

Un état de l'art sur les fonctions de croyance appliquées au traitement de l'information

Patrick Vannoorenberghe*

Perception, Systèmes, Information, FRE 2645 CNRS
Université de Rouen, UFR des Sciences
Place Emile Blondel, 76821 Mont Saint-Aignan Cedex
Patrick.Vannoorenberghe@univ-rouen.fr

Résumé

Un des principaux challenges des Technologies de l'Information et de la Communication concerne l'accès à une information plus fiable, à sa représentation informatique ainsi qu'à son traitement. Ainsi, dans le domaine de l'intelligence artificielle, nombreuses sont les recherches qui ont permis d'étendre les formalismes mis en place pour le développement de systèmes intelligents au traitement de l'information imprécise et incertaine. Parmi les modèles étudiés, on retrouve les différentes théories des mesures de confiance qui inclut notamment un formalisme capable d'appréhender à la fois imprécision et incertitude, la théorie des fonctions de croyance. Dans cet article, un état de l'art sur l'application de ce cadre théorique au traitement de l'information est proposé. Via la description du modèle des croyances transférables introduit par Ph. Smets, les principaux outils associés à la théorie sont présentés. Cette description permet au lecteur néophyte de se familiariser avec les éléments mathématiques utiles à la représentation et à la manipulation de l'information. Celle-ci peut prendre plusieurs formes (mesures, données ou connaissances) selon le type de problèmes abordés (reconnaissance des formes, fusion d'informations). Ce modèle est suffisamment souple pour être appliqué dans de nombreux domaines des Sciences et Technologies de l'Information et de la Communication tels que l'analyse de données, le diagnostic, l'aide à la décision, la perception multicapteurs et le traitement d'images.

Mots-clés : Information, Imprécision, Incertitude, Théorie des fonctions de croyance, Reconnaissance des formes, Fusion

1 INTRODUCTION

Dans les nombreux secteurs applicatifs liés aux Sciences et Technologies de l'Information et de la Communication (STIC), il s'agit de classer, décider, surveiller, commander à partir d'informations observées sur un système ou préalablement modélisées [52]. De manière générale, les finalités de ces applications consistent à aider un opérateur dans sa tâche, transmettre et enregistrer des savoir-faire, gérer des connaissances, améliorer les performances ou l'adaptabilité de systèmes artificiels. Dans [52], les auteurs dégagent quatre grandes classes de tâches de la problématique scientifique du traitement de l'information :

1. Mettre l'information sous une forme utile,
2. Stocker, retrouver, expliciter l'information,
3. Exploiter l'information pour décider et agir,
4. Communiquer l'information.

Pour ce faire, les systèmes développés se basent sur un flot d'informations qui est soit préalablement collecté (apprentissage), soit recueilli en ligne (observation). Le niveau d'abstraction de cette information diffère selon son origine et peut se catégoriser en trois familles :

- une **mesure** : c'est le cas notamment pour les systèmes de perception multicapteurs. Dans ce contexte, il s'agit alors de fusionner les mesures provenant d'un ensemble de capteurs (signaux et/ou images) pour la commande d'un système ou la prise d'une décision. Le problème consiste généralement à fiabiliser les mesures (filtrer, débruiter la mesure, lui associer un degré d'incertitude) pour les rendre exploitables *a posteriori*.
- une **donnée** : dans le cadre de la gestion des bases. Suivant le type de données (financières, commerciales, médicales,...), il peut s'agir d'extraire des variables discriminantes, d'identifier des structures sous-jacentes, d'extraire des informations utiles de manière générale, c'est le concept de fouille de données [33, 60, 69]. Une partie de la communauté scientifique se regroupe autour de cette problématique connue sous le vocable *KDD* pour *Knowledge Discovery in Databases*. Dans le cadre d'un système d'indexation, le problème peut consister à sélectionner les variables pertinentes de la base, catégoriser les données pour les indexer, fusionner les données avec la requête de l'utilisateur au niveau de la boucle de pertinence.

- une **connaissance** : on parle de modélisation de connaissances ou d'ingénierie des connaissances. Dans ce dernier cas, les difficultés proviennent de la capture des connaissances, de leur explicitation et de leur interprétation sans introduction d'un biais. Un exemple peut provenir de la structuration de connaissances en vue de leur formalisation logique.

Pour appréhender les difficultés liées au traitement de ces informations, voyons à quelles notions elles peuvent être attachées [52]. La mesure d'une grandeur physique est généralement entachée d'une imprécision liée au choix et à la technologie du capteur employé. Cette notion d'imprécision est généralement bien maîtrisée mais est souvent liée à un second concept qui se différencie par le fait qu'il ne fait pas référence au contenu de l'information mais à sa qualité. Par exemple, la phrase "Il va peut-être pleuvoir demain", concerne le manque de certitude quant à la prévision faite par la météorologie. Il s'agit dans ce cas d'une incertitude quant à la réalisation d'un événement. Enfin, un des nombreux problèmes couramment rencontré dans de nombreuses situations concerne le manque d'informations. C'est le cas lors du dysfonctionnement d'un capteur ; on parle alors d'incomplétude. Modéliser (représenter, accéder et traiter) des informations aussi imparfaites devient alors rapidement un problème crucial et constitue l'un des principaux challenges de la société de l'information¹.

Nombreuses sont les recherches qui se sont attachées à développer des outils théoriques pour la gestion de ces informations imparfaites [105, 154]. Ainsi, la théorie des sous-ensembles flous (Fuzzy Sets Theory) développée par L. Zadeh [156] a été un cadre précurseur. Un autre cadre théorique très séduisant est certainement la théorie des fonctions de croyance. Elle permet dans un même formalisme d'encoder l'imprécision et l'incertitude contenue dans une information. Un des principaux ambassadeurs de ce formalisme est sans doute Ph. Smets qui, via le modèle des croyances transférables [134, 129], continue à populariser la théorie développée par A. Dempster [35] et G. Shafer [118]. En France, la communauté scientifique s'organise autour de quelques laboratoires et équipes. Sur le plan théorique, D. Dubois et H. Prade ont énormément contribué. Du point de vue applicatif, citons les travaux de T. Denœux en analyse de données, les travaux d'A. Appriou en fusion d'informations, et d'autres travaux que nous aurons

1. En réalité, nombreuses sont les causes de l'imperfection des informations. Nous en avons cité trois mais dans [52], les auteurs parlent d'information ambiguë, bruitée, biaisée, incomplète, imprécise, incertaine, incohérente, redondante.

l'occasion de citer par la suite comme ceux d'I. Bloch [23] ou M. Rombaut.

Dans cet article, nous proposons de brosser un état de l'art sur l'application des fonctions de croyance au traitement de l'information. Nous présentons tout d'abord quelques éléments mathématiques associés à ce formalisme (section 2) qui a permis d'appréhender de nombreux problèmes liés au traitement d'informations incertaines. Nous décrivons ces problématiques ainsi que les avantages liés à l'utilisation des fonctions de croyance dans la section 3. Enfin, nous passons en revue les domaines applicatifs tels que l'analyse de données, le diagnostic, l'aide à la décision, la perception multi-capteur ainsi que le traitement d'images dans la section 4. Nous tenterons de donner quelques perspectives d'avenir concernant ce cadre théorique dans la conclusion de cet article.

2 ÉLÉMENTS MATHÉMATIQUES

Les travaux d'A. Dempster sur les bornes inférieure et supérieure d'une famille de distributions de probabilités [35] ont permis à G. Shafer d'asseoir les bases de la théorie des fonctions de croyance [118]. Ce formalisme théorique peut revêtir plusieurs interprétations [84] et appellations (théorie des fonctions de croyance ou théorie de Dempster-Shafer). G. Shafer a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines. L'utilité de ces fonctions, comme une alternative aux probabilités subjectives, a été démontrée plus tard de manière axiomatique par Ph. Smets [124, 127] au travers du modèle des croyances transférables. Ce modèle fournit une interprétation non probabiliste cohérente et permet de clarifier le concept sous-jacent à la théorie. Dans cette section, nous abordons quelques éléments mathématiques liés aux fonctions de croyance. Le concept de fonction de croyance est tout d'abord présenté, permettant ainsi au lecteur de comprendre comment sont représentées les informations dans une structure de croyance. Quelques outils associés à ce formalisme sont ensuite détaillés. Il ne s'agit pas ici de présenter l'ensemble des outils disponibles pour manipuler (combiner, réviser, transformer) les fonctions de croyance. Nous nous contentons de décrire ceux qui sont utiles à la compréhension de la suite de cet article.

2.1 Représentation de l'information

Soit $\Omega = \{\omega_1, \dots, \omega_k, \dots, \omega_K\}$ un ensemble fini, généralement appelé cadre de discernement. Une fonction de croyance (une capacité de Choquet [32] monotone d'ordre infini) bel est une fonction de 2^Ω vers $[0,1]$ définie par :

$$\begin{aligned} bel(\emptyset) &= 0 \\ \forall n \geq 1, \forall i = 1, \dots, n, A_i &\subseteq \Omega \\ bel(\cup_{i=1, \dots, n} A_i) &\geq \sum_{I \subseteq \{1, \dots, n\}, I \neq \emptyset} (-1)^{|I|+1} bel(\cap_{i=1, \dots, n} A_i). \end{aligned} \quad (1)$$

Le cas particulier où $n = 2$ et $A_1 \cap A_2 = \emptyset$ illustre le fait que la croyance attribuée à l'union de deux sous-ensembles disjoints de Ω est supérieure ou égale à la somme des croyances attribuées à chacun des sous-ensembles². Si l'inégalité est transformée en égalité dans l'équation (1), la fonction résultante bel , appelée généralement fonction de croyance bayésienne, devient une fonction de probabilité classique. Une fonction de croyance peut également être définie mathématiquement par une fonction (ou allocation) de masse, notée m définie de 2^Ω dans $[0,1]$, qui vérifie :

$$\sum_{A \subseteq \Omega} m(A) = 1.$$

Chaque sous-ensemble $A \subseteq \Omega$ tel que $m(A) > 0$ est appelé élément focal de m . Ainsi, la masse $m(A)$ représente le degré de croyance attribué à la proposition A et qui n'a pas pu, compte tenu de l'état de la connaissance, être affectée à un sous-ensemble plus spécifique que A . Une fonction telle que $m(\emptyset) = 0$ est dite normale. Dans le modèle des croyances transférables, la condition $\sum_{\emptyset \neq A \subseteq \Omega} m(A) = 1$ n'est pas supposée et $m(\emptyset) > 0$ est acceptée. Cette spécificité permet d'introduire la notion de *monde ouvert* en supposant que la croyance ne peut être placée sur un sous-ensemble de Ω . Dans ce contexte, l'ensemble vide \emptyset peut être interprété comme une hypothèse non clairement définie dans l'ensemble de discernement par opposition au *monde fermé* où l'ensemble des hypothèses est exhaustif. Etant donnée une fonction de masse m , on peut définir bel comme :

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad \forall A \subseteq \Omega. \quad (2)$$

2. $bel(A_1 \cup A_2) \geq bel(A_1) + bel(A_2) \quad \forall A_1, A_2 \subseteq \Omega$ si $A_1 \cap A_2 = \emptyset$.

Les fonctions m et bel représentent deux facettes de la même information³. De plus, on peut retrouver m à partir de bel en utilisant la transformée de Möbius [77, 76, 75]. L'implémentation de cette transformation peut être obtenue par de simples calculs matriciels [132]. Issues des travaux de G. Shafer [118], les fonctions de croyance sont de nos jours reconnues pour la modélisation d'informations incertaines.

Une des principales difficultés consiste à modéliser la connaissance disponible en initialisant de manière adéquate les fonctions de croyance. Supposons, par exemple, qu'il faille modéliser la connaissance concernant la culpabilité de trois suspects $\Omega = \{\text{Philippe, Arthur, Laurence}\}$ exprimée par un témoin dans une affaire de meurtre. Sans connaissance particulière sur la culpabilité des trois suspects, la fonction de croyance est caractérisée par l'allocation de masse vide $m(\Omega) = 1$ qui représente l'ignorance totale. Supposons maintenant que le témoin a aperçu la chevelure blonde du meurtrier qui ne peut correspondre qu'à Philippe ou Arthur. L'allocation de masse prend alors la forme $m(\{\text{Philippe, Arthur}\}) = 1$. Si le témoin a vu le meurtrier, il peut formaliser sa connaissance sous la forme d'une fonction attribuant la totalité de la masse sur un des singletons de Ω . L'ignorance partielle peut évidemment être codée elle aussi sous la forme d'une allocation de masse [53]. Modéliser le plus fidèlement l'information disponible confère aux fonctions de croyance l'avantage de ne pas introduire d'information *a priori* contrairement aux fonctions de probabilités. En effet, le concept d'équi-probabilité n'est pas identique au concept d'ignorance totale mais constitue déjà une forme précise de connaissance.

Les liens complexes avec la théorie des probabilités [123] n'ont malheureusement pas aidé à faire connaître ce cadre théorique. Ceci est dû probablement au fait qu'il est difficile de clairement définir une fonction de croyance et certaines confusions proviennent généralement des différences de compréhension que l'on peut donner à ce terme. Dans la littérature, on peut recenser trois familles de modèles qui font apparaître le terme « fonctions de croyance » :

- le «Upper and Lower Probability Model» [145],
- le «Dempster's Model» et le «Hint Model» de Kholas et Monney [84],
- le «Transferable Belief Model» introduit par Ph. Smets [134].

Dans cet article, nous ne détaillons pas ces familles de modèle qui sont dé-

3. Ph. Smets utilise la notation $m^\Omega[data]$ pour représenter une fonction de masse définie sur Ω basée sur l'observation des données $[data]$.

crites dans [131] mais présenterons uniquement le point de vue du modèle des croyances transférables. Dans ce modèle, deux niveaux peuvent être distingués : le *niveau crédal* où les croyances sont modélisées et révisées, et le *niveau pignistique* dans lequel les fonctions de croyance sont transformées en fonctions de probabilité pour la prise de décision [133].

Lorsque la source d'information dont la fonction de croyance est extraite n'est pas totalement fiable, il est possible d'introduire une opération d'affaiblissement. Dans ce cas, un coefficient α , qui représente une sorte de méta-connaissance concernant la fiabilité de la source, permet de transférer une partie de la croyance vers l'ensemble Ω . Ainsi, une fonction de masse affaiblie, notée m_α , peut se déduire de m par :

$$m_\alpha(A) = \alpha m(A) \quad \forall A \subset \Omega, A \neq \Omega \quad (3)$$

$$m_\alpha(\Omega) = 1 - \alpha + \alpha m(\Omega). \quad (4)$$

Ainsi, une fonction de croyance est capable de représenter plusieurs types de connaissance et constitue par ce biais un cadre riche et flexible pour la représentation d'informations incertaines. Dans la section suivante, nous introduisons quelques-uns des outils qui permettent de manipuler l'information dans une chaîne de traitement.

2.2 Outils associés

Un des premiers outils que nous choisissons de décrire concerne la combinaison de deux fonctions de croyance. Supposons que nous disposons de deux allocations de masse m_1 et m_2 définies sur le même référentiel Ω . Ces deux fonctions peuvent être agrégées par un opérateur de combinaison conjonctif noté \odot . Le résultat de cette opération conduit à une fonction de croyance unique à laquelle correspond une fonction de masse, notée m_{\odot} , qui peut être définie par :

$$m_{\odot}(A) = (m_1 \odot m_2)(A) \triangleq \sum_{B \cap C = A} m_1(B) m_2(C) \quad \forall A \subseteq \Omega. \quad (5)$$

Cette règle conjonctive est parfois dénommée règle de combinaison de Dempster non normalisée. Si nécessaire, l'hypothèse de normalisation $m_{\odot}(\emptyset) = 0$ peut être retrouvée en divisant chaque masse par un coefficient adéquat. L'opérateur résultant, qui est connu sous le nom de règle de Dempster et

dénoté m_{\oplus} , est défini par :

$$(m_1 \oplus m_2)(A) \triangleq \frac{(m_1 \odot m_2)(A)}{1 - m_{\odot}(\emptyset)} \quad \forall \emptyset \neq A \subseteq \Omega. \quad (6)$$

La quantité $m_{\odot}(\emptyset)$ est le degré de conflit entre les fonctions m_1 et m_2 et peut être calculée par :

$$m_{\odot}(\emptyset) = (m_1 \odot m_2)(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B)m_2(C). \quad (7)$$

L'utilisation de la règle de Dempster est possible si les fonctions de masse m_1 et m_2 ne sont pas en conflit total c'est-à-dire s'il existe deux éléments focaux B et C respectivement de m_1 et m_2 tels que $B \cap C \neq \emptyset$. Par ce fait, cette règle, qui permet de combiner des informations incertaines extraites sous forme de fonctions de croyance, a été très discutée [157, 153, 122]. Elle a cependant été justifiée en introduisant des axiomes particuliers [79] bien que certains auteurs ont envisagé d'autres solutions [152, 98] comme par exemple des règles disjonctives [125]. Nous reviendrons sur ce problème en détail dans la section 3.2 en présentant une approche originale [88, 89] qui permet de retrouver la plupart des règles décrites dans la littérature.

Dans le modèle des croyances transférables, la fonction de masse a une interprétation naturelle puisque la masse $m(A)$ représente le degré de croyance attribué à la proposition $A \subseteq \Omega$ qui n'a pas pu être assignée à un sous-ensemble plus spécifique. Si maintenant, on apprend que la croyance allouée au sous-ensemble $B \subseteq \Omega$ est totale alors la masse $m(A)$ initialement allouée à A est *transférée* au sous-ensemble $A \cap B$ d'où le nom du modèle introduit par Ph. Smets. La fonction de masse résultante est celle obtenue par l'application de la règle dite de conditionnement qui n'est qu'un cas particulier de la règle de combinaison de Dempster. Ainsi, le résultat de la combinaison d'une fonction de masse m conditionnellement à B produit une fonction de croyance, notée $bel(.|B)$ qui peut se calculer $\forall A \subseteq \Omega$ par :

$$bel(A|B) = bel(A \cup \overline{B}) - bel(\overline{B}). \quad (8)$$

Au niveau pignistique, une fonction de croyance unique, sorte de résumé exhaustif de l'information disponible au niveau crédal, est utilisée pour la prise de décision. En basant son raisonnement sur des arguments de rationalité développés dans le modèle des croyances transférables, Ph. Smets [134]

propose de transformer une fonction de masse m en une fonction de probabilité $BetP_m$ définie sur Ω (appelée fonction de probabilité *pignistique*) qui se formalise pour tout $\omega_k \in \Omega$ par :

$$BetP_m(\omega_k) = \frac{1}{1 - m(\emptyset)} \sum_{A \ni \omega_k} \frac{m(A)}{|A|} \quad (9)$$

où $|A|$ représente la cardinalité du sous-ensemble $A \subseteq \Omega$. Dans cette transformation, la masse de croyance $m(A)$ est uniformément distribuée parmi les éléments de A . À partir de cette distribution de probabilité, il est alors possible d'utiliser les outils classiques de la théorie de la décision statistique pour la prise de décision [133]. On suppose ainsi qu'il existe une fonction de coût $\lambda : \mathcal{A} \times \Omega \rightarrow \mathbb{R}$ où \mathcal{A} désigne un ensemble fini d'actions définie telle que $\lambda(a, \omega)$ quantifie le coût de choisir l'action a quand la vérité est ω . A partir de la probabilité $BetP_m$, on peut associer à chaque action $a \in \mathcal{A}$ un risque $BetR$ défini comme le coût espéré si l'on choisit l'action a lorsque la vérité est ω :

$$BetR(a) = \sum_{\omega \in \Omega} BetP_m(\omega) \lambda(a, \omega). \quad (10)$$

Il suffit alors de choisir l'action qui minimise ce risque, aussi appelé risque pignistique. D'autres règles de décision basées sur les fonctions de croyance elles-mêmes sont néanmoins envisageables [71, 72, 37, 138, 113]. Nous développons ici les stratégies basées sur les concepts de risque inférieur $InfR$ et supérieur $SupR$ présentées dans [37]. Ces deux risques peuvent être respectivement définis par les équations suivantes :

$$InfR(a) = \sum_{A \subseteq \Omega} m(A) \min_{\omega \in A} \lambda(a, \omega), \quad (11)$$

$$SupR(a) = \sum_{A \subseteq \Omega} m(A) \max_{\omega \in A} \lambda(a, \omega). \quad (12)$$

Ils conduisent à des prises de décision tantôt optimiste dans le cas de la minimisation du risque inférieur ou pessimiste pour le risque supérieur. Avec des coûts 0-1, définis par $\lambda(a_k, \omega_r) = 1 - \delta_{k,r}$ pour $k, r \in \{1, \dots, K\}$, il a été montré [37] que la minimisation du risque pignistique conduit à choisir la classe ω de probabilité pignistique maximale tandis que les risques inférieur et supérieur conduisent respectivement à choisir la classe de crédibilité respectivement plausibilité maximale.

2.3 Mesures

Les fonctions de croyances sont un outil précieux pour la gestion de l'incertain et nombreux sont les travaux qui tentent de quantifier l'incertitude liée à une fonction particulière [121, 151, 51, 80]. Dans une étude récente [81], G.J. Klir remarque qu'une fonction de croyance est capable de modéliser deux types d'incertitude : la non-spécificité et le conflit. La non-spécificité peut être interprétée comme une absence de connaissance tandis que le conflit est généralement vu comme la connaissance de plusieurs solutions possibles. Une mesure de non-spécificité qui a le mérite de généraliser la mesure de R.V.L. Hartley [68] aux fonctions de croyance a été introduite par D. Dubois et H. Prade [50]. Elle est définie par :

$$N(m) = \sum_{A \subseteq \Omega} m(A) \log_2 |A|. \quad (13)$$

Puisque les éléments focaux des mesures de probabilité sont des singletons, la non-spécificité est nulle pour les fonctions de probabilité, et est maximale ($\log_2 |\Omega|$) pour la fonction de croyance vide. Plusieurs mesures de conflit, interprétées comme la mesure d'entropie de Shannon, ont également été introduites [82, 3]. Une mesure définie dans le cadre du modèle des croyances transférables est le *désaccord*, définie par :

$$D(m) = - \sum_{A \subseteq \Omega} m(A) \log_2 \text{Bet}P(A) \quad (14)$$

qui est maximale ($\log_2 |\Omega|$) pour la distribution de probabilité pignistique uniforme sur Ω . Il est à noter qu'en admettant l'hypothèse du monde ouvert, la quantité $m(\emptyset)$ peut être également vue comme une mesure du conflit. Une mesure d'incertitude totale U_λ peut également être définie en utilisant une combinaison linéaire de N et de D :

$$U(m) = (1 - \lambda)N(m) + \lambda D(m) \quad (15)$$

où $\lambda \in [0,1]$ est un coefficient. Cette combinaison a l'avantage de conserver les propriétés respectives des deux mesures N et D en donnant plus ou moins d'importance à l'une d'entre elles. Le lecteur désireux d'avoir une vision plus détaillée sur les mesures d'incertitude, notamment sur leurs propriétés et leurs spécificités respectives est invité à consulter [101, 102].

3 PROBLÈMES ABORDÉS

De part sa généralité et sa grande capacité à gérer des informations incertaines, la théorie des fonctions de croyances a permis d'aborder un large éventail de problèmes. Dans le cadre général de la discrimination [54] (Reconnaissance des Formes), elle a permis de généraliser certaines méthodes classiques à des problèmes d'apprentissage complexes. Nous présentons dans le paragraphe 3.1 trois familles de modèles ; les modèles basées sur l'exploitation des fonctions de vraisemblance, une extension de la méthode des k plus proches voisins, ainsi qu'une approche basée sur les arbres de décision. Dans le cadre de la fusion d'informations incertaines, une attention particulière sera apportée à l'étape de combinaison des fonctions de croyance dans laquelle le choix de l'opérateur reste délicat (cf. section 3.2). Les avantages liés à ce formalisme sont présentés à la section 3.3.

3.1 Reconnaissance des Formes

De nombreux travaux sur les fonctions de croyance abordent le problème complexe de la discrimination ou Reconnaissance des Formes (RdF). Le problème consiste à assigner un vecteur d'entrée \mathbf{x} à une classe dans Ω , étant donné la connaissance d'un ensemble d'apprentissage \mathcal{L} composé de n formes x^i de classes connues. Chaque exemple dans \mathcal{L} est représenté par un vecteur de dimension p et sa variable de classe ω^i correspondante.

Plusieurs approches à ce problème ont été introduites au cours des dix dernières années. Les avantages de ces méthodes qui exploitent la théorie des fonctions de croyance sont nombreux. En particulier, on peut accéder à une description de l'incertitude sur la prédiction, avoir la possibilité de rejeter un vecteur ou enfin de détecter une classe inconnue. Ces classifieurs sont bien adaptés aux applications où l'information disponible provient de nombreuses sources d'information imparfaites. Nous présentons dans la suite trois familles de construction de modèles de fonctions de croyance à partir d'un ensemble d'apprentissage \mathcal{L} : une approche basée sur les fonctions de vraisemblance (§ 3.1.1), une méthode dite des distances pouvant être vue comme une extension de l'algorithme des plus proches voisins (§ 3.1.2), et une méthode inspirée des arbres de décision (§ 3.1.3). Ces trois familles de modèles diffèrent selon la façon de coder l'incertitude sur la classe d'appartenance des exemples de l'ensemble d'apprentissage. L'extension des deux dernières familles à la gestion d'étiquettes incertaines est présentée au para-

graphe § 3.1.4.

3.1.1 Approche vraisemblance

Une première famille de modèles est basée sur les fonctions de vraisemblances [7, 146] où l'on y retrouve notamment le modèle consonant proposé initialement par G. Shafer dans son livre [118]. L'approche la plus séduisante dans cette famille a été proposée indépendamment par Ph. Smets [125] via le théorème de Bayes généralisé et A. Appriou [7]. Soient $f(\mathbf{x}|\omega_k)$ les densités de probabilité conditionnelles des classes que nous supposons connues. En ayant observé \mathbf{x} , une fonction de vraisemblance est une fonction de Ω dans $[0, +\infty[$ définie par $L(\omega_k|\mathbf{x}) = f(\mathbf{x}|\omega_k)$, pour tout $k \in \{1, \dots, K\}$. En utilisant cette fonction de vraisemblance construite à partir des données d'apprentissage contenues dans \mathcal{L} , il est possible de définir une fonction de masse m_k associée à la classe ω_k compte-tenu de différents axiomes que nous ne détaillerons pas dans cet article. Les éléments focaux de cette fonction sont le complémentaire $\overline{\omega_k}$ du singleton $\{\omega_k\}$ et l'ensemble Ω lui-même. Cette fonction de masse est définie de la manière suivante :

$$m_k(\{\omega_k\}) = 0 \quad (16)$$

$$m_k(\overline{\omega_k}) = \alpha_k(1 - R.L(\omega_k|\mathbf{x})) \quad (17)$$

$$m_k(\Omega) = 1 - \alpha_k(1 - R.L(\omega_k|\mathbf{x})). \quad (18)$$

Dans ces équations, α_k est un coefficient d'affaiblissement associé à la classe ω_k tandis que R est un coefficient de normalisation positif inférieur ou égal à $[\sup_{\mathbf{x}} \max_k (L(\omega_k|\mathbf{x}))]^{-1}$. Le choix du paramètre R reste arbitraire mais le principe de maximum d'incertitude conduit généralement à choisir la plus grande valeur possible dans l'intervalle ce qui a pour effet d'avoir la fonction de croyance la moins spécifique possible. A partir de ces K fonctions de croyance et en utilisant la règle de combinaison, une fonction de masse unique m peut être obtenue. Elle quantifie le degré d'appartenance du vecteur \mathbf{x} à chacune des classes dans Ω .

3.1.2 Méthode des distances

La seconde famille de modèles se base sur des informations de distance. Dans cette dernière, citons comme exemple l'extension de l'algorithme des k plus proches voisins qui a été introduit par T. Deneux [36]. Dans cette méthode, une fonction de croyance m^i est directement construite en utilisant

les informations apportées par les vecteurs x^i situés dans le voisinage du vecteur inconnu \mathbf{x} par :

$$m^i(\{\omega_k\}) = \alpha^i \phi^i(d^i) \quad (19)$$

$$m^i(\Omega) = 1 - \alpha^i \phi^i(d^i) \quad (20)$$

$$m^i(A) = 0 \quad \forall A \in 2^\Omega \setminus \{\{\omega_k\}, \Omega\} \quad (21)$$

où d^i est la distance euclidienne au vecteur x^i , α^i un paramètre associé au i -ème voisin et $\phi^i(d^i) = \exp[-\gamma^i(d^i)^2]$ avec γ^i un paramètre positif. En utilisant les k plus proches voisins, on obtient k fonctions de croyance à agréger par la règle de combinaison pour la prise de décision. Notons que différentes versions de cet algorithme ont été développées soit pour proposer une version neuronale de l'algorithme [40], soit pour permettre l'optimisation des paramètres [158]. Dans cette dernière, les paramètres Θ de l'algorithme⁴ sont optimisés en minimisant un critère d'erreur défini par :

$$E(\Theta) = \sum_{i=1}^n \sum_{k=1}^K (BetP^i(\omega_k) - u_k^i)^2. \quad (22)$$

Dans cette équation, u_k^i est l'indicateur de classe de l'exemple x^i ($u_k^i = 1$ si $\omega^i = \omega_k$) et $BetP^i(\omega_k)$ est la probabilité pignistique de ω_k pour le vecteur x^i . Cette technique d'optimisation des coefficients peut également s'appliquer dans le cadre des autres modèles.

Une comparaison de ces deux dernières familles de modèles a été réalisée dans [141]. Elle montre que les performances de celles-ci appliquées à des problèmes classiques ne diffèrent pas de manière importante et que le choix reste un sujet délicat. D'autres algorithmes basés sur la méthode des distances sont également détaillés dans [103].

3.1.3 Arbres de décision

Une approche radicalement différente utilise le mécanisme des arbres de décision pour la discrimination. L'induction d'un arbre de décision suit une stratégie descendante pour diviser les nœuds en se basant sur une mesure d'impureté calculée à partir des exemples qui atteignent chaque nœud t de l'arbre. Lorsque l'arbre classe tous les exemples contenus dans l'ensemble

4. Les paramètres à optimiser sont les coefficients d'affaiblissement α^i ainsi que les paramètres γ^i .

d'apprentissage \mathcal{L} , le processus est arrêté. Cette procédure conduit malheureusement à sur-apprendre (apprendre par cœur) si on n'y adjoint pas une méthode pour limiter la complexité de l'arbre, la stratégie d'élagage [29]. Après l'apprentissage de l'arbre, à chaque feuille est associée une étiquette qui permet de classer de nouveaux vecteurs. Plusieurs approches sont dérivées de ce mécanisme en y intégrant la souplesse des fonctions de croyance pour la gestion d'informations incertaines. Ainsi, dans [20, 41, 142], la méthode proposée s'appuie sur des travaux récents relatifs à l'inférence crédale et construit pour chaque nœud de l'arbre une fonction de croyance. Le choix de la sélection des variables au nœud t se base sur l'examen d'un critère d'entropie. Ce critère est déduit des exemples atteignant le nœud t par le calcul d'une mesure d'incertitude totale (cf. équations (13), (14) et (15)). Une autre approche a été introduite par Z. Elouedi et al. dans [55] puis améliorée en y intégrant notamment une stratégie d'élagage [56]. Enfin, une dernière méthode a été proposée dans [12] où l'allocation de masse ne permet d'obtenir que des arbres de décision probabilistes basés sur des variables linguistiques.

3.1.4 Extension aux étiquettes incertaines

Suivant la connaissance disponible sur les données d'apprentissage, la règle de discrimination peut être plus ou moins difficile à construire. Ainsi, si l'ensemble d'apprentissage contient des données étiquetées et des données non étiquetées, il faut mettre en œuvre une technique capable de s'affranchir de ce type d'information. L'avantage des techniques manipulant les fonctions de croyance par rapport à d'autres approches réside dans le fait qu'elles peuvent formaliser ces données d'apprentissage sous la forme :

$$\mathcal{L} = \{(x^i, m^i), i = 1, \dots, n\}, \quad (23)$$

où l'allocation de masse m^i représente le degré de croyance concernant la classe d'appartenance de l'exemple i dans Ω . Ce concept permet en outre de gérer des ensembles d'apprentissage divers qui peuvent inclure :

- des étiquettes certaines dans le cas où la fonction m^i possède un seul élément focal singleton dans Ω ,
- des étiquettes probabilistes qui sont exprimées sous la forme d'une fonction de croyance bayésienne;
- des étiquettes possibilistes; dans ce cas la distribution de possibilité peut être transformée en fonction de croyance dont les éléments focaux sont emboîtés [45];

- des exemples non étiquetés pouvant s'exprimer par la fonction de masse vide $m^i(\Omega) = 1$.

La méthode des distances a été l'une des premières approches à être capable de gérer ces ensembles d'apprentissage [38, 45]. Une comparaison avec une approche par modèle de mélange a été réalisée dans [5]. L'extension de la méthode des arbres de décision aux étiquettes incertaines a été également menée dans [142].

3.2 Fusion d'informations

La fusion d'information a depuis une dizaine d'années suscité un intérêt certain dans la communauté scientifique [1, 2, 21]. En utilisant la redondance et la complémentarité des informations disponibles, elle permet d'accéder à une information plus fiable et donc d'améliorer la prise de décision. La fusion permet également la prise en compte d'informations hétérogènes (numériques ou symboliques) imparfaites (imprécises, incertaines et incomplètes) modélisées sous forme de sources qu'il s'agit de combiner, agréger, fusionner. L'intérêt de la théorie des fonctions de croyance pour la fusion d'informations n'est plus à démontrer [9, 131] et nous avons tenu à la différencier de la problématique de la reconnaissance des formes même si certains modèles (notamment ceux d'A. Appriou) sont utilisés dans ce cadre. Chaque source d'information est codée sous forme d'une fonction de croyance et peut provenir d'origines très diverses comme un robot ou un capteur⁵. Evidemment, selon la nature et la provenance des informations, les modèles à mettre en œuvre peuvent être de nature très différentes. Ainsi, les modèles qui se basent sur les fonctions de vraisemblance sont largement utilisés. Une application de ces modèles a été réalisée dans [58] où les auteurs ajustent les coefficients d'affaiblissement en y ajoutant des informations contextuelles relatives à l'application. Une méthode légèrement différente pour l'optimisation des coefficients d'affaiblissement est présentée dans [140]. Elle se base sur la minimisation du critère défini à l'équation (22). L'utilisation d'autres modèles est largement étudiée notamment dans le cadre d'applications [4, 57, 28].

Il faut signaler que, dans l'étape de fusion, il peut apparaître un conflit entre les différentes sources d'information à agréger. Ce conflit peut provenir d'origines très diverses comme par exemple :

- la mesure aberrante issue d'un capteur (mauvais fonctionnement lors

5. On peut parler de "belief holder" (Ph. Smets).

de son acquisition, mauvaise évaluation de la plage de fonctionnement du capteur lors de l'apprentissage) ;

- une modélisation imprécise des fonctions de croyance (mauvais choix de distance dans les équations (19)-(20), mauvaise estimation des fonctions de vraisemblance) ;
- le fait que les sources d'information n'observent pas le même phénomène [17, 117, 11, 94] ;
- un nombre élevé de fonctions de croyance impliquées dans le processus de fusion.

Dans [88, 89], les auteurs proposent de réécrire l'étape de fusion de fonctions de croyance sous la forme d'une somme de deux termes qui fait intervenir la conjonction et une redistribution du conflit $m_{\odot}(\emptyset)$. Cette formulation permet en outre de retrouver les principales règles préalablement proposées et ainsi d'avoir une formulation unique pour l'étape d'agrégation des fonctions de croyance. Ainsi, la fusion de n fonctions de croyance définies par leurs allocations de masse $\mathbf{m} = \{m_i, i = 1, \dots, n\}$ peut se mettre sous la forme :

$$m_{\oplus}(\emptyset) \triangleq w(\emptyset, \mathbf{m}).m_{\odot}(\emptyset) \quad (24)$$

$$m_{\oplus}(A) \triangleq m_{\odot}(A) + w(A, \mathbf{m}).m_{\odot}(\emptyset) \quad \forall A \subseteq \Omega \setminus \emptyset \quad (25)$$

où $w(A, \mathbf{m})$ représentent les poids associés à chacune des propositions $A \subseteq \Omega$ et respectent la contrainte :

$$\sum_{A \subseteq \Omega} w(A, \mathbf{m}) = 1. \quad (26)$$

La notation \oplus a été utilisée de manière à mettre en évidence l'aspect conjonctif de l'opérateur et la redistribution pondérée du conflit parmi les sous-ensembles du cadre de discernement. Pour définir les poids $w(.,.)$, une première solution consiste à redistribuer le conflit de manière proportionnelle parmi les sous-ensembles du cadre de discernement [118]. Une autre solution a été introduite par Ph. Smets [122] et consiste à laisser le conflit tel quel et donc de considérer que le cadre de discernement doit être étoffé d'une nouvelle hypothèse qui couvre tous les états possibles du monde. Ainsi, \emptyset peut être interprété comme cette hypothèse dans un monde ouvert. Une approche similaire a été proposée par R.R. Yager dans [152]. D'autres solutions ont été introduites en cas de conflit [98, 19, 73] mais notons que l'opérateur \oplus proposé dans [88, 89] permet de retrouver l'ensemble de ces opérateurs. Un article de synthèse proposant une comparaison entre plusieurs stratégies est proposée dans [74].

3.3 Gains attendus et étapes de la mise en œuvre

Dans les différents problèmes abordés, l'un des principaux avantages des fonctions de croyance provient du fait que la connaissance est modélisée le plus fidèlement possible. En reconnaissance des formes, un exemple non étiqueté dans l'ensemble d'apprentissage est codé par la fonction de croyance vide qui formalise l'ignorance totale quant à l'appartenance de cet exemple dans Ω . Cette modélisation n'introduit pas une information précise comme l'équi-probabilité en théorie des fonctions de probabilité. Pour la fusion d'informations, l'avantage provient essentiellement de l'étape d'agrégation qui permet généralement (suivant l'opérateur choisi) de diminuer l'incertitude sur une prédiction en utilisant la redondance et la complémentarité des informations. Dans cet article, nous nous intéressons au cadre théorique des fonctions de croyance mais il est à noter que certains formalismes [126] permettant de modéliser imprécision et incertitude peuvent être plus adéquats à résoudre certains problèmes particuliers⁶.

La mise en œuvre de la théorie pour une application spécifique en traitement de l'information requiert plusieurs étapes de modélisation. La première consiste à définir le cadre de discernement qui dépend de l'application envisagée. Celui-ci peut correspondre au nombre de régions à discerner pour la segmentation d'images, au nombre de classes pour les problèmes de discrimination. Cette étape est généralement établie de manière supervisée par un expert bien que certains auteurs utilisent des algorithmes de classification non supervisée. La modélisation des fonctions de masse est un problème crucial où il n'y a pas, à notre connaissance, de méthode universelle. Il s'agit de choisir avec soin les éléments focaux pour obtenir une modélisation riche et flexible capable d'appréhender une large gamme de situations. Mis à part les modèles que nous avons cité dans les paragraphes 3.1.1 à 3.1.3, on trouve essentiellement dans la littérature des méthodes heuristiques. Celles-ci calculent d'abord les masses sur les singletons puis répartissent ensuite le reste de la croyance sur certains sous-ensembles dans une démarche simplificatrice. Une phase d'apprentissage peut être nécessaire pour le choix des fonctions de masse ou l'optimisation des paramètres des modèles. Cette étape s'appuie généralement sur la minimisation d'un critère (équation (22)) à partir de données observées. Vient ensuite l'étape de fusion où l'opérateur de combinaison conjonctif reste une valeur sûre dans le cadre de la théorie des fonctions de croyance. Le choix de la règle de décision constitue bien souvent

6. For someone whose only tool is a hammer, the world looks like a nail!

la dernière étape de la mise en œuvre. Les critères les plus souvent utilisés sont les critères qui maximisent la probabilité pignistique ou la plausibilité. Selon la modélisation des fonctions de masse établie, il se peut qu'une règle soit plus adéquate qu'une autre. Dans la suite, nous proposons quelques applications des fonctions de croyance en traitement de l'information où nous tenterons de détailler ces différentes étapes.

4 DOMAINES D'APPLICATIONS

Les domaines d'application de la théorie des fonctions de croyance sont nombreux et il serait trop ambitieux d'en dresser une liste exhaustive. Nous proposons de catégoriser ces différentes applications en cinq parties qui sont l'analyse de données, le diagnostic, l'aide à la décision, la perception multi-capteur et le traitement d'images. Bien que ces domaines soient assez restrictifs, ils permettent néanmoins d'englober la plus grande majorité des applications rencontrées dans la littérature. Il est à noter qu'un guide sur l'utilisation pratique du modèle des croyances transférables a été proposé dans [130].

4.1 Analyse de données

Les travaux les plus importants sur les fonctions de croyance appliquées à l'analyse de données sont probablement ceux de T. Denœux [45, 40, 43, 158]. Ils traitent de plusieurs problématiques englobant la régression, la reconnaissance des formes mais aussi la classification. Dans le domaine de la discrimination, les modèles utilisés sont les modèles de distance présentés dans les équations (19) à (21). Ils permettent d'étendre les méthodes classiques à des problèmes d'apprentissage plus complexes comme par exemple la gestion d'étiquettes incertaines (c.f. section 3.1.4). En reconnaissance des formes, les données les plus fréquemment rencontrées sont extraites des objets eux-mêmes (*feature-based*). Une autre description concerne les données de proximité (relationnelles) qui représentent des dissimilarités entre objets. Une approche de classification à partir de ces données récapitulées dans un tableau de distances a été proposée dans [42]. La théorie des fonctions de croyance permet dans ce contexte de gérer des données bruitées, imprécises ou non-euclidiennes. Citons également une extension des fonctions de croyance au cas continu avec les travaux de Petit-Renaud en régression [106, 107].

Comme nous l'avons vu dans la section 3.1, il est possible de construire des classifieurs crédibilistes à partir d'un ensemble d'apprentissage \mathcal{L} . Une technique, appelée *bagging* (**bootstrap aggregating**) en théorie de l'apprentissage, consiste à construire B échantillons de l'ensemble d'apprentissage en tirant aléatoirement n exemples uniformément (avec remplacement) et peut être appliquée à ce type de classifieurs. Après optimisation des B classifieurs sur chaque échantillon de bootstrap, une décision peut être prise en combinant les sorties des différents classifieurs ainsi construits. L'avantage consiste à combiner au niveau de la fonction de croyance [62, 61, 63]. Toujours dans le cadre de la combinaison de classifieurs, il a été montré que l'utilisation simultanée et conjointe de plusieurs classifieurs peut améliorer nettement les performances du système de reconnaissance. Ainsi, l'idée de construire des fonctions de croyance à partir des sorties de divers classifieurs de manière à en combiner leurs décisions a été proposée par plusieurs auteurs [92, 150, 112, 100, 16, 104]. Les modèles construits sont très hétérogènes et il serait intéressant de poursuivre et d'étendre les recherches en ce sens.

Les fonctions de croyance sont parfois utilisées pour des tâches de validation de données [47]. Ainsi, lorsqu'un conflit apparaît après l'étape de combinaison et si les modèles ont été préalablement validés, c'est que la mesure qui provient d'un des capteurs est erronée. Par l'analyse du conflit, il est ainsi possible de détecter toute anomalie dans le fonctionnement des capteurs. Cette étape de validation est cruciale dans des applications de surveillance notamment dans le domaine nucléaire [47].

4.2 Diagnostic

On a souvent tendance à associer Diagnostic et Reconnaissance des Formes car les démarches associées à ces problématiques sont très similaires [54]. Ainsi, Ph. Smets a largement contribué à développer le modèle des croyances transférables pour le diagnostic [128]. Une autre approche qui privilégie une interprétation un peu différente des fonctions de croyance a été proposée par certains auteurs [85, 83]. Dans le domaine du diagnostic médical, citons les modèles développés dans [136, 137]. Ces modèles ont l'originalité d'être construits à partir de deux types de symptômes ; les symptômes plaçant plutôt pour la maladie et les signes rencontrés chez le patient en bonne santé. Ils sont basés sur l'observation des fréquences d'occurrence des symptômes qui sont considérés comme des éléments focaux de la fonction de croyance.

Ainsi, les auteurs obtiennent deux fonctions de masse pour la prise de décision. Enfin, toujours dans le domaine du diagnostic mais pour la prévision des ressources en eau, les fonctions de croyance sont utilisées pour des prises de décision en environnement incertain [91].

4.3 Aide à la décision

Enfin, nous citons la problématique de la décision multi-critères [18]. La méthode consiste à construire un corpus d'évidence pour chacun des critères à examiner. Ces fonctions de croyance sont construites en tenant compte de l'importance relative de chacun des critères. Les différents corpus d'évidence sont ensuite agrégés par la règle de combinaison de Dempster. On trouvera une comparaison de différentes méthodes de gestion de l'incertain dans le cadre du jugement de la valeur d'un candidat dans [49].

Il serait illusoire de dresser une liste exhaustive des applications des fonctions de croyance à l'analyse de données, au diagnostic et à l'aide à la décision car celles-ci sont nombreuses et variées. Le lecteur intéressé pourra encore trouver d'autres types d'application dans [135, 65, 116]. Dans la suite, nous tentons de recenser quelques approches utilisées dans le domaine de la perception multi-capteur et du traitement d'images.

4.4 Perception multi-capteur

Dès le début des années 1990, la problématique de la perception multi-capteur est abordée par le biais de la théorie des fonctions de croyance par différents auteurs [25] mais notamment par A. Appriou [7, 8]. Il s'agit dans un premier temps de mettre au point des systèmes d'identification de cibles dans le domaine militaire. Les différentes sources d'information sont modélisées, après un apprentissage préalable, via les signaux délivrés par des capteurs proprioceptifs. Les modèles utilisés sont ceux basés sur les fonctions de vraisemblance présentés dans la section 3.1 (équations (16) à (18)) auxquels sont généralement ajoutés des coefficients d'affaiblissement dont les valeurs dépendent de la fiabilité des capteurs en présence [14]. De nombreuses comparaisons entre les différentes théories des mesures de confiance ont été réalisées dans le cadre de détections de cibles aériennes [30] ou de reconnaissance d'objets de manière générale [26]. Elles montrent très souvent que les fonctions de croyance permettent de quantifier l'incertitude et d'améliorer généralement les performances du système de détection, mais

aussi de répondre au problème du suivi des cibles [10].

Une autre problématique concerne l'association de données pour la détection de cibles [11]. Dans le cadre du modèle des croyances transférables, cette problématique peut être résolue en analysant le conflit (c'est-à-dire la masse assignée à l'ensemble vide). Une procédure qui tente de minimiser le conflit permet d'associer les cibles et les mesures provenant des différents capteurs. Citons dans le même esprit les applications décrites dans [117, 94, 95, 97, 96].

La navigation des robots mobiles en environnement extérieur est un problème délicat. En effet, le robot en déplacement est constamment dans un environnement inconnu. A partir d'un capteur image couleur et d'une carte de distance, les fonctions de croyance sont utilisées pour le discernement des régions libres dans l'image et des régions occupées par des obstacles [148]. La technique est utilisée pour la construction d'une grille 2D de l'environnement du robot. Toujours dans le domaine de la robotique mobile, plusieurs systèmes de détection d'obstacles à l'avant des véhicules intelligents ont été proposés [66, 115]. Les véhicules sont désormais dotés d'un système de positionnement géographique (GPS) qui leur permet d'avoir une bonne estimation de leur position relative. L'idée d'améliorer cette précision et de positionner le véhicule à partir d'une cartographie a été introduite dans [99]. Là encore, les fonctions de croyance permettent de quantifier l'incertitude relative aux informations imparfaites fournies par le GPS ou l'atlas cartographique.

4.5 Traitement d'images

Dans le domaine du traitement d'images, la problématique s'apparente plutôt à un problème de fusion d'informations [24]. Les techniques d'acquisition d'images utilisent désormais plusieurs bandes du spectre électromagnétique de l'infra-rouge jusqu'aux rayons X, mais aussi différentes approches physiques ; de la résonance magnétique aux ultrasons : on parle ainsi d'imagerie multimodale. Dans ce contexte, il s'agit alors de fusionner les images acquises sous les différentes modalités (voir par exemple [59] et [48]). Le but de ce traitement est généralement la mise en évidence de régions d'intérêt dans l'image. Il s'agit alors d'identifier les régions semblables au sens d'un certain critère (homogénéité, texture, ...).

Dans le cas particulier de l'imagerie couleur, on distingue trois acquisitions dans la bande spectrale du visible. Une image couleur possède ainsi

trois composantes Rouge, Vert et Bleu qu'il s'agit de fusionner [90]. Une autre approche consiste à considérer chaque voisin du pixel en cours de traitement comme une source d'information qui permet de quantifier l'appartenance du pixel à une classe [144]. A partir d'un voisinage autour de chaque pixel est extrait un certain nombre de fonctions de croyance qu'il suffit d'agréger pour avoir une opinion pondérée sur l'appartenance du pixel à chaque région en présence dans l'image. Ce genre de technique a été appliqué dans le domaine de l'imagerie médicale [22, 64] ou en segmentation non supervisée [114]. Les nombreux travaux qui utilisent les fonctions de croyance dans le domaine de l'imagerie satellitaire attestent qu'il s'agit d'un problème délicat. Les informations disponibles proviennent de capteurs situés à de grandes distances des objets et sont donc par nature imprécises et incertaines [46, 27, 108, 147, 110]. Dans [70], l'introduction d'informations spatiales permet de quantifier de façon plus fine l'ignorance quant à la présence de nuages dans les images aériennes et les imprécisions dues au bruit dans les images radar. Une fonction de masse est définie pour chaque pixel proportionnellement au nombre de pixels voisins possédant la même étiquette permettant ainsi de fiabiliser la décision. Une étude qui utilise les modèles consonants proposés initialement par G. Shafer est présentée dans [78] dans le cadre de fusion d'images multispectrales. Dans [139], une méthode pour l'interprétation d'images radar a pour objectif de fusionner les résultats d'opérateurs de traitement d'images afin d'obtenir une cartographie de l'image en classes sémantiques. La fusion se situe au niveau des régions et non pas des pixels en introduisant une classe de rejet. Chaque opérateur étant dédié à la détection d'une classe, les éléments focaux sont le singleton correspondant à la classe et son complémentaire. Les fonctions de masse choisies de façon heuristique linéaires par morceaux sont déduites de l'observation des fréquences normalisées sur des échantillons d'apprentissage. Enfin, des comparaisons entre le cadre probabiliste et les fonctions de croyance appliquées à l'interprétation d'images satellite ont été menées dans [87, 31].

Dans le domaine de la vision par ordinateur, une problématique assez courante concerne la perception de l'environnement (d'un robot mobile par exemple) à l'aide de capteurs image. La tâche consiste à segmenter les images fournies généralement par des capteurs complémentaires (réflectance 2D, profondeur, 3D). Une telle approche utilisant les fonctions de croyance a été proposée dans [93]. On se retrouve confronté à des problèmes de cadre de discernement différents mais compatibles [72]. La relaxation, un processus itératif qui se base sur des décisions locales, qui est généralement utilisé

pour la mise en correspondance d'un motif (pattern) à un ensemble de données. Une extension d'un algorithme de relaxation utilisant les fonctions de croyance a été menée dans [34]. Cette application permet notamment de faire évoluer la confiance accordée à une décision locale entre deux itérations du processus de relaxation tout en quantifiant son incertitude. De part leur nature, les méthodes de relaxation trouvent beaucoup d'applications en traitement d'images : segmentation, détection de contours, appariement d'images.

Une fois encore, il ne s'agit pas de donner de manière exhaustive toutes les applications des fonctions de croyance au traitement d'images. Cependant, les problèmes rencontrés dans ce cadre applicatif sont nombreux (choix et utilisation des modèles, temps de calcul, stockage mémoire) et constituent à eux seuls des pistes de recherche importantes [143].

5 CONCLUSION

Dans cet article, un état de l'art sur l'application des fonctions de croyance à la gestion d'informations imparfaites a été présenté. Cette présentation a volontairement été limitée au modèle des croyances transférables proposé par Ph. Smets. Les principaux outils associés à ce modèle ont été brièvement présentés. Les problématiques de la Reconnaissance des Formes ainsi que de la fusion d'informations ont été décrites via notamment les grandes familles de modèles qui cohabitent. Enfin, en brossant un état de l'art, il a été montré que ce cadre théorique est suffisamment souple pour être appliqué dans de nombreux domaines applicatifs des Sciences et Technologies de l'Information et de la Communication tels que l'analyse de données, le diagnostic, l'aide à la décision, la perception multicapteurs et le traitement d'images.

Il a été souvent reproché à la théorie des fonctions de croyance d'engendrer des temps de calcul particulièrement importants lorsque la cardinalité du cadre de discernement devient grande. Certains travaux pourtant tentent de simplifier les fonctions en présence [67], soit lorsqu'elles ont des propriétés remarquables [13], soit en éliminant les éléments focaux dont la masse est faible [15] soit en réalisant une pré-classification des éléments focaux qui peuvent être regroupés [39, 44]. De nombreux efforts restent à faire dans ce domaine et constituent de ce fait une des perspectives de la théorie des fonctions de croyance telle qui avait déjà été mentionné dans [155]. Bien que beaucoup de travaux aient déjà été entrepris dans le domaine, une autre perspective pourrait consister à associer un modèle de représentation

de la connaissance (par exemple un modèle graphique) avec la souplesse des fonctions de croyance pour la gestion de l'incertitude (un cas particulier des travaux de P.P. Shenoy [119, 120]). Cette idée avait été proposée dans [149] mais les problèmes liés à la mise en œuvre d'algorithmes simples et rapides sont si compliqués que les applications de cette méthodologie sont peu développées [111]. Enfin, en ce qui concerne les applications de la théorie des fonctions de croyance, il semble que les perspectives d'utilisation soient nombreuses notamment dans le domaine de la recherche d'informations [109], de l'indexation de documents [86] ou de la fouille de données [6].

6 REMERCIEMENTS

Ce travail a été réalisé dans le cadre d'une délégation de deux ans au sein du laboratoire HEUDIASYC (UMR 6599 CNRS) effectuée à partir de Septembre 2000. Ce projet a été financé par le CNRS. L'auteur souhaite particulièrement remercier T. Denœux pour ses conseils, sa disponibilité, son efficacité et sa rigueur ainsi que l'ensemble des membres du laboratoire. Enfin, ce travail n'aurait pas pu aboutir sans E. Lefevre et O. Colot.

RÉFÉRENCES

- [1] Numéro spécial: Fusion de données. *Revue Traitement du Signal*, 11(6), 1994.
- [2] Numéro spécial: Fusion de données. *Revue Traitement du Signal*, 14(5), 1997.
- [3] J. Abellan et S. Moral. Completing a total uncertainty measure in Dempster-Shafer theory. *International Journal of General Systems*, 28:299–314, 1999.
- [4] M.A. Abidi et R.C. Gonzalez. *Data Fusion in Robotics and Machine Intelligence*, chapitre 4 : Multisensor strategies using Dempster-Shafer belief accumulation, pages 165–210. Academic Press, INC, 1992.
- [5] C. Ambroise, T. Denœux, G. Govaert et Ph. Smets. Learning from an imprecise teacher: Probabilistic and evidential approaches. In *Proceedings of ASMDA'2001*, pages 100–105, Compiègne, France, 2001.

- [6] S.S. Anand, D.A. Bell et J.G. Hughes. EDM: A general framework for data mining based on evidence theory. *Data and Knowledge Engineering*, 18:189–223, 1996.
- [7] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, 11:27–40, 1991.
- [8] A. Appriou. Formulation et traitement de l'incertain à l'analyse multi-senseurs. In *14ème Colloque GRETSI*, pages 951–954, 1993.
- [9] A. Appriou. *Aggregation and Fusion of Imperfect Information*, chapitre Uncertain Data Aggregation in Classification and Tracking Processes, pages 231–260. Physica-Verlag, 1998.
- [10] A. Appriou. Multisensor signal processing in the framework of the theory of evidence. In *Application of Mathematical Signal Processing Techniques to Mission Systems*, pages (5–1)(5–31). Research and Technology Organization (Lecture Series 216), November 1999.
- [11] A. Ayoun et Ph. Smets. Data association in multi-target detection using the transferable belief model. *International Journal of Intelligent Systems*, 16:1167–1182, 2001.
- [12] J.F. Baldwin, J. Lawry et T.P. Martin. Mass assignment based induction of decision trees on words. In *Proceedings of IPMU'1998*, pages 524–531, Paris, France, 1998.
- [13] J.A. Barnett. Calculating Dempster-Shafer plausibility. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):599–602, 1991.
- [14] A. Bastière. Methods for multisensor classification of airborne targets integrating evidence theory. *Aerospace Science and Technology*, 2(6):401–411, 1998.
- [15] M. Bauer. Approximation algorithms and decision making in the Dempster-Shafer theory of evidence - an empirical study. *International Journal of Approximate Reasoning*, 17:217–237, 1997.
- [16] F. Belloir. *Système intelligent de reconnaissance de formes développé autour d'un capteur à courants de Foucault*. PhD thesis, Université de Reims Champagne-Ardenne, 1999.
- [17] M. Bengtsson et J. Schubert. Dempster-Shafer clustering using Potts spin mean field theory. *Soft Computing*, 5(3):215–228, 2001.
- [18] M. Beynon, B. Curry et P. Morgan. The Dempster-Shafer theory of evidence: An alternative approach to multicriteria decision model-

- ling. *Omega, International Journal of Management Science*, 28:37–50, 2000.
- [19] P. Bhattacharya. On the Dempster-Shafer evidence theory and non-hierarchical aggregation of belief structures. *IEEE Transactions on Systems, Man and Cybernetics Part A*, 30(5):526–536, 2000.
- [20] M. Skarstein Bjanger. Induction of decision trees from partially classified data using belief functions. Master's thesis, Norwegian Univ. of Science and Technology, Dpt of Computer and Information Science, Feb. 2000. Available at <http://www.hds.utc.fr/~tdenoeux>.
- [21] I. Bloch. Incertitude, imprécision et additivité en fusion de données : point de vue historique. *Revue Traitement du Signal*, 13(4):267–288, 1996.
- [22] I. Bloch. Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern recognition Letters*, 17:905–919, 1996.
- [23] I. Bloch. *Fusion d'informations en traitement du signal et des images*. Traité IC2, Série Traitement du signal et de l'image. Editions Hermès, Paris, 2003.
- [24] I. Bloch et H. Maître. Fusion de données en traitement d'images: Modèles d'informations et décisions. *Traitement du Signal*, Numéro Spécial: Fusion de données(11):435–446, 1994.
- [25] P. L. Bogler. Shafer-Dempster reasoning with applications to multi-sensor target identification systems. *IEEE Transactions on Systems, Man and Cybernetics*, 17(6):968–977, 1987.
- [26] H. Borotschnig, L. Paletta, M. Prantl et A. Pinz. A comparison of probabilistic, possibilistic and evidence theoretic fusion schemes for active object recognition. *Computing*, 62:293–319, 1999.
- [27] H. Bracker. *Utilisation de la théorie de Dempster-Shafer pour la classification d'images satellitaires à l'aide de données multi-sources et multi-temporelles*. PhD thesis, Université de Rennes I, 1996.
- [28] J.J. Braun. Dempster-Shafer theory and bayesian reasoning in multisensor data fusion. In Belur V. Dasarathy, éditeur, *Sensor Fusion: Architectures, Algorithms and Applications IV*, volume 4051, pages 255–266. SPIE, 2000.
- [29] L. A. Breslow et D. W. Aha. Simplifying decision trees: A survey. *Knowledge Engineering Review*, 12(1):1–40, 1997.

- [30] D.M. Buede et P. Girardi. A target identification comparison of bayesian and Dempster-Shafer multisensor fusion. *IEEE Transactions on Systems, Man and Cybernetics Part A*, 27(5):569–577, 1997.
- [31] S. Chauvin. *Evaluation des théories de la décision appliquées à la fusion de capteurs en imagerie satellitaire*. PhD thesis, Université de Nantes, 1995.
- [32] G. Choquet. Théorie des capacités. *Annales de l'Institut Fourier*, 5:131–295, 1954.
- [33] G. Cooper et E. Herskovits. A bayesian method for constructing bayesian belief networks from databases. In *Proceedings of Seventh Conference on Uncertainty in Artificial Intelligence*, pages 86–94, Los Angeles, 1991. Morgan Kaufman.
- [34] P. Cucka et A. Rosenfeld. Evidence-based pattern-matching relaxation. *Pattern Recognition*, 26(9):1417–1427, 1993.
- [35] A. Dempster. Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics*, AMS-38:325–339, 1967.
- [36] T. Denœux. A k-nearest neighbour classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems Man and Cybernetics*, 25(5):804–813, 1995.
- [37] T. Denœux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7):1095–1107, 1997.
- [38] T. Denœux. Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning*, 20:79–111, 1999.
- [39] T. Denœux. Inner and outer clustering approximations of belief structures. In *Proceedings of IPMU'2000*, volume I, pages 125–132, Madrid, Spain, July 2000.
- [40] T. Denœux. 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, 2000.
- [41] T. Denœux et M. Skarstein Bjanger. Induction of decision trees from partially classified data using belief functions. In *Proceedings of SMC'2000*, pages 2923–2928, Nashville, USA, 2000. IEEE.
- [42] T. Denœux et M. Masson. Clustering of proximity data using belief functions. In *Proceedings of IPMU'2002*, Annecy, France, 2002.
- [43] T. Denœux, M. Masson et B. Dubuisson. Advanced pattern recognition techniques for system monitoring and diagnosis : A survey. *Journal européen des systèmes automatisés*, 31(9-10):1509–1539, 1998.

- [44] T. Denœux et A. Ben Yaghlane. Approximating the combination of belief functions using the fast Möbius transform in a coarsened frame. *International Journal of Approximate Reasoning*, 31(1-2):77–101, October 2002.
- [45] T. Denœux et L.M. Zouhal. Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3):47–62, 2001.
- [46] J. Desachy, L. Roux et E. Zahzah. Numeric and symbolic data fusion: A soft computing approach to remote sensing images analysis. *Pattern Recognition Letters*, 17:1361–1378, 1996.
- [47] A. F. Dragoni et P. Giorgini. Sensor data validation for nuclear power plants through Bayesian conditioning and Dempster’s rule of combination. *Computers and Artificial Intelligence*, 17:151–168, 1998.
- [48] A. Dromigny-Badin, S. Rossato et Y.M. Zhu. Fusion de données radioscopiques et ultrasonores via la théorie de l’évidence. *Traitement du Signal*, 14(5):499–510, 1997.
- [49] D. Dubois, M. Grabisch, H. Prade et Ph. Smets. Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: The assessment of the value of a candidate. *International Journal of Intelligent Systems*, 16:1245–1272, 2001.
- [50] D. Dubois et H. Prade. A note on measures of specificity for fuzzy sets. *International Journal of General Systems*, 10:279–283, 1985.
- [51] D. Dubois et H. Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy Sets and Systems*, 124:161–182, 1987.
- [52] D. Dubois et H. Prade. La problématique scientifique du traitement de l’information. *Information-Interaction-Intelligence*, 1(2), 2001.
- [53] D. Dubois, H. Prade et Ph. Smets. Representing partial ignorance. *IEEE Systems, Machine and Cybernetics*, 26:361–377, 1996.
- [54] B. Dubuisson. *Diagnostic et Reconnaissances de Formes*. Hermes, 1990.
- [55] Z. Elouedi, K. Mellouli et Ph. Smets. Belief decision trees: Theoretical foundations. *International Journal of Approximate Reasoning*, 28:91–124, 2001.
- [56] Z. Elouedi, K. Mellouli et Ph. Smets. A pre-pruning method in belief decision trees. In *Proceedings of IPMU’2002*, pages 579–586, Annecy, France, July 2002.

- [57] A.M. Erkmen et H.E. Stephanou. Information fractals for evidential pattern classification. *IEEE Transactions on Systems, Man and Cybernetics*, 20(5):1103–1114, September/October 1990.
- [58] S. Fabre, A. Appriou et X. Briottet. Presentation and description of two classification methods using data fusion based on sensor management. *Information Fusion*, 2:49–71, 2001.
- [59] S. Fabre, A. Appriou, X. Briottet et P. Marthon. Pixel fusion : contribution of contextual physical data for the a priori database construction. In *Proceedings of PSIP'99*, Versailles, France, 1999.
- [60] U.M. Fayyad et P. Smyth. Image database exploration: Progress and challenge. In *Proceedings of the 1993 AAAI Workshop on Knowledge Discovery in Database*, pages 14–27, Washington, DC, USA, 1993.
- [61] J. François. *La fusion des connaissances expérimentales et expertes pour le diagnostic. Application à un procédé agro-alimentaire*. PhD thesis, Université de Technologie de Compiègne, 2000.
- [62] J. Francois, Y. Grandvalet, T. Denœux et J.-M. Roger. Bagging belief structures in Dempster-Shafer K-NN rule. In *Proceedings of IP-MU'2000*, pages 111–118, Madrid, Spain, July 2000.
- [63] J. François, Y. Grandvalet, T. Denœux et J.-M. Roger. Resample and combine: An approach to improving uncertainty representation in evidential pattern classification. *Information Fusion*, 4(2):75–85, June 2003.
- [64] L. Gautier. *Validation de la segmentation d'images par la théorie des croyances. Application aux séquences d'images IRM du rachis lombaire*. PhD thesis, Université du Littoral Côte -d'Opale, 2000.
- [65] P.R. Gillett. Monetary unit sampling : A belief-function implementation for audit and accounting applications. *International Journal of Approximate Reasoning*, 25:43–70, 2000.
- [66] D. Gruyer. *Etude du traitement de données imparfaites pour le suivi multi-objets : Application aux situations routières*. PhD thesis, Université de Technologie de Compiègne, 1999.
- [67] D. Harmanec. Faithful approximations of belief functions. In Kathryn Blackmond Laskey et Henri Prade, éditeurs, *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, Stockholm, Sweden, 1999. Morgan Kaufmann Publishers.
- [68] R.V.L. Hartley. Transmission of information. *The Bell System Technical Journal*, 7(3):535–563, 1928.

- [69] D. Heckerman, D. Geiger et D. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197–243, 1995.
- [70] S. Le Hégarat-Masclé, I. Bloch et D. Vidal-Madjar. Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover. *Pattern recognition*, 31(11):1811–1823, November 1998.
- [71] T. Horiuchi. Decision rule for pattern classification by integrating interval feature values. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):440–448, April 1998.
- [72] F. Janez et A. Appriou. Theory of evidence and non-exhaustive frames of discernment : Plausibilities correction methods. *International Journal of Approximate Reasoning*, 18:1–19, 1998.
- [73] A. Josang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 9(3):279–311, 2001.
- [74] A. Jøsang, M. Daniel et P. Vannoorenberghe. Strategies for combining conflicting dogmatic beliefs. In *Sixth International Conference on Information Fusion, FUSION'2003*, pages 1133–1140, Cairns, Australia, July 2003.
- [75] R. Kennes. Computational aspects of the Möbius transform of a graph. *IEEE Transactions on Systems, Man and Cybernetics*, 22:201–223, 1992.
- [76] R. Kennes et Ph. Smets. Fast algorithms for Dempster-Shafer theory. In B. Bouchon-Meunier, R.R. Yager et L. A. Zadeh, éditeurs, *Uncertainty in knowledge bases*, pages 14–23. Springer Verlag, Berlin, 1990.
- [77] R. Kennes et Ph. Smets. Computational aspects of the Möbius transform. In P.P. Bonissone, M. Henrion, L.N. Kanal et J.F. Lemmer, éditeurs, *Uncertainty in Artificial Intelligence 6*, pages 401–416. North Holland, Amsterdam, 1991.
- [78] H. Kim et P.H. Swain. Evidential reasoning approach to multisource-data classification in remote sensing. *IEEE Transactions on Systems, Man and Cybernetics*, 25(8):1257–1265, 1995.
- [79] F. Klawonn et E. Schwecke. On the axiomatic justification of Dempster's rule combination. *International Journal of Intelligent Systems*, 7:469–478, 1992.

- [80] G.J. Klir. *Measures of uncertainty in the Dempster-Shafer theory of evidence*, pages 35–49. Advances in the Dempster-Shafer Theory of Evidence. John Wiley and Sons, New York, R.R. Yager and M. Fedrizzi and J. Kacprzyk edition, 1994.
- [81] G.J. Klir et R.M. Smith. Recent developments in generalized information theory. *International Journal of Fuzzy Systems*, 1(1), 1999.
- [82] G.J. Klir et M.J. Wierman. *Uncertainty-Based Information*. Physica-Verlag, Heidelberg, Germany, 1998.
- [83] J. Kohlas, B. Anrig, R. Haenni et P.A. Monney. Model-based diagnostics and probabilistic assumption-based reasoning. *Artificial Intelligence*, 104:71–106, 1998.
- [84] J. Kohlas et P. A. Monney. *A Mathematical Theory of Hints: An Approach to Dempster-Shafer Theory of Evidence*. Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag, 1995.
- [85] J. Kohlas, P.A. Monney, R. Haenni et N. Lehmann. Model-based diagnostics using hints. In Ch. Fridevaux et J. Kohlas, éditeurs, *Symbolic and Quantitative Approaches to Uncertainty, European Conference ECSQARU'95, Fribourg*, pages 259–266. Springer, 1995.
- [86] M. Lalmas et I. Ruthven. Representing and retrieving structured documents using Dempster-Shafer theory of evidence: Modelling and evaluation. *Journal of documentation*, 54(5):529–565, 1998.
- [87] T. Lee, J.A. Richards et P.H. Swain. Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience Remote Sensing*, 25(3):283–293, 1987.
- [88] E. Lefevre, O. Colot et P. Vannoorenberghe. Belief functions combination and conflict management. *Information Fusion*, 3(2):149–162, June 2002.
- [89] E. Lefevre, O. Colot et P. Vannoorenberghe. Reply to the comments of R. Haenni on the paper : Belief functions combination and conflict management. *Information Fusion*, 4(1):63–65, 2003.
- [90] E. Lefevre, P. Vannoorenberghe et O. Colot. About the use of Dempster-Shafer theory for color image segmentation. In *CGIP'2000, First International Conference on Color in Graphics and Image Processing*, pages 164–169, October 1-4, Saint-Etienne, France, September 2000.

- [91] W. B. Luo et B. Caselton. Using Dempster-Shafer theory to represent climate change. *Journal of Environmental Management*, 49:73–93, 1997.
- [92] E. Mandler et J. Schurman. Combining the classification results of independent classifiers based on Dempster-Shafer theory of evidence. *International Journal of Pattern Recognition and Artificial Intelligence*, pages 381–393, 1988.
- [93] S. Mathevet, L. Trassoudaine, P. Checchin et J. Alizon. Application de la théorie de l'évidence à la combinaison de segmentations en régions. In *Dix-Septième Colloque GRETSI (Vannes)*, pages 635–638, September 1999.
- [94] N. Milisavljevic. *Analysis and Fusion Using Belief Functions Theory of Multisensor Data for Close-Range Humanitarian Mine Detection*. PhD thesis, Ecole Normale Supérieure de Télécommunications, Paris, France, 2001.
- [95] N. Milisavljevic, I. Bloch et M. Acheroy. Characterization of mine detection sensors in terms of belief function and their fusion. In *Proc. 3rd Int. Conf. on Information Fusion (FUSION 2000)*, pages ThC3.15–ThC3.22, 2000.
- [96] N. Milisavljevic, I. Bloch et M. Acheroy. Modeling, combining and discounting mine detection sensors within Dempster-Shafer framework. In *Detection Technologies for Mines and Minelike Targets*, volume 4038, pages 1461–1472. SPIE Press, Orlando, USA, 2000.
- [97] N. Milisavljevic, S. van den Broek, I. Bloch, P. B. W. Schwering, H. A. Lensen et M. Acheroy. Comparison of belief function and voting method for fusion of mine detection sensors. In *Detection and Remediation Technologies for Mines and Minelike Targets VI*, volume 4394. SPIE Press, Orlando, USA, 2001.
- [98] C. K. Murphy. Combining belief functions when evidence conflicts. *Decision Support Systems*, 29:1–9, 2000.
- [99] Maan E. El Najjar et P. Bonnifait. A road reduction method using multi-criteria fusion. In *Proceedings of IEEE Intelligent Vehicle Symposium*, Versailles, France, 2002.
- [100] G.S. Ng et H. Singh. Data equalisation with evidence combination for pattern recognition. *Pattern Recognition Letters*, 19:227–235, 1998.

- [101] N.R. Pal, J.C. Bezdek et R. Hemasinha. Uncertainty measures for evidential reasoning i: A review. *International Journal of Approximate Reasoning*, 7:165–183, 1992.
- [102] N.R. Pal, J.C. Bezdek et R. Hemasinha. Uncertainty measures for evidential reasoning ii: A new measure of total uncertainty. *International Journal of Approximate Reasoning*, 8:1–16, 1993.
- [103] N.R. Pal et S. Ghosh. Some classification algorithms integrating Dempster-Shafer theory of evidence with the rank nearest neighbor rules. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Applications*, 31(1):59–66, 2001.
- [104] C.R. Parikh, M.J. Pont et N.B. Jones. Application of Dempster-Shafer theory in condition monitoring applications: A case study. *Pattern recognition letters*, 22:777–785, 2001.
- [105] S. Parsons. Current approaches to handling imperfect information in data and knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 8(3):353–372, 1996.
- [106] S. Petit-Renaud. *Application des systèmes flous et de la théorie des croyances à l'estimation fonctionnelle en présence d'informations incertaines ou imprécises*. PhD thesis, Université de Technologie de Compiègne, 1999.
- [107] S. Petit-Renaud et T. Denœux. Application de la théorie des fonctions de croyance en régression. In *Rencontres francophones sur la Logique Floue et ses Applications*, pages 169–176, Valenciennes, 1999. Cépadués.
- [108] M. Petrou et A. Stassopoulou. Advanced techniques for fusion of information in remote sensing : An overview. In *EUROPTO Conference on Image and Signal Processing for Remote Sensing*, volume 3870, pages 264–275, Florence, Italy, September 1999. SPIE.
- [109] J. Picard et R. Haenni. Modeling information retrieval with probabilistic argumentation systems. In *Proceedings of the 20th BCS-IRSG Annual Colloquium*, 1998.
- [110] L. Pigeon, B. Solaiman, T. Toutin et K.P.B. Thomson. Dempster-Shafer theory for multi-satellites remotely-sensed observations. In Belur V. Dasarathy, éditeur, *Sensor Fusion : Architectures, Algorithms and Applications IV*, volume 4051, pages 228–236. SPIE, 2000.
- [111] S. Populaire et T. Denœux. Combining expert knowledge with data based on belief function theory: an application in waste water treat-

- ment. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, Hammamet, Tunisie, October 6-9 2002.
- [112] G. Rogova. Combining the results of several neural networks classifiers. *Neural Networks*, 7(5):777–781, 1994.
- [113] M. Rombaut et V. Berge-Cherfaoui. Decision making in data fusion using Dempster-Shafer's theory. In *3th IFAC Symposium on Intelligent Components and Instrumentation for Control Applications*, Annecy, France, 9-11 Juin 1997.
- [114] M. Rombaut et Y. M. Zhu. Study of Dempster-Shafer theory for image segmentation applications. *Image and Vision Computing*, 20:15–23, 2002.
- [115] C. Royère. *Théorie de l'évidence en détection d'obstacles*. PhD thesis, Université de Technologie de Compiègne, 2002.
- [116] J. Schubert. Creating prototypes for fast classification in Dempster-Shafer clustering. In *Proceedings of European Conference on Symbolic and Qualitative Approaches to Reasoning under Uncertainty 1997*, pages 525–535, 1997.
- [117] J. Schubert. Clustering belief functions based on attracting and conflicting metalevel evidence. In *Proceedings of IPMU'2002*, pages 571–578, Annecy, France, 2002.
- [118] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [119] P.P. Shenoy. *Using Dempster-Shafer's belief-function theory in expert systems*, pages 395–414. *Advances in the Dempster-Shafer Theory of Evidence*. John Wiley and Sons, New York, R.R. Yager and M. Fedrizzi and J. Kacprzyk edition, 1994.
- [120] P.P. Shenoy et G. Shafer. Propagating belief functions using local computations. *IEEE Expert*, 1:43–52, 1986.
- [121] Ph. Smets. Information content of an evidence. *International Journal of Machine Studies*, 19:33–43, 1983.
- [122] Ph. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.
- [123] Ph. Smets. Resolving misunderstandings about belief functions. *International Journal of Approximate Reasoning*, 6:321–344, 1990.
- [124] Ph. Smets. An axiomatic justification for the use of belief function to quantify beliefs. In IJCAI-93, éditeur, *International Joint Conference*

- on *Artificial Intelligence*, pages 598–603. Morgan Kaufman, San Mateo, Ca, 1993.
- [125] Ph. Smets. Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.
 - [126] Ph. Smets. Probability, possibility, belief: which for what? In G. De Cooman, Ruan Da et E.E. Kerre, éditeurs, *Foundation and applications of possibility theory*, pages 20–40. World Scientific, Singapore, 1995.
 - [127] Ph. Smets. The normative representation of quantified beliefs by belief functions. *Artificial Intelligence*, 92(1-2):229–242, 1997.
 - [128] Ph. Smets. The application of the transferable belief model to diagnostic problems. *International Journal of Intelligent Systems*, 13:127–157, 1998.
 - [129] Ph. Smets. The transferable belief model for quantified belief representation. In D. M. Gabbay et Ph. Smets, éditeurs, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 267–301. Kluwer, Dordrecht, The Netherlands, 1998.
 - [130] Ph. Smets. Practical uses of belief functions. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Stockholm, Sweden, 1999.
 - [131] Ph. Smets. Data fusion in the Transferable Belief Model. In *Proceedings of FUSION'2000*, pages PS 21–33, Paris, France, 2000.
 - [132] Ph. Smets. The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31(1-2):1–30, October 2002.
 - [133] Ph. Smets. Decision making in a context where uncertainty is represented by belief functions. In R. P. Srivastava et T.J. Mock, éditeurs, *Belief Functions in Business Decisions*, pages 17–61. Physica-Verlag, Heidelberg, Germany, 2002.
 - [134] Ph. Smets et R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66(2):191–234, 1994.
 - [135] R. P. Srivastava. Audit decisions using belief functions: a review. *Control and Cybernetics*, 26 (2):135–160, 1997.
 - [136] E. Straszcka et J. Straszcka. Dempster-Shafer theory and basic probability assignment in diagnosis support. In J. Fodor B. De Baets et L.T.Koczy, éditeurs, *Proceedings of EUROFUSE'99*, pages 395–398,

- Budapest, Hungary, 1999. University of Veterinary Science, Technical university of Budapest.
- [137] E. Straszeka et J. Straszeka. An interpretation of a medical score test as a basic probability assignment. In *Proceedings of the 7th European Congress on Intelligent Techniques on Soft Computing, EU-FIT'99*, page on CDROM, Aachen, Germany, September 1999.
 - [138] T. Strat. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, 4:391–417, 1990.
 - [139] F. Tupin. *Reconnaissance des formes et analyse de scènes en imagerie radar à ouverture synthétique*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1997.
 - [140] P. Vannoorenberghe et T. Denœux. Diagnostic de la pollution atmosphérique par une approche RDF utilisant les fonctions de croyance. In *Automatique et Environnement 2001*, St-Etienne, France, Juillet 2001.
 - [141] P. Vannoorenberghe et T. Denœux. Likelihood-based vs. distance-based evidential classifiers. In *Proceedings of FUZZ-IEEE'2001*, pages 320–323, Melbourne, Australia, December 2001. IEEE.
 - [142] P. Vannoorenberghe et T. Denœux. Handling uncertain labels in multiclass problems using belief decision trees. In *Proceedings of IP-MU'2002*, pages 1919–1926, Annecy, France, 2002.
 - [143] P. Vannoorenberghe, E. Lefevre et O. Colot. Traitement d'images et théorie des fonctions de croyance. In *Rencontres Francophones sur la Logique Floue et Ses Applications, LFA'2003*, pages 287–294. Cépaduès, 26-27 Novembre 2003.
 - [144] A. Verikas, E. Malmqvist et M. Bacauskiene. Combining neural networks, fuzzy sets and evidence theory based approaches for analysing colour images. In *Proceedings of the International Joint Conference on Neural Networks*, volume II, pages 297–302, Como, Italy, 2000.
 - [145] P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.
 - [146] P. Walley et S. Moral. Upper probabilities based only on the likelihood function. *Journal of Royal Statistical Society B*, 61(4):831–847, 1999.
 - [147] G.G. Wilkinson et J. Megier. Evidential reasoning in a pixel classification hierarchy - a potentiel method for integrating image classifiers and expert system rules based on geographic context. *International Journal of Remote Sensing*, 11(10):1963–1968, 1990.

- [148] Y.-G. Wu, J.-Y. Yang et K. Liu. Obstacle detection and environment modeling based on multisensor fusion for robot navigation. *Artificial Intelligence in Engineering*, 10:323–333, 1996.
- [149] H. Xu, Y.-T. Hsia et Ph. Smets. The transferable belief model for decision making in the valuation-based system. *IEEE Transactions on Systems, Man and Cybernetics*, 26A:698–707, 1996.
- [150] L. Xu, A. Krzyzak et C.Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435, 1992.
- [151] R.R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9:249–260, 1983.
- [152] R.R. Yager. Hedging in the combination of evidence. *Journal of Information and Optimization Sciences*, 4(1):73–81, 1983.
- [153] R.R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41:93–138, 1987.
- [154] R.R. Yager. Uncertainty representation using fuzzy measures. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 32(1):13–20, February 2002.
- [155] R.R. Yager, J. Kacprzyk et M. Fedrizzi, éditeurs. *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, New York, 1994.
- [156] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.
- [157] L. Zadeh. A simple view of the Dempster-Shafer framework and its implication for the rule of combination. *AI Magazine*, 7:85–90, 1986.
- [158] L.M. Zouhal et T. Denceux. An evidence-theoretic k-nn rule with parameter optimization. *IEEE Transactions on Systems, Man and Cybernetics-Part C*, 28(2):263–271, May 1998.