aim of estimating sources' independence is to guide the choice of combination rules to use when combining belief functions provided by that sources; or to integrate degrees of independence in a new combination rule; or to discount belief functions with their source's degree of independence, positive and negative dependence. Uses of independence measure, detailed in this chapter, will be proposed in the next one.

A source is assumed to be cognitively independent on another one when the knowledge of beliefs of that source, does not affect beliefs of the other one. However, two sources are dependent when they are either communicating or having the same knowledge. Information on the independence of sources guides the choice of the type of combination rules to use. For example, when belief information are completely dependent only cautious or bold combinations can be applied (Denœux, 2008; Boubaker et al., 2013). In another hand, if evidential information are completely independent, another set of combination rules can be applied (Yager, 1987; Dubois and Prade, 1988; Smets and Kennes, 1994; Murphy, 2000; Martin and Osswald, 2007a; Lefèvre and Elouedi, 2013).

In the sequel, Section 4.2 is a state of art of variables independence in the theory of belief functions. In Section 4.3, we detail the notion of correlation between mass functions. In Section 4.4, clustering algorithm detailed in Chapter 3 will be used in the first step of the independence measure process. Independence measure is estimated in four steps; in the first step the clustering algorithm is applied; second a mapping between clusters is performed; then independence of clusters and sources are deduced in the last two steps. Independence is learned for only two sources and then generalized for a greater number of sources. In the case of dependent sources, type of this dependence is then estimated in Section 4.5. The proposed method is illustrated on random mass functions in Section 4.6. Finally, conclusions are drawn.

4.2 Independence concepts in the theory of belief functions

In the theory of probabilities, two hypotheses H_1 and H_2 are assumed to be statistically independent if $P(H_1 \cap H_2) = P(H_1) \times P(H_2)$ or $P(H_1|H_2) = P(H_1)$. In the context of the theory of belief functions, (Shafer, 1976) defined cognitive and evidential independence. Ben Yaghlane et al. defined in (Ben Yaghlane et al., 2002a; Ben Yaghlane et al., 2002b; Ben Yaghlane, 2002) variables' doxastic independence as well as non-interactivity and irrelevance.

4.2.1 Cognitive independence: weak independence

According to (Shafer, 1976), two variables are assumed to be cognitively independent with respect to a belief function if any new evidence that appears on only one of them does not change the evidence of the other variable. (Shafer, 1976) proposed the following definition:

Definition 4.1 *"Two frames of discernment may be called cognitively independent with respect to the evidence if new evidence that bears on only one of them will not change the degree of support for propositions discerned by the other" ((Shafer, 1976), page 149).*

The cognitive independence is a *weak independence*; two variables are independent with respect to a mass function if new evidence that bears on only one of the two variables does not change propositions discerned by the other one. For two variables X and Y such that Ω_X and Ω_Y their domains (frames of discernment) and $\Omega_X \times \Omega_Y$ the product space of domains Ω_X and Ω_Y. Variables X and Y are cognitively independent with respect to $m^{\Omega_X \times \Omega_Y}$ if and only if: $\forall x \subseteq \Omega_X$ and $\forall y \subseteq \Omega_Y$:

$$\mathrm{pl}^{\Omega_X \times \Omega_Y}(x, y) = \mathrm{pl}^{\Omega_X \times \Omega_Y \downarrow \Omega_X}(x) \times \mathrm{pl}^{\Omega_X \times \Omega_Y \downarrow \Omega_Y}(y) \tag{4.1}$$

Note that $\Omega_X \times \Omega_Y \downarrow \Omega_X$ is the marginalization of $\Omega_X \times \Omega_Y$ in Ω_X as detailed in Section 2.3.2.

Example 4.1 Let $\Omega_S = \{S, \bar{S}\}$, S for smoking and $\bar{S}$ for not smoking, a domain of a variable "Smoking attitude" describing whether a person is smoking or not. The frame of discernment $\Omega_A = \{A, T\}$, A for adult and T for teenager, defines possible age categories of a person. The product space $\Omega_S \times \Omega_A = \{a,\ b,\ c,\ d\}$, defining smoking and not smoking persons for each category of age; it can be schematically represented by figure 4.1. For short, we note $a = (S, T)$, $b = (S, A)$, $c = (\bar{S}, T)$ and $d = (\bar{S}, A)$.

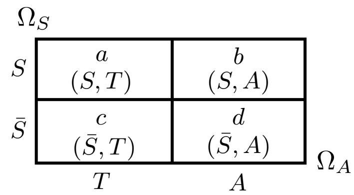

Figure 4.1: Product space $\Omega_S \times \Omega_A$

Suppose a mass function $m^{\Omega_S \times \Omega_A}$ defined on the product space $\Omega_S \times \Omega_A$ such that: $m^{\Omega_S \times \Omega_A}(a) = 0.26$, $m^{\Omega_S \times \Omega_A}(c) = 0.16$ and $m^{\Omega_S \times \Omega_A}(a \cup c) = 0.58$.
In table 4.1, $\mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_S}$ and $\mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_A}$ are computed, table 4.2 shows that variables "Smoking attitude" and "Age category" are cognitively independent according to $m^{\Omega_S \times \Omega_A}$ as the following equalities are verified:

$$\begin{cases} \mathrm{pl}^{\Omega_S \times \Omega_A}(a) = \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_A}(T) \times \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_S}(S) \\ \mathrm{pl}^{\Omega_S \times \Omega_A}(b) = \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_A}(A) \times \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_S}(S) \\ \mathrm{pl}^{\Omega_S \times \Omega_A}(c) = \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_A}(T) \times \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_S}(\bar{S}) \\ \mathrm{pl}^{\Omega_S \times \Omega_A}(d) = \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_A}(A) \times \mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_S}(\bar{S}) \end{cases} \tag{4.2}$$

Table 4.1: $\mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_S}$ and $\mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_A}$

Ω_S	$\mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_S}$	Ω_A	$\mathrm{pl}^{\Omega_S \times \Omega_A \downarrow \Omega_A}$
$\emptyset$	0	$\emptyset$	0
S	0.84	T	1
$\bar{S}$	0.74	A	0
$S \cup \bar{S}$	1	$T \cup A$	1

Table 4.2: Variables cognitively independent according to $m^{\Omega_S\times\Omega_A}$

$\Omega_S\times\Omega_A$	$m^{\Omega_S\times\Omega_A}$	$\Omega_S\times\Omega_A$	$\mathrm{pl}^{\Omega_S\times\Omega_A}$	Requirement
$\emptyset$	0	a	0.84	$\mathrm{pl}^{\Omega_S}(S)\times\mathrm{pl}^{\Omega_A}(T)=0.84\times 1=0.84$
a	0.26	b	0	$\mathrm{pl}^{\Omega_S}(S)\times\mathrm{pl}^{\Omega_A}(A)=0.84\times 0=0$
c	0.16	c	0.74	$\mathrm{pl}^{\Omega_S}(\bar{S})\times\mathrm{pl}^{\Omega_A}(T)=0.74\times 1=0.74$
$a\cup c$	0.58	d	0	$\mathrm{pl}^{\Omega_S}(\bar{S})\times\mathrm{pl}^{\Omega_A}(A)=0.74\times 0=0$

4.2.2 Evidential independence: strong independence

(Shafer, 1976) defines also a *strong independence* called evidential independence as follows:

Definition 4.2 *"Two frames of discernment are evidentially independent with respect to a support function if that support function could be obtained by combining evidence that bears on only one of them with evidence that bears on only the other" ((Shafer, 1976), page 149).*

According to Shafer, two variables are evidentially independent if their joint mass function can be obtained by combining marginal mass functions that bears on each one of them. Variables X and Y are evidentially independent with respect to $m^{\Omega_X\times\Omega_Y}$ if:

$$\begin{cases} \mathrm{pl}^{\Omega_X\times\Omega_Y}(x,y)=\mathrm{pl}^{\Omega_X\times\Omega_Y\downarrow\Omega_X}(x)\times\mathrm{pl}^{\Omega_X\times\Omega_Y\downarrow\Omega_Y}(y) \\ \mathrm{bel}^{\Omega_X\times\Omega_Y}(x,y)=\mathrm{bel}^{\Omega_X\times\Omega_Y\downarrow\Omega_X}(x)\times\mathrm{bel}^{\Omega_X\times\Omega_Y\downarrow\Omega_Y}(y) \end{cases} \tag{4.3}$$

Cognitive independence is weaker than evidential independence; evidential independence requires constraints on both pl and bel but cognitive independence requires only one constraint on only pl. Therefore, if two variables are evidentially independent according to a mass function then they are also cognitively independent according to that mass function. Whereas, if variables are cognitively independent according to a joint mass function they are not necessarily evidentially independent according to that mass function. Two variables X and Y are evidentially independent if and only if:

$$m^{\Omega_X\times\Omega_Y}(x,y)=m^{\Omega_X\times\Omega_Y\downarrow\Omega_X}(x)\times m^{\Omega_X\times\Omega_Y\downarrow\Omega_Y}(y) \tag{4.4}$$

Note that m, pl, and bel are normalized mass, plausibility and belief functions.

Example 4.2 We will show that variables "Smoking attitude" and "Age category" are evidentially independent according to the mass function of example 4.1. Requirement on pl is already checked in tables 4.1 and 4.2. In table 4.3, $\mathrm{bel}^{\Omega_X\times\Omega_Y\downarrow\Omega_X}$ and $\mathrm{bel}^{\Omega_X\times\Omega_Y\downarrow\Omega_Y}$ are computed. Requirement on bel is then checked in table 4.4. Therefore, "Smoking attitude" and "Age category" are evidentially independent according to $m^{\Omega_S \times \Omega_A}$.

Table 4.3: $\text{bel}^{\Omega_S \times \Omega_A \downarrow \Omega_S}$ and $\text{bel}^{\Omega_S \times \Omega_A \downarrow \Omega_A}$

Ω_S	$\text{bel}^{\Omega_S \times \Omega_A \downarrow \Omega_S}$	Ω_A	$\text{bel}^{\Omega_S \times \Omega_A \downarrow \Omega_A}$
$\emptyset$	0	$\emptyset$	0
S	0.26	T	1
$\bar{S}$	0.16	A	0
$S \cup \bar{S}$	1	$T \cup A$	1

Table 4.4: Variables evidentially independent according to $m^{\Omega_S \times \Omega_A}$

$\Omega_S \times \Omega_A$	$m^{\Omega_S \times \Omega_A}$	$\Omega_S \times \Omega_A$	$\text{bel}^{\Omega_S \times \Omega_A}$	Requirement
$\emptyset$	0	a	0.26	$\text{bel}^{\Omega_S}(S) \times \text{bel}^{\Omega_A}(T) = 0.26 \times 1 = 0.26$
a	0.26	b	0	$\text{bel}^{\Omega_S}(S) \times \text{bel}^{\Omega_A}(A) = 0.26 \times 0 = 0$
c	0.16	c	0.16	$\text{bel}^{\Omega_S}(\bar{S}) \times \text{bel}^{\Omega_A}(T) = 0.16 \times 1 = 0.16$
$a \cup c$	0.58	d	0	$\text{bel}^{\Omega_S}(\bar{S}) \times \text{bel}^{\Omega_A}(A) = 0.16 \times 0 = 0$

Accordingly, variables "Smoking attitude" and "Age category" are evidentially and so cognitively independent. Cognitive independence does not imply evidential independence but evidential independence implies cognitive independence.

4.2.3 Non-interactivity of variables

Non-interactivity illustrates the *compositional independence*. Two variables X and Y are non-interactive according to a joint mass function $m^{\Omega_X \times \Omega_Y}$ if it can be retrieved by combining variables marginal mass functions using Dempster's rule. Variables X and Y are non-interactive with respect to $m^{\Omega_X \times \Omega_Y}$ noted $X \perp_m Y$ if:

$$m^{\Omega_X \times \Omega_Y} = m^{(\Omega_X \times \Omega_Y \downarrow \Omega_X) \uparrow \Omega_X \times \Omega_Y} \oplus m^{(\Omega_X \times \Omega_Y \downarrow \Omega_Y) \uparrow \Omega_X \times \Omega_Y} \tag{4.5}$$

That implies the following equalities:

$$\begin{cases} \text{pl}^{\Omega_X \times \Omega_Y}(x, y) = \text{pl}^{\Omega_X \times \Omega_Y \downarrow \Omega_X}(x) \times \text{pl}^{\Omega_X \times \Omega_Y \downarrow \Omega_Y}(y) \\ \text{bel}^{\Omega_X \times \Omega_Y}(x, y) = \text{bel}^{\Omega_X \times \Omega_Y \downarrow \Omega_X}(x) \times \text{bel}^{\Omega_X \times \Omega_Y \downarrow \Omega_Y}(y) \end{cases} \tag{4.6}$$

Note that non-interactivity and evidential independence are equivalent.

Example 4.3 As shown in example 4.2, "Smoking attitude" and "Age category" are evidentially independent and non-interactive according to the mass function $m^{\Omega_S \times \Omega_A}$ detailed in example 4.1.
To check that $m^{\Omega_S \times \Omega_A} = m^{(\Omega_S \times \Omega_A \downarrow \Omega_A) \uparrow \Omega_S \times \Omega_A} \oplus m^{(\Omega_S \times \Omega_A \downarrow \Omega_S) \uparrow \Omega_S \times \Omega_A}$, marginal mass functions
$m^{\Omega_S \times \Omega_A \downarrow \Omega_A}$ and $m^{\Omega_S \times \Omega_A \downarrow \Omega_S}$ are computed in table 4.5.
To combine marginal mass functions $m^{\Omega_S \times \Omega_A \downarrow \Omega_A}$ and $m^{\Omega_S \times \Omega_A \downarrow \Omega_S}$, they must be defined on a common space $\Omega_A \times \Omega_S$ thus we proceed to a vacuous extension as shown in table 4.6. The combination of $m^{(\Omega_S \times \Omega_A \downarrow \Omega_A) \uparrow (\Omega_S \times \Omega_A)}$ and $m^{(\Omega_S \times \Omega_A \downarrow \Omega_S) \uparrow (\Omega_S \times \Omega_A)}$ with Dempster's rule is illustrated in table 4.7. Note that:
$m^{\Omega_S \times \Omega_A} = m^{(\Omega_S \times \Omega_A \downarrow \Omega_A) \uparrow \Omega_S \times \Omega_A} \oplus m^{(\Omega_S \times \Omega_A \downarrow \Omega_S) \uparrow \Omega_S \times \Omega_A}$.
Therefore, attributes "Smoking attitude" and "Age category" are non-interactive according to $m^{\Omega_S \times \Omega_A}$ because $m^{\Omega_S \times \Omega_A} = m^{(\Omega_S \times \Omega_A \downarrow \Omega_A) \uparrow \Omega_S \times \Omega_A} \oplus m^{(\Omega_S \times \Omega_A \downarrow \Omega_S) \uparrow \Omega_S \times \Omega_A}$.

Table 4.5: Marginal mass functions $m^{\Omega_S \times \Omega_A \downarrow \Omega_A}$ and $m^{\Omega_S \times \Omega_A \downarrow \Omega_S}$

$\Omega_S \times \Omega_A$	$m^{\Omega_S \times \Omega_A}$	Ω_S	$m^{\Omega_S \times \Omega_A \downarrow \Omega_S}$	Ω_A	$m^{\Omega_S \times \Omega_A \downarrow \Omega_A}$
a	0.26	S	0.26	T	1
c	0.16	$\bar{S}$	0.26	A	0
$a \cup c$	0.58	$S \cup \bar{S}$	0.58	$A \cup T$	0

Table 4.6: Vacuous extension of $m^{\Omega_S \times \Omega_A \downarrow \Omega_A}$ and $m^{\Omega_S \times \Omega_A \downarrow \Omega_S}$

$\Omega_S \times \Omega_A$	$m^{(\Omega_S \times \Omega_A \downarrow \Omega_S) \uparrow (\Omega_S \times \Omega_A)}$	$\Omega_S \times \Omega_A$	$m^{(\Omega_S \times \Omega_A \downarrow \Omega_A) \uparrow (\Omega_S \times \Omega_A)}$
$a \cup b$	0.26	$a \cup c$	1
$c \cup d$	0.16		
$a \cup b \cup c \cup d$	0.58		

Table 4.7: Combination of $m^{(\Omega_S \times \Omega_A \downarrow \Omega_A) \uparrow (\Omega_S \times \Omega_A)}$ and $m^{(\Omega_S \times \Omega_A \downarrow \Omega_S) \uparrow (\Omega_S \times \Omega_A)}$

$\Omega_S \times \Omega_A$	$m^{(\Omega_S \times \Omega_A \downarrow \Omega_A) \uparrow (\Omega_S \times \Omega_A)} \oplus m^{(\Omega_S \times \Omega_A \downarrow \Omega_S) \uparrow (\Omega_S \times \Omega_A)}$
a	0.26
c	0.16
$a \cup c$	0.58

4.2.4 Irrelevance of variables

Independence can also be defined in terms of *irrelevance*. Two variables are irrelevant if the knowledge of the value of one variable does not change the belief on the other one. In the theory of belief functions, irrelevance is based on the conditioning. Variables X and Y are irrelevant with respect to m, noted $\mathrm{IR}_m(X, Y)$ if the marginal mass function on X is obtained by conditioning the joint mass function on values y of Y and marginalizing this conditioned joint mass function on X:

$$m^{\Omega_X \times \Omega_Y}[y]^{\downarrow \Omega_X}(x) = m^{\Omega_X \times \Omega_Y \downarrow \Omega_X}(x) \tag{4.7}$$

Note that this equality is replaced by proportionality $\propto$ when $m^{\Omega_X \times \Omega_Y}[y]^{\downarrow \Omega_X}$ and $m^{\Omega_X \times \Omega_Y \downarrow \Omega_X}$ are not normalized.

Example 4.4 Suppose that the following mass function $m^{\Omega_S \times \Omega_A}$ is the joint mass function for variables "Smoking attitude" and "Age category":
$m^{\Omega_S \times \Omega_A}(a \cup c) = 0.75$, $m^{\Omega_S \times \Omega_A}(b \cup d) = 0.13$ and $m^{\Omega_S \times \Omega_A}(\Omega_S \times \Omega_A) = 0.12$.
The mass function $m^{\Omega_S \times \Omega_A}$ is conditioned on $\{a \cup b\}$ and $\{c \cup d\}$, then marginalized on Ω_A as shown in tables 4.8 and 4.9.
Table 4.9 shows that:

$$\begin{cases} m^{\Omega_S \times \Omega_A}[a \cup b]^{\downarrow \Omega_A} = m^{\Omega_S \times \Omega_A \downarrow \Omega_A} \\ m^{\Omega_S \times \Omega_A}[c \cup d]^{\downarrow \Omega_A} = m^{\Omega_S \times \Omega_A \downarrow \Omega_A} \end{cases}$$

Thus, variables "Smoking attitude" and "Age category" are irrelevant.

Table 4.8: Conditioned mass function on $\{a \cup b\}$ and marginalized on Ω_A

$\Omega_S \times \Omega_A$	$m^{\Omega_S \times \Omega_A}$	$\Omega_S \times \Omega_A$	$m^{\Omega_S \times \Omega_A}[a \cup b]$	Ω_A	$m^{\Omega_S \times \Omega_A}[a \cup b]^{\downarrow \Omega_A}$
$a \cup c$	0.75	a	0.75	T	0.75
$b \cup d$	0.13	b	0.13	A	0.13
$a \cup b \cup c \cup d$	0.12	$a \cup b$	0.12	$T \cup A$	0.12

4.2.5 Doxastic independence of variables

Doxastic independence is especially proposed in the theory of belief functions by (Ben Yaghlane et al., 2000; Ben Yaghlane et al., 2002a; Ben Yaghlane et al., 2002b; Ben Yaghlane, 2002).

Table 4.9: Conditioned mass function on $\{c \cup d\}$ and marginalized on Ω_A

$\Omega_S \times \Omega_A$	$m^{\Omega_S\times\Omega_A}$	$\Omega_S \times \Omega_A$	$m^{\Omega_S\times\Omega_A}[c \cup d]$	Ω_A	$m^{\Omega_S\times\Omega_A}[c \cup d]^{\downarrow\Omega_A}$
$a \cup c$	0.75	c	0.75	T	0.75
$b \cup d$	0.13	d	0.13	A	0.13
$a \cup b \cup c \cup d$	0.12	$c \cup d$	0.12	$T \cup A$	0.12

Table 4.10: Marginalized mass function $m^{\Omega_S\times\Omega_A}$ on Ω_A

$\Omega_S \times \Omega_A$	$m^{\Omega_S\times\Omega_A}$	Ω_A	$m^{\Omega_S\times\Omega_A^{\downarrow}\Omega_A}$
$a \cup c$	0.75	T	0.75
$b \cup d$	0.13	A	0.13
$a \cup b \cup c \cup d$	0.12	$T \cup A$	0.12

Definition 4.3 *"Two variables are considered as doxastically independent only when they are irrelevant and this irrelevance is preserved under Dempster's rules of combination" (Ben Yaghlane et al., 2002a; Ben Yaghlane, 2002).*

In other words, two variables X and Y are doxastically independent if they are irrelevant with respect to $m \oplus m_0$ when they are irrelevant with respect to m and m_0. Thus, if $\mathrm{IR}_m(X,Y)$, $\mathrm{IR}_{m_0}(X,Y)$ and $\mathrm{IR}_{m\oplus m_0}(X,Y)$ are verified then X and Y are doxastically independent. We do not focus on variables independence (Shafer, 1976; Ben Yaghlane et al., 2002a; Ben Yaghlane et al., 2002b; Ben Yaghlane, 2002) but on *sources independence*. Variables marginal and conditional independencies are checked with respect to marginal and/or joint belief functions even if according frames of discernment are almost the same. Sources independence is computed according to a set of different belief functions provided by each source separately. Sources are dependent when all their beliefs are correlated; there is a link between all mass functions they provide. This problem is not tackled till now, we noticed a lack of references treating this problem.

4.3 Correlation of belief functions

Cognitive, evidential and doxastic independencies as well as non-interactivity and irrelevance defines variables' independencies from a joint mass function. Correlation can also inform about any relation between pieces of evidence.

Suppose m_A and m_B two updates of a mass function m_0 that is interpreted as the correlation between m_A and m_B. When m_0 is a vacuous mass function, there is no correlation between evidences that induced m_A and m_B. (Smets, 1992a) defines distinctness to illustrate independence of pieces of evidence as follows:

Definition 4.4 *Two pieces of evidence are distinct if and only if the mass function common to mass functions they induce is vacuous.*

The mass function m_{AB} is the combination of all pieces of evidence that have induced m_0, m_A and m_B, thus $m_{AB} = m_A \oplus m_B$. If $m^{\star}_{AB}$ induced by the conjunction of pieces of evidence that induced m_A and m_B individually, the correlation m_0 can be deduced by comparing m_{AB} and $m^{\star}_{AB}$.
When $m^{\star}_{AB}$ is known, the computation of the correlation is easy and the commonality m_0 of the mass function m_0 is given by:

$$\mathrm{q}_0(A) = \frac{\mathrm{q}_1(A) \times \mathrm{q}_2(A)}{\mathrm{q}^{\star}_{AB}} \qquad \forall A \subseteq \Omega \tag{4.8}$$

with $\mathrm{q}^{\star}_{AB}$ is the commonality function of the mass function $m^{\star}_{AB}$ induced by the conjunction of pieces of evidences that induced m_A and m_B. Unfortunately, $m^{\star}_{AB}$ is almost unknown.

Cosine function is also an indicator about correlation of mass functions, it is given as follows:

$$\cos(m_1, m_2) = \frac{m_1' W m_2}{||m_1||_w.||m_2||_w} \tag{4.9}$$

Where W is a weighting matrix that is required to be symmetric, square and positive definite. Cosine function computes the angle between two mass functions considered as vectors in a $2^{|\Omega|}$ dimensional space. If Υ denotes the angle between two vectors, $\cos(\Upsilon) = -1$ when vectors are opposite and thus mass functions are negatively correlated. Also $\cos(\Upsilon) = 1$ in the case of collinear vectors leading to correlated mass functions and $\cos(\Upsilon) = 1$ means that vectors are orthogonal. Other cosine values represent intermediate correlation values.

4.4 Learning sources independence degree

In this section, we detail a statistical approach to learn sources' independence degree as we proposed in (Chebbah et al., 2012a; Chebbah et al., 2013). To study sources' independence, we propose a method based on a great number of mass functions provided by both sources. This set of mass functions must be defined on the same frame of discernment according to the same problem. For example, two distinct doctors provide N diagnoses in the examination of the same N patients. In that case, the frame of discernment contains all diseases and is the same for both doctors. We define sources' independence as follows:

Definition 4.5 *Two sources are cognitively independent if they do not communicate and if their evidential corpora are different.*

According to definition 4.5, not only communicating sources are considered dependent but also sources having the same background of knowledge since their beliefs are correlated. The aim of estimating sources independence is either to guide the choice of combination rules when aggregating their beliefs, or to integrate this degree of independence in a new combination rule as detailed in Chapter 5. When sources have the same evidential corpus, the same background of knowledge and the same reasoning, they are considered dependent.

In the following, we propose a measure of independence $\mathrm{I_d}$, $(\mathrm{I_d}(s_1, s_2))$, as the independence of a source s_1 one another one s_2[1] verifying the following axioms:

1. Non-negativity: The independence of a source s_1 on another source s_2, $\mathrm{I_d}(s_1, s_2)$ cannot be negative, it is either positive or null.

2. Normalization: The degree of independence $\mathrm{I_d}$ is a degree over $[0, 1]$, it is null when the first source is dependent on the second one, equal to 1 when it is completely independent and a degree from $[0, 1]$ otherwise.

3. Non-symmetry: In the case where s_1 is independent on s_2, s_2 is not necessarily independent on s_1. Even if s_1 and s_2 are mutually independent, degrees of independence are not necessarily equal.

4. Identity: Any source is completely dependent on itself and $\mathrm{I_d}(s_1, s_1) = 0$.

If s_1 and s_2 are independent, there will be no correlation between their mass functions. The main idea is: First, classify mass functions provided by each source separately. Then, study similarities between cluster repartitions to reveal any dependence between sources. By using clustering algorithm, sources overall behavior is studied.

The proposed method is in three steps: First, mass functions of each source are classified using a clustering technique. Then, similar clusters are matched. Finally, weights of linked clusters and sources independence are quantified. In a case of dependent sources a degree of positive and negative dependence is also assessed.

[1]Reciprocally, $\mathrm{I_d}(s_2, s_1)$ is the independence of s_2 on s_1

4.4.1 Clustering of belief functions

Clustering algorithm detailed in Section 3.5.3 is used to classify two sets of N mass functions respectively provided by sources s_1 and s_2. Clustering algorithm is performed on all mass functions of s_1 independently of the clustering performed on those of s_2. We remind that all mass functions of both sources are defined on the same frame of discernment and so considered as values of only one attribute when classifying their corresponding objects. Mass functions can be stored in evidential databases introduced in Section 3.4.1.

For the same example of doctors, patients are objects to classify according to an attribute *disease*. Values of this attribute are mass functions defined on the frame of discernment Ω_d enumerating all possible diseases. Distance given by equation (3.42) can be simplified as we have only one attribute. However, we define a distance D of an object o_l and a cluster Cl_k is the mean of distances between m_i^Ω, the mass function value of the object o_l, and all N_k mass functions values of objects o_q classified into cluster Cl_k as follows:

$$D(o_l, Cl_k) = \frac{1}{N_k} \sum_{\mathrm{q}=1}^{N_k} d(m_i^\Omega, m_q^\Omega) \tag{4.10}$$

We fixed the number of clusters to the number of hypotheses in the frame of discernment[2]. In a classification point of view, number of hypotheses is the number of possible classes. For example, the frame of discernment of the attribute disease enumerates all possible diseases. Hence, when a doctor examines a patient, he gives a mass function as a classification of the patient in some possible diseases. Number of solutions in the frame of discernment is the number of possible classes (clusters). Other methods for determining the number of clusters are reviewed in (Masson and Denœux, 2008).

4.4.2 Cluster matching

Clustering algorithm groups similar mass functions into the same cluster. We mean by similar mass functions, near mass functions using Jousselme distance defined by equation (3.24). As this distance uses Jaccard coefficient, similar mass functions are those having non-contradictory or even similar focal elements.

After clustering technique, mass functions provided by s_1 are distributed on C clusters and mass functions of s_2 are also distributed on C clusters. Note that $C = n = |\Omega|$. We try to find a mapping between clusters in order to link those containing the same objects. If clusters are perfectly linked, meaning all objects are classified similarly for both sources, we can conclude that sources are dependent as they are choosing similar

[2] $C = \mid \Omega \mid$

focal elements (not contradictory at least) when providing mass functions for the same objects. If clusters are weakly linked[3], sources choose similar focal elements for different objects and so they are independent. Clusters' independence degree is proportional to the number of objects similarly classified. More clusters contain the same objects, more they are dependent more they are correlated.

Once clustering algorithm performed, the most similar clusters have to be linked; then a cluster matching is performed on clusters of s_1 and those of s_2. We note $Cl^1_{k_1}$ where $1 \leq k_1 \leq n$ for clusters of s_1 and $Cl^2_{k_2}$ where $1 \leq k_2 \leq n$ for those of s_2. The similarity between two clusters $Cl^1_{k_1}$ and $Cl^2_{k_2}$ is a proportion of objects simultaneously classified into $Cl^1_{k_1}$ and $Cl^2_{k_2}$:

$$\beta^i_{k_i,k_j} = \beta^i(Cl^i_{k_i}, Cl^j_{k_j}) = \frac{|Cl^i_{k_i} \cap Cl^j_{k_j}|}{|Cl^i_{k_i}|} \tag{4.11}$$

with $i, j \in \{1, 2\}$ and $i \neq j$.

The value $\beta^1_{k_1,k_2}$ quantifies a proportion of objects classified simultaneously in clusters $Cl^1_{k_1}$ and $Cl^2_{k_2}$ with regard to objects in $Cl^1_{k_1}$, analogically $\beta^2_{k_2,k_1}$ is a proportion of objects simultaneously in $Cl^1_{k_1}$ and $Cl^2_{k_2}$ with regard to those in $Cl^2_{k_2}$. In general case, we have $\beta^1_{k_1,k_2} \neq \beta^2_{k_2,k_1}$ since the number of objects classified into $Cl^1_{k_1}$ and $Cl^2_{k_2}$ can be different ($|Cl^1_{k_1}| \neq |Cl^2_{k_2}|$).
The similarity between clusters is the proportion of objects simultaneously classified into that clusters. In other words, we suppose that two sources s_1 and s_2 that provide evidential values for N objects. Evidential values are values of attributes assessed by sources. Once clustering algorithm is performed the similarities between clusters are deduced from the number of objects commonly classified into clusters.

We remind that β^1 are similarities towards s_1 and β^2 are those towards s_2. It is obvious that $\beta^i(Cl^i_{k_i}, Cl^j_{k_j}) = 0$ when $Cl^i_{k_i}$ and $Cl^j_{k_j}$ do not contain any common object; therefore they are completely different. The similarity $\beta^i(Cl^i_{k_i}, Cl^j_{k_j}) = 1$ is reached when clusters $Cl^i_{k_i}$ and $Cl^j_{k_j}$ are strongly similar; thus they contain the same objects.

Example 4.5 Suppose two sources s_1 and s_2 providing (each one) c evidential values for N objects; i.e. each source provides a mass function for each attribute from the c evidential values for all the N objects. Figure 4.2 illustrates the output of the clustering algorithm. Figure 4.3 illustrates the similarities of the cluster Cl^1_2 of the source s_1 with all clusters of s_2 (according to s_1) and also similarities of all clusters of s_2 with Cl^1_2

[3]The link between clusters is quantified from their similarities. Dissimilar clusters are weakly linked.

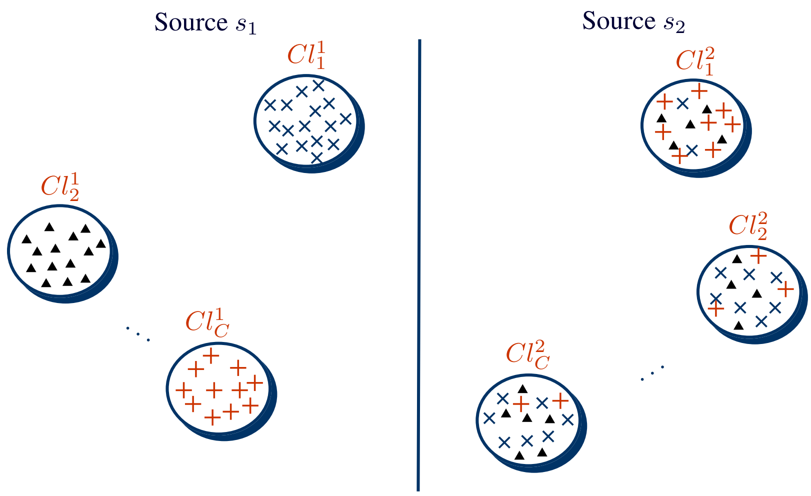

Figure 4.2: Clustering of N objects for which c values are assessed by s_1 and s_2

(according to s_2). Finally, figure 4.4 emphasizes pairwise similarities between clusters of s_1 and s_2. Once pairwise similarities are computed, the most similar clusters will be linked.

A similarity matrix M_1 containing clusters similarities of s_1 according to those of s_2 (β^1), and M_2 the similarity matrix between clusters of s_2 and those of s_1 (β^2) are defined as follows:

$$M_1 = \begin{pmatrix} \beta^1_{1,1} & \beta^1_{1,2} & \dots & \beta^1_{1,n} \\ \dots & \dots & \dots & \dots \\ \beta^1_{k,1} & \beta^1_{k,2} & \dots & \beta^1_{n,n} \\ \dots & \dots & \dots & \dots \\ \beta^1_{n,1} & \beta^1_{n,2} & \dots & \beta^1_{n,n} \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} \beta^2_{1,1} & \beta^2_{1,2} & \dots & \beta^2_{1,n} \\ \dots & \dots & \dots & \dots \\ \beta^2_{k,1} & \beta^2_{k,2} & \dots & \beta^2_{n,n} \\ \dots & \dots & \dots & \dots \\ \beta^2_{n,1} & \beta^2_{n,2} & \dots & \beta^2_{n,n} \end{pmatrix} \quad (4.12)$$

We note that M_1 and M_2 are almost different since $\beta^1_{k_1,k_2} \neq \beta^2_{k_2,k_1}$. Clusters of s_1 are matched to those of s_2 according to maximum of β^1 such that each cluster $Cl^1_{k_1}$ is linked to only one cluster $Cl^2_{k_2}$ and each cluster $Cl^2_{k_2}$ has only one cluster $Cl^1_{k_1}$ linked to it. The idea is to link iteratively clusters having the maximal β^1 in M_1 then eliminate these clusters and the corresponding line and column from the matrix till reaching a bijective cluster matching. For example, in the first iteration the maximum of M_1 is in line k_1 and column k_2. Cluster $Cl^1_{k_1}$ is linked to $Cl^2_{k_2}$, then line k_1 and column k_2 are eliminated from M_1. Algorithm 5 details cluster matching process.

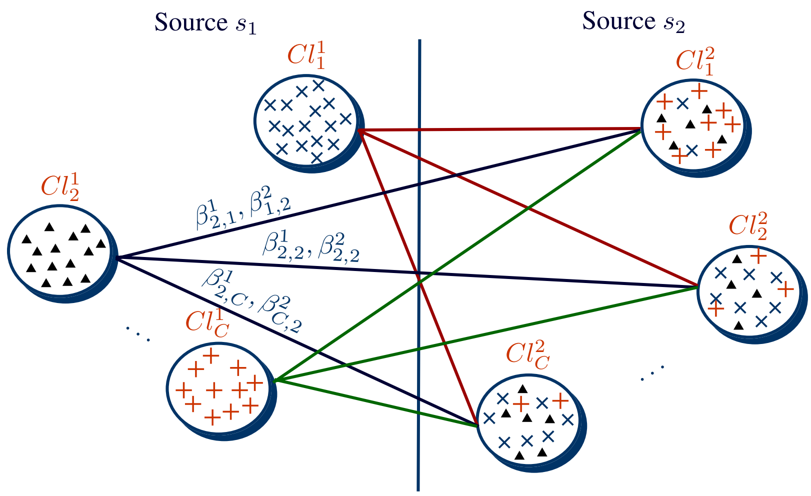

Figure 4.3: Similarities of Cl_2^1 and all clusters of s_2

Algorithm 5 Cluster matching

Require: Similarity matrix M_i.

while M_i is not empty **do**
    Find $\max(M_i)$ and indexes k_i and k_j of clusters having this maximal similarity.
    Map clusters k_i and k_j.
    Delete line k_i and column k_j from M_i.
end while
return Cluster matching.

This algorithm is iterative and the number of iteration is equal to the number of clusters C. Even if this algorithm is quite simple, it provides a matching of clusters in order to compare evidential information provided by both sources. The assignment algorithm proposed by (Munkres, 1957) for square matrices and that for rectangular matrices (Bourgeois and Lassalle, 1971) can also be used to minimize the dissimilarity between matched clusters. Other methods for cluster matching (Wemmert and Gançarski, 2002) and (Gançarski and Wemmert, 2005) can also be used.

Example 4.6 Assume two matched clusters Cl_1^1 and Cl_1^2 according to s_1, such that Cl_1^1 contains 25 objects from which 20 objects[4] are commonly classified with Cl_1^2. The cluster Cl_1^2 can be linked to Cl_3^1 according to s_2 as 25 objects from 45 are simultaneously classified into Cl_1^2 and Cl_3^1 according to s_2. Other methods for cluster matching detailed

[4]Objects can be records of evidential databases. Their values can be evidential for evidential attributes. The clustering is performed according to their evidential values but the matching takes into account the proportion of objects commonly classified into clusters.

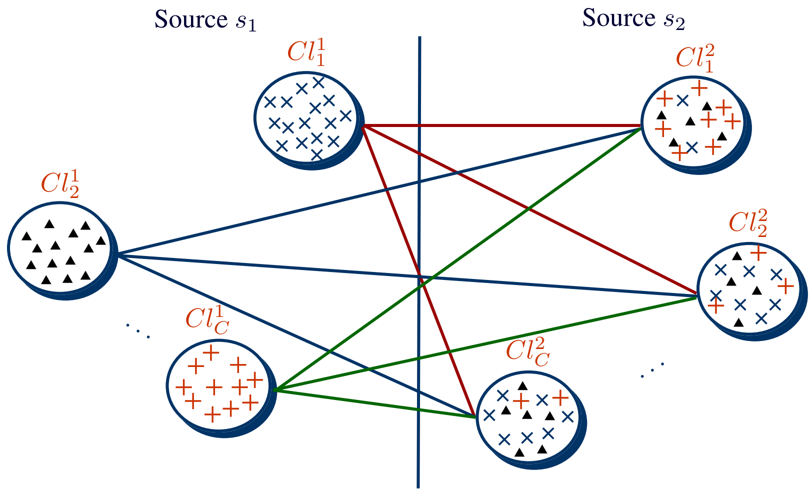

Figure 4.4: Pairwise similarities between clusters of s_1 and clusters of s_2

in (Wemmert and Gançarski, 2002) and (Gançarski and Wemmert, 2005) can also be used.

Note that different matchings are obtained for s_1 and s_2 because M_1 and M_2 are different.

Example 4.7 Suppose the following matrix M_1 a similarity matrix of s_1 according to s_2. Clusters of s_2 are indexed in rows and those of s_1 in lines. Number of clusters is 3, thus M_1 is given as follows:

$$M_1 = \left(\begin{array}{ccc|c} Cl_1^2 & Cl_2^2 & Cl_3^2 & \\ \hline 0.31 & 0.36 & 0.33 & Cl_1^1 \\ 0.48 & 0.3 & 0.22 & Cl_2^1 \\ 0.32 & 0.2 & 0.48 & Cl_3^1 \end{array} \right)$$

1. Iteration 1: maximum of M_1 is 0.48 either in cells $(2,1)$ or $(3,3)$. We choose randomly $(2,1)$, thus cluster Cl_1^2 is linked to Cl_2^1. Line 2 and row 1 are eliminated from M_1 as follows:

$$M_1 = \left(\begin{array}{cc|c} Cl_2^2 & Cl_3^2 & \\ \hline 0.36 & 0.33 & Cl_1^1 \\ 0.2 & 0.48 & Cl_3^1 \end{array} \right)$$

2. Iteration 2: maximum of M_1 is 0.48 in cell $(3,3)$, thus cluster Cl_3^2 is linked to Cl_3^1; line 2 and row 1 are eliminated from M_1 as follows:

$$M_1 = \left(\begin{array}{c|c} Cl_2^2 & \\ \hline 0.36 & Cl_1^1 \end{array} \right)$$

3. Iteration 3: finally clusters Cl_2^1 and Cl_1^1 are linked. Therefore Cl_1^1 is linked to Cl_2^2; Cl_2^1 is linked to Cl_1^2 and Cl_3^1 is linked to Cl_3^2 according to s_1.

Note that the maximum is chosen randomly when more than two clusters can be matched

4.4.3 Mass functions of clusters' independence

Once cluster matching is obtained, a degree of independence/dependence of matched clusters is quantified in this step. A set of matched clusters is obtained for both sources and a mass function quantifies each couple of clusters' independence. Assume that cluster $Cl_{k_1}^1$ is matched to $Cl_{k_2}^2$, a mass function $m^{\mathcal{I}}$ defined on the frame of discernment $\mathcal{I} = \{\bar{I},\ I\}$, such that $\bar{I}$ for dependent and I for independent hypotheses, describes degree of independent or dependent of this couple of clusters as follows:

$$\begin{cases} m_{k_i k_j}^{\mathcal{I},i}(\bar{I}) = \alpha_{k_i}^i \beta_{k_i,k_j}^i \\ m_{k_i k_j}^{\mathcal{I},i}(I) = \alpha_{k_i}^i (1 - \beta_{k_i,k_j}^i) \\ m_{k_i k_j}^{\mathcal{I},i}(\bar{I} \cup I) = 1 - \alpha_{k_i}^i \end{cases} \tag{4.13}$$

A mass function quantifies degree of independence of each couple of clusters according to each source; $m_{k_i k_j}^{\Omega_I,i}$ is a mass function for the independence of each linked clusters $Cl_{k_i}^i$ and $Cl_{k_j}^j$ according to s_i with $i, j \in \{1, 2\}$ and $i \neq j$. More a couple of matched clusters contains similar objects, more they are dependent.

Coefficient $\alpha_{k_i}^i$ is used to take into account of number of mass functions in each cluster Cl_{k_i} of the source i. Reliability factor $\alpha_{k_i}^i$ is not the reliability of any source but it can be seen as the reliability of the clusters independence estimation. Consequently, independence estimation is more reliable when clusters contain enough mass functions. For example, assume two clusters; one containing only one mass function and the second one containing 100 mass functions. It is obvious that the independence estimation of the second cluster is more precise and significant than the independence estimation of the first one.

Reliability factors $\alpha_{k_i}^i$ are functions of number of hypotheses in the frame of discernment $\mid \Omega \mid$, and number of objects classified in $Cl_{k_i}^i$ as follows:

$$\alpha_{k_i}^i = f(\mid \Omega \mid, \mid Cl_{k_i}^i \mid) \tag{4.14}$$

The bigger $\mid \Omega \mid$ is, the more mass functions are needed to have a reliable cluster independence estimation. For example, if $\mid \Omega \mid= 5$ then there are 2^5 possible focal elements, however independence estimation of a cluster containing 20 objects cannot be precise. No existing method to define such function f. Hence, we use simple heuristics as follows:

$$\alpha^i_{k_i} = 1 - \frac{1}{\mid Cl^i_{k_i} \mid^{\frac{1}{|\Omega|}}} \tag{4.15}$$

As shown in figure 4.5, if $\mid \Omega \mid$ and number of mass functions in a cluster are big enough, cluster independence mass function is almost not discounted. Proposed reliability factors are increasing with the increase of $\mid \Omega \mid$ and the increase of the number of mass functions in clusters. Reliability factor is an increasing function of $\mid \Omega \mid$ and $\mid Cl^i_{k_i} \mid$ which favors big clusters[5].

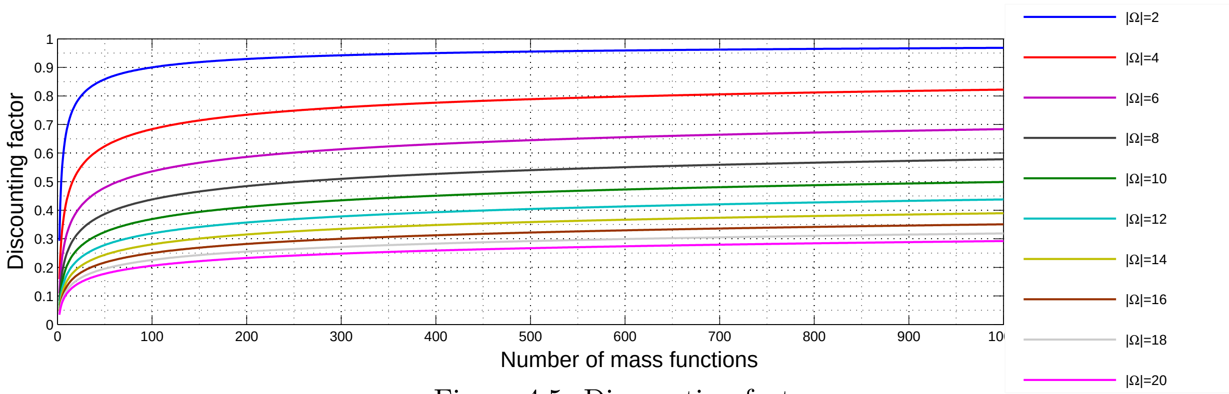

Figure 4.5: Discounting factors α_i

4.4.4 A measure of sources' independence

Obtained mass functions quantify each matched clusters independence according to each source. Therefore, n mass functions are obtained for each source such that each mass function quantifies the independence of each couple of matched clusters. The combination of n mass functions for each source using the mean, defined by equation (2.33), is a mass function $m^{\mathcal{I}}$ defining the whole independence of one source on another one:

$$m^{\mathcal{I},i}(A) = \frac{1}{n}\sum_{k_j=1}^{n} m^{\mathcal{I},i}_{k_i k_j}(A) \quad \forall A \subseteq 2^{\mathcal{I}} \tag{4.16}$$

with k_j is the cluster of s_j matched to k_i according to s_i and $m^{\mathcal{I},i}_{k_i k_j}$ are given by equation 4.13. The mean combination is chosen because it is idempotent and the combined

[5] Big clusters are those containing enough mass functions according to $\mid \Omega \mid$.

mass function is normalized if all mass functions are normalized. We note that any idempotent combination rule that can combine all types of mass functions can also be used (the cautious rule is limited to the combination of non-dogmatic mass functions). Two different mass functions $m^{\mathcal{I},1}$ and $m^{\mathcal{I},2}$ are obtained for s_1 and s_2 respectively. We note that $m^{\mathcal{I},1}$ is the combination of n mass functions representing the independence of matched clusters according to s_1 defined in equation (4.13) as follows:

$$\begin{cases} m^{\mathcal{I},i}(\bar{I}) = \dfrac{1}{n}\displaystyle\sum_{k_i=1}^{n} m^{\mathcal{I},i}_{k_i k_j}(\bar{I}) \\ m^{\mathcal{I},i}(I) = \dfrac{1}{n}\displaystyle\sum_{k_i=1}^{n} m^{\mathcal{I},i}_{k_i k_j}(I) \\ m^{\mathcal{I},i}(\bar{I}\cup I) = \dfrac{1}{n}\displaystyle\sum_{k_i=1}^{n} m^{\mathcal{I},i}_{k_i k_j}(\bar{I}\cup I) \end{cases} \tag{4.17}$$

or equivalently:

$$\begin{cases} m^{\mathcal{I},i}(\bar{I}) = \dfrac{1}{n}\displaystyle\sum_{k_i=1}^{n} \alpha^i_{k_i}\beta^i_{k_i,k_j} \\ m^{\mathcal{I},i}(I) = \dfrac{1}{n}\displaystyle\sum_{k_i=1}^{n} \alpha^i_{k_i}(1-\beta^i_{k_i,k_j}) \\ m^{\mathcal{I},i}(\bar{I}\cup I) = \dfrac{1}{n}\displaystyle\sum_{k_i=1}^{n} 1-\alpha^i_{k_i} \end{cases} \tag{4.18}$$

Mass functions $m^{\mathcal{I},1}$ and $m^{\mathcal{I},2}$ are almost always different since cluster matchings are different which verifies the axiom of non-symmetry. Proportions $\beta^1_{k_1,k_2}, \beta^2_{k_2,k_1} \in [0,1]$ verify the non-negativity and the normalization axioms. Finally, pignistic probabilities are computed from these mass functions in order to decide about sources independence $\mathrm{I_d}$ such that:

$$\begin{cases} \mathrm{I_d}(s_1,s_2) = \mathrm{BetP}(I) \\ \overline{\mathrm{I_d}}(s_1,s_2) = \mathrm{BetP}(\bar{I}) \end{cases} \tag{4.19}$$

If $\mathrm{I_d}(s_1,s_2) > \overline{\mathrm{I_d}}(s_1,s_2)$ we claim that s_1 is independent on s_2; if $\mathrm{I_d}(s_2,s_1) > \overline{\mathrm{I_d}}(s_2,s_1)$ we conclude that s_2 is independent on s_1.

4.4.5 General case of sources' independence

The method detailed above estimates the independence of one source on another one. Independence measure is non-symmetric because if a source s_1 is independent on a source s_2 then s_2 is not necessarily independent on s_1 and even if it is the case, degrees of independence are not necessarily the same. When combining mass functions provided by both sources, degrees of independence are needed to choose the appropriate type of combination rules. Combination rule using conjunctive and/or disjunctive combinations needs a strong hypothesis on sources independence. The question is what degree of independence do we attribute to a couple of sources (s_1, s_2) if degrees of independence of each one according to the other are different $(\mathrm{I_d}(s_1, s_2) \neq \mathrm{I_d}(s_2, s_1))$. It is wise to choose the minimum independence from $\mathrm{I_d}(s_1, s_2)$ and $\mathrm{I_d}(s_2, s_1)$ as the overall independence. Consequently, if at least one of two sources is dependent on the other, then sources are considered dependent. In other words, two sources are independent only if they are mutually independent. Hence, overall independence that is denoted $\mathrm{Ind}(s_1, s_2)$ is given by:

$$\mathrm{Ind}(s_1, s_2) = \min(\mathrm{I_d}(s_1, s_2), \mathrm{I_d}(s_2, s_1)) \tag{4.20}$$

We note that $\mathrm{Ind}(s_1, s_2)$ is non-negative, normalized, symmetric and identical.
We define an independence measure, noted Ind, generalizing the independence for M $(M \geq 2)$ sources verifying the following axioms:

1. Non-negativity: Many sources independence $\{s_1, s_2, s_3, \dots, s_M\}$, $\mathrm{Ind}(s_1, s_2, \dots, s_M)$ cannot be negative, it is either positive or null.

2. Normalization: Sources independence Ind is a degree in $[0, 1]$. The minimum 0 is reached when sources are completely dependent and the maximum 1 is reached when they are completely independent.

3. Symmetry: $\mathrm{Ind}(s_1, s_2, s_3, \dots, s_M)$ is sources overall independence and $\mathrm{Ind}(s_1, s_2, s_3, \dots, s_M) = \mathrm{Ind}(s_2, s_1, s_3, \dots, s_M) = \mathrm{Ind}(s_3, s_1, s_2, \dots, s_M) = \mathrm{Ind}(s_M, s_1, s_2, s_3, \dots)$.

4. Identity: $\mathrm{Ind}(s_1, s_1, s_1) = 0$. It is obvious that any source is completely dependent on itself.

5. Increasing with inclusion: $\mathrm{Ind}(s_1, s_2) \leq \mathrm{Ind}(s_1, s_2, s_3)$, more there are sources, more they are likely to be independent.

To compute the overall independence of M sources $\{s_1, s_2, \dots, s_M\}$, independencies of pairs of sources are computed and the maximum independence is the sources overall independence:

$$\mathrm{Ind}(s_1, s_2, \dots, s_M) = \max(\mathrm{Ind}(s_i, s_j)), \qquad \forall i \in [1, M] \ \ , j \in]i, M] \tag{4.21}$$

or equivalently:

$$\mathrm{Ind}(s_1, s_2, \dots, s_M) = \max(\min(I_d(s_i, s_j), I_d(s_j, s_i))), \qquad \forall i, j \in [1, M] \ \ i \neq j \tag{4.22}$$

Note that *the max is chosen to insure the property of increasing with inclusion.*

For example the overall independence of three sources is given by:

$$\mathrm{Ind}(s_1, s_2, s_3) = \max(\mathrm{Ind}(s_1, s_2), \mathrm{Ind}(s_1, s_3), \mathrm{Ind}(s_2, s_3)) \tag{4.23}$$

or

$$\begin{aligned} \mathrm{Ind}(s_1, s_2, s_3) = \max(&\min(\mathrm{I_d}(s_1, s_2), \mathrm{I_d}(s_2, s_1)), \\ &\min(\mathrm{I_d}(s_1, s_3), \mathrm{I_d}(s_3, s_1)), \\ &\min(\mathrm{I_d}(s_2, s_3), \mathrm{I_d}(s_3, s_2))) \end{aligned} \tag{4.24}$$

Sources independence degree is then integrated in the combination step using the mixed combination rule detailed in Chapter 5.

4.5 Positive and negative dependence for two sources

Mass functions of equation (4.17) quantifies sources' degrees of independence; unfortunately, that mass functions does not reflect if that dependence is either positive or negative. In the case of dependent sources, this dependence can be positive meaning that beliefs of one source are positively dependent on beliefs of the other one, thus both sources have either the same corpus, they are communicating or evidences they observe are either correlated or not distinct. In the case of negative dependence, the knowledge of one source is the opposite of the other one. In (Chebbah et al., 2012b), we detailed a method for learning sources positive and negative dependence; that method is detailed in the following.

Definition 4.6 *A source is positively dependent on another source when it is dependent on it and their beliefs are positively correlated.*

If a source s_1 is negatively dependent on s_2, beliefs of s_1 are almost opposing to beliefs of s_2.

Definition 4.7 *A source is negatively dependent on another source when it is dependent on it and their beliefs are negatively correlated.*

Sources are assumed to be dependent if they are choosing similar focal elements for similar objects, this is checked with the clustering algorithm detailed in Sections 4.4.1 and 3.5.3. Therefore, sources dependence is conditioned on the proportion of objects similarly classified since it is the proportion of objects where each source choose similar focal elements. Sources are dependent when similar focal elements are chosen by each source separately for similar cases (objects). If focal elements chosen by both sources for similar objects are almost similar, we can claim that sources are positively dependent but if focal elements chosen by each source for similar objects are different sources are negatively dependent.

As previously said, matched clusters are those having maximal similarity. Similarity of two clusters is the proportion of objects simultaneously classified into that clusters. In a case of positive dependent clusters, mass functions they contain are conflicting. In case of positive dependent clusters, the mass functions they contain are conflicting. In case of negative dependent clusters, the mass functions classified in that clusters are conflicting. Thus, we define the conflict between two clusters $Cl^i_{k_i}$ and $Cl^j_{k_j}$ ($\{i, j\} \in \{1, 2\}$ and $i \neq j$) as the mean of distances between objects simultaneously classified into $Cl^i_{k_i}$ and $Cl^j_{k_j}$ as follows:

$$\begin{cases} \mathrm{Conf}(Cl^i_{k_i}, Cl^j_{k_j}) = \dfrac{1}{|Cl^i_{k_i} \cap Cl^j_{k_j}|} \displaystyle\sum_{l \in E(Cl^i_{k_i}, Cl^j_{k_j})} d(m_l^{\Omega,i}, m_l^{\Omega,j}) & \text{if } |Cl^i_{k_i} \cap Cl^j_{k_j}| \neq 0 \\ 1 & \text{otherwise} \end{cases} \tag{4.25}$$

with

$$E(Cl^i_{k_i}, Cl^j_{k_j}) = \{k \in [1, g], g = |Cl^i_{k_i} \cap Cl^j_{k_j}|, m_k^{\Omega,i} \in Cl^i_{k_i} and\ m_k^{\Omega,j} \in Cl^j_{k_j}\} \tag{4.26}$$

Clusters' conflict is the mean of conflicts between objects commonly classified into clusters $Cl^i_{k_i}$ and $Cl^j_{k_j}$ as defined in Section 3.4.2. Conflict measure considers only common objects because conflict does not exist between sources only if mass functions according to the same problems and objects are compared. This conflict is computed for each source for each couple of matched clusters. A mass function defined on a frame of discernment [6] $\mathcal{P} = \{I,\ P,\ \bar{P}\}$ describes matched clusters $(Cl^i_{k_i}, Cl^j_{k_j})$ positive or negative dependence according to a sources s_i as follows:

$$\begin{cases} m^{\mathcal{P},i}_{k_i k_j}[\bar{I}](P) = 1 - \mathrm{Conf}(Cl^i_{k_i}, Cl^j_{k_j}) \\ m^{\mathcal{P},i}_{k_i k_j}[\bar{I}](\bar{P}) = \mathrm{Conf}(Cl^i_{k_i}, Cl^j_{k_j}) \end{cases} \tag{4.27}$$

These mass functions reflect contradiction between matched clusters. Conflict between clusters reflects degree of negative dependence. Dependent and conflicting clusters are negatively dependent whereas dependent and not conflicting clusters are positively dependent.

Mass functions of equation (4.27) are defined for each linked clusters according to each source. Note that this mass function is conditional to clusters dependence. Clusters positive or negative dependence is conditional to clusters dependence. Furthermore, for each couple of matched clusters $(Cl^i_{k_i}, Cl^j_{k_j})$, we have two mass functions; the first one is about clusters dependence and is defined by equation (4.13); the second one is a mass function on the matched clusters $(Cl^i_{k_i}, Cl^j_{k_j})$ positive and negative

[6] I for independence hypothesis, P for positive dependence and $\bar{P}$ for Negative dependence

dependence conditioned on clusters dependence as defined by equation (4.27).

To combine equations (4.27) and (4.13), they have to be defined on a common frame of discernment $\mathcal{P}$. The hypothesis on a high dependence of matched clusters in the mass functions of equations (4.27) have to be removed thus the use of the deconditioning detailed in Section 2.5.2. Deconditioning mass functions of equation (4.27) is given as follows:

$$\begin{cases} m_{k_ik_j}^{\mathcal{P},i}(P \cup I) = 1 - \mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j) \\ m_{k_ik_j}^{\mathcal{P},i}(\bar{P} \cup I) = \mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j) \end{cases} \tag{4.28}$$

The frame of discernment $\mathcal{P}$ is a refinement $\mathcal{I}$ such that $\bar{I} = P \cup \bar{P}$. Mass functions of equations (4.13) are refined as follows:

$$\begin{cases} m_{k_ik_j}^{\mathcal{P},i}(I) = \alpha_{k_i,k_j}^i\,(1 - \beta_{k_i,k_j}^i) \\ m_{k_ik_j}^{\mathcal{P},i}(P \cup \bar{P}) = \alpha_{k_i,k_j}^i\,\beta_{k_i,k_j}^i \\ m_{k_ik_j}^{\mathcal{P},i}(I \cup P \cup \bar{P}) = 1 - \alpha_{k_i,k_j}^i \end{cases} \tag{4.29}$$

Mass functions of equations (4.28) and (4.29) can be combined with the conjunctive rule of combination as they are defined on the common frame of discernment $\mathcal{P}$. The combined mass function is defined as follows:

$$\begin{cases} m_{k_ik_j}^{\mathcal{P},i}(I) = \alpha_{k_i,k_j}^i\,(1 - \beta_{k_i,k_j}^i) \\ m_{k_ik_j}^{\mathcal{P},i}(P) = \alpha_{k_i,k_j}^i\,\beta_{k_i,k_j}^i(1 - \mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j)) \\ m_{k_ik_j}^{\mathcal{P},i}(\bar{P}) = \alpha_{k_i,k_j}^i\,\beta_{k_i,k_j}^i\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j) \\ m_{k_ik_j}^{\mathcal{P},i}(I \cup P) = (1 - \alpha_{k_i,k_j}^i)\,(1 - \mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j)) \\ m_{k_ik_j}^{\mathcal{P},i}(I \cup \bar{P}) = (1 - \alpha_{k_i,k_j}^i)\,\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j) \end{cases} \tag{4.30}$$

Mass functions on the dependence of sources s_i and s_j are then obtained by the mean combination of all mass functions of equation (4.30) as follows:

$$m_i^{\mathcal{P}}(A) = \frac{1}{n}\sum_{k_i=1}^{n} m_{k_i,k_j}^{\mathcal{P},i}(A), \qquad \forall A \subseteq \mathcal{P} \tag{4.31}$$

That mass function quantifies degrees of independence, positive and negative independence of one source according to the other one. Thus for two sources s_1 and s_2, $m_1^{\mathcal{P}}$ is a belief function on the independence, positive dependence of s_1 on s_2; $m_2^{\mathcal{P}}$ is that of s_2 according to s_1. In the next chapter, these mass functions will be used to discount evidential information provided by sources. If one need to make a decision on the type of the dependence or independence of one source, pignistic transformation detailed in Section 2.2.3 can be used. Pignistic probabilities of hypotheses $\bar{P}$, P and I help to decide about the type of dependence according to the principal of maximal pignistic probabilities.

4.6 Experiments

Because of the lack of real evidential databases, we use generated mass functions to test the method detailed above. Moreover, it is difficult to simulate all situations with all possible combinations of focal elements for several degrees of independence between sources. At first two sets of mass functions are generated for two sources s_1 and s_2, then we illustrate for three sources.

4.6.1 Generated data depiction

Generating sets of N mass functions for several sources depends on sources independence. We discern two possible cases:

1. Independent sources: In general, to generate mass functions some information are needed: the number of hypotheses in the frame of discernment, $|\Omega|$, and the number of mass functions. We note that number of focal elements, and masses are chosen randomly. In a case of independent sources, masses can be anywhere and focal elements of both sources are chosen independently. Mass functions of s_1 and s_2 are generated following algorithm 1. We note that focal elements, their number and mass functions are chosen randomly according to the universal law. Algorithm 1 generates random mass functions for one, two or several sources independently. In that case, sources are considered independent as focal elements, their number and masses are independently and randomly chosen for both sources according to the universal law.

2. Dependent sources: The case of dependent sources is a bit difficult to simulate as several scenarios can occur. In this section, we will try to illustrate the most common situations.
   Generated mass functions for dependent sources are supposed to be consistent and do not enclose any internal conflict (Daniel, 2010). Consistent mass functions contain at least one focal element common to all focal sets.

   Algorithm 6 generates a set of N consistent mass functions defined on a frame of discernment of size $|\Omega|$. In a case of dependent sources, they are almost consistent and at least one of them is dependent on the other. To simulate a case where one source is dependent on another one, consistent mass functions of the first one are generated following algorithm 6 then those of the second source are generated knowing decisions of the first one. In a case of one source dependent on another one, it knows at least decisions of the other source. In a case of extreme dependence it may know also number of focal elements and their mass functions. Algorithm 7 generates a set of mass functions that are dependent on another set of mass functions. Dependence is due to the knowledge of other source's decisions.

Algorithm 6 Consistent mass functions generating

Require: $|\Omega|$, N : number of mass functions
for $i = 1$ to N **do**
  Choose randomly a focal set A_i (it can be a single point) from Ω.
  Find the set S of all focal sets including A_i.
  Choose randomly $|\mathcal{F}|$, the number of focal elements on $[1, |S|]$.
  Choose randomly $|\mathcal{F}|$ focal elements from S noted $\mathcal{F}$.
  Divvy the interval $[0, 1]$ into $|F|$ continuous sub-intervals.
  Masses of focal elements are intervals sizes.
end for
return N consistent mass functions

Algorithm 7 Dependent mass functions generating

Require: $|\Omega|$, N : number of mass functions, ω decision classes of another source
for $i = 1$ to N **do**
  Find the set S of all focal sets including ω_i.
  Choose randomly $|\mathcal{F}|$, the number of focal elements on $[1, |S|]$.
  Choose randomly $|\mathcal{F}|$ focal elements from S noted $\mathcal{F}$.
  Divvy the interval $[0, 1]$ into $|\mathcal{F}|$ continuous sub-intervals.
  Masses of focal elements are intervals sizes.
end for
return N consistent mass functions

4.6.2 Tests results

Algorithms detailed in the previous section are used to test some cases of sources dependence and independence. We note that in extreme cases where mass functions are certain or even when focal elements do not intersect; maximal values of independence are obtained. In a case of perfect dependence; mass functions have the same focal elements; however clusters contain mass functions with consistent focal elements. Clustering is performed according to focal elements and clusters are perfectly linked.

- Independent sources: In this paragraph, mass functions are independent. Focal elements and mass functions are randomly chosen ensuing algorithm 1. For tests, we choose $|\Omega| = 5$ which is considered as medium-sized frame of discernment and $N = 100$. Table 4.11 illustrates the mean of 100 tests in the case of independent sources. As a matter of fact, 100 independent mass functions[7] are generated for

[7] We are talking about mass functions and not objects because we consider only one evidential attribute. Indeed, we generate 100 mass functions values of one evidential attribute for 100 objects; thus, we will talk about mass functions rather then objects because the example is quite simple.

two sources 100 times and the mean of sources independence is illustrated on table 4.11. The mean of 100 tests for two dependent sources yields to a degree of independence Ind $= 0.67$, thus sources are independent. The independence of three sources is the maximum of degrees of independence of all couples of sources. To illustrate the case of three independent sources, three sets of 100 independent mass functions are generated following algorithm 1 with $|\Omega| = 5$. The mean of 100 tests are illustrated in table 4.12.

Table 4.11: Mean of 100 tests on 100 generated mass functions for two sources

Dependence type	**Degree of independence**	**Overall independence**
Independence	$I_d(s_1, s_2) = 0.68$ $\bar{I}_d(s_1, s_2) = 0.32$	$\text{Ind}(s_1, s_2) = 0.67$
	$I_d(s_2, s_1) = 0.67$ $\bar{I}_d(s_2, s_1) = 0.33$	
Dependence	$I_d(s_1, s_2) = 0.34$ $\bar{I}_d(s_1, s_2) = 0.66$	$\text{Ind}(s_1, s_2) = 0.34$
	$I_d(s_2, s_1) = 0.35$ $\bar{I}_d(s_2, s_1) = 0.65$	

Table 4.12: Mean of 100 tests on 100 generated mass functions for three independent sources

Sources	**Degree of independence**	**Pairwise independence**	**Overall independence**
s_1-s_2	$I_d(s_1, s_2) = 0.65$ $\bar{I}_d(s_1, s_2) = 0.35$	$\text{Ind}(s_1, s_2) = 0.65$	
	$I_d(s_2, s_1) = 0.66$ $\bar{I}_d(s_2, s_1) = 0.34$		
s_1-s_3	$I_d(s_1, s_3) = 0.68$ $\bar{I}_d(s_1, s_3) = 0.32$	$\text{Ind}(s_1, s_3) = 0.68$	$\text{Ind}(s_1, s_2, s_3) = 0.68$
	$I_d(s_3, s_1) = 0.69$ $\bar{I}_d(s_3, s_1) = 0.31$		
s_2-s_3	$I_d(s_2, s_3) = 0.64$ $\bar{I}_d(s_2, s_3) = 0.36$	$\text{Ind}(s_2, s_3) = 0.64$	
	$I_d(s_3, s_2) = 0.65$ $\bar{I}_d(s_3, s_2) = 0.35$		

- Dependent sources: In a case of dependent sources, mass functions are generated ensuing algorithms 6 and 7. For tests, we choose $|\Omega| = 5$ and $N = 100$. We generate 100 mass functions of both s_1 and s_2 for 100 times and then compute the average of $\mathrm{I_d}(s_1, s_2)$, $\mathrm{I_d}(s_2, s_1)$ and $\mathrm{Ind}(s_1, s_2)$. Table 4.11 summarizes obtained results. Table 4.11 illustrates the mean of 100 independence degrees of two dependent sources providing each one 100 randomly generated mass functions. These sources are dependent with a degree $1 - \mathrm{Ind} = 0.66$.

To illustrate a case of three dependent sources, three sets of 100 dependent mass functions are generated following algorithms 6 and 7 when $|\Omega| = 5$. The mean of 100 degrees of independence are illustrated in table 4.13. In other words, 100 dependent mass functions defined on $|\Omega| = 5$ are generated using algorithms 6 and 7 for 100 times. The mean of 100 pairs of independent degrees are illustrated on table 4.13.

Table 4.13: Mean of 100 tests on 100 generated mass functions for three dependent sources

Sources	**Degree of independence**	**Pairwise independence**	**Overall independence**
s_1-s_2	$I_d(s_1, s_2) = 0.35$ $\bar{I}_d(s_1, s_2) = 0.65$	$\mathrm{Ind}(s_1, s_2) = 0.34$	$\mathrm{Ind}(s_1, s_2, s_3) = 0.68$
	$I_d(s_2, s_1) = 0.34$ $\bar{I}_d(s_2, s_1) = 0.66$		
s_1-s_3	$I_d(s_1, s_3) = 0.68$ $\bar{I}_d(s_1, s_3) = 0.68$	$\mathrm{Ind}(s_1, s_3) = 0.68$	
	$I_d(s_3, s_1) = 0.31$ $\bar{I}_d(s_3, s_1) = 0.69$		
s_2-s_3	$I_d(s_2, s_3) = 0.68$ $\bar{I}_d(s_2, s_3) = 0.64$	$\mathrm{Ind}(s_2, s_3) = 0.68$	
	$I_d(s_3, s_2) = 0.35$ $\bar{I}_d(s_3, s_2) = 0.65$		

In this second part of illustrations, we used generated random mass functions with keeping some control on number of focal elements in the two sets of mass functions generated for s_1 and s_2. In fact, we generated two sets of mass functions for both sources such that the number of focal elements is know and is the same for both. From that focal elements, we fixed also the number of focal elements that are exactly the same with same masses, the number of not conflicting focal elements and the number of conflicting focal elements. Algorithm 8 details steps of generating two sets of N mass functions with keeping control on number of not conflicting focal elements focal elements, number of the same focal elements with exactly the same masses and the number of conflicting focal elements. We recall that not conflicting focal elements are those which are intersecting. Suppose $\mathcal{F} = \{A_1, A_2, \ldots, A_{|\mathcal{F}|}\}$ the set of focal elements of a given mass function; A_i are not conflicting if they are discerning at least one same hypothesis (leading to $\bigcup A_i \neq \emptyset$).

Algorithm 8 Generating mass functions when focal elements are controlled

Require: $|\Omega|$, M mass functions for s_1 and s_2, $|\mathcal{F}|$: number of focal elements, $|\mathcal{F}_{NC}|$: number of not contradicting focal elements, $|\mathcal{F}_S|$: number of the same focal elements with the same masses
  Choose randomly a set $\mathcal{F}_C$ of not conflicting focal elements.
  Choose randomly $|\mathcal{F}_S|$ focal elements noted $\mathcal{F}_S$.
  for $i = 1$ to M **do**
    Choose randomly $|\mathcal{F}_{NC}|$ focal elements noted $\mathcal{F}_C \setminus \mathcal{F}_S$.
    Choose randomly $|\mathcal{F}_{CO}| = |\mathcal{F}| - (|\mathcal{F}_S| + |\mathcal{F}_{NC}|$ from $2^\Omega \setminus \mathcal{F}_C$ noted $\mathcal{F}_{CO}$.
    Focal elements $\mathcal{F}$ of m_i are $\mathcal{F}_S$ and $\mathcal{F}_{CO}$.
  end for
  Generate randomly masses like in algorithm 1 such that masses of F_S are the same for both mass functions.
  return M mass functions

Note that $|\mathcal{F}_{NC}| + |\mathcal{F}_S| + |\mathcal{F}_{CO}| = |\mathcal{F}|$ such that $|\mathcal{F}|$ is the number of focal elements, $|\mathcal{F}_S|$ is the number of the same focal elements that have the same masses, $|\mathcal{F}_{NC}|$ is the number of not conflicting focal elements and $|\mathcal{F}_{CO}|$ is the number of conflicting focal elements. The set $\mathcal{F}_{NC}$ of not conflicting focal elements is generated by choosing randomly any hypothesis and enumerating all subsets of 2^Ω such that their intersection with the selected focal element is not empty. For example, if the randomly chosen hypothesis is ω, $\mathcal{F}_{NC}$ is composed of subsets of 2^Ω supporting ω.

Algorithm 8 is used to generate two sets of N mass functions for two sources s_1 and s_2 such that the number of not conflicting focal elements is the same for mass functions of the same index for both sources. From that focal elements, some of them are exactly the same with the same mass for that mass functions. The rest of focal elements are randomly chosen for both sources.

Plots of figures 4.6, 4.7, 4.8, 4.9 and 4.10 emphasizes the overall independence of sources s_1 and s_2, $\mathrm{Ind}(s_1, s_2)$. In each plot, number of not conflicting focal elements and number of the same focal elements with exactly the same masses are the same for mass functions with the same index for both sources. In figure 4.6, we generated 100 mass functions with algorithm 8 for two sources s_1 and s_2. Mass functions are defined on a frame of discernment $|\Omega| = 3$; there are 4 focal elements and in each plot the number of not contradictory focal elements is fixed. For the blue plot, there are 4 focal elements and all these 4 focal elements are not contradicting. The number of the same focal elements with the same masses is variating in $[0, 4]$. The blue plot shows that whatever the number of the same focal elements with the same masses,

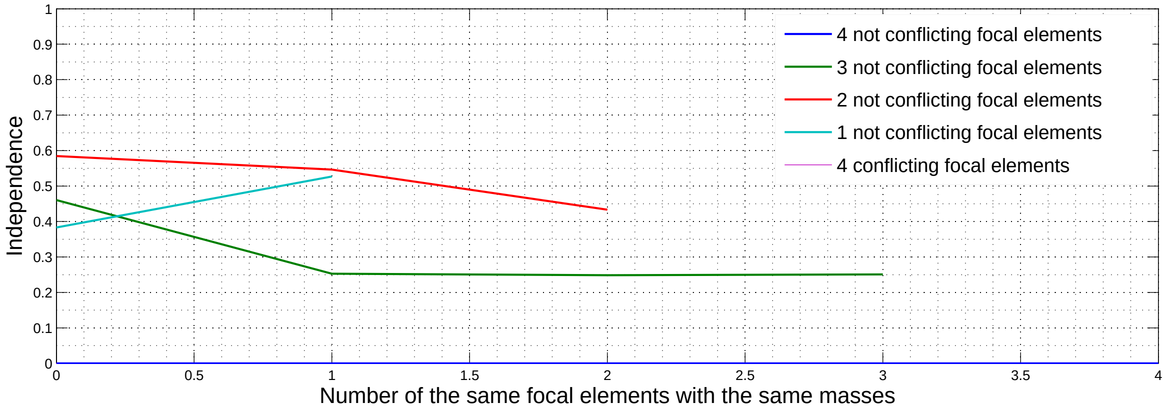

Figure 4.6: Independence degree when the number of focal elements is 4 and $|\Omega| = 3$

sources are dependent, $\text{Ind}(s_1, s_2) = 0$, because all focal elements are not conflicting. In that case the clustering of mass functions of both sources gives similar results. Mass functions with not conflicting focal elements are grouped together; as not conflicting focal elements are chosen for the same mass functions, clusters of both sources are similar leading to totally dependent sources.

In the green plot, mass functions of s_1 and s_2 have 4 focal elements from which 3 focal elements are not conflicting and only 1 conflicting focal element. The number of the same focal elements is variating on $[0, 3]$. Similarly for the red and blue plots, there are also 4 focal elements from respectively 2 and 1 focal elements are not conflicting and number of the same focal elements with the same masses are variating. Figure 4.8

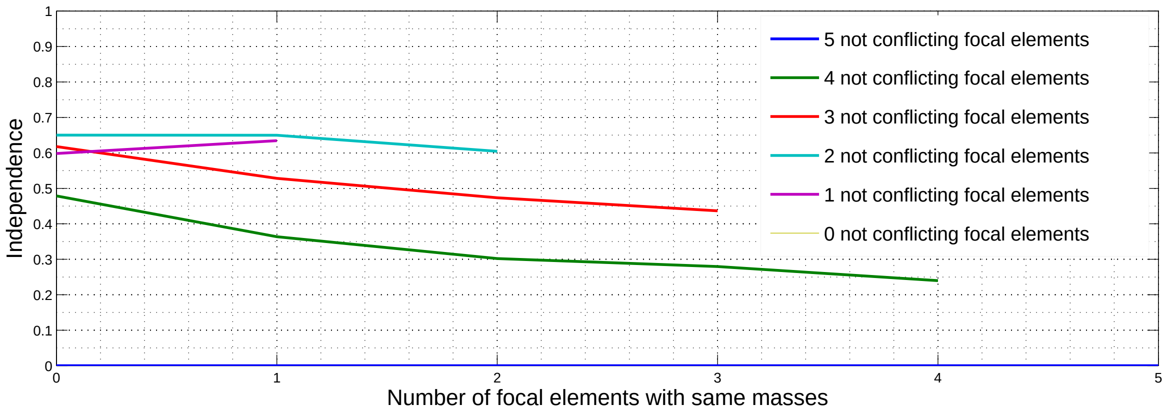

Figure 4.7: Independence degree when the number of focal elements is 5 and $|\Omega| = 4$

emphasizes the overall independence of s_1 and s_2, $\mathrm{Ind}(s_1, s_2)$, when $|\Omega| = 4$ with 8 focal elements. In each plot the number of not conflicting focal elements is fixed from which the same focal elements with the same masses is variating. Also figure 4.9 emphasizes

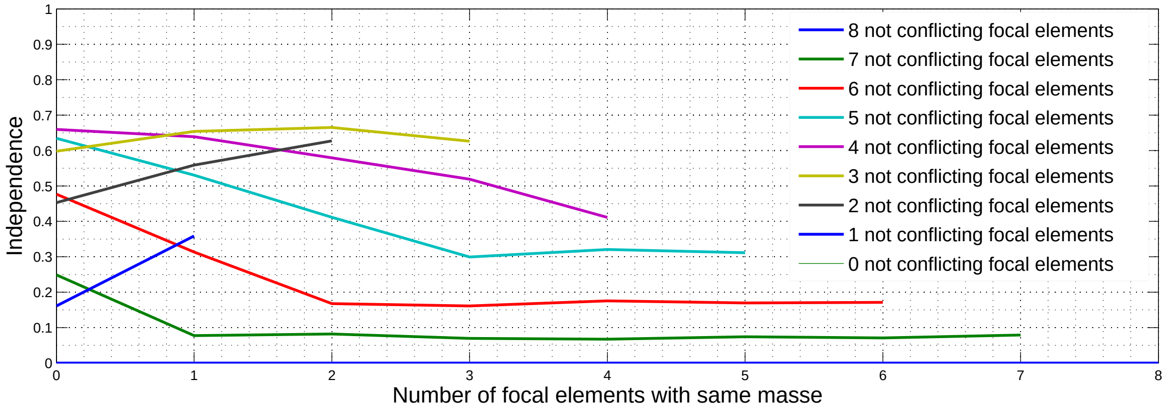

Figure 4.8: Independence degree when the number of focal elements is 8 and $|\Omega| = 4$

the overall independence of s_1 and s_2, $\mathrm{Ind}(s_1, s_2)$, when $|\Omega| = 5$ with 8 focal elements. In each plot the number of not conflicting focal elements is fixed from which the same focal elements with the same masses is variating. Finally, figure 4.10 emphasizes the

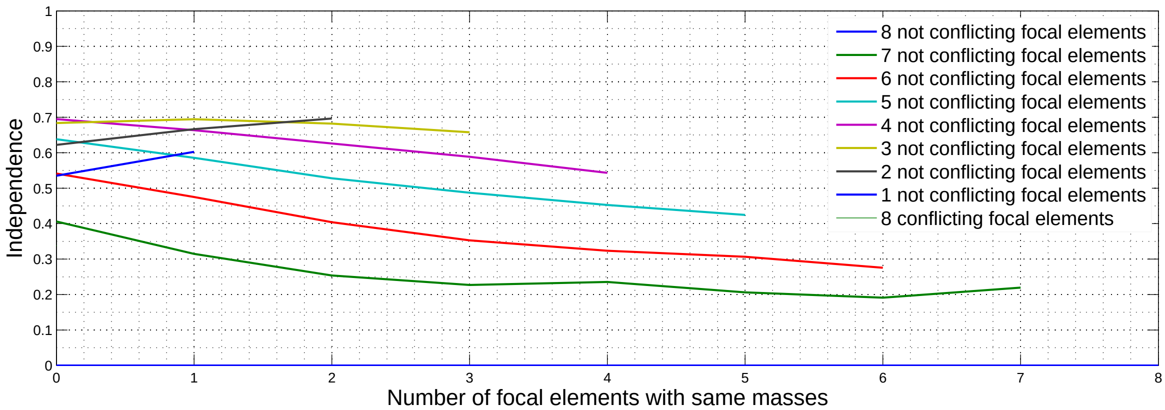

Figure 4.9: Independence degree when the number of focal elements is 8 and $|\Omega| = 5$

overall independence of s_1 and s_2, $\mathrm{Ind}(s_1, s_2)$, when $|\Omega| = 5$ with 16 focal elements. In each plot the number of not conflicting focal elements is fixed from which the same focal elements with the same masses is variating.

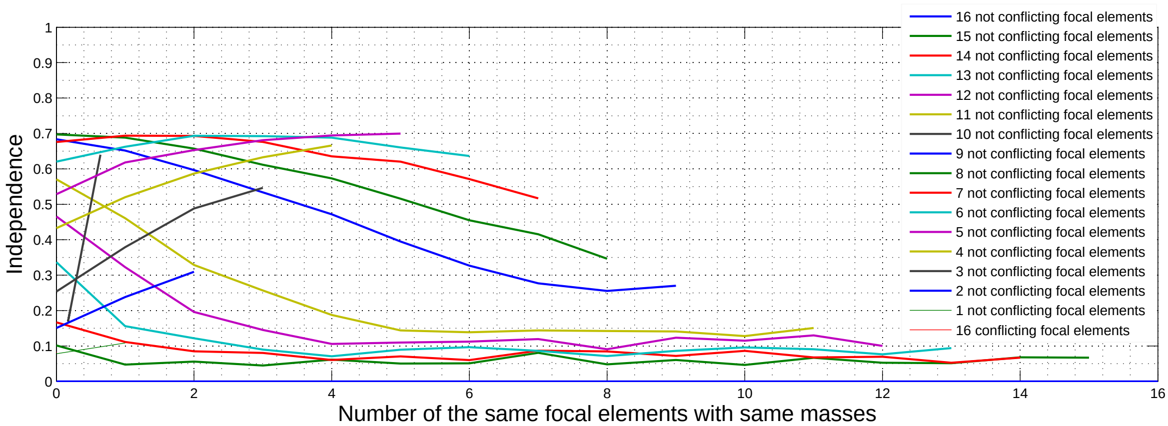

Figure 4.10: The independence when the number of focal elements is 16 and $|\Omega| = 5$

Figures 4.6, 4.7, 4.8, 4.9 and 4.10 shows that sources are dependent when the majority of focal elements are not conflicting or when the majority and conflicting. However, when almost all focal elements are not conflicting clusters of both sources are almost the same as mass functions are grouped according to the similar focal elements (not conflicting). In the case of almost all focal elements are conflicting they are also grouped according to the conflicting focal elements that are in this case similar. In the case of variating number of not conflicting, similar and conflicting focal elements sources are independent as there are no link between their focal elements.

4.7 Conclusion

In this chapter, we proposed a method to learn sources' cognitive independence in order to use the appropriate combination rule either when sources are cognitively dependent or independent. We proposed also to learn sources' positive and negative dependence from the positive or negative correlation of their mass functions. Sources are cognitively independent if they are different; not communicating and they have distinct evidential corpora. The proposed statistical approach is based on a clustering algorithm applied to mass functions provided by several sources. A pair of sources independence is deduced from weights of linked clusters after a matching of their clusters. Mass functions provided by both sources are classified into several clusters, then clusters of both sources are matched in order to estimate their independence, positive and negative dependence. Clusters positive and negative dependence is conditional to sources dependence. Mass functions of clusters dependence/independence and that of clusters positive and negative dependence are written in the same common frame of discernment in order to be combined. The sources' overall dependence is the combination of all mass functions on clusters positive/negative dependence.

In the next chapter, independence degree of sources is used either guide the choice of the combination rule if it is either 1 or 0; when it is a degree over $]0,1[$, a new combination rule that weights the conjunctive and cautious combinations with sources' independence degree will be proposed. The proposed combination rule takes into account independence degree of sources. The next chapter is about uses of dependence degree.

5
On the use of independence measure

Contents

Summary

Once sources' degrees of independence are learned, one can decide if sources are independent, positively or negatively dependent according to the principal of the maximum of pignistic probability. Information on sources' independence can guide the choice of the type of combination rules to use. Thus, in a case of dependent sources only combination rules tolerating redundant information can be used; otherwise if sources are independent combination rules using the conjunctive and/or disjunctive combination can be applied. Another solution is to integrate sources' degrees of independence in the provided mass functions. In this chapter, we recall some combination rules proposed in the framework of the theory of belief functions; we detail a new combination rule that weights the conjunctive and cautious combinations with sources' degrees of independence. Finally, we propose a tool for integrating sources' degrees of independence in their mass functions.

5.1 Introduction

In previous chapters, we was interested on sources' cognitive independence and especially on estimating sources' degrees of independence, positive and negative dependence. We proposed a statistical approach for learning sources' independence from all mass functions they provide. In our method, we used a clustering algorithm minimizing the conflict into clusters. The purpose of that algorithm is to group together similar mass functions or more precisely not conflicting mass functions.

After clustering, links between clusters are quantified in order to reveal any dependence between clusters. When two clusters contain mass functions values of the same objects, they are linked: The purpose is to link the most similar clusters containing mass functions values of the same objects. Two different matching of clusters are obtained according to each source, and a weight is learned for each couple of linked clusters reflecting their degree of dependence or independence. A set of mass functions for each couple of linked clusters is obtained for each source.

A mass function reflects the dependence or independence of each linked clusters. Another mass function can inform if the dependence is positive or negative in the case of dependent clusters. If mass functions of dependent clusters are not conflicting then this dependence is positive. In a such case, clusters contain same objects and their values are not conflicting. Thus, sources are choosing similar focal elements when reporting about same objects. Therefore, sources are almost dependent. In the case when mass functions are conflicting, linked clusters are negatively dependent because they contain mass functions values of same objects but these mass functions are conflicting. In that case sources are choosing conflicting focal elements when reporting about the same objects.

Eventually, two mass function are obtained for each couple of matched clusters; one mass function for the dependence or independence of that couple of clusters, and the second is for the positive or negative dependence for the same clusters. The two mass functions can be easily defined on a common frame of discernment in order to be combined. The combined mass function illustrates each couple of linked clusters independence, positive and negative dependence. The combination of obtained mass functions yields to a unique mass function on the source independence, positive or negative dependence on the other one.

Learning sources independence, positive or negative dependence has three aims: First, information about sources independence may guide the choice of the more appropriate combination rule to use in the combination of mass functions provided by these sources. Second, that information is used in a new combination rule that is a weighted average of the conjunctive and cautious combination. We propose a mixed combination rule that calibrates the conjunctive and cautious combination with sources' degree of independence. Finally, information on sources' degree of independence may be considered as a meta-knowledge that can be integrated in mass functions provided by that source. In fact, mass function on sources independence and mass function provided by that source are combined leading to a modified discounting operator.

The sequel of this chapter is organized as follows: In the second section, we present an overview of combination rules proposed in the framework of the theory of belief functions. In the third section, we detail our combination rule that is a weighted average of the conjunctive and cautious combinations. Finally before illustrating and concluding, we propose the justification of the combination of mass functions provided by sources and that mass function on sources independence.

5.2 Idempotent and non-idempotent combination rules

Combination rules merge several sources' beliefs in order to stress common hypotheses on which they agree. Let s_1 and s_2 be two distinct and independent sources providing two different mass functions m_1^Ω and m_2^Ω defined on the same frame of discernment Ω. Combining these belief functions induces a third one m_{12}^Ω defined on the same frame of discernment Ω summarizing m_1^Ω and m_2^Ω. To combine these mass functions, several rules are proposed. Some combination rules work under the strong assumption that sources are distinct and independent. That rules are not idempotent and do not tolerate redundancy. Other rules do not impose such condition because they are idempotent. Note that mass functions used in this chapter are defined on a frame of discernment Ω and combined mass functions are defined on the same frame Ω.

5.2.1 Non-idempotent combination rules

(Dempster, 1967) proposed the orthogonal sum as a combination rule. The orthogonal sum includes a uniform distribution of the conflict among all focal elements; thus it is hidden and does not appear. The orthogonal sum is defined for two distinct mass functions m_1 and m_2 as follows:

$$m_{1\oplus 2}^\Omega(A) = (m_1^\Omega \oplus m_2^\Omega)(A) = \begin{cases} \dfrac{\displaystyle\sum_{B\cap C=A} m_1^\Omega(B)\times m_2^\Omega(C)}{1-\displaystyle\sum_{B\cap C=\emptyset} m_1^\Omega(B)\times m_2^\Omega(C)} & \forall A\subseteq\Omega,\ A\neq\emptyset \\ 0 & \text{if } A=\emptyset \end{cases} \tag{5.1}$$

Note that Dempster interpreted in (Dempster, 1967) the amount k as conflict.

The basic belief mass of the empty set is null, therefore this rule verifies normality condition and works under the closed world assumption.
We note that $m(\emptyset) \neq 1$, thus combining contradictory certain mass functions must be avoided. The orthogonal sum cannot combine mass functions which cores does not intersect.

(Zadeh, 1984) noticed a counter example where the orthogonal sum provided unsatisfactory results. The counter example enlightened by Zadeh appears in the case of a high disagreement between two sources. Suppose a frame of discernment $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and two mass functions m_1 and m_2 provided by two distinct and independent sources. Mass functions m_1, m_2 and $m_{1\oplus 2}$ are in table 5.1. In that case the

Table 5.1: Zadeh counter example

$\mathbf{2^\Omega}$	$\mathbf{m_1}$	$\mathbf{m_2}$	$\mathbf{m^\Omega_{1\oplus 2}}$
$\emptyset$	0	0	0
ω_1	0.9	0	0
ω_2	0	0.9	0
$\omega_1 \cup \omega_2$	0	0	0
ω_3	0.1	0.1	1
$\omega_1 \cup \omega_3$	0	0	0
$\omega_2 \cup \omega_3$	0	0	0
Ω	0	0	0

least probable hypothesis ω_3 becomes the most probable one after the combination due to the high disagreement between m_1 and m_2. In order to solve the problem enlightened by Zadeh's counter example in (Zadeh, 1984) where the orthogonal sum produced unsatisfactory results, many combination rules appeared.

One solution, proposed by Dubois and Prade in (Dubois and Prade, 1986), consists on eliminating the mass on the empty set by the use of the disjunctive combination. The disjunctive combination of two mass functions m_1 and m_2 is given as follows:

$$m_{1 \bigcirc\!\!\!\!\cup\, 2}(A) = (m_1 \bigcirc\!\!\!\!\cup\, m_2)(A) = \sum_{B \cup C = A} m_1(B) \times m_2(C) \tag{5.2}$$

(Yager, 1987) interpreted $m(\emptyset)$ as the amount of total ignorance, therefore it is affected to Ω in order to have a normalized mass function. Yager's rule of combination is defined for two mass functions m_1 and m_2 as follows:

$$\begin{cases} m_Y(A) = m_{1 \textcircled{\cap} 2}(A) & \forall A \in 2^\Omega,\ A \neq \Omega \text{ and } A \neq \emptyset \\ m_Y(\Omega) = m_{1 \textcircled{\cap} 2}(\Omega) + m_{1 \textcircled{\cap} 2}(\emptyset) & \\ m_Y(\emptyset) = 0 & \end{cases} \tag{5.3}$$

(Dubois and Prade, 1988) proposed a suitable distribution of the mass on the empty set. It is not the mass of the empty set, interpreted as global conflict, that is distributed but the conflict issued from the combination of each conflicting focal elements. That partial conflict, issued from the combination of conflicting focal elements, is attributed to the disjunction of these elements. The combination of two mass functions m_1 and m_2 with Dubois and Prade's rule is given as follows:

$$m_{DP}(A) = \begin{cases} m_{1 \textcircled{\cap} 2}(A) + \sum\limits_{\substack{B \cap C = \emptyset, \\ B \cup C = A}} m_1(C) m_2(B) & A \neq \emptyset,\ \forall B, C \subseteq \Omega,\ B, C \neq \emptyset \\ 0 & A = \emptyset \end{cases} \tag{5.4}$$

(Smets, 1990) proposed to use an open world assumption where a non null mass can be affected to the empty set representing the degree of belief that the true hypothesis is not enumerated in Ω. The conjunctive rule of combination does not redistribute $m(\emptyset)$ but it is kept on the contradictory state $\emptyset$. The conjunctive rule of combination for two mass functions m_1 and m_2 is defined as follows:

$$m_{1 \textcircled{\cap} 2}(A) = (m_1 \textcircled{\cap} m_2)(A) = \sum_{B \cap C = A} m_1(B) \times m_2(C) \tag{5.5}$$

Smets proposed to use the conjunctive rule of combination only when both sources are known to be reliable; the disjunctive rule of combination can be used when one of the sources is unreliable.

(Lefèvre et al., 2002; Lefèvre et al., 2003) proposed a general framework in order to unify several classical rules of combination and suggested other combination rules allowing an arbitrary or adapted distribution of the conflict among subsets. The idea of the general formulation of the combination rules consists on assigning the conflict to subsets proportionally to a weighting function W. The general formula is given as follows:

$$m(A) = m_{\textcircled{\cap}}(A) + W(A) m_{\textcircled{\cap}}(\emptyset) \tag{5.6}$$

The weighting function detailed in equation 5.7 is that leading to the conjunctive rules of combination; that of equation 5.8 leads to Yager's rule.

$$W(A) = \begin{cases} 1 & \text{if } A = \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

$$W(A) = \begin{cases} 1 & \text{if } A = \Omega \\ 0 & \text{otherwise} \end{cases} \tag{5.8}$$

(Lefèvre et al., 2002) proposed a learning of weighting factors by minimizing the mean square error between pignistic probability and the membership indicator of a training set.

(Smarandache and Dezert, 2005) proposed a *Proportional Conflict Redistribution (*PCR5*)* rule distributing the partial conflict on conflicting focal elements. The PCR5 was reformulated by Martin and Osswald who proposed a PCR6 for the combination of two or many mass functions (Martin and Osswald, 2006b; Martin and Osswald, 2006a; Martin and Osswald, 2007b). The PCR6 combines two mass functions m_1 and m_2 as follows:

$$m_{\mathrm{PCR6}}(A) = m_{1 \bigcirc\!\!\!\!\cap\, 2}(A) + \sum_{\substack{B \in 2^\Omega, \\ A \cap B = \emptyset}} \left(\frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right) \tag{5.9}$$

In order to combine M mass functions provided by M independent and distinct sources, the PCR6 (Martin and Osswald, 2006b; Martin and Osswald, 2006a; Martin and Osswald, 2007b) is given as follows:

$$\begin{aligned} m_{\mathrm{PCR6}}(A) &= m_{\bigcirc\!\!\!\!\cap}(A) + \\ &\sum_{j=1}^{M} m_j(A)^2 \sum_{\substack{\bigcap_{j'=1}^{M-1} B_{\sigma_j(j')} \cap A = \emptyset \\ (B_{\sigma_j(1)}, \ldots, B_{\sigma_j(M-1)}) \in (2^\Omega)^{M-1}}} \left(\frac{\prod_{j'=1}^{M-1} m_{\sigma_j(j')}(B_{\sigma_j(j')})}{m_j(A) + \sum_{j'=1}^{M-1} m_{\sigma_j(j')}(B_{\sigma_j(j')})} \right) \end{aligned} \tag{5.10}$$

And σ_j is in $[1, M]$ according to j as follows:

$$\begin{cases} \sigma_j(j') = j' & \text{if } j' < j \\ \sigma_j(j') = j' + 1 & \text{if } j' \geq j \end{cases} \tag{5.11}$$

Dubois and Prade's combination rule of equation (5.4) is a mixture of the conjunctive and disjunctive combinations that distributes the partial conflict among conflicting focal elements. Florea et al. proposed in (Florea et al., 2006; Florea, 2007) a mixed combination rule that is a weighted average of the disjunctive and conjunctive combinations. That combination rule distributes the global conflict $k = m(\emptyset)$ on focal elements independently on their partial conflict. The combination of two mass functions m_1 and m_2 with combination rule proposed by (Florea et al., 2006; Florea, 2007) is given as follows:

$$m_{Flo} = W_1(k)m_{1 \bigcirc\!\!\!\!\cup\, 2} + W_2(k)m_{1 \bigcirc\!\!\!\!\cup\, 2} \tag{5.12}$$

With:

$$\begin{aligned} W_1(k) &= \frac{k}{1-k+k^2} \\ W_2(k) &= \frac{1-k}{1-k+k^2} \end{aligned} \tag{5.13}$$

(Martin and Osswald, 2007a) proposed also a mixed combination rule that weights the disjunctive and conjunctive combinations with the similarity between focal elements. The combination of two mass functions m_1 and m_2 is given as follows:

$$m_{Mar}(A) = \sum_{B\cup C=A} W_1'(B,C)m_1(B)m_2(C) + \sum_{B\cap C=A} W_2'(B,C)m_1(B)m_2(C) \tag{5.14}$$

Note that W_1' and W_2' are dissimilarity and similarity of B and C. For example:

$$W_1'(B,C) = \frac{|B\cap C|}{min(|B|,|C|)} \tag{5.15}$$

and $W_2'(B,C) = 1-W_1'(B,C)$. Note that when $W_2' = 1$ the conjunctive rule is obtained and $W_1' = 1$ corresponds to the disjunctive rule.
(Lefèvre and Elouedi, 2013) proposed a mixed combination rule that weights the conjunctive and Dempster's rules of combination. the proposed rule, named Combination With Adapted conflict ($CWAC$) and noted $\bigcirc\!\!\!\!\leftrightarrow$, is given as follows:

$$m_{\bigcirc\!\!\!\!\leftrightarrow}(A) = \gamma_1 m_{\bigcirc\!\!\!\!\cap}(A) + \gamma_2 m_{\oplus}(A) \tag{5.16}$$

With:

$$\gamma_1 = d(m_1,m_2) \tag{5.17}$$

$$\gamma_2 = 1-d(m_1,m_2) \tag{5.18}$$

Note that $d(m_1,m_2)$ is Jousselme distance given in equation (3.24).

Example 5.1 In table 5.2, we present two mass functions given by two different sources, these mass functions are combined using different combination rules.

5.2.2 Idempotent combination rules

All combination rules detailed in the previous section work under a strong assumption of cognitive independence since they are used to combine mass functions induced by two distinct sources. This strong assumption assumed for most of rules but never verified. (Denœux, 2008), proposed a family of conjunctive and disjunctive rules based on triangular norms and conorms. Cautious and bold rules are members of that family and

Table 5.2: Combining two mass functions with several combination rules

	$\mathbf{m_1}$	$\mathbf{m_2}$	$\mathbf{m_{1\oplus2}}$	$\mathbf{m_{1\bigcirc\!\!\!\!\cap 2}}$	$\mathbf{m_Y}$	$\mathbf{m_{DP}}$	$\mathbf{m_{Mean}}$
$\emptyset$	0	0	0	0.17	0	0	0
P	0	0.3	0.2771	0.23	0.23	0.23	0.15
H	0.3	0.4	0.494	0.41	0.41	0.41	0.35
$P \cup H$	0.2	0.1	0.1084	0.09	0.09	0.18	0.15
M	0	0	0	0	0	0	0
$M \cup P$	0.2	0	0.00482	0.04	0.04	0.04	0.1
$H \cup M$	0	0	0	0	0	0	0
Ω	0.3	0.2	0.0723	0.06	0.23	0.14	0.25

combine mass functions for which independence assumption is not checked. Cautious combination of two mass functions m_1 and m_2 issued from probably dependent sources is defined as follows:

$$m_1 \owedge m_2 = \bigcirc\!\!\!\!\cap_{A\subset\Omega} A^{w_1(A)\wedge w_2(A)} \tag{5.19}$$

Where $A^{w_1(A)}$ and $A^{w_2(A)}$ are simple support functions focused on A with weights w_1 and w_2 (cf. Section 2.2.1) issued from the canonical decomposition (cf. Section 2.4.2) of m_1 and m_2 respectively, note also that $\wedge$ is a *min operator* of simple support functions weights. Cautious rule of equation (5.19) is unnormalized because of the use of the conjunctive combination that has the empty set as absorbing element. By the use of Dempster's rule, the normalized cautious rule is obtained as follows:

$$m_1 \owedge^{\star} m_2 = \oplus_{A\subset\Omega, A\neq\emptyset} A^{w_1(A)\wedge w_2(A)} \tag{5.20}$$

When the min operator $\wedge$, is replaced by a *max operator* $\vee$, the *bold combination rule* is obtained. Both cautious and bold rules combine mass functions issued from dependent sources, but the cautious rule is more fitted to reliable sources, otherwise the bold rule fits unreliable ones. Both bold and cautious combination rules are commutative, associative and idempotent. Cautious and bold rule do not have a neutral element.

(Boubaker et al., 2013) proposed a mixed combination rule that is a weighting of the unnormalized and normalized cautious rule. The Cautious Combination with Adapted Conflict ($CCAC$), noted $\odot$, is given as follows:

$$m_1 \odot m_2(A) = d(m_1, m_2) m_{\owedge}(A) + (1 - d(m_1, m_2)) m_{\owedge^{\star}}(A), \quad \forall A \subseteq \Omega \text{ and } m_{\owedge}(A) \neq 1 \tag{5.21}$$

Note that $d(m_1, m_2)$ is Jousselme distance given by equation (3.24).

Finally, the mean is also a combination rule, thus a combined mass function is the average of a set of N mass functions. Therefore, for each focal element A of N mass functions, the combined one is defined as follows:

$$m_{\text{Mean}}(A) = \frac{1}{N}\sum_{i=1}^{N} m_i(A) \tag{5.22}$$

The choice of the combination rule is based on the dependence of mass functions sources. Combination rules like (Dubois and Prade, 1988; Martin and Osswald, 2007a; Murphy, 2000; Smets and Kennes, 1994; Yager, 1987) combine mass functions which sources are independent, whereas cautious, bold, mean and CCAC rules are the most fitted to combine mass functions issued from dependent sources. In Chapter 4, we suggested a new learning of sources' degrees of independence helping the choice of the type of combination rules to be used. In the next section, we detail our new combination rule integrating sources' degrees of independence in the combination.

5.3 Mixed combination rule

In the combination step, sources dependence or independence hypothesis is intuitively made without any possibility of check. Sources independence degree is not either 0 or 1 but a level over $[0, 1]$. The main question is "which combination rule to use when combining partially independent/dependent mass functions?"

In this section, we propose a new mixed combination rule using conjunctive and cautious rules detailed in equations (5.5) and (5.19). Combined mass function is a weighted average of conjunctive and cautious combinations. When sources are dependent, combined mass function is similar to the cautious combination; when they are independent, combined mass function is similar to the conjunctive combination. In a case of totally dependent sources (where independence is 0), the cautious and proposed mixed combination rules are similar; whereas in a case of totally independent sources (independence is 1), the conjunctive and proposed combination rules are similar. In a case of an independence degree in $]0, 1[$, combined mass function is the average of conjunctive and cautious combinations weighted by independence degree of sources.

Assume that two sources s_1 and s_2 are independent with a degree γ such that $\gamma = \text{Ind}(s_1, s_2)$; m_1 and m_2 are mass functions provided by s_1 and s_2. The proposed mixed combination rule is defined as follows:

$$m_{\text{Mixed}}(A) = \gamma * m_{\text{\textcircled{$\cap$}}}(A) + (1-\gamma) * m_{\text{\textcircled{$\wedge$}}}(A), \quad \forall A \subseteq \Omega \tag{5.23}$$

Note that the combined mass of a focal element A, $m_{\text{Mixed}}(A)$, is the mean of its combined mass using the conjunctive and cautious rules weighted by independence degree of sources.

Degree of independence of a set of sources is given by equation (4.21), and the mixed combination of a set of mass functions $\{m_1, m_2, \ldots, m_M\}$ provided by sources $\{s_1, s_2, \ldots, s_M\}$ is also a weighted average such that:

$$\gamma = \text{Ind}(s_1, s_2, \ldots, s_M) \tag{5.24}$$

Properties of the proposed mixed combination rule

- Commutativity: Conjunctive and cautious rules are commutative. Independence measure is symmetric because sources degree of independence is the same for a set of sources. Then the proposed rule is commutative.
- Associativity: Conjunctive and cautious rule are associative but the proposed rule is not because independence degree of M sources and $M+1$ ones is not necessarily the same.
- Idempotent: Degree of independence of one source to itself is 0, in that case the proposed rule is equivalent to the cautious rule. As the cautious rule is idempotent, it is the case of the proposed mixed rule.
  Therefore, $m_{\text{Mixed}}(A) = m(A)$, $\forall A \subseteq \Omega$.
- Neutral element: Mixed combination rule does not have any neutral element.
- Absorbing element: No absorbing element also.

Example Assume a frame of discernment $\Omega = \{a, b, c\}$ and two sources s_1 and s_2 providing two mass functions m_1 and m_2. Table 5.3 illustrates conjunctive and cautious combinations as well as mixed combination in the cases where $\gamma = 0$, $\gamma = 0.3$, $\gamma = 0.6$ and $\gamma = 1$. When $\gamma = 0$, mixed and cautious combinations are equivalent; when $\gamma = 1$, mixed and conjunctive combinations are equivalent, otherwise it is a weighted average by $\gamma \in]0, 1[$.

5.4 Integrating independence, positive and negative dependence in mass functions

In the aim to take into account sources' degrees of independence in mass functions they provide, we proposed in (Chebbah et al., 2014) a method for combining mass functions provided by sources and mass functions on sources' degrees of independence.

Two different mass functions are combined in order to take account of sources' degrees of independence. The first one is that of equation (4.31) which defines the independence, positive and negative dependence of one source according to another one. The second one is any mass function m^Ω provided by that source. Combining those

Table 5.3: Combination of two mass functions

2^Ω	m_1	m_2	$m_{\textcircled{$\wedge$}}$	$m_{\textcircled{$\cap$}}$	m_{Mixed} $\gamma = 0$	m_{Mixed} $\gamma = 0.3$	m_{Mixed} $\gamma = 0.6$	m_{Mixed} $\gamma = 1$
$\emptyset$	0	0	0.1071	0.06	0.1071	0.093	0.0789	0.06
a	0.3	0.3	0.2679	0.45	0.2679	0.3225	0.3771	0.45
b	0	0	0	0	0	0	0	0
$a \cup b$	0	0	0	0	0	0	0	0
c	0.2	0	0.1786	0.14	0.1786	0.167	0.1554	0.14
$a \cup c$	0.2	0.4	0.2551	0.26	0.2551	0.2566	0.2580	0.26
$b \cup c$	0	0	0	0	0	0	0	0
$a \cup b \cup c$	0.3	0.3	0.1913	0.09	0.1913	0.1609	0.1305	0.09

mass functions is similar to the combination of two different mass functions, a mass function provided by a source with a mass function on that source's reliability. The justification of the discounting in Section 2.5.3 inspired us to justify the integration of sources degrees of independence in mass functions they provide.

When combining with the conjunctive rule, an explicit hypothesis on sources' total independence is made. When sources are not so (independent), one can claim that mass functions do not have to be combined with that rule and a more fitted rule has to be used. However, sources are not either dependent or independent but they are independent, positive and negative dependent with some degree in $[0, 1]$.

In the case of independent sources, their mass functions can be combined with the conjunctive rule; whereas, when sources are positively dependent, the mass function provided by the dependent one must have no influence in the combination; it has to be assimilated to a vacuous mass function. In the case of negative dependent sources, the mass function of the negatively dependent one has to be reduced to a contradictory mass function.

Consequently, mass functions of independent sources are taken as such; mass functions of positively dependent source are considered vacuous (neutral element of the conjunctive combination) and mass functions of negatively dependent sources are converted to a contradictory mass function (high conflict due to the negative dependence).

To do this discounting schema, mass function of equation (4.31) that is defined on $\mathcal{P}$ is combined with a mass function m^Ω. We recall that in Section 4.5, two mass functions defined on the frame of discernment $\mathcal{P}$ are obtained. Each mass function informs about the independence, positive and negative dependence of one source on the other. We recall that mass functions of (4.31) are given as follows:

$$\begin{cases} m_i^{\mathcal{P}}(I) = \frac{1}{n}\sum_{k_i=1}^{n}\alpha_{k_i,k_j}^i\,(1-\beta_{k_i,k_j}^i) \\ m_i^{\mathcal{P}}(P) = \frac{1}{n}\sum_{k_i=1}^{n}\alpha_{k_i,k_j}^i\,\beta_{k_i,k_j}^i(1-\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j)) \\ m_i^{\mathcal{P}}(\bar{P}) = \frac{1}{n}\sum_{k_i=1}^{n}\alpha_{k_i,k_j}^i\,\beta_{k_i,k_j}^i\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j) \\ m_i^{\mathcal{P}}(I\cup P) = \frac{1}{n}\sum_{k_i=1}^{n}(1-\alpha_{k_i,k_j}^i)\,(1-\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j)) \\ m_i^{\mathcal{P}}(I\cup \bar{P}) = \frac{1}{n}\sum_{k_i=1}^{n}(1-\alpha_{k_i,k_j}^i)\,\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j) \end{cases} \tag{5.25}$$

When number of mass functions in clusters provide a good estimation of clusters independence, $\alpha_{k_i,k_j}^i = 1$, $\forall\{i,j\} \in \{1,2\}$ and $\forall\{k_i,k_j\} \in [1,n]$. In that case, equation (5.25) is given as follows:

$$\begin{cases} m_i^{\mathcal{P}}(I) = \frac{1}{n}\sum_{k_i=1}^{n}1-\beta_{k_i,k_j}^i \\ m_i^{\mathcal{P}}(P) = \frac{1}{n}\sum_{k_i=1}^{n}\beta_{k_i,k_j}^i(1-\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j)) \\ m_i^{\mathcal{P}}(\bar{P}) = \frac{1}{n}\sum_{k_i=1}^{n}\beta_{k_i,k_j}^i\mathrm{Conf}(Cl_{k_i}^i, Cl_{k_j}^j) \end{cases} \tag{5.26}$$

Note that α_{k_i,k_j}^i is given by equation (4.14).

Mass function on sources dependence of equation (5.25) is defined on the frame of discernment $\mathcal{P}$ but the mass function provided by a source is defined on any frame of discernment noted Ω. To combine both mass functions, they must be defined on a common frame of discernment. To do that, we will use the vacuous extension, deconditioning and marginalization detailed respectively in Sections (2.3.2), (2.5.2) and (2.3.2). We assume a mass function provided by a source m^Ω according to the source's independence or dependence on another source as follows:

$$\begin{cases} m^\Omega[I](A) = m^\Omega(A) \\ m^\Omega[\bar{P}](A) = m^\Omega(A) \quad m^\Omega(A) = 1 \text{ if } A = \emptyset, \text{ 0 else} \\ m^\Omega[P](A) = m^\Omega(A) \quad m^\Omega(A) = 1 \text{ if } A = \Omega, \text{ 0 else} \end{cases} \tag{5.27}$$

Equation (5.27) illustrates cases of an independent, positive and negative dependent

source. When the source is independent, its mass function is unchanged; when it is negatively dependent, its mass function is assimilated to a vacuous mass function and finally when the source is positively dependent, its mass function is assimilated to a contradictory mass function.
That mass functions $m^{\Omega}[I]$, $m^{\Omega}[\bar{P}]$ and $m^{\Omega}[P]$ are deconditioned as follows:

$$\begin{cases} m^{\Omega}[I]^{\Uparrow\Omega\times\mathcal{P}}((A\times I)\cup(\Omega\times\bar{I})) = m^{\Omega}[I](A), \quad A\subseteq\Omega \\ m^{\Omega}[\bar{P}]^{\Uparrow\Omega\times\mathcal{P}}((A\times\bar{P})\cup(\Omega\times\{I\cup P\})) = m^{\Omega}[\bar{P}](A), \quad A\subseteq\Omega \\ m^{\Omega}[P]^{\Uparrow\Omega\times\mathcal{P}}((A\times P)\cup(\Omega\times\{I\cup\bar{P}\})) = m^{\Omega}[P](A), \quad A\subseteq\Omega \end{cases} \tag{5.28}$$

The last two lines of equation (5.28) correspond to the deconditioning of contradictory and vacuous mass functions.

The mass function $m^{\mathcal{P}}$ is extended to the frame $\Omega\times\mathcal{P}$ with the vacuous extension as follows:

$$m^{\mathcal{P}\uparrow\Omega\times\mathcal{P}}(B) = \begin{cases} m^{\mathcal{P}}(A) & \text{if } B=\Omega\times A, \quad A\subseteq\mathcal{P} \\ 0 & \text{else} \end{cases} \tag{5.29}$$

Then, both mass functions $m^{\mathcal{P}\uparrow\Omega\times\mathcal{P}}$ and $m^{\Omega\Uparrow\Omega\times\mathcal{P}}$ are defined on a common frame $\Omega\times\mathcal{P}$ and can be combined with the conjunctive rule of combination:

$$m^{\Omega\times\mathcal{P}}_{\mathrm{Conj}}(B) = m^{\mathcal{P}\uparrow\Omega\times\mathcal{P}} \circledcirc m^{\Omega}[I]^{\Uparrow\Omega\times\mathcal{P}} \circledcirc m^{\Omega}\left[\bar{P}\right]^{\Uparrow\Omega\times\mathcal{P}}(B), \quad \forall B\subset\Omega\times\mathcal{P} \tag{5.30}$$

Note that $m^{\Omega}[P]^{\Uparrow\Omega\times\mathcal{P}}$ is a vacuous mass function.
Finally, $m^{\Omega\times\mathcal{I}}_{\mathrm{Conj}}$ is marginalized on Ω:

$$m^{\Omega\times\mathcal{P}\downarrow\Omega}(A) = \sum_{\{B\subseteq\Omega\times\mathcal{P}\,|\,Proj(B\downarrow\Omega)=A\}} m^{\Omega\times\mathcal{P}}_{\mathrm{Conj}}(B) \tag{5.31}$$

That process is done for both mass functions of s_1 and s_2 according to their independence degrees. In other words, $m_1^{\mathcal{P}}$ is a mass function on the independence, positive and negative dependence of s_1 according to s_2, $m_2^{\mathcal{P}}$ is that of s_2 according to s_1. The mass function $m_1^{\mathcal{P}}$ is combined with m_1^{Ω} provided by s_1 and $m_2^{\mathcal{P}}$ is combined with m_2^{Ω} provided by s_2.

5.5 Experiments

Finally, to illustrate the proposed mixed combination rule and compare it to other combination rules, three mass functions are generated randomly using algorithm 1. These mass functions are combined with conjunctive, Dempster, Yager, disjunctive, cautious and mean combination rules. They are also combined with the mixed combination rule with different independence levels.

Figure 5.1 illustrates distances[1] between the mixed combination with several degrees of independence and combined mass functions using several rules: conjunctive, Dempster, Yager, disjunctive, cautious and mean combination rules. Distances between mixed combination with several independence degrees; and Yager, disjunctive, mean and Dempster's rules are linear and decreasing proportionally to γ. When sources' degree of independence tends to 1, the mixed combination tends to the conjunctive combination. On the other hand, when sources are dependent; mixed combination is similar to cautious combination. When $\gamma = 0.5$, mixed combination is equally distant to the cautious and conjunctive combinations.

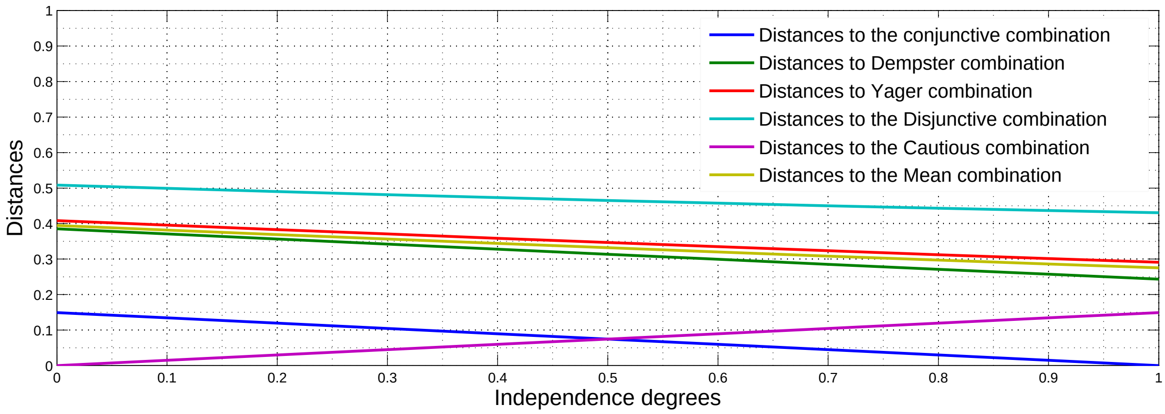

Figure 5.1: Distances between combined mass functions

We proposed that combination rule in order to take consideration of sources' degrees of independence in the combination step. The proposed combination rule is a mixture of the conjunctive and cautious combinations weighted with sources' degrees of independence. Another solution to take consideration of sources degrees of independence is to integrate that degrees into mass functions provided by that sources.

In the previous section, we detailed the process of discounting mass functions with their sources' degree of independence. Assume a frame of discernment containing three hypotheses: $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and suppose that two sources s_1 and s_2 provided mass functions detailed in table 5.4. The third row of table 5.4 is the conjunctive combination of m_1^Ω and m_2^Ω under a strong assumption of cognitive independence of s_1 and s_2. Suppose that the strong assumption of cognitive independence of sources cannot be made because sources' independence degrees are learned with methods detailed in

[1]Jousselme distance detailed in equation (3.24).

Table 5.4: Mass functions provided by s_1 and s_2

2^Ω	m_1^Ω	m_2^Ω	$m_{1 \circledcirc 2}^\Omega$
$\emptyset$	0	0	0.02
ω_1	0.2	0	0.18
ω_2	0	0.1	0.08
$\omega_1 \cup \omega_2$	0.5	0.6	0.63
ω_3	0	0	0
$\omega_1 \cup \omega 3$	0	0	0
$\omega_2 \cup \omega_3$	0	0	0
$\omega_1 \cup \omega_2 \cup \omega_3$	0.3	0.3	0.09

Chapter 4. If s_1 is positively dependent on s_2 and the mass function on the dependence of s_1 on s_2 is given as follows:

$$\begin{cases} m^{\mathcal{P}}(I) = 0.26 \\ m^{\mathcal{P}}(P) = 0.56 \\ m^{\mathcal{P}}(\bar{P}) = 0.18 \end{cases} \tag{5.32}$$

Two possible solutions to take consideration of this mass function of the dependence of s_1 on s_2:

- Use of the mixed combination rule: The first solution is to use the mixed combination rule to combine m_1^Ω and m_2^Ω. Note that $I_d(s_1, s_2) = 0.26$ and if we had $I_d(s_2, s_1)$, $\mathrm{Ind}(s_1, s_2)$ would be the minimum of $I_d(s_1, s_2)$ and $I_d(s_2, s_1)$. For this example $\mathrm{Ind} = 0.26$ and table 5.5 illustrates the combination of m_1^Ω and m_2^Ω such that $\gamma = \mathrm{Ind} = 0.26$.

- Integrating $m^{\mathcal{P}}$ in m^Ω: Mass functions $m^{\mathcal{P}}$ and m^Ω are combined leading to the discounting process detailed in Section 5.4. Table 5.6 details steps of integrating degrees of independence detailed in equation (5.32) into mass function provided by s_2. In the first step (first column), $m^{\mathcal{P}}$ is extended to $m^{\Omega \times \mathcal{P}}$ with the vacuous extension; Then (in the second and third columns) $m^\Omega[I]$ and $m^\Omega[\bar{P}]$ are deconditioned on $\Omega \times \mathcal{P}$ to have $m^\Omega[I]^{\Uparrow \Omega \times \mathcal{P}}$ and $m^\Omega[\bar{P}]^{\Uparrow \Omega \times \mathcal{P}}$. Finally, mass functions are combined with the conjunctive rule. Once mass functions $m^{\mathcal{P}}$ and m_1^Ω are combined, discounted mass function $m_1^{\Omega \times \mathcal{P} \downarrow \Omega}$ can be combined with m_2^Ω as detailed in table 5.7.

Table 5.8 compares the combined mass function using the proposed mixed rule and the mass function obtained by the discounting process proposed to integrate $m^{\mathcal{P}}$ in m^Ω. The proposed rule generates less ignorance and Dempster's conflict than the discounting

Table 5.5: Mixed combination of m_1^Ω and m_2^Ω

2^Ω	m_1^Ω	m_2^Ω	m_{Mixed}^Ω $\gamma = 0.26$
$\emptyset$	0	0	0.02
ω_1	0.2	0	0.18
ω_2	0	0.1	0.08
$\omega_1 \cup \omega_2$	0.5	0.6	0.519
ω_3	0	0	0
$\omega_1 \cup \omega 3$	0	0	0
$\omega_2 \cup \omega_3$	0	0	0
$\omega_1 \cup \omega_2 \cup \omega_3$	0.3	0.3	0.201

process. The proposed discounting process induces higher masses on the empty set and on the ignorance because it takes consideration about degrees of independence, positive and negative dependence and not only the independence degree (the proposed rule uses only the independence degree).

Table 5.8: Comparison of the proposed mixed rule and the approach of integrating independence degree in the mass function

2^Ω	$m_1^{\Omega\times\mathcal{P}\downarrow\Omega} \circledcirc m_2^\Omega$	m_{mixed}^Ω
$\emptyset$	0.25432	0.02
ω_1	0.0468	0.18
ω_2	0.00768	0.08
$\omega_1 \cup \omega_2$	0.15528	0.519
Ω	0.53592	0.201

In table 5.9, we illustrate the combination of mass functions m_1^Ω and m_2^Ω detailed in table 5.4 with different degrees of independence, positive and negative dependence given in tables 5.10 and 5.11. We notice that in the case of independent sources, the combination of discounted mass functions is equivalent to the conjunctive combination. When one source is negatively dependent on the other, a higher mass is attributed to the empty set and finally, when one source is positively dependent on the other, some degree of belief is transferred to the frame of discernment. Thus, the mass on the empty set alarms about the contradiction between sources.

Table 5.6: Discounting m^{Ω} with the degree of independence of s_1

2^{Ω}	$m^{\mathcal{P}\uparrow\Omega\times\mathcal{P}}$	$m^{\Omega}[I]^{\Uparrow\Omega\times\mathcal{P}}$	$m^{\Omega}[\bar{P}]^{\Uparrow\Omega\times\mathcal{P}}$	$m_{\textcircled{\cap}}^{\Omega\times\mathcal{P}}$
$\emptyset$				0.18
$\omega_1 \times I$				0.052
$(\omega_1 \cup \omega_2) \times I$				0.13
$\Omega \times I$	0.26			0.078
$\Omega \times P$	0.56			0.56
$(\omega_1 \times I) \cup (\Omega \times P)$				
$((\omega_1 \cup \omega_2) \times I) \cup (\Omega \times P)$				
$\Omega \times \bar{P}$	0.18			
$\Omega \times (I \cup P)$			1	
$(\omega_1 \times I) \cup (\Omega \times (P \cup \bar{P}))$		0.2		
$((\omega_1 \cup \omega_2) \times I) \cup (\Omega \times (P \cup \bar{P}))$		0.5		
$\Omega \times \mathcal{P}$		0.3		

Table 5.7: Combining discounted $m_1^{\Omega\times\mathcal{P}\downarrow\Omega}$ and m_2^{Ω}

2^{Ω}	$m_1^{\Omega\times\mathcal{P}\downarrow\Omega}$	m_2^{Ω}	$m_1^{\Omega\times\mathcal{P}\downarrow\Omega} \textcircled{\cap} m_2^{\Omega}$
$\emptyset$	0.18		0.25432
ω_1	0.052		0.0468
ω_2		0.1	0.00768
$\omega_1 \cup \omega_2$	0.13	0.6	0.15528
Ω	0.638	0.3	0.53592

Table 5.9: Combination results under several hypotheses on sources s_1 et s_2 independence

s_1 (table 5.10)	Focal elements	$m_1^{\Omega\times\mathcal{P}\downarrow\Omega}$	s_2 (table 5.11) Case 1 $m_2^{\Omega\times\mathcal{P}\downarrow\Omega}$	Case 1 $m_{1\circledcap 2}^{\Omega\times\mathcal{P}\downarrow\Omega}$	Case 2 $m_2^{\Omega\times\mathcal{P}\downarrow\Omega}$	Case 2 $m_{1\circledcap 2}^{\Omega\times\mathcal{P}\downarrow\Omega}$	Case 3 $m_2^{\Omega\times\mathcal{P}\downarrow\Omega}$	Case 3 $m_{1\circledcap 2}^{\Omega\times\mathcal{P}\downarrow\Omega}$
Case 1	$\emptyset$	0.09	0.1	0.19558	0.85	0.86512	0.1	0.18262
	ω_1	0.162	0	0.13122	0	0.02268	0	0.14418
	ω_2	0	0.09	0.06732	0.01	0.00748	0.01	0.00748
	$\omega_1 \cup \omega_2$	0.405	0.54	0.51327	0.06	0.07728	0.06	0.38103
	Ω	0.343	0.27	0.09261	0.08	0.2744	0.83	0.28468
Case 2	$\emptyset$	0.11	0.1	0.20062	0.85	0.86668	0.1	0.19918
	ω_1	0.018	0	0.01458	0	0.00252	0	0.01602
	ω_2	0	0.09	0.07848	0.01	0.00872	0.01	0.00872
	$\omega_1 \cup \omega_2$	0.045	0.54	0.48303	0.06	0.05592	0.06	0.08967
	Ω	0.27	0.27	0.22329	0.08	0.06616	0.83	0.68641
Case 3	$\emptyset$	0.7	0.1	0.7336	0.85	0.9554	0.1	0.7304
	ω_1	0.04	0	0.0324	0	0.0056	0	0.0356
	ω_2	0	0.09	0.0234	0.01	0.0026	0.01	0.0026
	$\omega_1 \cup \omega_2$	0.1	0.54	0.1674	0.06	0.0236	0.06	0.0986
	Ω	0.16	0.27	0.0432	0.08	0.0128	0.83	0.1328

Table 5.10: Cases of independence, positive and negative dependence degrees of s_1

$2^{\mathcal{P}}$	Case 1	Case 2	Case 3
I	0.81	0.09	0.2
P	0.1	0.8	0.1
$\bar{P}$	0.09	0.11	0.7

Table 5.11: Cases of independence, positive and negative dependence degrees of s_2

$2^{\mathcal{P}}$	Case 1	Case 2	Case 3
I	0.9	0.1	0.1
P	0	0.05	0.8
$\bar{P}$	0.1	0.85	0.1

Finally, proposed discounting schema changes decision of mass functions. The discounting operator detailed in Section 2.5.3 integrates sources reliability degrees into mass functions they provide without changing decisions. In other words, after discounting a mass function, the most probable hypothesis does not change. Fortunately, discounting a mass function with sources' degree of independence change decision. The discounting schema enhances faiths on the empty set and the frame of discernment according to the source's positive and negative degrees of dependence.

5.6 Conclusion

In this chapter we detailed some uses of independence, positive and negative degrees. Thus, previous chapters were focused on learning sources' degrees of independence and this chapter details some uses of these degrees. A simple use of information on sources degrees of independence is to guide the choice of the appropriate type of combination rules to use. Hence, when sources are completely independent, combination rules detailed in Section 5.2.1 can be applied; but if sources are completely dependent, combination rules of Section 5.2.2 fit more. When sources degrees of independence are degrees over $[0, 1]$, we proposed two uses of these degrees: The first use is in a combination rule that weights to the conjunctive and cautious combinations with sources degrees of independence. The second solution consists on integrating independence, positive and negative degrees on mass function. The third solution leads to a discounting operator that redistributes beliefs on the empty set and the frame of discernment in cases of negative and positive dependence. After the discounting process the independence assumption can be assumed.

6
Conclusion

Managing uncertainty is a vast domain of research, many theories are used in this context such as the theory of belief functions. This theory is a strong tool used for representing and managing uncertainty and also to combine several uncertain information in order to limit uncertainty or to reduce it. Evidential databases are used to store both certain and evidential information. The main problem appearing when handling several evidential databases is the conflict which may appear reflecting the disagreement between their sources.

This conflict can be managed in the combination by using an appropriate combination rule and can also be eliminated or reduced before combination by taking into account sources' reliability degrees. Discounting evidential information before combining reduces the conflict which may appear after combining but this operator needs an *a prior* knowledge on source's reliability degree.

Many researches are focused on estimating or learning source's reliability degree. In this context, we proposed a method estimating a source's reliability degree from their conflict. Assuming all evidential information is provided by several sources and this information is stored in an evidential databases, sources' degrees of reliability are estimated from a conflict measure computed from all that evidential information. This method estimates the conflict of a source with all other available sources and uses that conflict to compute the reliability of that source.
Source's reliability estimation using its evidential database is done in the purpose of discounting all evidential information stored in this evidential database before integrating them with evidential databases of other sources.

The conflict defined for evidential databases was also used in a clustering algorithm. The clustering technique was proposed in the purpose of grouping together similar objects of several sources in order to compare sources' overall dependence and estimate their independence. The conflict measure defined in evidential databases is used in the clustering algorithm to minimize the conflict into resulting clusters and maximize that conflict between clusters. In a case of several sources providing evidential information according to same objects; objects are classified separately in order to group situations where a source have the same behavior. After clustering, sources' clusters are compared to find a clusters' matching. By matching clusters, most similar clusters are linked in order to compare cases where sources have roughly same behavior.

Once clusters are matched, the link between these clusters is quantified in order to have a mass function on clusters independence or dependence. A first mass function is obtained for each couple of linked clusters. When clusters are dependent, that dependence can be positive or negative. Clusters' positive and negative dependencies are estimated from the conflict between them. Thus, matched clusters contains similar objects; if that objects are not conflicting then we can claim that sources choose similar focal elements for same objects. Thus, they are not conflicting and they are positively dependent. In the case where objects are conflicting, sources are choosing conflicting mass functions for same objects. Thus clusters are negatively dependent.

Two mass functions are obtained for each couple of matched clusters: A first mass function for clusters' independence and dependence. The second mass function is for matched clusters' positive and negative dependence; this mass function is a conditional to the one on clusters dependence. To combine both mass functions, they are transformed to a common frame in order to have only one mass function for matched clusters' independence, positive and negative dependence. All mass functions on matched clusters' independence, positive and negative independence are combined in order to have only one mass function on sources' independence, positive and negative dependence.

The proposed method of learning sources' cognitive independence guides the choice of the appropriate combination rule either when sources are cognitively dependent or independent. Sources are cognitively independent if they are different; not communicating and they have distinct evidential corpora. The proposed statistical approach is based on a clustering algorithm applied to mass functions provided by several sources. Sources' independence is deduced from weights of linked clusters after a matching of their clusters.

The mass function on sources' independence, positive and negative dependence can be integrated into mass functions provided by that sources. In that case, two different mass functions defined on different and not compatible frames of discernment have to be combined. Mass functions are transfered to a common frame of discernment and then combined. This combination, integrates sources' degrees of independence into mass functions they provide. As mass functions are discounted proportionally to sources' degrees of independence, they can be combined by assuming the hypothesis on their sources independence.

Sources' independence degree can either be integrated in mass functions provided by that sources or guide the choice of the combination rule if it is either 1 or 0. When sources independence is 1, they are assumed independent and combination rules using the conjunctive and/or disjunctive combinations can be applied. When sources independence is 0, sources are dependent and an idempotent rule is needed for the combination.

When sources' degrees of independence is not either 0 or 1, the choice of the combination rule is quite difficult. In extreme cases of independence, the choice of combination rules is easy but when independence degree is over $]0,1[$, the choice is not enough justified. Therefore, when sources' independence degree is over $]0,1[$, we propose a new combination rule that weights the conjunctive and cautious combinations with sources independence degree. The proposed combination rule takes into account independence degree of sources.

In future works, we will tend to investigate and improve our researches in some fields such as the combination, social networks, evidential databases and sources' relevance and truthfulness as follows:

- In the combination field, we will attack the dynamic part of the combination for belief revision and updating (Smets, 2007). In the dynamic combination, mass functions according to the same object are combined. That mass functions are induced by distinct evidences by not necessarily provided by two distinct bodies of evidences. Mass functions correspond to an evolving object. It is interesting to study the combination in that context to propose a decision rule to choice the more fitted combination rule.
- In the social networks field:
  - We will try to collect real data in medical area or from social networks to test the proposed learning method. In fact, users of social networks are subjective sources who are communicating and providing dependent and independent information. Testing the method on real data where sources are known and their degrees of independence are known may validate the proposed method. Testing the proposed clustering on real data is quite interesting especially for comparison with other clustering algorithms.
  - Some researches are about trust inference in social networks (Levien, 2002; Kuter and Golbeck, 2007; Kuter and Golbeck, 2010) can motivate dependence inference in such networks. The proposed method may be adapted to social networks to detect dependencies between users. In fact, when dependencies between users of social networks are known reliability of propagated information may also be learned. It can also be used in marketing (Kempe et al., 2003) for influence propagation to promote new products and define

new marketing strategies. Indeed, a company wishing to launch a marketing campaign or a new product can use relations of dependencies to speed up the propagation. It is interesting to investigate the combination of propagated messages in social networks and study the combination in that case.

- For evidential databases, we proposed a conflict measure to solve the conflict that may appear when integrating these databases. Till now, some researches are focused on possibilistic (Bosc et al., 2003), probabilistic (Cheng et al., 2003) and fuzzy queries (Bosc and Pivert, 1995); however, there are not enough researches on evidential databases and evidential queries; there is no model for such databases even it seems promising (Anand et al., 1996). Thus, it is interesting to investigate and propose a new model for evidential databases and also a query language that supports uncertainty in such databases. In uncertain query languages, the combination and conflict solving have also to be tackled.

- Finally, recent researches (Pichon et al., 2012) are focused on relevance and truthfulness of information sources. It may be interesting to propose a learning of truthfulness and relevance of sources because such information is useful when coping with evidential information provided by that sources.

Bibliography

Abellán, J. and Moral, S. (2000). A non-specificity measure for convex sets of probability distributions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(3):357–368.

Anand, S. S., Bell, D. A., and Hughes, J. G. (1996). Aspects of uncertainty handling for knowledge discovery in databases. Technical report, Department of Information Systems, University of Ulster (Jordanstown).

Antoine, V., Quost, B., Masson, M.-H., and Denœux, T. (2010). CECM: Adding pairwise constraints to evidential clustering. In *Proceedings of IEEE World Congress on Computational Intelligence*, pages 1–8, Barcelona, Spain.

Antoine, V., Quost, B., Masson, M.-H., and Denœux, T. (2011). CEVCLUS: Constrained evidential clustering of proximity data. In *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology*, pages 876–882, Aix-Les-Bains, France. Atlantis Press.

Antoine, V., Quost, B., Masson, M.-H., and Denœux, T. (2012). CECM: constrained evidential C-means algorithm. *Computational Statistics & Data Analysis*, 56(4):894–914.

Aregui, A. and Denœux, T. (2007). Consonant belief function induced by a confidence set of pignistic probabilities. In Mellouli, K., editor, *Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 4724 of *Lecture Notes in Computer Science*, pages 344–355, Hammamet, Tunisia. Springer Berlin Heidelberg.

Bach Tobji, M.-A. (2012). *Frequent itemset mining and maintenance in imperfect databases*. PhD thesis, Higher Institute of Management, University of Tunis, Tunisia.

Bach Tobji, M.-A., Ben Yaghlane, B., and Mellouli, K. (2008). A new algorithm for mining frequent itemsets from evidential databases. In Magdalena, L., Ojeda-Aciego, M., and Verdegay, J., editors, *Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 1535–1542, Málaga, Spain.

Ben Hariz, S. and Elouedi, Z. (2010a). DK-BKM: Decremental K belief K-modes method. In Deshpande, A. and Hunter, A., editors, *Proceedings of the 4th International Conference on Scalable Uncertainty Management*, volume 6379 of *Lecture Notes in Computer Science*, pages 84–97, Toulouse, France. Springer Berlin Heidelberg.

Ben Hariz, S. and Elouedi, Z. (2010b). IK-BKM: An incremental clustering approach based on intra-cluster distance. In *Proceedings of ACS/IEEE International Conference on Computer Systems and Applications*, pages 1–8, Hammamet, Tunisia.

Ben Hariz, S., Elouedi, Z., and Mellouli, K. (2006). Clustering approach using belief function theory. In Euzenat, J. and Domingue, J., editors, *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology*, volume 4183 of *Lecture Notes in Computer Science*, pages 162–171, Varna, Bulgaria. Atlantis Press.

Ben Hariz, S., Elouedi, Z., and Mellouli, K. (2007). Selection initial modes for belief K-modes method. *International Journal of Applied Science, Engineering and Technology*, 4(4):233–242. ISSN 1307-4318.

Ben Yaghlane, A., Denœux, T., and Mellouli, K. (2006a). Constructing belief functions from qualitative expert opinions. *Information and Communication Technologies*, 1:1363–1368.

Ben Yaghlane, A., Denœux, T., and Mellouli, K. (2006b). Elicitation of expert opinions for constructing belief functions. In *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, volume 1, pages 403–411, Paris, France.

Ben Yaghlane, B. (2002). *Uncertainty representation and reasoning in directed evidential networks*. PhD thesis, Higher Institute of Management, University of Tunis, Tunisia.

Ben Yaghlane, B., Smets, P., and Mellouli, K. (2000). Independence concepts for belief functions. In *Proceedings of the 8th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, volume 1, pages 357–364, Madrid, Spain.

Ben Yaghlane, B., Smets, P., and Mellouli, K. (2002a). Belief function independence: I. The marginal case. *International Journal of Approximate Reasoning*, 29(1):47–70.

Ben Yaghlane, B., Smets, P., and Mellouli, K. (2002b). Belief function independence: II. The conditional case. *International Journal of Approximate Reasoning*, 31(1-2):31–75.

Blackman, S. and Popoli, R. (1999). *Design and Analysis of Modern Tracking Systems*. Artech House.

Bosc, P., Duval, L., and Pivert, O. (2003). An initial approach to the evaluation of possibilistic queries addressed to possibilistic databases. *Fuzzy Sets and Systems*, 140(1):151–166.

Bosc, P. and Pivert, O. (1995). SQLf: a relational database language for fuzzy querying. *IEEE Transactions on Fuzzy Systems*, 3(1):1–17.

Boubaker, J., Elouedi, Z., and Lefèvre, E. (2013). Conflict management with dependent information sources in the belief function framework. In *Proceedings of IEEE 14th International Symposium on Computational Intelligence and Informatics*, pages 393–398, Budapest, Hungary.

Bouchard, M., Jousselme, A.-L., and Doré, P.-E. (2013). A proof for the positive definiteness of the Jaccard index matrix. *International Journal of Approximate Reasoning*, 54(5):615–626.

Bourgeois, F. and Lassalle, J.-C. (1971). An Extension of the Munkers Algorithm for the Assignement Problem to Rectangular Matrices. *Communication of the ACM*, 12(14):802–804.

Bryson, N. and Mobolurin, A. (1999). A process for generating quantitative belief functions. *European Journal of Operational Research*, 115(3):624–633.

Burger, T. and Destercke, S. (2012). Random generation of mass functions: A short howto. In Denœux, T. and Masson, M.-H., editors, *Belief Functions: Theory and Applications, Proceedings of the 2nd International Conference on Belief Functions*, volume 164 of *Advances in Intelligent and Soft Computing*, pages 145–152, Compiègne, France. Springer Berlin Heidelberg.

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

Chebbah, M., Kharoune, M., Martin, A., and Ben Yaghlane, B. (2014). Considérant la dépendance dans la théorie des fonctions de croyance. *Revue des Nouvelles Technologies de l'Information*, To appear.

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

Chebbah, M., Martin, A., and Ben Yaghlane, B. (2011). Estimation de la fiabilité des sources des bases de données évidentielles. *Revue des Nouvelles Technologies de l'Information, Numéro spécial : Fouille de données complexes - Complexité liée aux données multiples*, RNTI-E-21:193–209.

Chebbah, M., Martin, A., and Ben Yaghlane, B. (2012a). About sources dependence in the theory of belief functions. In Denœux, T. and Masson, M.-H., editors, *Belief Functions: Theory and Applications, Proceedings of the 2nd International Conference on Belief Functions*, volume 164 of *Advances in Intelligent and Soft Computing*, pages 239–246, Compiègne, France. Springer Berlin Heidelberg.

Chebbah, M., Martin, A., and Ben Yaghlane, B. (2012b). Positive and negative dependence for evidential database enrichment. In Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., and Yager, R. R., editors, *Advances in Computational Intelligence, Proceedings of the14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, volume 299, pages 575–584, Catania, Italy.

Chebbah, M., Martin, A., and Ben Yaghlane, B. (2013). Mesure de dépendance positive et négative de sources crédibilistes. In Vrain, C., Péninou, A., and Sèdes, F., editors, *Extraction et Gestion des Connaissances*, Toulouse, France. Hermann-Éditions.

Cheng, R., Kalashnikov, D. V., and Prabhakar, S. (2003). Evaluating probabilistic queries over imprecise data. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 551–562, San Diego, California.

Cuzzolin, F. (2008). A geometric approach to the theory of evidence. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 38(4):522–534.

Cuzzolin, F. (2009). Consistent approximations of belief functions. In *6th International Symposium on Imprecise Probability: Theories and Applications*, Durham, United Kingdom.

Daniel, M. (2010). Conflicts within and between belief functions. In Hllermeier, E., Kruse, R., and Hoffmann, F., editors, *Computational Intelligence for Knowledge-Based Systems Design, Proceedings of the 13th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, volume 6178 of *Lecture Notes in Computer Science*, pages 696–705, Dortmund, Germany. Springer Berlin Heidelberg.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339.

Deng, Y., Wang, D., and Li, Q. (2008). An improved combination rule in fault diagnosis based on Dempster-Shafer theory. In *Proceedings of the 7th International Conference on Machine Learning and Cybernetics*, volume 1, pages 212–216, Kunming, China.

Denœux, T. (2000). A neural network classifier based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(2):131–150.

Denœux, T. (2001). Inner and outer approximation of belief structures using a hierarchical clustering approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(4):437–460.

Denœux, T. (2006a). The cautious rule of combination for belief functions and some extensions. In *Proceedings of the 9th International Conference on Information Fusion*, pages 1–8, Firenze, Italy.

Denœux, T. (2006b). Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, 42(3):228–252.

Denœux, T. (2008). Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2-3):234–264.

Denœux, T. and Masson, M.-H. (2003). Clustering of proximity data using belief functions. In Bouchon-Meunier, B., Foulloy, L., and Yager, R., editors, *Intelligent Systems for Information Processing: From Representation to Application*, pages 291–302, Amsterdam, Kingdom of the Netherlands. Elsevier.

Denœux, T. and Masson, M.-H. (2004). EVCLUS: evidential clustering of proximity data. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 34(1):95–109.

Destercke, S. and Burger, T. (2012). Revisiting the notion of conflicting belief functions. In Denœux, T. and Masson, M.-H., editors, *Belief Functions: Theory and Applications, Proceedings of the 2nd International Conference on Belief Functions*, volume 164 of *Advances in Intelligent and Soft Computing*, pages 153–160, Compiègne, France. Springer Berlin Heidelberg.

Diaz, J., Rifqi, M., and Buchon-Meunier, B. (2006). A similarity measure between basic belief assignments. In *Proceedings of the 9th International Conference on Information Fusion*, pages 1–6, Firenze, Italy.

Dubois, D. and Prade, H. (1986). A set-theoretic view of belief functions - logical operations and approximation by fuzzy sets. *International journal of General Systems*, 12(3):193–226.

Dubois, D. and Prade, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264.

Dubois, D. and Yager, R. R. (1987). The principle of minimum specificity as a basis for evidential reasoning. In Bouchon, B. and Yager, R., editors, *Uncertainty in Knowledge-Based Systems, International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, volume 286 of *Lecture Notes in Computer Science*, pages 75–84, Pais, France. Springer Verlag.

Elouedi, Z., Lefèvre, E., and Mercier, D. (2010). Discountings of a belief function using a confusion matrix. In *Proceedings of the 22nd IEEE International Conference on Tools with Artificial Intelligence*, volume 1, pages 287–294, Arras, France.

Elouedi, Z., Mellouli, K., and Smets, P. (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. (2007). *Combinaison d'informations hétérogènes dans le cadre unificateur des ensembles aléatoires : Approximations et robustesse.* PhD thesis, Université Laval, Québec, Québéc.

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

Florea, M. C., Bossé, E., and Jousselme, A.-L. (2009a). Metrics, distances and dissimilarity measures within Dempster-Shafer theory to characterize sources' reliability. In *COGnitive systems with Interactive Sensors*, Paris, France.

Florea, M. C., Dezert, J., Valin, P., Smarandache, F., and Jousselme, A.-L. (2006). Adaptative combination rule and proportional conflict redistribution rule for information fusion. In *COGnitive systems with Interactive Sensors*, Paris, France.

Florea, M. C., Jousselme, A.-L., and Bossé, E. (2010). Dynamic estimation of evidence discounting rates based on information credibility. *RAIRO - Operations Research*, 44(4):285–306.

Florea, M. C., Jousselme, A.-L., Bossé, E., and Grenier, D. (2009b). Robust combination rules for evidence theory. *Information Fusion*, 10(2):183–197.

Frikha, A. (2014). On the use of a multi-criteria approach for reliability estimation in belief function theory. *Information Fusion*, 18:20–32.

Gançarski, P. and Wemmert, C. (2005). Collaborative multi-strategy classification: Application to per-pixel analysis of images. In *Proceedings of the 6th International Workshop on Multimedia Data Mining: Mining Integrated Media and Complex Data*, pages 15–22, Chicago, Illinois, USA.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states.* MIT Press, Cambridge, MA.

Guan, X., Yi, X., and He, Y. (2008). Research on conflicting evidences combination strategies. In *Proceedings of the 7th International Conference on Machine Learning and Cybernetics*, volume 1, pages 110 – 114, Kunming, China.

Harmanec, D. (1999). Faithful approximations of belief functions. In Laskey, K. and Prade, H., editors, *Proceedings of the 15th conference on Uncertainty in Artificial Intelligence*, pages 271–278, Stockholm, Sweden. Morgan Kaufmann.

Hewawasam, K., Premaratne, K., Subasingha, S., and Shyu, M.-L. (2005). Rule mining and classification in imperfect databases. In *Proceedings of the 8th International Conference on Information Fusion*, volume 1, pages 661–668, Philadelphia, USA.

Hsia, Y.-T. (1991). Characterizing belief with minimum commitment. In Mylopoulos, J. and Reiter, R., editors, *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 1184–1189, Sydney, Australia. Morgan Kaufmann.

Huynh, V.-N. (2009). Discounting and combination scheme in evidence theory for dealing with conflict in information fusion. In Torra, V., Narukawa, Y., and Inuiguchi, M., editors, *Proceedings of 6th International Conference on Modeling Decisions for Artificial Intelligence*, volume 5861 of *Lecture Notes in Computer Science*, pages 217–230, Awaji Island, Japan. Springer Berlin Heidelberg.

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

Jousselme, A.-L. and Maupin, P. (2010). On some properties of distances in evidence theory. In *Workshop on the theory of belief functions*, Brest, France.

Jousselme, A.-L. and Maupin, P. (2012). Distances in evidence theory: Comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 53(2):118–145.

Kempe, D., Kleinberg, J., and Tardos, E. (2003). Maximizing the spread of influence through a social network. In Doningos, P. and Faloutsos, C., editors, *Proceedings* *of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, Washington, USA. ACM Press.

Kendall, D. G. (1974). Foundation of a theory of random sets. *Stochastic Geometry*, pages 322–376.

Klir, G. J. and Wierman, M. J. (1999). *Uncertainty-Based Information: Elements for Generalized Information Theory*, volume 15. Physica-Verlag Heidelberg, New york, 2nd edition.

Kuter, U. and Golbeck, J. (2007). SUNNY: A new algorithm for trust inference in social networks using probabilistic confidence models. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*, pages 1377–1382, Vancouver, British Columbia, Canada. The AAAI Press, Menlo Park, California.

Kuter, U. and Golbeck, J. (2010). Using probabilistic confidence models for trust inference in webbased social networks. *ACM Transactions on Internet Technology*, 10(2):Article 8.

Lefèvre, E., Colot, O., and Vannoorenberghe, P. (2002). Belief function combination and conflict management. *Information Fusion*, 3(2):149–162.

Lefèvre, E., Colot, O., and Vannoorenberghe, P. (2003). Reply to the comments of R. Haenni on the paper "Belief function combination and conflict management". *Information Fusion*, 4(1):63–65.

Lefèvre, E. and Elouedi, Z. (2013). How to preserve the conflict as an alarm in the combination of belief functions? *Decision Support Systems*, 56:326–333.

Lefèvre, E., Elouedi, Z., and Mercier, D. (2011). Towards an alarm for opposition conflict in a conjunctive combination of belief functions. In Liu, W., editor, *Proceedings of the 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 6717 of *Lecture Notes in Computer Science*, pages 314–325, Belfast, UK. Springer Berlin Heidelberg.

Levien, R. (2002). *Computing with Social Trust*, chapter Attack-Resistant Trust Metrics, pages 121–132. Human-Computer Interaction Series. Springer London.

Liu, W. (2006). Analyzing the degree of conflict among belief functions. *Artificial Intelligence*, 170(11):909–924.

Liu, W., Hughes, J. G., and McTear, M. F. (1992). Representing heuristic knowledge in D-S theory. In *Proceedings of the Eighth Annual Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 182–190, Stanford University, Stanford, CA, USA.

Liu, Z.-g., Dezert, J., Pan, Q., and Mercier, G. (2011). Combination of sources of evidence with different discounting factors based on a new dissimilarity measure. *Decision Support Systems*, 52(1):133–141.

Martin, A. (2009). *Advances and Applications of DSmT for Information Fusion*, volume 3, chapter Implementing general belief function framework with a practical codification for low complexity, pages 217–274. American Research Press Rehoboth.

Martin, A. (2012). About conflict in the theory of belief functions. In Denœux, T. and Masson, M.-H., editors, *Belief Functions: Theory and Applications, Proceedings of the 2nd International Conference on Belief Functions*, volume 164 of *Advances in Intelligent and Soft Computing*, pages 161–168, Compiègne, France. Springer Berlin Heidelberg.

Martin, A., Jousselme, A.-L., and Osswald, C. (2008). Conflict measure for the discounting operation on belief functions. In *Proceedings of the 11th International Conference on Information Fusion*, pages 1–8, Cologne, Germany.

Martin, A. and Osswald, C. (2006a). *Advances and Applications of DSmT for Information Fusion (Collected works)*, volume 2, chapter A new generalization of the proportional conflict redistribution rule stable in terms of decision, pages 69–88. American Research Press Rehoboth.

Martin, A. and Osswald, C. (2006b). Human experts fusion for image classification. *Information & Security: An International Journal, Special issue on Fusing Uncertain, Imprecise and Paradoxist Information (DSmT)*, 20:122–143.

Martin, A. and Osswald, C. (2007a). Toward a combination rule to deal with partial conflict and specificity in belief functions theory. In *Proceedings of the 10th International Conference on Information Fusion*, pages 1–8, Québec, Canada.

Martin, A. and Osswald, C. (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.

Masson, M.-H. and Denœux, T. (2004). Clustering interval-valued proximity data using belief functions. *Pattern Recognition Letters*, 25(2):163–171.

Masson, M.-H. and Denœux, T. (2008). ECM: an evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384–1397.

Mercier, D. (2006). *Fusion d'informations pour la reconnaissance automatique d'adresses postales dans le cadre de la théorie des fonctions de croyance*. PhD thesis, Université de Technologie de Compiègne.

Mercier, D., Denœux, T., and Masson, M.-H. (2006). Refined sensor tuning in the belief function framework using contextual discounting. In *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, volume 2, pages 1443–1450, Paris, France.

Mercier, D., Quost, B., and Denœux, T. (2005). Contextual discounting of belief functions. In Godo, L., editor, *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 3571 of *Lecture Notes in Computer Science*, pages 552–562, Barcelona, Spain. Springer Berlin Heidelberg.

Mercier, D., Quost, B., and Denœux, T. (2008). Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246–258.

Munkres, J. (1957). Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38.

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

Osswald, C. and Martin, A. (2006). Understanding the large family of Dempster-Shafer theory's fusion operators - a decision-based measure. In *Proceedings of the 9th International Conference on Information Fusion*, pages 1–7, Firenze, Italy.

Perry, W. L. and Stephanou, H. E. (1991). Belief function divergence as a classifier. In *Proceedings of the 1991 IEEE International Symposium on Intelligent Control*, pages 280–285, Arlington, Virginia, USA.

Pichon, F., Dubois, D., and Denœux, T. (2012). Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53(2):159–175.

Ristic, B. and Smets, P. (2006). The TBM global distance measure for the association of uncertain combat ID declarations. *Information Fusion*, 7(3):276–284.

Schubert, J. (2003). Clustering belief functions based on attracting and conflicting metalevel evidence. In *Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, volume 3, pages 571–578, Annecy, France. Elsevier Science.

Schubert, J. (2004). Clustering belief functons based on attracting and conflicting metalevel evidence using potts spin mean field theory. *Information Fusion*, 5(4):309–318.

Schubert, J. (2008). Conflict management in Dempster-Shafer theory by sequential discounting using the degree of falsity. In Magdalena, L., Ojeda-Aciego, M., and Verdegay, J., editors, *Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 298–305, Málaga, Spain.

Schubert, J. (2011). Conflict management in Dempster-Shafer theory using the degree of falsity. *International Journal of Approximate Reasoning*, 52(3):449–460.

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

Smarandache, F. and Dezert, J. (2005). Information fusion based on new proportional conflict redistribution rules. In *Proceeding of 8th International Conference on Information Fusion*, volume 2, pages 907–914, Philadelphia, USA.

Smets, P. (1988). Belief functions. In Smets, P., Mamdani, A., Dubois, D., and Prade, H., editors, *Non standard logics for automated reasoning*, pages 253–286, London, England. Academic Press.

Smets, P. (1990). The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458.

Smets, P. (1992a). The concept of distinct evidence. In Bouchon-Meunier, B., Valverde, L., and Yager, R. R., editors, *Advanced Methods in Artificial Intelligence, Proceedings 4th International Conference on Processing and Management of Uncertainty in Knowledge-Based Systems*, Lecture Notes in Computer Science, pages 789–794, Palma de Mallorca, Spain. Springer Berlin Heidelberg.

Smets, P. (1992b). The nature of the unnormalized beliefs encountered in the transferable belief model. In Dubois, D. and Wellman, M. P., editors, *Proceedings of the 8th international conference on Uncertainty in Artificial Intelligence*, pages 292–297, Stanford, California, USA. Morgan Kaufmann. ISBN 1-55860-258-5.

Smets, P. (1993). Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1–35.

Smets, P. (1995). The canonical decomposition of a weighted belief. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, volume 2, pages 1896–1901, Montréal, Québec, Canada. Morgan Kaufman.

Smets, P. (1997). *Uncertainty Management in Information Systems*, chapter Imperfect Information: Imprecision and Uncertainty, pages 225–254. Springer US.

Smets, P. (2005). Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasonning*, 38(2):133–147.

Smets, P. (2007). Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412.

Smets, P. and Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, 66(2):191–234.

Smets, P. and Kruse, R. (1997). *Uncertainty Management in Information Systems: From Needs to Solutions*, chapter The Transferable Belief Model for Belief Representation, pages 343–368. Springer US, Boston.

Smithson, M. (1989). *Ignorance and uncertainty: Emerging paradigms.* Springer-Verlag, New York, USA.

Stephanou, H. E. and Lu, S.-Y. (1988). Measuring consensus effectiveness by a generalized entropy criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):544–554.

Tessem, B. (1993). Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61(2):315–329.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London.

Wemmert, C. and Gançarski, P. (2002). A multi-view voting method to combine unsupervised classifications. In *Proceedings of the 2nd IASTED International Conference on Artificial Intelligence and Applications*, pages 447–453, Málaga, Spain.

Wong, S. K. M. and Lingras, P. (1994). Representation of qualitative user preference by quantitative belief functions. *IEEE Transactions on Knowledge and Data Engineering*, 6(1):72–78.

Yager, R. R. (1983). Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9(4):249–260.

Yager, R. R. (1986). The entailment principile for Dempster-Shafer granules. *International Journal of Intelligent Systems*, 1(4):247–262.

Yager, R. R. (1987). On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41(2):93–137.

Yang, Y., Han, D., and Han, C. (2013). Discounted combination of unreliable evidence using degree of disagreement. *International Journal of Approximate Reasoning*, 54(8):1197–1216.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

Zadeh, L. A. (1984). A mathematical theory of evidence *(book review). AI magazine*, 5(3):81–83.

Zadeh, L. A. (1999). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100(1):9–34.

Zeng, C. and Wu, P. (2007). A reliability discounting strategy based on plausibility function of evidence. In *Proceedings 10th International Conference on Information Fusion*, pages 1–6, Québec, Canada.

Zouhal, L. M. and Denœux, T. (1998). An evidence-theoric k-NN rule with parameter optimization. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 28(2):263–271.