N° : 01 ISAM 0007

Fusion adaptée d'informations conflictuelles dans le cadre de la théorie de l'évidence

Application au diagnostic médical

Thèse de doctorat
présentée en vue de l'obtention du titre de
Docteur de l'Institut National des Sciences Appliquées de Rouen
Discipline : Physique

présentée par

Eric LEFEVRE

Soutenue le 22 Novembre 2001 devant le jury composé de :

T. DENOEUX	Professeur à l'Université de Compiègne, Rapporteur
P. SMETS	Professeur à l'Université Libre de Bruxelles, Rapporteur
A. APPRIOU	Directeur de recherche à l'ONERA, Examinateur
B. DUBUISSON	Professeur à l'Université de Compiègne, Président du jury
P. VANNOORENBERGHE	Maître de conférences à l'Université de Rouen, Examinateur
D. de BRUCQ	Professeur à l'Université de Rouen, Directeur de thèse
O. COLOT	HDR à l'Université de Rouen, co-Directeur de thèse

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE ROUEN
LABORATOIRE PERCEPTION SYSTÈMES INFORMATION

Remerciements

J'exprime ma gratitude à Monsieur Thierry Denœux, Professeur à l'Université de Technologie de Compiègne (UTC) d'avoir accepté de rapporter ma thèse. Je le remercie également pour les conseils pertinents qu'il m'a prodigué.

Pour les mêmes raisons, je remercie vivement Monsieur Philippe Smets, Professeur à l'Université Libre de Bruxelles, pour son intérêt quant à mes travaux.

Je remercie beaucoup Monsieur Bernard Dubuisson, Professeur à l'Université de Technologie de Compiègne, de m'avoir fait l'honneur de présider le jury. Je remercie également Monsieur Alain Appriou, Directeur de Recherche à l'Office National d'Etudes et de Recherches en Aérospatiales (ONERA) d'avoir bien voulu participer à ce jury.

Je souhaite exprimer ma reconnaissance à Denis de Brucq, Professeur à l'Université de Rouen, pour m'avoir accepté dans son équipe et avoir assumé la direction de ma thèse.

Un grand merci également à Olivier Colot, Maître de Conférences (HDR) à l'Université de Rouen, de m'avoir si bien encadré et supporté tout au long de cette thèse.

Merci à Patrick Vannoorenberghe, Maître de Conférences à l'Université de Rouen, qui a pensé à moi, après mon année sabbatique sous les drapeaux, pour démarrer cette thèse et m'a épaulé pendant les différentes étapes de celle-ci.

Je n'oublie pas tous ceux du laboratoire Perception Systèmes et Information qui ont rendu ces années de thèse agréables et sympathiques.

Enfin, je remercie ma famille et mes amis qui m'ont soutenu au long de ces années, qu'ils sachent que je les aime et que je leur dois aussi beaucoup.

Table des matières

Liste des figures

Notations

Reconnaissance des formes

$\mathcal{X}$: espace des caractéristiques

L : nombre de caractéristiques

l : indice des caractéristiques

$\mathcal{T}$: base de test

Θ : cadre de discernement

N : nombre d'hypothèses

H_n : hypothèse singleton

$\mathcal{L}$: base d'apprentissage

I : nombre de vecteur d'apprentissage

$x^{(i)}$: $i^{\text{ème}}$ vecteur d'apprentissage

u : vecteur d'étiquettes

q_k^i : nombre d'experts classant l'exemple i dans la classe k

q : nombre d'experts

u_k^i : degré de possibilité de l'exemple i d'appartenir à la classe k

Théorie de l'évidence

A : proposition

m : fonction de croyance

$\mathcal{F}(m)$: ensemble des éléments focaux de m

Bel : fonction de crédibilité

Pl : fonction de plausibilité

Q : fonction de communalité

S_j : $j^{\text{ème}}$ source d'information

m_j : fonction de croyance issue de la source S_j

$m_\cap$: fonction de croyance résultante de la combinaison conjonctive

$m_\oplus$: fonction de croyance résultante de la combinaison de Dempster

K : masse de croyance conflictuelle

α_j : coefficient d'affaiblissement de la source S_j

$m_{\alpha_j,j}$: jeu de masses de la source d'information S_j affaibli par α_j

$BetP$: probabilité pignistique

Autres mesures floues

$f(x|H_n)$: densité de probabilité conditionnelle

$f(x)$: densité de probabilité non conditionnelle (densité mélange)

$P(H_n|x)$: probabilité *a posteriori*

$P(H_n)$: probabilité *a priori*

$L(H_n|x)$: fonction de vraisemblance

$\mathcal{P}r$: ensemble des probabilités compatibles avec la mesure de probabilité P

g : mesure floue

N : mesure de nécessité

Π : mesure de possibilité

Décision

E_{BetP} : espérance mathématique relative à $BetP$

a_n : action de décider l'hypothèse H_n

$\mathcal{A}$: ensemble des actions possibles

λ : fonction de coûts

$R_{Bet}(a_i|x)$: risque conditionnel de décider a_i sachant x

$R_*(a_i|x)$: risque conditionnel inférieur de décider a_i sachant x

$R^*(a_i|x)$: risque conditionnel supérieur de décider a_i sachant x

Mesures d'incertitude et mesures de conflit

U : fonction de Hartley

NS : mesure de non-spécificité

S : entropie de Shannon

$Confl_m$: conflit existant au sien d'une fonction de croyance m

E : mesure de dissonance

D : mesure de discordance

C : mesure de confusion

ST : mesure *strife*

MC : forme générale d'une mesure de conflit

$D(A, B)$: distance entre les éléments focaux A et B

AU : mesure d'incertitude globale

Inc : mesure d'incertitude

Modélisation

ϕ_i : fonction décroissante monotone

d_i : distance eucliedienne entre le vecteur x et $i$ème prototype

k : nombre de voisins

γ_i : paramètre associé à la modélisation proposé par Denoeux

R : facteur de normalisation

$\mathcal{C}_{j,H_n}$: partition optimale obtenue pour la source S_j sous l'hypothèse H_n

C_{j,H_n} : nombre de classes de la partition $\mathcal{C}_{j,H_n}$

$\hat{\lambda}^a_{C_{j,H_n}}$: estimation de la loi de la base d'apprentissage sur la partition $\mathcal{C}_{j,H_n}$

$\hat{\lambda}^v_{C_{j,H_n}}$: estimation de la loi de la base de validation sur la partition $\mathcal{C}_{j,H_n}$

$Hell$: distance de Hellinger entre deux lois de probabilité

E_{MS} : erreur quadratique moyenne

Combinaison de fonctions de croyance

m_S : fonction de croyance résultante de la combinaison de Smets

m_Y : fonction de croyance résultante de la combinaison de Yager

m_D : fonction de croyance résultante de la combinaison de Dubois et Prade

$m^c(A)$: partie de la masse conflictuelle K attribuée à A

$w(A, \mathbf{m})$: poids associé à l'ensemble A pour l'affectation de la masse conflictuelle K

$m^{(i)}_{\mathcal{L}}$: étiquette imprécise du $i^{\text{ème}}$ vecteur de la base d'apprentissage $\mathcal{L}$

Introduction

La problématique générale de tout processus fondé sur un *diagnostic* automatique est tout d'abord d'apprendre les *situations* ou les *états* (notions d'apprentissage) dans lesquels le système observé peut se trouver. Un diagnostic peut alors être établi à partir des informations, des mesures ou des données présentes à l'entrée du processus. Toutefois, l'obtention de ce diagnostic nécessite une modélisation de l'information. Se pose alors le problème du type de modèle retenu et de la *confiance* que l'on peut ou doit accorder aux données, aux mesures ou de façon plus générale aux informations délivrées par les sources (par exemple les capteurs) observant le système à diagnostiquer. En effet, dans la plupart des systèmes, les informations issues d'observations sont *imprécises* et *incertaines*. Le problème peut alors se résumer en cette question : "Quelle confiance puis-je accorder aux informations nécessaires au diagnostic de mon système ?". En outre, la plupart des diagnostics automatiques ne se fondent pas sur une information unique, mais sur un ensemble d'informations qu'il faut synthétiser préalablement au diagnostic, ceci afin de déterminer de façon plus fiable l'état du système. Malheureusement, dans certains cas les informations peuvent être contradictoires, ce que l'on qualifiera alors d'informations conflictuelles. Le problème posé peut alors se résumer en ces termes : "Comment synthétiser les informations lorsqu'elles sont contradictoires ?". Ces deux questions résument à elles seules toute la problématique de la représentation et de la gestion de l'incertain. On retrouve cette problématique dans plusieurs systèmes faisant appel à des diagnostics automatiques. Toutefois, dans le cadre de cette étude, nous nous sommes intéressé au problème de diagnostic médical et plus particulièrement au diagnostic du mélanome malin.

Le mélanome malin est le plus grave des cancers cutanés, bien qu'il ne représente que 10% de ce type de cancer. De plus, c'est le cancer dont la progression est la plus forte en France et dans la plupart des pays occidentaux. En Suisse, le mélanome malin est en tête de toutes les tumeurs malignes diagnostiquées avec une croissance de 3 à 7% par an [108]. Dans ce même pays, la fréquence de cette forme de cancer cutané est d'environ 10 cas pour 100.000 habitants. En Australie, le mélanome est le troisième cancer en fréquence, et représente la première cause de mortalité par cancer chez l'adulte jeune (20-40 ans). Dans notre pays, sa fréquence est en augmentation constante depuis 50 ans et son incidence est de 7 à 10 nouveaux cas pour 100.000 habitants. Ainsi, chaque année plus de 5.000 personnes sont atteintes de ce cancer et 1.000 en meurent tous les ans. Enfin, la fréquence de ce cancer en France double tous les 10 ans. Malgré ces

chiffres, le mélanome reste la forme de cancer la plus facile à prévenir et se prête ainsi idéalement à une politique de dépistage. Le mélanome est une tumeur facilement accessible à l'examen. Le suivi de patients atteints de nævi et le dépistage des mélanomes sont actuellement assurés de façon visuelle par les dermatologues. Si environ 5000 mélanomes malins sont enlevés chaque année en France, une enquête menée auprès des laboratoires d'anatomopathologie de Seine-Maritime permet d'estimer à au moins 10 fois plus le nombre de lésions nævіques bénignes (grains de beauté) enlevées. La gravité potentielle d'un mélanome enlevé trop tardivement justifie actuellement l'exérèse (ablation chirurgicale) de ces lésions nævіques. Toutefois, le coût d'une exérèse et d'un examen histologique est d'environ 400 francs. Outre les problèmes d'ordre esthétique et traumatique qu'engendre toute intervention chirurgicale, on peut estimer à environ 20 Millions de francs le surcoût annuel lié à l'exérèse inutile de lésions nævіques. Ainsi, la mise au point de méthodes permettant d'améliorer le diagnostic de mélanomes malins paraît donc hautement justifié, tant du point de vue médical qu'économique.

Les premiers travaux [125] ont permis de proposer des caractéristiques pertinentes à partir d'images couleur. Toutefois, l'acquisition des images traitées n'est pas contrôlée. De plus, ces images ne possèdent pas de référence métrique ni de référence pour la couleur. Ainsi, aucune information sur la taille ou l'épaisseur de la lésion n'est disponible. Enfin, certaines informations développées lors de l'étude préliminaire s'avèrent être sensibles aux conditions de prise de vue, à l'agrandissement des images, aux propriétés de couleur et de texture de la peau ou encore à l'échantillonnage spatial.

Partant de ce constat, le but du travail exposé ici est de concevoir une méthode de diagnostic automatique qui s'appuie sur des outils nécessaires et suffisants, en supprimant, si besoin est, les informations peu fiables et contradictoires.

Le présent mémoire est composé de cinq chapitres.

Le premier chapitre est consacré à la présentation des différentes approches du raisonnement automatique permettant de poser un diagnostic. Ensuite, nous abordons plus précisément les problèmes rencontrés dans le cadre du diagnostic médical. Les difficultés reposent pour l'essentiel sur les connaissances utiles au diagnostic. En effet, les connaissances dont on dispose sont généralement peu fiables, contradictoires ou manquantes. Ainsi, différentes formes d'imperfections peuvent être énumérées. Les principales sont :

- l'incertitude, qui correspond à la véracité de l'information,
- l'imprécision qui est relative au contenu de l'information.

L'approche probabiliste, par exemple, permet de manipuler facilement des informations incer-

taines. Cependant, le cadre purement probabiliste de la gestion de l'incertain n'offre pas toujours une solution "raisonnable" au problème posé. En particulier, en présence de populations sous-représentées ou mal connues, la théorie des probabilités n'offre pas un cadre *ad hoc*. On se tournera alors vers d'autres solutions comme la *théorie des sous-ensembles flous* [145], la *théorie des possibilités* [45, 50, 146], ou bien vers la *théorie des fonctions de croyance* appelée également *théorie de l'évidence* [109] ou encore *théorie de Dempster-Shafer* en référence à ses pères.
Ces différentes théories, ou plutôt ces différentes approches destinées à représenter et gérer l'incertain n'ont pas à être opposées et ne sont nullement contradictoires. Il faut, à notre sens, plutôt choisir en fonction du *cadre de travail* (entendons par là l'application visée).

Le second chapitre détaille les concepts de base de la théorie de l'évidence que nous avons choisi d'utiliser dans le cadre de notre application. Ce formalisme permet de modéliser et de combiner des informations imparfaites de manière plus "naturelle" que ne le permet la théorie des probabilités[1].

Dans le troisième chapitre, nous abordons plus particulièrement les différentes approches de modélisation des connaissances dans le cadre de la reconnaissance crédibiliste des formes. Nous apportons des éléments de réponse quant à la question que tout praticien se pose : "Quel modèle pour représenter l'information ?" Si différents modèles sont rappelés, c'est à la notion de fiabilité des informations que nous nous attachons particulièrement. Deux méthodes permettant de déterminer la fiabilité associée à chaque information sont présentées. La première approche repose sur le calcul, obtenu à l'aide de critères d'information, de la dissemblance entre des informations de même nature. La seconde méthode permet, à partir de la minimisation d'un critère d'erreur, de déterminer les coefficients de fiabilité de chaque source d'information afin d'optimiser le taux de reconnaissance. Ces deux approches sont étudiées dans le cadre de deux méthodes de modélisation crédibiliste des connaissances. Une illustration du comportement de ces deux approches, dans le cadre d'un problème d'évolution de contexte, est présentée.

Le quatrième chapitre traite des problèmes liés à la combinaison des connaissances dans le cadre de la théorie de l'évidence. Disposer de plusieurs sources d'information implique de mettre en œuvre une méthodologie de fusion des informations *ad hoc*, afin de synthétiser un état de connaissance avant d'interpréter et éventuellement décider. Cela nécessite également de gérer d'éventuels conflits, qui peuvent être liés à l'inconsistance des sources. Cet aspect de fusion d'informations incertaines est exploré depuis quelques années [1, 2, 40]. Il soulève un certain nombre de questions, dont l'une centrale est de savoir "comment gérer le conflit engendré par plusieurs sources d'information ?". Dans le cadre de la théorie de l'évidence, le but de la fusion

[1] Notons cependant que la théorie de probabilités n'est qu'un cas particulier de la théorie de l'évidence.

est d'obtenir un jeu de masses, combinaison des jeux de masses initiaux établis sur les différentes sources d'information, synthétisant la connaissance. La gestion du conflit lors de la combinaison d'informations contradictoires est un aspect essentiel lors du processus de fusion. Les différentes règles d'agrégation d'informations, qui sont utilisées dans le cadre de la théorie de l'évidence et qui permettent de résoudre ces contradictions, sont présentées. Un cadre générique unifiant les opérateurs de combinaison classiques est introduit. La définition de ce cadre permet d'obtenir une famille d'opérateurs de combinaison permettant une redistribution de la masse conflictuelle plus ou moins adaptée selon l'objectif visé (taux de bonne classification maximum, taux de non-détection minimum,...).

Enfin, le cinquième et dernier chapitre permet l'illustration du propos dans le cadre de la détection du mélanome malin. Le protocole de l'analyse de l'image ainsi que l'extraction des caractéristiques de cette image sont présentés avant d'aborder la phase de diagnostic.

Chapitre 1

Diagnostic, reconnaissance des formes et fusion de données

L'implémentation du raisonnement humain sur une machine occupe depuis de nombreuses années une place importante dans le domaine scientifique. Les premières approches d'automatisation du raisonnement reposent sur l'imitation du comportement humain par le traitement de l'information réalisé par la machine. Ces approches ont donné naissance au terme d'*intelligence artificielle* [12]. Ainsi, dans le cas de problèmes simples, les méthodes d'obtention des solutions peuvent souvent prendre la forme d'algorithmes. Cependant, dans le cas de problèmes plus complexes, c'est-à-dire des problèmes pour lesquels les résolutions sont difficiles à décrire, l'automatisation devient alors plus ardue. C'est pourquoi, les recherches actuelles, moins ambitieuses qu'à l'origine, s'orientent vers la construction de solutions appropriées à un problème donné par le biais de moyens informatiques.

1.1 Le diagnostic

1.1.1 Introduction au diagnostic

Dans le cadre de l'intelligence artificielle, le diagnostic [52] peut être défini comme la reconnaissance de l'état d'un système. Cette définition peut s'appliquer soit à un système industriel avec la reconnaissance de son mode de fonctionnement, soit à la détermination d'une pathologie chez un patient. Le diagnostic peut être abordé par des approches différentes généralement classées en deux catégories.

La première catégorie regroupe les approches qui tentent de modéliser le système. Une fois les différents états du système modélisés, on reconnaît l'état dans lequel le système se trouve en fonction de l'écart qui existe entre l'état mesuré et l'état théorique. Ce type d'approches est pratique lorsque le système est simple. L'utilisation de ces approches devient en revanche impossible lorsque le système est trop complexe.

La seconde catégorie rassemble les méthodes connues sous le nom de *reconnaissance des formes*. Bien que ces méthodes souffrent de l'absence de raisonnement propre au domaine, elles permettent une mise œuvre rapide d'un système de diagnostic.

1.1.2 Les connaissances nécessaires au diagnostic

Afin d'établir un diagnostic, il faut avoir un minimum de connaissances sur le système à diagnostiquer. En effet, l'efficacité de la reconnaissance de la pathologie d'un patient (ou de façon plus générale de l'état d'un système) dépend des connaissances acquises. On peut alors distinguer deux types de connaissances nécessaires pour établir un diagnostic.

Le premier type de connaissance appelé connaissance *a priori* donne une représentation des états possibles de fonctionnement du système en leur associant un certain nombre d'exemples appelés *historique* (ou *apprentissage*).

Le second type de connaissance dite "instantanée" regroupe les éléments qui permettent de déterminer le mode de fonctionnement du système à un instant donné. Ces connaissances sont issues d'observations numériques et/ou symboliques. Le diagnostic sera alors établi en associant, à partir des connaissances *a priori*, la connaissance instantanée à un mode de fonctionnement.

1.1.3 Particularité du diagnostic médical

Approche du diagnostic médical

Comme nous l'avons vu dans la section précédente, le diagnostic peut être abordé par différentes approches : approche avec modèle ou sans modèle. Dans le cadre du diagnostic médical, comme dans tous les domaines liés au vivant, il s'avère utopique de vouloir modéliser un patient afin d'établir un diagnostic tant la "machine" à modéliser reste complexe.

Étant donnée cette complexité propre au diagnostic médical, nous avons retenu un système de diagnostic fondé sur *la reconnaissance des formes*. Il existent différentes approches pour la reconnaissance des formes : *la reconnaissance des formes structurelles* qui représentent les formes à l'aide de grammaire et *la reconnaissance des formes statistiques* qui utilise une représentation numérique des formes. C'est cette dernière approche que nous avons retenu et que nous développons dans la section 1.2.

Les connaissances du diagnostic médical

Plusieurs remarques sur l'utilisation des connaissances *a priori* dans le cadre du diagnostic médical peuvent être faites. Dans un premier temps, la détermination de manière exhaustive de l'ensemble des pathologies, afin d'établir un diagnostic, reste impossible. En effet, même si l'on

répertorie l'ensemble des pathologies rencontrées dans le passé, d'autres maladies, non encore identifiées, peuvent exister. Le diagnostic médical sera donc établi sur une connaissance *a priori* incomplète. Il est alors souhaitable d'avoir la possibilité d'associer les symptômes d'un patient à une pathologie inconnue ou non recensée dans l'historique car rare ; on parle alors de *rejet*. De plus, la constitution d'un historique nécessite idéalement un grand nombre d'exemples. Cela s'avère difficile (nous serions tentés de dire heureusement !) en ce qui concerne les pathologies. En effet, dans la constitution d'un historique, les patients associés à une pathologie sont toujours moins nombreux que les patients sains. Ainsi, prenons l'exemple du mélanome malin. En Suisse, la fréquence de cette forme de cancer cutané est de 10 pour 100.000 habitants. De la même manière, en France, environ 5000 lésions diagnostiquées mélanome malin sont enlevées par an. Ces chiffres montrent la difficulté de mettre en place un historique dans le cadre de pathologies relativement rares. Enfin, il arrive parfois que dans l'historique, certains patients ne soient pas associés précisément à une pathologie mais à un ensemble de pathologies. Cela peut être le cas lorsque que le patient souffre, par exemple, de plusieurs pathologies (notion d'étiquettes multiples). D'une manière différente, un patient peut être associé à des pathologies différentes car les médecins ont diagnostiqué des maladies différentes ; on parle alors d'*apprentissage imprécis*.

En ce qui concerne les connaissances "instantanées", celles-ci sont issues d'appareils qui ne mesurent pas toujours précisément la grandeur que l'on souhaite. De plus, ces mesures ne sont pas toujours reproductibles (exemple : éclairage non contrôlé lors de l'acquisition d'image, système d'acquisition non calibré, ...). Ces différents phénomènes font alors intervenir la notion d'imprécision sur les données. Enfin, il se peut que l'une de ces connaissances, présente normalement, soit manquante car le test clinique fournissant cette connaissance n'a pas pu être effectué sur le patient compte tenu de son état de santé par exemple. On parle alors de *données manquantes* ou *d'imcomplétude*.

A partir de ces différents facteurs (imprécision, incertitude, imcomplétude), il apparaît alors essentiel d'utiliser plusieurs capteurs ou informations redondantes et/ou complémentaires pour réduire les incertitudes afin d'établir un diagnostic plus pertinent. C'est l'objet de la combinaison de sources d'information aussi appelée fusion d'informations. Il existe deux grandes classes de méthodes de fusion d'informations disponibles en diagnostic : la *synthèse de décisions* et la *fusion de données*. La première classe de combinaisons utilise indépendamment les différents systèmes de diagnostic existant et combine leurs décisions. La stratégie employée dans la seconde classe repose sur la combinaison de ce que l'on connaît avant de décider, au contraire de la méthode précédente où l'on combine des décisions déjà existantes. Cette solution, en gardant l'ensemble

des informations et ensuite en les utilisant conjointement avant de décider, paraît plus prudente. C'est cette approche, dans l'objectif du diagnostic médical, que nous privilégions et qui fait l'objet de la section 1.3.

1.2 La reconnaissance des formes

Nous avons vu, dans la section précédente, que le but du diagnostic était de définir le mode de fonctionnement d'un système à partir de mesures réalisées sur celui-ci. Dans le cadre de la reconnaissance des formes, les modes de fonctionnement sont appelés *classes* et les mesures observées sur le système constituent le *vecteur forme* noté x.

1.2.1 Introduction

Un problème de reconnaissance des formes, ou *discrimination*, a pour objectif de partager des formes à reconnaître en classes, dans le but de satisfaire un critère, en général celui d'approcher le partage subjectif de ces formes réalisé par un être humain. Pour cela, il faut exploiter et interpréter un certain nombre d'indices (ou caractéristiques) extraits d'une forme inconnue, de manière à produire une hypothèse sur sa classe d'appartenance. Ainsi, chaque vecteur forme pourra être représenté dans l'espace des L caractéristiques noté $\mathcal{X}$, pour lequel chacune des composantes l correspond à l'une des caractéristiques mesurées sur le système.

En théorie, on peut supposer que la conception d'un système de reconnaissance des formes consiste simplement à développer un dispositif dans lequel l'être humain injecte un ensemble de règles permettant par la suite une interprétation automatique des indices extraits de la forme à reconnaître pour trouver sa classe d'appartenance.

En pratique, d'énormes difficultés se présentent. Les problèmes concernent, en premier lieu, le choix des caractéristiques qui doivent être représentatives de l'ensemble des formes à reconnaître. Ce choix est important car il conditionne toute la méthodologie mise en œuvre pour la reconnaissance. Enfin, en supposant ces caractéristiques connues, il est souvent difficile de définir un ensemble de règles permettant de les exploiter. En d'autres termes, il est souvent impossible de définir dans un système une méthodologie *a priori* permettant d'exploiter ces indices, en vue de procéder à une reconnaissance.

1.2.2 Définition

D'un point de vue général, on peut formaliser la discrimination de la manière suivante.

Soient N classes connues, notées $H_1, \ldots, H_N$, auxquelles on associe généralement un *ensemble d'apprentissage* $\mathcal{L}$. Cet ensemble est constitué de I exemples ou échantillons. Chaque

exemple dans $\mathcal{L}$ est représenté par deux grandeurs : un vecteur de caractéristiques $x^{(i)} = [x_1^{(i)}, \ldots, x_l^{(i)}, \ldots, x_L^{(i)}]^t$, de dimension L, et un vecteur d'étiquettes u indiquant son appartenance à l'une des classes[1]. Ce vecteur est composé d'éléments $u_n^i \in \{0, 1\}$ qui indiquent l'appartenance du vecteur $x^{(i)}$ à chaque hypothèse H_n. Par exemple $u_n^i = 1$ si le vecteur $x^{(i)}$ appartient à la classe H_n, et $u_{n'}^i = 0$ pour tout $n' \neq n$. On dispose ainsi d'un ensemble d'apprentissage $\mathcal{L} = \{(x^{(1)}, u^{(1)}), \ldots, (x^{(I)}, u^{(I)})\}$ relatif à I individus. Le problème de la discrimination consiste à analyser les ressemblances et les dissemblances entre ces différentes données pour caractériser de manière plus ou moins directe chacune des classes H_n. Soit une forme $x \in \mathcal{X}$ "inconnue", c'est-à-dire sans étiquette, issue d'un ensemble fini $\mathcal{T}$ appelé *base de test*. L'objectif est ensuite de pouvoir classer la forme x dans l'une des N classes. En outre, un ensemble de données étiquetées, différentes de celles rencontrées dans la base d'apprentissage, appelé *base de validation* permet de valider ou d'optimiser les paramètres susceptibles d'exister au sein de l'algorithme de discrimination.

Il existe de nombreuses méthodes de discrimination qui sont à la base de la conception des systèmes de reconnaissance des formes. Les approches classiques sont aujourd'hui bien recensées et on pourra se référer à [14, 52, 53, 90] pour en avoir une description approfondie.

Il existe deux types de discrimination :

- la *discrimination paramétrique* lorsque l'on dispose d'informations régissant les observations,
- ou *discrimination non paramétrique* dans le cas contraire.

Les méthodes de discrimination paramétrique font appel à des connaissances préalables sur la forme des classes (e.g. : distributions gaussiennes). Toutefois, il est rarement possible de définir ces connaissances préalables. L'emploi des méthodes de discrimination paramétrique nécessitent alors de faire des suppositions sur l'existence de ces prérequis ou d'utiliser un modèle en sachant qu'il est inexact.

Les méthodes supervisées non paramétriques, quant à elles, ne nécessitent pas de connaissance préalable sur les formes des classes. Ainsi, ces méthodes peuvent être appliquées à des problèmes très mal connus.

L'objectif des sections suivantes n'est pas de présenter un résumé exhaustif des méthodes de discrimination mais simplement d'évoquer les méthodes les plus couramment utilisées.

[1] Bien que, dans le cadre du diagnostic médical, nous avons vu dans la section précédente que les exemples pouvaient être affectés (ou appartenir) à plusieurs classes, nous considérons ici uniquement le cas d'un apprentissage précis.

1.2.3 Estimateur bayésien

Soit $\Theta = \{H_1, \ldots, H_N\}$ l'ensemble des hypothèses possibles (état du monde, état du système). Soit x un vecteur L-dimensionnel appartenant à une classe inconnue.

On suppose que pour chacune des hypothèses $H_n \in \Theta$, le vecteur forme x est caractérisé par une densité de probabilité conditionnelle notée $f(x|H_n)$. Si l'on sait estimer ou si l'on connaît la probabilité *a priori* $P(H_n)$ de chacune des classes, alors la probabilité *a posteriori* $P(H_n|x)$ peut être obtenue à l'aide de la règle de Bayes :

$$P(H_n|x) = \frac{f(x|H_n)P(H_n)}{f(x)} \quad n = 1, \ldots, N \tag{1.1}$$

avec :

$$f(x) = \sum_{n=1}^{N} f(x|H_n)P(H_n). \tag{1.2}$$

Dans l'équation (1.2), $f(x)$ représente la densité de probabilité non conditionnelle.

Soit x un vecteur appartenant à l'hypothèse $H_k \in \Theta$ inconnue et $\mathcal{A} = \{a_1, \ldots, a_A\}$ un ensemble de A actions possibles. Une fonction de coût $\lambda : \mathcal{A} \times \Theta \to \mathbb{R}$ est définie de telle manière que $\lambda(a_i|H_n)$ représente le coût encouru si l'on choisit l'action a_i alors que la vraie classe est H_n [53]. Dans le cadre de la théorie bayésienne de la décision, le critère de décision repose sur la minimisation du risque moyen. L'espérance mathématique du coût si l'on choisit l'action a_i est appelée risque conditionnel et noté $R(a_i|x)$. Ce risque est défini à l'aide des probabilités *a posteriori* et des fonctions de coûts par la relation suivante :

$$R(a_i|x) = \sum_{n=1}^{N} \lambda(a_i|H_n)P(H_n|x). \tag{1.3}$$

En notant D la fonction de décision, le risque moyen de décision R s'exprime alors par :

$$R = \int R(D(x)|x)f(x)dx. \tag{1.4}$$

La minimisation du risque moyen s'obtient en minimisant $R(D(x)|x)$ pour tout x, c'est-à-dire en choisissant l'action a_i dont le risque conditionnel est le plus faible. Nous obtenons alors :

$$D_B(x) = a_i \Leftrightarrow R(a_i|x) < R(a_j|x) \quad \forall j \neq i. \tag{1.5}$$

La fonction de décision D_B est appelée règle de décision de Bayes.

Coûts $\{0, 1\}$

Dans le domaine du diagnostic, les hypothèses du monde $\Theta = \{H_1, \ldots, H_N\}$ sont souvent associées à l'ensemble des actions possibles. Dans ce cas, l'ensemble des actions $\mathcal{A}$ est constitué

de N éléments $\{a_1, \ldots, a_N\}$ qui correspondent respectivement à l'affectation du vecteur x à chacune des hypothèses H_n de Θ. En règle générale, si aucune information n'est donnée pour la définition des coûts, on considère alors qu'aucune erreur d'affectation n'est plus dommageable qu'une autre. Dans ce cas, on peut alors définir les coûts de la manière suivante :

$$\lambda(a_k|H_k) = 0 \tag{1.6}$$

$$\lambda(a_i|H_k) = 1 \qquad \forall\, i \in \{1, \ldots, N\}, \quad i \neq k. \tag{1.7}$$

Le risque conditionnel obtenu à partir de ces poids peut alors s'écrire de la manière suivante :

$$\begin{aligned} R(a_i|x) &= \sum_{n=1}^{N} \lambda(a_i|H_n) P(H_n|x) \\ &= \sum_{n \neq i} P(H_n|x) \\ &= 1 - P(H_i|x). \end{aligned} \tag{1.8}$$

Dans ce cas, la règle de décision de Bayes revient à choisir l'hypothèse de plus grande probabilité *a posteriori* :

$$D_B(x) = a_i \Leftrightarrow P(H_i|x) = \max_{n=1,\ldots,N} P(H_n|x). \tag{1.9}$$

Coûts $\{0, 1\}$ avec rejet

Chow [23] a proposé d'introduire une action supplémentaire a_0 dans $\mathcal{A}$. Cette action, appelé action de rejet, consiste à refuser de prendre une décision lorsque le risque de mauvaise classification est trop grand. Si l'on note λ_0 le coût associé à l'action a_0, la règle de Bayes avec rejet s'écrit alors :

$$\begin{aligned} D(x) = a_0 \quad \text{si} \qquad & \max_{n=1,\ldots,N} P(H_n|x) < 1 - \lambda_0 \\ D(x) = a_i \quad \text{si} \qquad & P(H_i|x) = \max_{n=1,\ldots,N} P(H_n|x) \geq 1 - \lambda_0. \end{aligned} \tag{1.10}$$

Ainsi, plus λ_0 est petit plus la probabilité de rejet est importante, jusqu'au rejet total pour $\lambda_0 = 0$, où l'on ne commet plus aucune erreur, c'est-à-dire que l'on rejette systématiquement tout vecteur à classer. A l'inverse, lorsque $\lambda_0 \leq 1 - 1/N$ le rejet est impossible puisqu'au moins l'une des probabilités *a posteriori* est supérieure à $1/N$. Cette règle de décision sépare l'espace de représentation en $N+1$ régions de décision. Le rejet, ainsi introduit par Chow, est qualifié de rejet d'ambiguïté par Dubuisson [52] car la zone qui lui est attribuée dans l'espace de représentation se situe systématiquement entre les classes.

Toutefois, nous avons vu, dans le cadre du diagnostic, que de nouvelles classes pouvaient apparaître (exemple : apparition d'une nouvelle pathologie). En effet, un nouveau vecteur à classer peut être très éloigné dans l'espace $\mathcal{X}$ des classes existantes. Les vecteurs éloignés peuvent traduire l'apparition d'une nouvelle classe. Dans ce cas, la notion de rejet d'ambiguïté n'est pas suffisante.

Il faut alors introduire une nouvelle possibilité de rejet désigné par rejet de distance [52]. L'action associée à ce rejet est noté a_d. A partir des probabilités *a posteriori* et de la densité de probabilité, l'utilisation du rejet de distance nous amène à la règle de décision suivante :

$$\begin{array}{lll} D(x) = a_d & \text{si} & f(x) < \lambda_d \\ D(x) = a_i & \text{si} & f(x) \geq \lambda_d \quad \text{et} \quad P(H_i|x) = \max\limits_{n=1,\ldots,N} P(H_n|x). \end{array} \tag{1.11}$$

La règle de décision incluant les deux types de rejet est définie de la manière suivante :

$$\begin{array}{lll} D(x) = a_d & \text{si} & f(x) < \lambda_d \\ D(x) = a_0 & \text{si} & f(x) \geq \lambda_d \quad \text{et} \quad \max\limits_{n=1,\ldots,N} P(H_n|x) < 1 - \lambda_0 \\ D(x) = a_i & \text{si} & f(x) \geq \lambda_d \quad \text{et} \quad P(H_i|x) = \max\limits_{n=1,\ldots,N} P(H_n|x) \geq 1 - \lambda_0. \end{array} \tag{1.12}$$

Cette dernière règle de décision sépare l'espace des caractéristiques $\mathcal{X}$ en $N + 2$ régions.

1.2.4 Les 'k Plus Proches Voisins' (k-ppv)

La méthode des plus proches voisins est une méthode qui cherche à déterminer directement la partition de l'espace de représentation en classes, sans faire d'hypothèse sur la distribution des points, ni sur la nature des surfaces séparatrices. Le principe de cette méthode est très simple : si une forme inconnue est proche d'une autre dans l'espace de représentation $\mathcal{X}$, elles sont certainement de la même classe. Ainsi un vecteur x d'étiquette inconnue est associé à la classe majoritairement représentée parmi ses k plus proches voisins. De la même manière que pour l'estimateur bayésien, il existe des règles de décision fondées sur le rejet d'ambiguïté [41, 69] et/ou le rejet de distance [27, 52].

Cependant, dans sa forme initiale, cette méthode nécessite un nombre d'échantillons important afin de bien remplir l'espace des caractéristiques, avec comme autre défaut, un gros volume de calcul engendré par la recherche des k plus proches voisins dans la base d'apprentissage. Des variantes ont été développées afin de diminuer le temps de calcul (prototypes [67], "pavage" de l'espace des caractéristiques [31], ...).

Soulignons que le choix de la distance et du nombre k de plus proches voisins ont une influence directe sur les performances du classifieur.

1.2.5 Conclusion

Dans cette section, nous avons présenté l'approche diagnostic par reconnaissance des formes. Cette approche du diagnostic permet de minimiser les informations relatives au système. Au contraire, les approches basées sur l'obtention de modèles demandent une connaissance rigoureuse et exhaustive du système, ce qui, dans le cas particulier du diagnostic médical s'avère rare et difficile. C'est pourquoi, nous avons arrêté notre choix sur un diagnostic à base de discrimination.

Toutefois, se pose encore le problème de la gestion de l'incertitude des connaissances (Cf. section 1.1.3). Les techniques de fusion d'informations gèrent souvent l'incertitude, les conflits, les incohérences et les redondances présentes dans les informations issues des mesures du système. Ceci fait l'objet de la section suivante.

1.3 La fusion de données

1.3.1 Les informations

L'un des principaux enjeux de la société de l'information concerne la gestion des *informations imparfaites*. En effet, dans de nombreux secteurs applicatifs, il s'agit de classer, décider, surveiller, commander à partir d'informations observées sur le système ou préalablement modélisées [19, 42, 61, 95]. Ces informations peuvent revêtir plusieurs aspects :

- l'aspect *donnée* dans le cadre de la gestion des bases (concept de fouille de données ou data mining : données financières, commerciales, médicales,...). Ainsi, il peut s'agir d'extraire des variables discriminantes, d'identifier des structures sous-jacentes, d'extraire de la connaissance de manière générale,
- l'aspect *mesure* pour les systèmes multicapteurs. Dans ce contexte, il s'agit alors de fusionner les mesures pour la commande d'un système ou la prise d'une décision.

Le contenu des informations peut être de type *numérique* : attributs de forme, de radiométrie, de position (altitude), de mouvement (vitesse, accélération), de texture, etc... Ils pourront être également de type *symbolique* afin de prendre en compte les relations pouvant exister entre objets ; on pourra avoir par exemple des relations de type structurel (décomposition d'un aéroport en objets plus simples) ou de connexité (un pont se situe toujours au dessus d'une route, d'une voie ferrée ou d'une rivière).

L'imperfection des informations fait appel à plusieurs concepts. Le premier, généralement bien maîtrisé, concerne l'imprécision des informations. L'incertitude est un second concept à différencier de l'imprécision par le fait qu'il ne fait pas référence au contenu de l'information mais à sa "qualité". Décrivons, de manière plus détaillée, ces deux notions :

- *L'imprécis* - Les imprécisions correspondent à une difficulté dans l'énoncé de la connaissance, soit parce que des connaissances numériques sont mal connues, soit parce que des termes de langage naturel sont utilisés pour qualifier certaines caractéristiques du système de façon vague. Le premier cas est la conséquence d'une insuffisance des instruments d'observation (2000 à 3000 manifestants), d'erreurs de mesure (poids à 1% près) ou encore de connaissance flexibles (la taille d'un adulte est environ entre 1,5 mètres et 2 mètres). Le

second provient de l'expression verbale de connaissances (températures douces, proche de la plage) ou de l'utilisation de catégories aux limites mal définies (enfant, adulte, vieillard).

– *L'incertain* - Les incertitudes concernent un doute sur la validité d'une connaissance. Celles-ci peuvent provenir de la fiabilité relative à l'observation faite par un système, celui-ci pouvant être peu sûr, susceptible de commettre des erreurs ou de donner intentionnellement des informations erronées, ou encore d'une difficulté dans l'obtention ou la vérification de la connaissance (l'affirmation d'une forte douleur par un patient). Des incertitudes sont également présentes dans le cas des prévisions (en météorologie par exemple). Pour une proposition incertaine, c'est la vérité même de la proposition qui est en cause.

Il existe d'autres sortes d'imperfections plus ou moins dépendantes de l'imprécision et de l'incertitude, telles que l'incomplétude et l'indétermination.

Traiter correctement un problème posé avec des connaissances aussi hétérogènes et imparfaites devient alors rapidement un problème crucial. De façon générale, l'utilisation d'une solution telle que la *fusion de données* peut permettre de résoudre de tels problèmes.

1.3.2 La fusion de données

La fusion de données a depuis peu suscité un intérêt certain dans la communauté scientifique [1, 2, 16, 17]. Elle consiste à mettre à profit le maximum d'information sur les données afin de réduire les faiblesses de certaines à l'aide des autres. En effet, il est intéressant de pouvoir utiliser conjointement plusieurs sources d'information. Les améliorations en terme de qualité de décision sont généralement reconnues [6]. Une décision fondée sur un grand nombre d'informations d'origines et de natures variées est généralement plus fiable et plus précise qu'une décision ne dépendant que d'un seul type ou d'une seule source d'information. Ces conséquences positives proviennent de :

– la redondance des informations qui s'obtient lorsque différentes sources exploitent les mêmes paramètres. La fusion de ces sources permet alors de diminuer l'incertitude globale des informations fournies relativement à ces paramètres. Elle offre également une plus grande robustesse de l'information en permettant de faire face à la défaillance de l'une des sources.

– la complémentarité qui s'obtient lorsque les sources exploitent des paramètres différents. Elle permet alors de déduire une information globale plus complète concernant certains aspects du problème qu'une source, opérant individuellement, serait incapable de saisir.

Ainsi les techniques de fusion gèrent souvent l'incertitude, les redondances, les conflits ou les incohérences présentes dans les informations fournies par les sources. Généralement, la fusion de données s'appuie sur la théorie des mesures de confiance qui inclut les mesures probabilistes,

possibilistes et crédibilistes.

Cadre probabiliste : approche bayésienne

Les méthodes les plus utilisées pour la gestion de l'incertitude et de la fusion de données ont tout d'abord été envisagées sous l'approche bayésienne. La mise à jour des informations (modélisées par des distributions de probabilité) se fait à l'aide du théorème de Bayes.

Cette théorie repose sur des bases solides et des axiomes connus. L'un des inconvénients majeurs de cette technique réside dans l'exigence de la connaissance parfaite des probabilités, et plus particulièrement de la probabilité *a priori*. Malheureusement, lorsque les connaissances sur le problème sont imparfaites, ces probabilités ne sont pas connues avec "certitude". Ces limitations, maintenant bien identifiées [15], ont donné naissance à de nombreuses extensions ou nouvelles propositions afin de gérer l'incertitude.

Cadre possibiliste

La théorie des possibilités a été introduite en 1978, par Zadeh [146] pour permettre de manipuler des incertitudes de nature non probabiliste, donc auxquelles les moyens classiques de la théorie des probabilités n'apportent pas de solution. Ainsi, dans le cadre de cette théorie, les connaissances imprécises et les connaissances incertaines peuvent coexister et être traitées conjointement.

La théorie des possibilités considère certaines situations plus ou moins *possibles* par rapport à d'autres. Elle modélise, non pas un degré de croyance ou de vérité, mais plutôt la préférence que l'on a pour une hypothèse.

De plus, le cadre possibiliste offre un choix d'opérateurs de combinaison relativement important (conjonction, disjonction, adaptative, ...) [40], ce qui peut être un avantage bien qu'une sélection s'avère obligatoire.

Cadre crédibiliste

La dernière approche, proposée par Shafer [109] prenant appui sur les bases formulées par Dempster [32], est appelée théorie de l'évidence. Cette approche permet de gérer les situations d'ignorance ce qui n'est pas le cas dans le cadre de la théorie des probabilités. La modélisation des informations se fait à l'aide de fonctions de croyance. Une fois les fonctions de croyance obtenues, la fusion est réalisée par l'intermédiaire de la règle de combinaison de Dempster.

Dans la section suivante, nous présentons les intérêts d'une telle approche par rapport à l'approche bayésienne classique.

1.3.3 Intérêts et domaines d'application de la théorie de l'évidence

Au regard de la section précédente, deux points de vue se distinguent.

Le point de vue défendu par les probabilistes reposent sur le fait que toutes les approches vues dans la section précédente mènent à des résultats qui auraient pu être obtenus avec un modèle probabiliste à condition que ce modèle soit suffisamment bien adapté au problème, c'est-à-dire suffisamment représentatif de l'état du monde.

Les adeptes de la théorie des possibilités et de la théorie de l'évidence, qui constituent le second point de vue, préfèrent modéliser le plus fidèlement possible les informations disponibles.

Toutefois, la théorie des croyances propose des avantages par rapport à la théorie des probabilités.

Intérêts de la théorie de l'évidence

Bien que la théorie des probabilités repose sur des bases mathématiques solides, elle souffre de plusieurs inconvénients qui ne sont pas rencontrés dans le cadre de la théorie de l'évidence.

L'approche bayésienne nécessite une initialisation qui prend la forme d'une affectation de probabilité *a priori* à chacune des classes. L'estimation de ces probabilités est souvent délicate. Si les probabilités conditionnelles peuvent être souvent bien estimées par des approches fréquentistes, comme par exemple dans des applications de traitement d'images, ce n'est pas le cas pour les probabilités *a priori*. Leur évaluation sort du cadre fréquentiste et fait appel à des approches plus subjectives. L'utilisation de la théorie de l'évidence ne nécessite pas, elle, de connaissance *a priori* sur le problème à traiter.

La modélisation probabiliste raisonne uniquement sur les singletons, qui représentent les différentes classes. Mais la prise en compte uniquement des singletons ne permet pas de résoudre des problèmes complexes (exemple : appartenance d'un vecteur à plusieurs classes à la fois). A la différence de l'approche bayésienne, la théorie de l'évidence permet de répartir de la croyance non seulement sur les hypothèses élémentaires mais aussi sur des compositions d'hypothèses.

Enfin, la théorie des probabilités, de par sa contrainte d'additivité, suppose qu'un objet existe ou n'existe pas. Ainsi, la probabilité d'appartenance à une classe détermine la probabilité dans sa négation et la somme des probabilités vaut l'unité. De cette manière, la représentation de l'ignorance dans le cadre probabiliste est difficilement modélisable sauf dans le cas de l'ignorance totale où l'on suppose alors l'équiprobabilité de l'ensemble des événements, ce qui constitue un *a priori* fort[2]. La théorie de l'évidence n'impose aucune relation entre un événement et sa

[2]La loi uniforme est une loi, et donc une représentation du monde.

négation. Cette approche ne modélise que la croyance créditée à une classe, sans influencer la croyance attribuée aux autres classes.

A l'étude de ces différences et avantages de la théorie de l'évidence par rapport à la théorie bayésienne, nous avons retenu cette approche pour la suite de notre étude.

Application de la théorie de l'évidence

La théorie de l'évidence est appliquée dans des domaines divers. Sans constituer une liste exhaustive, on peut citer un certain nombre de domaines d'application exploitant les avantages de cette théorie :

- la fusion multi-capteurs [5, 20, 61, 72, 75],
- la fusion de classifieurs [87, 104, 138],
- la reconnaissance des formes [37, 39, 126, 148],
- la surveillance de l'environnement [85, 131, 130],
- le traitement d'image [18, 42, 127, 133].

Plus récemment, une application de fusion de données multi-sensorielles pour la détection de mines anti-personnelles a mis en avant l'avantage de la théorie de l'évidence par rapport à l'approche bayésienne [91, 92].

D'autres travaux ont été menés afin de convertir des outils existants en approche crébibiliste pour la régression [98, 106] et pour les arbres de décision [3, 38, 55, 56].

1.4 Conclusion

Ce chapitre nous a servi d'introduction générale aux différents domaines liés à notre étude.

Nous avons, dans un premier temps, défini le diagnostic en présentant les différentes approches possibles (modèle ou reconnaissance des formes). A partir de l'étude des particularités du diagnostic médical et plus particulièrement de la complexité de mise en œuvre d'un modèle, nous avons choisi l'approche du diagnostic par reconnaissance des formes . Nous avons présenté différentes méthodes classiques, paramétrique ou non paramétrique, de la reconnaissance des formes.

Ensuite, nous avons introduit les principes de la fusion d'informations qui permettent de résoudre le problème de la gestion des connaissances imparfaites. Nous avons abordé les différents formalismes intégrant la fusion de données : la théorie des probabilités, la théorie des possibilités et la théorie de l'évidence. Ce dernier cadre permet de représenter naturellement l'incertitude, de gérer des hypothèses composées et de minimiser la part d'*a priori* pour la modélisation du problème. C'est pourquoi nous avons porté notre choix sur la théorie de l'évidence.

Ainsi, pour la suite de notre étude, nous travaillerons sur le diagnostic médical avec une approche par reconnaissance des formes fondée sur la théorie de l'évidence.

Le chapitre suivant est consacré à une présentation des concepts et des interprétations de la théorie de l'évidence. Dans, le troisième chapitre, nous présentons les différentes approches de modélisation des connaissances dans le cadre de la théorie de l'évidence pour la reconnaissance des formes, ainsi que l'approche que nous proposons pour déterminer la fiabilité des sources. Le chapitre 4 concerne la combinaison employée dans le cadre de la théorie de l'évidence et le formalisme que nous proposons, avant d'aborder, dans le cinquième chapitre, les applications traitées.

Chapitre 2

La théorie de l'évidence

La gestion des données imparfaites (incertaines, imprécises et incomplètes) en reconnaissance de formes a été abordée à l'aide de nombreuses techniques et théories. Parmi celles-ci, on trouve la théorie des fonctions de croyance, qui permet de gérer aussi bien les incertitudes que les imprécisions. Cette théorie aussi appelée théorie de l'évidence a pour origine les travaux d'Arthur Dempster [32] sur les bornes inférieure et supérieure d'une famille de distributions de probabilité. Néanmoins, l'élaboration du formalisme de la théorie est imputable à Glenn Shafer [109]. Shafer a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines. De plus, elle permet de combiner des connaissances sur un phénomène obtenues au travers de différentes sources de manière plus souple que le formalisme probabiliste. En effet, la représentation des informations imparfaites ou de l'absence d'information, est mal prise en compte par la théorie des probabilités. Dans ce chapitre, nous décrivons, dans un premier temps, les différentes fonctions de croyance développées dans le cadre de la théorie de l'évidence (Section 2.1). Nous abordons, dans la section 2.2, la combinaison des croyances, aussi appelée révision. Dans la section 2.3, nous présentons les différentes interprétations des fonctions de croyance et plus particulièrement celle introduite par Smets sous le mon de modèle des croyances transférables [114, 116, 123] avant d'évoquer (Section 2.4) la prise de décision dans le cadre de la théorie de l'évidence. Enfin, la section 2.5 décrit différentes opérations possibles sur les fonctions de croyance.

2.1 Les fonctions de croyance

2.1.1 Masse de croyance élémentaire

Soit Θ l'ensemble des N hypothèses solutions du problème posé. L'ensemble Θ, appelé *cadre de discernement*, est défini de la manière suivante :

$$\Theta = \{H_1, \cdots, H_n, \cdots, H_N\}. \tag{2.1}$$

On suppose que le cadre de discernement est exhaustif et que les hypothèses sont exclusives. Cette notion est aussi appelée monde fermé (*closed-world*). Toutefois, il est possible de s'affranchir de cette condition en admettant que l'ensemble Θ est un cadre de discernement non exhaustif. Cette approche est alors appelée hypothèse du monde ouvert (*open world*) [114]. Nous préférons garder l'hypothèse d'exhaustivité de Θ, quitte à introduire explicitement dans Θ un élément supplémentaire représentant l'ensemble des valeurs inconnues de H_n. A partir de ce cadre de discernement, on définit un ensemble noté 2^Θ qui désigne l'ensemble des 2^N parties A de Θ. Cet ensemble est défini de la manière suivante :

$$2^\Theta = \{A \mid A \subseteq \Theta\} = \{\emptyset, \{H_1\}, \ldots, \{H_N\}, \{H_1, H_2\}, \ldots, \Theta\}. \tag{2.2}$$

Cet ensemble contient les hypothèses singletons de Θ, toutes les disjonctions possibles de ces hypothèses ainsi que l'ensemble vide. Par la suite, nous noterons H_n une hypothèse singleton, et A une proposition désignant indifféremment une hypothèse ou une disjonction d'hypothèses.

Soit une information (qui peut être issue d'un capteur, d'un agent, d'un expert, ...) traduisant une opinion sur l'état d'un système par exemple. Cette opinion est alors caractérisée par des degrés de croyance dans les différentes hypothèses. Ces degrés de croyance peuvent être décrits par une *fonction de croyance* notée m. La fonction m est alors définie par :

$$m : 2^\Theta \rightarrow [0, 1] \tag{2.3}$$

et vérifie les propriétés suivantes :

$$\begin{cases} m(\emptyset) = 0 \\ \sum\limits_{A \subseteq \Theta} m(A) = 1. \end{cases} \tag{2.4}$$

La quantité $m(A)$ s'interprète comme la part de croyance placée strictement sur A. Cette quantité se différencie d'une probabilité par le fait que la totalité de la croyance est répartie non seulement sur les hypothèses singletons H_n mais aussi sur les hypothèses composites A. On peut alors accorder une partie de la croyance à une proposition, et ainsi affecter à l'ensemble des hypothèses contenues dans la proposition une croyance à la réalisation de chacune d'entre-elles sans prendre partie pour l'une d'elles précisément. Cela signifie qu'en l'état actuel des connaissances, la masse $m(A)$ ne peut pas être affectée à un sous-ensemble de A. Toutefois, dans l'éventualité d'un apport de nouvelles informations, la masse de croyance pourra alors être allouée plus précisément. Les sous-ensembles A dont la masse est non nulle sont appelés *éléments focaux*. Nous noterons $\mathcal{F}(m)$ l'ensemble des éléments focaux de m. De plus, l'union des éléments focaux est appelée *noyau*. Nous présentons maintenant des fonctions de croyance particulières, définissant des connaissances bien précises, de manière à mieux illustrer la modélisation des connaissances.

Pour chacun des exemples qui suit nous présentons les masses de croyance accordées aux éléments focaux.

– *Fonction de croyance d'ignorance totale :*

$$m(\Theta) = 1. \tag{2.5}$$

C'est le cas de l'indétermination ou de l'ignorance totale. L'observateur sait que l'hypothèse solution se trouve dans le cadre de discernement mais il ne peut pas en dire plus. Il est incapable de répartir sa connaissance sur un ensemble plus petit que Θ. Nous verrons que cette fonction de croyance joue un rôle particulier dans le processus de combinaison.

– *Fonction de croyance de certitude totale :*

$$m(H_n) = 1. \tag{2.6}$$

Cette répartition de masse de croyance est concentrée sur une hypothèse singleton. L'observateur est sûr de connaître l'hypothèse solution et modélise sa connaissance par une masse de croyance totale en l'hypothèse H_n.

Comme pour la plupart des autres représentations, la principale difficulté repose sur l'affectation des masses de croyance à chacune des hypothèses. Toute méthode de définition d'une fonction de croyance est potentiellement acceptable. La plupart des modélisations existantes dépendent de l'application envisagée. Nous reviendrons sur les principales modélisations dans le chapitre suivant.

2.1.2 Autres mesures de croyance

A partir de l'affectation de masses de croyance (*basic belief assignment* ou *bba*), on peut calculer d'autres mesures de croyance : la crédibilité (Bel), la plausibilité (Pl) et la communalité (Q).

La fonction de Crédibilité

La fonction de crédibilité est définie de la manière suivante :

$$Bel(A) = \sum_{B \subseteq A} m(B) \qquad \forall A \subseteq \Theta. \tag{2.7}$$

$Bel(A)$ représente l'ensemble de la croyance apportée aux éléments de cette disjonction d'hypothèses. Elle correspond à la quantité d'information qui est toute entière contenue dans le

sous-ensemble considéré. Cette mesure peut être vue comme une capacité de Choquet d'ordre infini [109]. La fonction Bel vérifie les propriétés suivantes :

1. $Bel(\emptyset) = 0$,
2. $Bel(\Theta) = 1$,
3. $Bel(A_1 \cup \ldots \cup A_n) \geq \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i)$.

La distribution de masse de croyance m et la fonction de crédibilité Bel sont deux représentations équivalentes d'une même information. En effet, la transformation de Möbius [74, 100] permet de calculer la distribution de masses à partir de la distribution de crédibilité à l'aide de la relation suivante :

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B) \tag{2.8}$$

où $|A - B|$ est le cardinal de l'ensemble des éléments de A qui n'appartiennent pas à B.

La fonction de Plausibilité

La fonction de plausibilité Pl peut être définie à l'aide d'une fonction de croyance m de la manière suivante :

$$Pl(A) = \sum_{(A \cap B) \neq \emptyset} m(B) \qquad \forall A \subseteq \Theta. \tag{2.9}$$

$Pl(A)$ s'interprète comme la part de croyance qui pourrait potentiellement être allouée à A, compte tenu des éléments qui ne discréditent pas cette hypothèse, c'est-à-dire toute l'information contenue dans les sous-ensembles ayant une intersection avec A. La fonction de plausibilité Pl vérifie les propriétés suivantes :

1. $Pl(\emptyset) = 0$,
2. $Pl(\Theta) = 1$,
3. $Pl(A_1 \cap \ldots \cap A_n) \leq \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} Pl(\bigcup_{i \in I} A_i)$.

La plausibilité de A est également reliée à la crédibilité du complémentaire de A. Elle correspond à toute l'information ne créditant pas la véracité du complémentaire de A :

$$Pl(A) = 1 - Bel(\overline{A}) \tag{2.10}$$

avec $\overline{A}$ le complémentaire de A. Cette expression exprime la dualité des fonctions Bel et Pl. Ainsi, à l'aide de cette relation et de l'équation (2.8), il est possible de passer de façon unique d'une répartition de plausibilité ou de crédibilité à une répartition de masses.

Fonction de Communalité

Enfin, il existe une fonction appelée *fonction de communalité* déduite de la fonction de masses m. Cette fonction est définie de la manière suivante :

$$Q(A) = \sum_{A \subseteq B} m(B) \quad \forall A \subseteq \Theta. \tag{2.11}$$

Cette fonction est utilisée pour simplifier les calculs engendrés par la combinaison, nous reviendrons plus en détail sur ce point dans la section 2.2. La fonction de communalité vérifie les propriétés suivantes :

1. $Q(\emptyset) = 1$,
2. $Q(\Theta) = m(\Theta)$.

De plus, la communalité peut s'exprimer en fonction de la crédibilité Bel et inversement. Ainsi pour tout $A \subseteq \Theta$, nous avons les relations suivantes :

$$Q(A) = \sum_{B \subseteq A} (-1)^{|B|} Bel(\overline{B})$$
$$Bel(A) = \sum_{B \subseteq \overline{A}} (-1)^{|B|} Q(B).$$

Interprétation des fonctions de plausibilité et de crédibilité

On peut représenter ces deux mesures (crédibilité et plausibilité) par la figure FIG. 2.1. Cette figure explicite les appellations de vraisemblance minimale et maximale dont on affecte parfois la crédibilité et la plausibilité dans le cadre de la théorie de l'évidence. On visualise sur cette figure le fait que la crédibilité regroupe toutes les masses des éléments focaux (A_1, A_2 et A_3) inclus dans le sous-ensemble A_3, alors que la plausibilité correspond à toute les masses intersectant (A_1, A_2, A_3, A_4 et A_5) avec le sous-ensemble considéré A_3.

Si Bel et Pl sont respectivement la crédibilité et la plausibilité d'une même fonction de croyance, elles vérifient les relations suivantes :

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B), \tag{2.12}$$

$$Pl(A \cap B) \leq Pl(A) + Pl(B) - Pl(A \cup B), \tag{2.13}$$

$$0 \leq Bel(A) \leq Pl(A) \leq 1, \tag{2.14}$$

$$Bel(A) + Pl(\overline{A}) = Bel(\Theta) = Pl(\Theta) = 1. \tag{2.15}$$

Les équations (2.12) et (2.13) correspondent respectivement aux propriétés de sur-additivité de la crédibilité et de sous-addivité de la plausibilité. En outre, la quantité $Bel(A) - Pl(A)$ est une

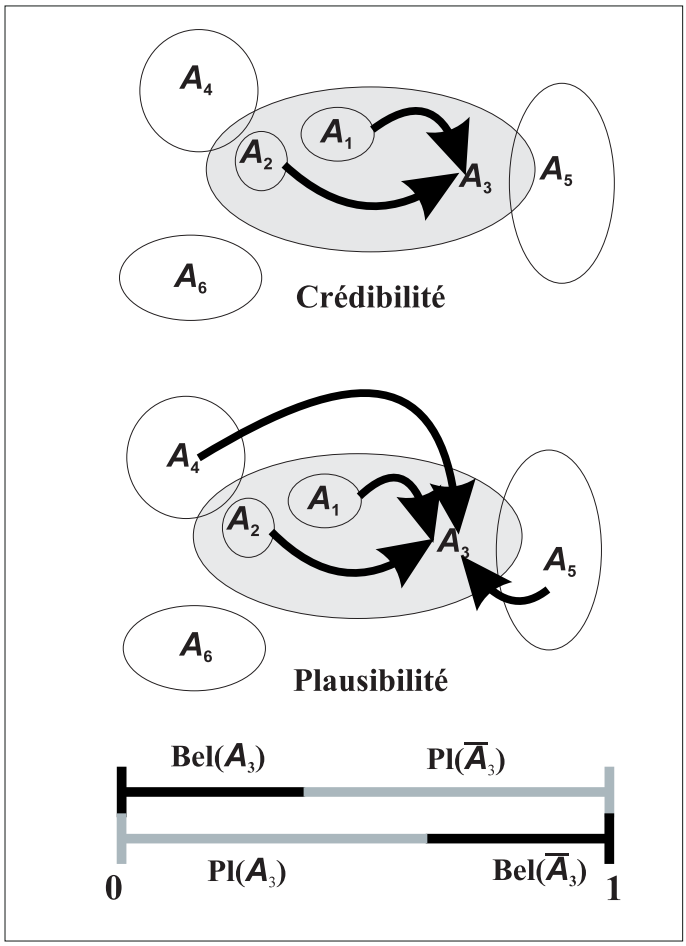

Fig. 2.1 : – Crédibilité et plausibilité d'un point de vue ensembliste.

mesure de l'ignorance relativement à A. Dans le cadre de la théorie de Demspter-Shafer, on peut interpréter l'intervalle défini par $[Bel(A), Pl(A)]$ comme un encadrement de la probabilité réelle de A. Dans ce cadre de travail, $Bel(A)$ et $Pl(A)$ peuvent être vues respectivement comme une *probabilité basse* et une *probabilité haute*. Il faut noter qu'il existe d'autres fonctions symétriques de la crédibilité et de la plausibilité, à savoir *l'incrédulité* (la crédibilité du contraire) et *le doute* (la quantité de masse de croyance restante). Les relations existantes entre ces différentes mesures sont représentées sur la figure FIG. 2.2.

2.1.3 Les fonctions de croyance parmi les mesures floues

Afin d'approfondir le parallélisme existant entre les théories classiques de la reconnaissance de formes, nous présentons un cadre plus général, celui des *mesures floues*, défini par Sugeno [124].

Une mesure floue attribue à tout événement ou sous-ensemble de Θ un coefficient compris entre 0 et 1 qui indique dans quelle mesure on peut penser que l'événement se réalisera. Une mesure floue est une fonction g des parties de Θ, à valeurs dans $[0, 1]$, telle que les axiomes

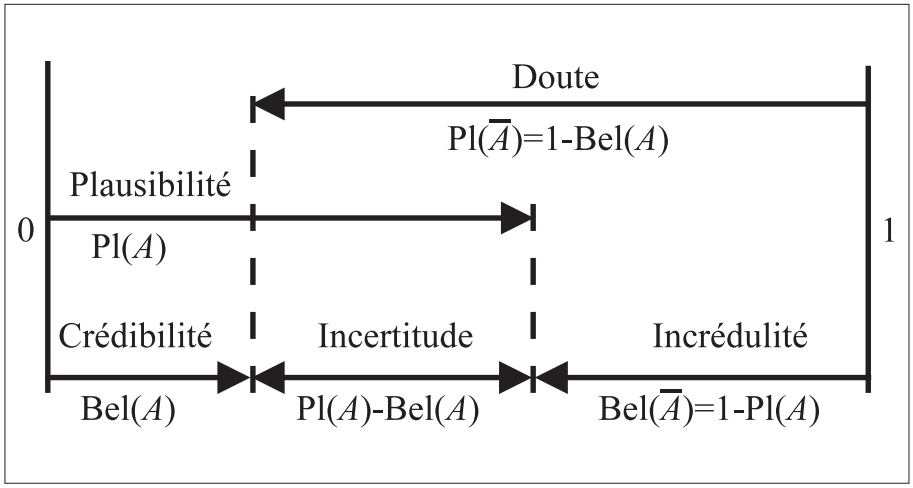

Fig. 2.2 : – Intervalle d'évidence d'un ensemble A.

suivants soient vérifiés :

$$g(\emptyset) = 0 \tag{2.16}$$

$$g(\Theta) = 1 \tag{2.17}$$

$$B \subseteq A \Rightarrow g(B) \leq g(A). \tag{2.18}$$

Le premier axiome (2.16) traduit que la confiance en l'élément vide est nulle. Le second axiome (2.17) définit l'hypothèse du référentiel exhaustif. Enfin, le dernier axiome (2.18) correspond à la propriété de croissance vis-à-vis de l'inclusion, c'est-à-dire que plus l'ensemble d'hypothèses est grand, plus la croyance qu'on lui accorde est grande. Cette propriété de monotonie implique les deux propriétés suivantes :

$$g(A \cup B) \geq max(g(A), g(B)) \tag{2.19}$$

$$g(A \cap B) \leq min(g(A), g(B)). \tag{2.20}$$

Les fonctions de plausibilité et de crédibilité sont des cas particuliers des mesures floues. L'ajoût de contraintes sur les deux propriétés précédentes permet de définir les mesures de probabilité, de nécessité et de possibilité.

En effet, si les mesures floues g_1 et g_2 vérifient les axiomes de monotonie suivants :

$$g_1(A \cup B) = max(g_1(A), g_1(B)) \qquad g_2(A \cap B) = min(g_2(A), g_2(B)) \tag{2.21}$$

alors g_1 est une mesure de possibilité Π et g_2 est une mesure de nécessité N. De plus, si g est une mesure floue qui vérifie l'axiome suivant :

$$\forall A, \quad B \subseteq \Theta \quad g(A \cup B) = g(A) + g(B) - g(A \cap B), \tag{2.22}$$

alors g est une mesure de probabilité P.

Nous allons présenter les relations existantes entre des fonctions de croyance particulières et ces mesures floues.

– *Fonction de croyance bayésienne :*

Une distribution de masses est dite bayésienne si tous ses éléments focaux sont des singletons. Lorsque la masse de croyance est équi-répartie sur les hypothèses singleton, on appelle alors cette distribution de masse *masse bayésienne uniforme*. Les fonctions de crédibilité et de plausibilité sont alors confondues avec la mesure de probabilité P :

$$\forall H_n \subseteq \Theta \quad Bel(\{H_n\}) = Pl(\{H_n\}) = P(\{H_n\}) \tag{2.23}$$

– *Fonction de croyance consonante :*

Une fonction de croyance m est consonante lorsque les éléments focaux sont emboîtés au sens de l'inclusion :

$$\mathcal{F}(m) = \{A_k\}_{k=1}^{K} \qquad A_1 \subseteq \ldots \subseteq \ldots \subseteq A_K. \tag{2.24}$$

Nous obtenons alors pour une fonction de crédibité Bel et de plausibilité Pl consonante :

$$\forall\, A, B \subseteq \Theta \quad Bel(A \cap B) = min(Bel(A), Bel(B)) \tag{2.25}$$

$$\forall\, A, B \subseteq \Theta \quad Pl(A \cup B) = max(Pl(A), Pl(B)). \tag{2.26}$$

Dans ce cas, les mesures de crédibilité et de plausibilité sont strictement équivalentes respectivement aux mesures de nécessité N et de possibilité Π. Notons qu'une explication plus détaillée des relations existantes entre ces mesures est présentée dans [122].

Les relations entre ces mesures floues sont illustrées dans la figure FIG. 2.3. Afin de présenter la mise en œuvre de ces mesures floues, nous pouvons reprendre un exemple issu de [8].

Considérons le jeu de roulette et soit un ensemble Θ comprenant l'ensemble fini des chiffres susceptibles de sortir. On attribue des masses de croyance en plaçant une proportion x des jetons sur une case ou un groupe de cases.

– *Situation n°1 :*

Si l'on croit que soit le 11, soit le 12 va sortir à coup sûr, alors on peut miser x de ses jetons sur le 11 et $(1-x)$ sur le 12. Nous avons alors, $m(\{11\}) = x$ et $m(\{12\}) = 1 - x$. Les éléments focaux sont alors des singletons et on se place dans le cadre de la théorie des

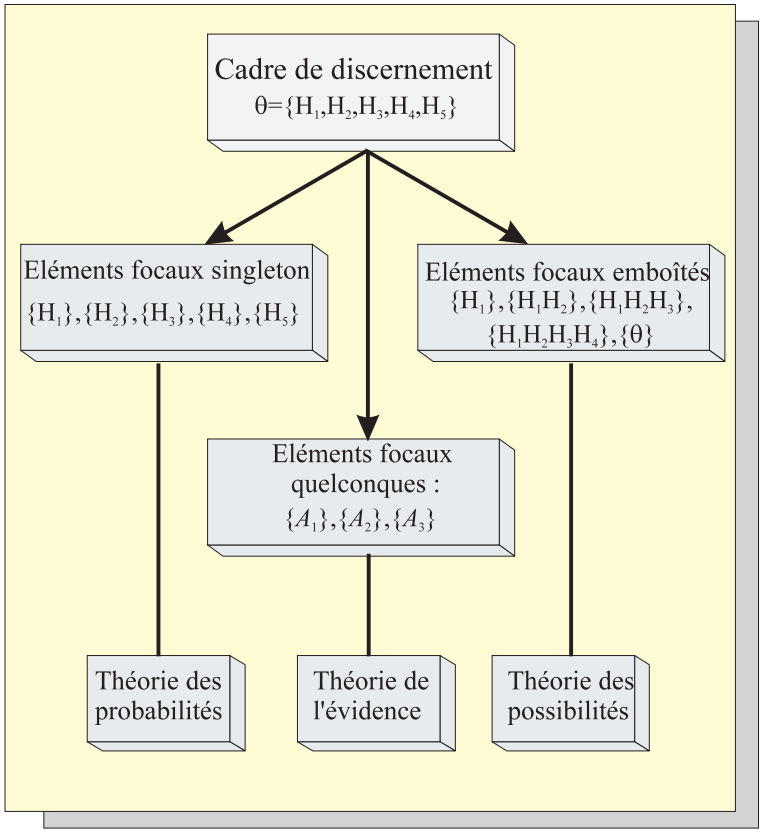

FIG. 2.3 : – Liens entre différentes mesures floues.

probabilités. Dans ce contexte, nous obtenons :

$$\begin{array}{c} P(\{11\}) = x \\ P(\{12\}) = 1 - x \\ P(\{Impairs\}) = P(\{1,3,5,\ldots\}) = x \\ P(\{Pairs\}) = P(\{2,4,6,\ldots\}) = 1 - x \\ P(\{11,12\}) = 1. \end{array} \tag{2.27}$$

– *Situation n°2 :*

Si l'on croit que le chiffre 13 va sortir, on peut attribuer x de ces jetons à la case 13 et le reste à la case "impair", donc $m(\{13\}) = x$ et $m(\{1,3,5,\ldots\}) = 1 - x$. Les éléments focaux sont alors emboîtés puisque $\{13\} \subseteq \{1,3,5,\ldots\}$ et on se trouve dans le cadre de la théorie des possibilités, avec :

$$\begin{array}{c} N(\{13\}) = Bel(\{13\}) = x \\ \Pi(\{13\}) = Pl(\{13\}) = 1 \\ N(A) = \Pi(A) = P(A) = Bel(A) = Pl(A) = 1 \quad avec \quad A = \{1,3,5,\ldots\} \\ x \leq P(\{13\}) \leq 1. \end{array} \tag{2.28}$$

– *Situation n°3 :*

Si l'on n'a que peu d'avis sur le chiffre qui va sortir, on peut miser par exemple 40% de ses jetons sur "impair", 10% sur "pair", 30% sur le 1 et 20% sur le 12, ce qui nous donne

$m(\{1,3,5,\ldots\}) = 0.4$, $m(\{2,4,6,\ldots\}) = 0.1$, $m(\{1\}) = 0.3$ et $m(\{12\}) = 0.2$. Nous avons alors des éléments focaux quelconques et nous pouvons déterminer la confiance dans la sortie de chacun des chiffres à l'aide de fonctions de crédibilité ou de plausibilité. Nous obtenons alors :

$$\begin{array}{ll} Bel(\{1\}) = 0.3 & Pl(\{1\}) = 0.7 \\ Bel(\{2\}) = 0 & Pl(\{2\}) = 0.1 \\ \vdots & \vdots \\ Bel(\{12\}) = 0.2 & Pl(\{12\}) = 0.3. \end{array} \tag{2.29}$$

Dans ce contexte, nous avons donc :

$$\begin{array}{c} 0.3 \leq P(\{1\}) \leq 0.7 \\ 0 \leq P(\{2\}) \leq 0.1 \\ \vdots \\ 0.2 \leq P(\{12\}) \leq 0.3. \end{array} \tag{2.30}$$

En conclusion, l'utilisation de ces théories suppose des niveaux de connaissances complètement différents sur les événements susceptibles de se produire. Pour la théorie des probabilités, les avis sont très nets et divergents. Dans ce cas, on peut attribuer des degrés de confiance en chacun des événements. Pour la théorie des possibilités, les avis sont cohérents mais imprécis, la répartition des degrés de confiance se fera sur des intervalles. Pour la théorie de l'évidence, les avis peuvent être incohérents et imprécis. L'affectation des degrés de confiance pourra alors se faire sur des ensembles quelconques. L'emploi de cette théorie se fait donc sans connaissance *a priori* sur les événements qui sont susceptibles de se produire.

2.2 Combinaison des croyances

La *combinaison*, ou *révision*, des croyances intervient lorsque l'on dispose de nouvelles informations codées sous forme de fonctions de croyance qu'il faut fusionner avec les fonctions de croyance déjà existantes. La fusion d'informations se présente comme une solution permettant d'accéder à une information plus fiable ou à une synthèse des connaissances dans un environnement multi-sources. De façon générale, elle offre de nombreux avantages parmi lesquels la complémentarité et la redondance de l'information.

Après avoir abordé la combinaison de fonctions de croyance, nous présenterons la règle de *conditionnement*.

2.2.1 Combinaison conjonctive de fonctions de croyance

Soient S_1 et S_2 deux sources d'informations indépendantes. Ces sources sont supposées totalement fiables. Les informations obtenues, via ces sources, sont modélisées à l'aide de fonctions

de croyance appelées respectivement m_1 et m_2. On note $m_\cap$ la fonction de croyance résultat de la combinaison conjonctive de m_1 et m_2 :

$$m_\cap = m_1 \cap m_2. \tag{2.31}$$

La combinaison conjonctive de ces deux fonctions de croyance est définie par :

$$m_\cap(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \quad \forall A \subseteq \Theta. \tag{2.32}$$

La règle de combinaison conjonctive (2.32) vérifie un certain nombre de propriétés intéressantes. Elle est associative, commutative et possède un élément neutre.

Par associativité, on peut généraliser cette règle à J sources d'information produisant J fonctions de croyance. En notant $m_\cap = \bigcap_{j=1}^{J} m_j$, on obtient :

$$m_\cap(A) = \sum_{A_1 \cap \ldots \cap A_J = A} \left(\prod_{j=1}^{J} m_j(A_j) \right) \quad \forall A \subseteq \Theta. \tag{2.33}$$

De façon générale, cette règle de combinaison produit une fonction de croyance non normalisée, c'est-à-dire que $m_\cap(\emptyset) \neq 0$. Cette hypothèse n'est pas possible dans le cas du monde fermé (cf. section 2.1). Il est donc nécessaire d'introduire une étape de normalisation. Cette loi de combinaison conjonctive normalisée est plus connue sous le nom de *règle de combinaison de Dempster*. La règle de Dempster, aussi appelée *somme orthogonale*, de deux fonctions de croyance m_1 et m_2 est notée de la manière suivante :

$$m_\oplus = m_1 \oplus m_2. \tag{2.34}$$

On définit la somme orthogonale de m_1 et m_2 par :

$$m_\oplus(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C) \quad \forall A \subseteq \Theta, A \neq \emptyset \tag{2.35}$$

où K est défini par :

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C). \tag{2.36}$$

Dans l'équation (2.35), K reflète la masse de croyance conflictuelle existante entre les deux fonctions de croyance à combiner. Cette masse varie dans $[0, 1]$ selon les fonctions m_1 et m_2. Lorsque K est nul, les sources sont en parfait accord. Au contraire, lorsque cette masse est égale à 1, les sources sont en conflit total et les informations ne peuvent pas être fusionnées. De la même manière que pour la combinaison conjonctive, nous pouvons généraliser la combinaison de

Dempster à J fonctions de croyance en notant $m_\oplus = \bigoplus_{j=1}^{J} m_j$. La fonction $m_\oplus$ est alors définie de la manière suivante :

$$m_\oplus(A) = \frac{1}{1-K} \sum_{A_1 \cap \ldots \cap A_J = A} \left(\prod_{j=1}^{J} m_j(A_j) \right) \quad \forall A \subseteq \Theta, A \neq \emptyset \tag{2.37}$$

avec :

$$K = \sum_{A_1 \cap \ldots \cap A_J = \emptyset} \left(\prod_{j=1}^{J} m_j(A_j) \right). \tag{2.38}$$

La règle de combinaison de Dempster vérifie un certain nombre de propriétés intéressantes comme l'associativité et la commutativité. De plus, cette combinaison possède un élément neutre qui est la fonction d'ignorance totale :

$$m_1 \oplus m_2 = m_1 \quad \text{si} \quad m_2(\Theta) = 1. \tag{2.39}$$

L'élément absorbant de la somme orthogonale est la fonction de certitude totale :

$$m_1 \oplus m_2 = m_2 \quad \text{si} \quad m_2(H_n) = 1. \tag{2.40}$$

Afin de faciliter le calcul de la combinaison, on utilise la fonction de *communalité*. Dans ce cas, le résultat de la combinaison de Dempster m de deux fonctions de croyance m_1 et m_2 peut s'exprimer à l'aide de leurs fonctions de communalité respectives Q_1 et Q_2 par la relation suivante :

$$Q = Q_1 Q_2. \tag{2.41}$$

A l'aide de la transformée de Möbius (cf. equation 2.8), on peut revenir à partir de la fonction de communalité à une distribution de masse :

$$m(A) = \sum_{A \subseteq B} (-1)^{|B-A|} Q(B). \tag{2.42}$$

Ainsi, on peut définir indifféremment une fonction de croyance à l'aide d'une distribution de masse, de plausibilité, de crédibilité et de communalité.

Enfin, la règle de combinaison de Dempster a fait l'objet de nombreuses critiques [117, 134, 147]. La plupart de ces critiques sont liées à la normalisation. Smets [120] propose même l'abandon de cette normalisation. Dans ce cas, on adopte l'hypothèse du monde ouvert et on admet ne pas connaître nécessairement toutes les hypothèses possibles, générant ainsi une masse sur l'ensemble vide qui ne pourra être redistribuée par la suite qu'à l'aide d'un conditionnement. Nous reviendrons sur ces critiques ainsi que sur les solutions qui ont été proposées dans la littérature dans le chapitre 4.

2.2.2 Conditionnement

Le conditionnement consiste à réviser une croyance m initiale quelconque lorsqu'une proposition $A \subseteq \Theta$ est devenue vraie. On note $m(.|A)$ le conditionnement de m par A. Le but du conditionnement est de définir la redistribution de la masse d'un sous-ensemble $B \subseteq \Theta$. Le conditionnement d'une fonction m sur $A \subseteq \Theta$ définie dans le monde fermé donne la fonction de croyance conditionnelle $m(B|A)$ ainsi définie :

$$\begin{cases} m(B|A) = k \sum\limits_{C \subseteq \overline{A}} m(B \cup C) & \forall B \subseteq A, \\ \\ m(B|A) = 0 & \forall B, \quad B \cap \overline{A} \neq \emptyset \end{cases} \tag{2.43}$$

avec :

$$k = \frac{1}{1 - \sum\limits_{B \subseteq A} m(B)}. \tag{2.44}$$

Ce conditionnement est aussi appelé règle de conditionnement de Dempster. La normalisation par le facteur k correspond à la redistribution de la masse incompatible avec A sur l'ensemble $B \subseteq A$. Idéalement, cette masse ne devrait pas exister car elle est en contradiction avec l'information nouvelle. Lorsque la masse de croyance accordée à l'élément inverse de A est égale à 1, c'est-à-dire lorsque $Bel(\overline{A}) = 1$, on ne peut pas appliquer la normalisation et le conditionnement n'est pas possible car l'information nouvelle est complètement incompatible avec les informations reçues précédemment. Les expressions des crédibilités et des plausibilités sont données par les relations suivantes :

$$Bel(B|A) = \frac{Bel(B \cup \overline{A}) - Bel(\overline{A})}{1 - Bel(\overline{A})} \tag{2.45}$$

$$Pl(B|A) = \frac{Pl(B \cup A)}{Pl(A)}. \tag{2.46}$$

On peut remarquer une certaine similarité entre la dernière équation et la règle de Bayes. La règle de Dempster généralise la règle de Bayes aux fonctions de croyance. Lorsque les fonctions de croyance sont bayésiennes alors les deux règles sont équivalentes.

2.3 Le modèle des croyances transférables

Des modifications et des généralisations ont été apportées à la théorie initiale de Shafer dans différentes directions [79, 142], occasionnant ainsi une certaine confusion dans la théorie dite de "Dempster-Shafer". Cette appellation regroupe en fait 3 points de vue distincts qui sont des extensions ou des variantes du modèle proposé par Shafer. Le premier repose sur la théorie des

probabilités imprécises appelée aussi théorie des probabilités inférieures [136]. Cette théorie suppose l'existence d'une mesure de probabilité P précise, mais non parfaitement connue. On note alors $\mathcal{P}r$ l'ensemble des probabilités compatibles avec l'information disponible. Le modèle peut alors être défini soit directement par $\mathcal{P}r$, soit par ses enveloppes inférieure et supérieure. Sous certaines conditions, la probabilité inférieure peut alors être considérée comme une fonction de croyance. Le second modèle proposé par Dempster [32, 33, 34] est un cas particulier des probabilités imprécises. A l'aide de ce modèle, on peut là aussi définir des probabilités inférieure et supérieure. Ce modèle peut être rapproché des *hints* présentés par Monney et Kohlas [79]. L'interprétation logique de la théorie de l'évidence proposé par Cholvy [22] peut être vue comme un cas particulier du modèle de Kohlas. Enfin le dernier modèle, présenté par Smets [114, 116, 123], est une extension des fonctions de croyance, proposée au travers du modèle des croyances transférables (*Transferable Belief Model*). L'approche proposée par Smets se distingue des approches précédentes de par son caractère fondamentalement non probabiliste. De plus, elle offre une justification axiomatique cohérente des principaux concepts de la théorie de l'évidence et a permis de clarifier le lien existant entre la représentation des croyances [119] et la prise de décision. Nous évoquons ici le point de vue non probabiliste de Smets.

La démarche proposé par Smets peut se décomposer en deux niveaux. Un niveau *crédal* dans lequel s'effectue la modélisation et la combinaison des croyances et un niveau *pignistique* utilisé lors de la prise de décision.

Le niveau crédal est composé principalement de deux étapes distinctes. La première, la modélisation, correspond à la partie statique du modèle. Cette étape permet la représentation des connaissances sous la forme de fonctions de croyance (section 2.1). La seconde étape représentant la partie dynamique, correspond à la combinaison des croyances aussi appelée révision. Smets [114], au contraire de Shafer [109], voit la combinaison comme un cas particulier du conditionnement de Dempster (section 2.2.2).

De plus, la règle de combinaison, notée $\oplus$ doit répond à un certain nombre d'axiomes primordiaux [47, 76, 114] :

- **Axiome** A_1 : $(m_1 \oplus m_2)(A)$ doit être une fonction de A, m_1 et m_2 seulement,
- **Axiome** A_2 : $\oplus$ doit être commutative,
- **Axiome** A_3 : $\oplus$ doit être associative,
- **Axiome** A_4 : symétrie : $m_1 \cap m_2 = m_2 \cap m_1$
- **Axiome** A_5 : si $m_2(A) = 1$ alors $m_{\oplus}(.) = m_1(.|A)$.

Ces axiomes sont définis de la manière suivante. L'axiome A_1, la composition, traduit l'indé-

pendance de sources. Les axiomes A_2 et A_3 traduisent la monotonie, c'est-à-dire que l'ordre de fusion des sources n'a pas d'importance. L'axiome A_4 indique que le résultat de la combinaison ne doit pas être modifié par une permutation des éléments focaux. Enfin, l'axiome A_5, basé sur le conditionnement, est une condition de compatibilité dans le cas particulier de la combinaison avec une fonction de croyance monofocale.

Ainsi, à partir de ces axiomes et de la règle de conditionnement, Smets [114] a montré l'unicité de la règle de combinaison de Dempster. De plus, on peut montrer que :

$$(m_1 \oplus m_2)(.) = (m_1 \cap m_2)(.|\Theta). \tag{2.47}$$

Ainsi la combinaison de Dempster correspond à une combinaison conjonctive suivie d'un conditionnement sur Θ.

2.4 La décision

Le modèle des croyances transférables proposé par Smets se différencie des autres approches par le cloisonnement existant entre la modélisation des connaissances et la prise de décision. En effet, au niveau *crédal* s'effectue la modélisation (partie statique) et la révision (partie dynamique) des connaissances en écartant toute approche probabiliste car seules les connaissances certaines sont prises en compte. La décision s'effectue au niveau dit *pignistique*[1], qui impose une transformation des fonctions de croyance. En effet, dans la théorie bayésienne, décider à l'aide de connaissances incertaines revient à un *pari* sur l'hypothèse la plus probable [30]. Il est alors nécessaire de transformer les fonctions de croyance en distributions de probabilité. Le résultat de cette transformation est appelé *probabilité pignistique*. Nous verrons, dans cette section, la construction de la probabilité pignistique ainsi que les règles de décision qui lui sont associées (Section 2.4.1). Nous présentons ensuite d'autres règles de décision utilisées dans le cadre de la théorie des croyances (Section 2.4.2).

2.4.1 Probabilité pignistique : le principe de raison insuffisante

Soit m, une fonction de croyance synthétisant l'ensemble des connaissances à l'instant où la décision doit être prise. Le problème, qui se pose, peut être présenté de la manière suivante. Comment, afin d'utiliser la règle de décision de Bayes, peut-on transformer m en une distribution de probabilité ?

Smets [115] se base alors sur le principe de raison insuffisante. C'est-à-dire qu'en l'absence de raison de privilégier une hypothèse plus qu'une autre, on suppose que les hypothèses sont

[1] Le terme *pignistique* a été introduit par Smets. Il provient du latin *pignus* signifiant pari.

équi-probables. Ainsi, pour tout élément focal $A \in \mathcal{F}(m)$, la masse $m(A)$ sera redistribuée uniformément sur les éléments de A. On obtient alors une distribution de probabilité particulière dite probabilité pignistique et notée $BetP$. Cette probabilité est obtenue de la manière suivante :

$$BetP(H_n) = \sum_{A \subseteq \Theta} \frac{|H_n \cap A|}{|A|} m(A) \qquad \forall H_n \in \Theta. \tag{2.48}$$

Ainsi, on associe une probabilité pignistique à une seule fonction de croyance. Inversement, on associe une infinité de fonctions de croyance à une distribution pignistique. Cela reflète la perte d'information occasionnée au passage entre le niveau crédal et le niveau pignistique. En effet, la transformation d'une fonction de croyance en distribution de probabilité s'accompagne de contraintes telles que :

- l'affectation d'une probabilité à chaque élément de Θ,
- le respect du principe d'additivité.

Ces contraintes font perdre les avantages liés aux fonctions de croyance.

Une fois la distribution de probabilité construite, on utilise, de manière classique, la décision bayésienne qui préconise l'action pour laquelle l'espérance du coût est la plus faible. Ainsi, on peut définir l'espérance pignistique d'une fonction $f : \Theta \to \mathbb{R}$ comme son espérance mathématique relativement à $BetP$:

$$E_{bet}(f) = \sum_{H_n \in \Theta} f(H_n) BetP(H_n). \tag{2.49}$$

Soit l'individu de vecteur forme x qui est à l'origine de la fonction de croyance m définie sur l'ensemble des hypothèses de Θ. Alors, de la même manière que dans le cadre de la théorie bayésienne de la décision (cf. Annexes), nous pouvons définir $\mathcal{A} = \{a_1, \ldots, a_N\}$ un ensemble de N actions possibles et une fonction de coût $\lambda : \mathcal{A} \times \Theta \to \mathbb{R}$ de telle manière que $\lambda(a_i|H_n)$ représente le coût encouru si l'on choisit l'action a_i alors que la vraie classe est H_n. Le risque conditionnel de décider a_i sachant x est défini par :

$$R_{Bet}(a_i|x) = \sum_{H_n \in \Theta} \lambda(a_i|H_n) BetP(H_n). \tag{2.50}$$

L'action $a \in \mathcal{A}$ qui minimise ce risque est celle qui sera retenue. Dans le cas de coûts $\{0, 1\}$, la minimisation du risque conditionnel revient à choisir l'hypothèse de plus grande probabilité pignistique. Notons, que les principes de coûts avec rejet, définis dans la décision bayésienne, peuvent être utilisés comme règle de décision.

2.4.2 Autre règles de décision

Bien que seule la transformation des fonctions de croyance en probabilité pignistique et l'utilisation de la théorie de la décision bayésienne aient été justifiées, d'autres principes de

décision ont vu le jour.

Parmi eux, le maximum de crédibilité ou le maximum de plausibilité qui sélectionnent respectivement l'hypothèse la plus crédible et l'hypothèse la plus plausible. Ce dernier critère est préconisé dans [5]. Une étude sur le choix entre une décision basée sur la probabilité pignistique et sur le maximum de plausibilité est présentée dans [96].

De plus, de par l'interprétation probabiliste de la théorie des croyances, des risques conditionnels autres que celui associé à la probabilité pignistique peuvent être employés. En effet, la définition d'enveloppes inférieure et supérieure, respectivement représentées par la crédibilité et la plausibilité, pour encadrer la véritable probabilité aboutit à la construction de risques inférieur et supérieur. Ces risques, étudiés dans [21, 36], sont définis par :

$$R_*(a_i|x) = \sum_{A \subseteq \Theta} m(A) \min_{H_n \in A} \lambda(a_i|H_n) \tag{2.51}$$

$$R^*(a_i|x) = \sum_{A \subseteq \Theta} m(A) \max_{H_n \in A} \lambda(a_i|H_n). \tag{2.52}$$

L'équation (2.51) définit le risque inférieur. La minimisation de ce risque correspond à une stratégie "optimiste". L'équation (2.52) définit le risque supérieur. La minimisation de ce risque correspond à une stratégie dite "pessimiste". La distribution pignistique étant une distribution particulière entre les enveloppes inférieure et supérieure, nous obtenons alors la relation d'inégalité suivante qui lie les équations (2.50), (2.51) et (2.52) :

$$R_*(a_i|X) \leq R_{Bet}(a_i|X) \leq R^*(a_i|X). \tag{2.53}$$

Dans le cas de coûts $\{0, 1\}$, la minimisation des risques inférieur et supérieur revient respectivement à une prise de décision basée soit sur le maximum de plausibilité soit sur le maximum de crédibilité.

Tous les critères de décision ont été définis de manière à prendre une décision sur une hypothèse simple, c'est-à-dire une hypothèse singleton H_n de l'ensemble Θ. Ceci est valable lorsqu'on souhaite prendre une décision qui soit précise. Malheureusement, il peut arriver que cette précision soit au détriment de la fiabilité de la décision. En effet, il peut parfois être préférable et plus prudent de choisir un ensemble d'hypothèses plutôt que choisir une hypothèse singleton. Il y a peu d'application où les critères autorisent les hypothèses composées [19]. Toutefois, si l'on s'autorise à prendre comme décision des hypothèses composées, quelques précautions doivent être prises. On peut utiliser les stratégies de décision définies précédemment, cependant les décisions seront toujours en faveur des hypothèses composées par rapport aux hypothèses singletons, car de manière générale, plus le cardinal de l'ensemble est grand, plus les fonctions de croyance

sont importantes. Ceci mène donc à imposer des contraintes supplémentaires, comme le cardinal maximal de la solution, ou un seuil sur le critère de décision.

2.5 Opérations sur les fonctions de croyance

2.5.1 Affaiblissement

On appelle *affaiblissement* d'une fonction de croyance m l'opération qui, pour tout $A \subseteq \Theta$, déplace une partie de la croyance de A vers un sous-ensemble $B \subseteq \Theta$ tel que $A \subseteq B$. Cela revient à une augmentation des plausibilités de chaque sous-ensemble et donc à une diminution de leur crédibilité, augmentant ainsi l'incertitude générale. On réduit ainsi le degré de certitude des éléments focaux. L'intérêt de l'affaiblissement est de maîtriser l'influence des sources selon leur fiabilité avant de les combiner.

L'affaiblissement le plus simple a été proposé par Shafer [109]. Il consiste à déplacer proportionnellement une part de la masse de croyance affectée aux éléments focaux vers l'ensemble Θ représentant l'incertitude. Soit une fonction de croyance m fournie par une source S et un coefficient α, avec $\alpha \in [0, 1]$, qui représente le degré de confiance que l'on accorde à la source S. Nous obtenons alors le formalisme suivant :

- $\alpha = 0$ signifie une remise en cause totale de l'efficacité de S,
- $\alpha = 1$ signifie une confiance absolue en la source S.

On note m_α la fonction de croyance m affaiblie par le facteur $(1 - \alpha)$. Cette fonction est définie ainsi :

$$\begin{cases} m_\alpha(A) &= \alpha m(A) \\ m_\alpha(\Theta) &= 1 - \alpha + \alpha m(\Theta). \end{cases} \tag{2.54}$$

D'autres types d'affaiblissement peuvent être mis en œuvre. Nous reviendrons sur cette notion d'affaiblissement et plus particulièrement sur la définition du coefficient α dans le chapitre 3.

2.5.2 Les mesures d'incertitude des fonctions de croyance

Les mesures d'incertitude relatives à un événement caractérisent la nature de l'information, celle-ci pouvant être imprécise, fragmentaire ou peu fiable. Les premières mesures d'incertitude (aussi appelée mesures d'information) ont été développées par Hartley [68] en théorie des ensembles, puis par Shannon [110] dans le cadre de la théorie des probabilités. Avec l'émergence de nouvelles théories (théorie des possibilités et théorie des croyances), le concept d'incertitude a été étendu. Il existe principalement deux types de mesures d'incertitude :

- la *non-spécificité* (ou *imprécision*), qui dépend de la taille des différentes propositions sur lesquelles la masse de croyance est répartie,

– la *discorde* qui quantifie le conflit entre les différentes propositions.

De nombreux travaux ont été réalisés sur les mesures d'incertitude dans le cadre de la théorie de l'évidence aussi bien au niveau de la définition de ces mesures [45, 70, 139] qu'au niveau des propriétés de ces mesures [49, 78]. Ces travaux ont abouti à la définition de nombreuses mesures d'incertitude. Dans cette étude, nous allons nous restreindre à présenter celles qui nous semblent avoir les propriétés les plus intéressantes.

Non-Spécificité : La non-spécificité a été introduite dans un premier temps par Hartley [68] pour mesurer l'imprécision d'un ensemble A. Cette imprécision est alors définie par :

$$U(A) = \log_2 |A| \tag{2.55}$$

avec $|A|$ le cardinal de l'ensemble A. Dans l'equation (2.55), la fonction U est appelé *fonction de Hartley*. Une généralisation, réalisée par Dubois et Prade [45], de cette fonction dans le cadre de la théorie de l'évidence a permis de définir la non-spécificité. La mesure de non-spécificité d'une fonction de croyance m peut alors s'écrire :

$$NS(m) = \sum_{A \in \mathcal{F}(m)} m(A) \log_2 |A|. \tag{2.56}$$

Dans [101], il est montré que la fonction NS est l'unique fonction, conditionnelle à certains axiomes, de non-spécificité possible dans le cadre de la théorie de l'évidence. Cette fonction représente une moyenne pondérée de la fonction de Hartley pour l'ensemble des éléments focaux. Les poids sont les masses de croyance accordées à chacun des éléments focaux. Pour chaque élément focal A, $m(A)$ indique le degré de croyance centré sur A, alors que $\log_2 |A|$ indique le manque de spécificité de cette part de croyance. Plus l'ensemble A est grand, moins la croyance accordée à cet ensemble est spécifique. Les valeurs de la fonction sont comprises dans l'intervalle $[0, \log_2 |\Theta|]$. Le maximum de cette fonction, $NS(m) = \log_2 |\Theta|$, est obtenu pour une fonction de croyance d'ignorance totale c'est-à-dire $m(\Theta) = 1$. Le minimum, $NS(m) = 0$, est atteint quand la fonction de croyance m est une fonction de croyance bayésienne quelconque, c'est-à-dire lorsque les éléments focaux sont des singletons.

De plus, soit une fonction de croyance m issue de la combinaison de Dempster de J fonctions de croyance :

$$m_{\oplus} = \bigoplus_{j=1}^{J} m_j. \tag{2.57}$$

Alors, la non-spécificité d'une fonction de croyance $m_{\oplus}$ issue d'une combinaison de Dempster est inférieure ou égale à la plus petite des non-spécificités associées à chacune des fonctions de

masses agrégées pour l'obtention de $m_{\oplus}$. Nous pouvons alors écrire :

$$NS(m_{\oplus}) \leq \min_{j=1,\ldots,J} NS(m_j). \tag{2.58}$$

Cette propriété est vraie uniquement lorsque la combinaison employée est basée sur un opérateur conjonctif. En effet, le but des opérateurs conjonctifs est de déplacer la masse sur l'intersection de sous-ensembles. Ces intersections étant plus petites que les ensembles d'origine, la fonction de masse résultante de la combinaison est alors au moins aussi spécifique que la plus spécifique des fonctions de masse.

Enfin, on peut remarquer que quelque soit la distribution de masse bayésienne, qu'elle soit uniforme ou certaine par exemple, la non-spécificité sera toujours nulle. Toutes les fonctions de croyance bayésiennes sont totalement spécifiques. Cette mesure ne permet donc pas de différencier les fonctions de masse bayésiennes. Pourtant, l'entropie de Shannon, applicable uniquement dans le cadre de la théorie des probabilités, définit une mesure d'incertitude. En fait, cette grandeur quantifie le conflit.

Mesures de conflit : L'entropie de Shannon S, qui est définie uniquement pour les mesures de probabilité, peut s'appliquer sur une fonction de croyance bayésienne m. Dans ce cas, cette entropie s'écrit :

$$S(m) = -\sum_{H_n \in \Theta} m(H_n) \log_2(m(H_n)). \tag{2.59}$$

L'entropie de Shannon peut s'écrire aussi sous le forme suivante :

$$S(m) = -\sum_{H_n \in \Theta} m(H_n) \log_2 \left[1 - Confl_m(H_n)\right] \tag{2.60}$$

avec :

$$Confl_m(H_n) = \sum_{\substack{H_{n'} \in \Theta \\ H_{n'} \neq H_n}} m(H_{n'}). \tag{2.61}$$

Dans l'équation (2.60), le terme $Confl_m(H_n)$ représente la croyance totale attribuée aux éléments focaux différents de l'élément focal H_n. Cette grandeur exprime toute la croyance en conflit total avec la masse soutenant l'événement H_n. L'entropie de Shannon est donc une moyenne pondérée, par les masses de croyance, du conflit existant entre les différentes propositions d'une fonction de croyance bayésienne. La généralisation de l'entropie de Shannon à la théorie des croyances a conduit à la définition de nombreuses mesures. Les mesures de conflit les plus couramment employées pour une fonction de croyance m sont : la dissonance $E(m)$ [139], la confusion

$C(m)$ [70, 71] et la discordance $D(m)$ [77]. Ces mesures sont définies de la manière suivante :

$$E(m) = -\sum_{A\in\mathcal{F}(m)} m(A)\log_2 Pl(A), \tag{2.62}$$

$$C(m) = -\sum_{A\in\mathcal{F}(m)} m(A)\log_2 Bel(A), \tag{2.63}$$

$$D(m) = -\sum_{A\in\mathcal{F}(m)} m(A)\log_2\left[\sum_{B\in\mathcal{F}(m)} m(B)\frac{|A\cap B|}{|B|}\right]. \tag{2.64}$$

$$\tag{2.65}$$

D'après les relations d'ordre, on peut montrer facilement que pour tout m, $E(m) \leq D(m) \leq C(m)$. Une autre alternative à la fonction de discordance a été présentée par Klir [78]. Cette fonction s'appelle *strife* et est définie par :

$$ST(m) = -\sum_{A\in\mathcal{F}(m)} m(A)\log_2\left[\sum_{B\in\mathcal{F}(m)} m(B)\frac{|A\cap B|}{|A|}\right]. \tag{2.66}$$

Ces différentes mesures de conflit peuvent être réécrites à l'aide d'une forme générale MC en fonction d'un degré de conflit existant entre les éléments focaux comme pour l'entropie de Shannon (2.60). Cette expression générale prend alors la forme suivante :

$$MC(m) = -\sum_{A\in\mathcal{F}(m)} m(A)\log_2(1 - Confl_m(A)). \tag{2.67}$$

La grandeur $Confl_m(A)$ représente le conflit existant entre la masse accordée à l'ensemble A et le reste de la croyance de m. Cette valeur prend plusieurs expressions selon les mesures de conflit utilisées. Le conflit $Confl_m(A)$ s'exprime, selon les mesures utilisées, par :

- Dissonance : $Confl_m(A) = \sum_{A\cap B=\emptyset} m(B)$,
- Confusion : $Confl_m(A) = \sum_{A\not\subseteq B} m(B)$,
- Discordance : $Confl_m(A) = \sum_{B\in\mathcal{F}(m)} m(B)\frac{|B-A|}{|B|}$,
- Strife : $Confl_m(A) = \sum_{B\in\mathcal{F}(m)} m(B)\frac{|A-B|}{|A|}$.

D'autres auteurs définissent une distance D entre les éléments focaux. Cette distance quantifie la différence existant entre les éléments focaux. Pour George et Pal [63], cette distance peut être définie de la manière suivante :

$$D(A,B) = 1 - \frac{|A\cap B|}{|A\cup B|} \quad \forall A, B \subseteq \Theta. \tag{2.68}$$

On peut alors définir une mesure de conflit au sein d'une structure telle que :

$$\Delta(m) = -\sum_{A\in\mathcal{F}(m)} m(A)\log_2\left(1 - \sum_{B\in\mathcal{F}(m)} m(B)D(A,B)\right). \tag{2.69}$$

On retrouve alors l'expression de l'équation (2.67), avec :

$$Confl_m(A) = \sum_{B \in \mathcal{F}(m)} m(B)D(A,B). \tag{2.70}$$

Il existe une relation d'ordre reliant les différentes mesures de conflit qui ont été présentées :

$$E(m) \leq D(m) \leq \Delta(m) \qquad et \qquad E(m) \leq ST(m) \leq \Delta(m). \tag{2.71}$$

Ces mesures sont bornées par des valeurs identiques : $MC(m) \in [0, \log_2 |\Theta|]$. Pour ces mesures, le minimum est atteint pour une fonction de croyance certaine, et le maximum pour une fonction de croyance bayésienne. Enfin, ces cinq mesures de conflit généralisent l'entropie de Shannon. Dans le cas particulier d'une fonction de croyance bayésienne, elles se réduisent à l'entropie de Shannon. De plus, le conflit mesuré par chacune de ces fonctions prend bien en compte le degré d'inclusion (et donc la différence) entre les éléments focaux présents au sein du noyau. Ces mesures semblent donc similaires du point de vue de leurs constructions. Il est donc difficile de justifier que l'une d'entre-elles est meilleure qu'une autre.

Nous avons, jusqu'a présent, défini les deux types de mesures d'incertitude employés dans le cadre de la théorie des fonctions de croyance, c'est-à-dire : la non-spécificité et le conflit. Plusieurs auteurs ont envisagé la construction d'une *mesure d'incertitude globale AU* (*Aggregate Uncertainty*) qui tiendrait compte de la non-spécificité et du conflit. Une première idée, afin de cumuler la non-spécificité et le conflit, fût d'additionner leurs mesures. Cette idée a été pour la première fois mise en œuvre par Lamata et Moral [81], qui ont suggéré d'additionner la fonction NS, définie dans l'equation (2.56) et la fonction E, définie dans l'équation (2.62). Plus tard, les sommes $NS + D$ et $NS + ST$ ont été étudiées [77, 102, 132]. Malheureusement ces fonctions de mesures d'incertitude totales ne répondent pas à une série d'axiomes souhaitables pour la fonction AU [78]. Une mesure d'incertitude globale pour la théorie des croyances satisfaisant l'ensemble des axiomes a été développée par plusieurs auteurs dans les années 90 [66, 86]. Cette mesure est définie de la manière suivante :

$$AU(m) = \max_{P \in \mathcal{P}r} S(P). \tag{2.72}$$

Cette mesure représente le maximum de l'entropie de Shannon parmi toutes les mesures de probabilité compatibles avec la fonction de croyance m.

Enfin, une autre mesure d'incertitude $Inc(m)$ a été proposée par Smets [112]. Cette mesure d'incertitude est construite à l'aide de la fonction de communalité Q :

$$Inc(m) = -\sum_{A \in \mathcal{F}(m)} \log_2 Q(A). \tag{2.73}$$

Cette mesure présente l'avantage d'être additive lors de la combinaison conjonctive :

$$m = m_1 \cap m_2 \quad \Rightarrow \quad Inc(m) = Inc(m_1) + Inc(m_2). \tag{2.74}$$

2.6 Conclusion

Dans ce chapitre, nous avons présenté la théorie de l'évidence. Cette théorie présente l'avantage de modéliser d'une manière plus adaptée les informations incertaines. Le problème de représentation des fonctions de croyance a donné naissance à plusieurs interprétations [32, 79]. Parmi ces modèles, celui des croyances transférables proposé par Smets [114] repose sur une interprétation non probabiliste des croyances. De plus, ce modèle offre une justification axiomatique cohérente des différents concepts fondamentaux de la théorie de l'évidence et permet de clarifier le lien existant entre la modélisation des croyances (niveau crédal) et la prise de décision (niveau pignistique). Néanmoins, l'approche proposée pour la prise de décision repose sur la construction d'une fonction de probabilité à partir d'une fonction de croyance. Enfin, le modèle des croyances transférables généralise des théories telles que la théorie des probabilités et la théorie des possibilités.

Malheureusement, l'utilisation de la théorie de l'évidence pose plusieurs problèmes :

- Il n'existe pas de méthode générique pour modéliser les fonctions de croyance. Si la plupart des travaux traitant des applications de la théorie de l'évidence définissent des fonctions de croyance adaptées au problème spécifique, très peu tentent de proposer une méthode ou une approche plus générale.
- La gestion du conflit lors de la combinaison de fonctions de croyance reste un problème non totalement traité. Les contradictions entre les sources, qui peuvent provenir d'un problème mal posé ou de sources non fiables, engendrent une masse conflictuelle lors de la combinaison. La normalisation proposée par Dempster afin de conserver une fonction de croyance unitaire n'est pas justifiable lorsque le conflit est élevé.

Ces différents points représentent les travaux qui seront abordés dans les prochains chapitres. Dans le chapitre suivant, nous présenterons en détail des méthodes génériques d'obtention des fonctions de croyance à partir des connaissances du système.

Chapitre 3

Modélisation des fonctions de croyance

Dans le cadre de la théorie de l'évidence pour la reconnaissance des formes, il n'existe pas de méthodes génériques pour l'obtention des fonctions de croyance. Dans la plupart des cas, la modélisation utilisée dépend de l'application envisagée [42, 64]. Cependant, on peut regrouper les principales méthodes de modélisation en deux approches :

- les approches fondées sur le calcul d'une distance,
- les approches fondées sur le calcul de vraisemblance.

Ce sont ces deux approches que nous allons détailler dans les sections suivantes.

3.1 Méthodes de modélisation

3.1.1 Approche basée sur la distance

Plusieurs travaux ont été menés sur la modélisation des fonctions de croyance à l'aide d'une distance [18, 93, 94, 133]. Dans cette section, nous présentons uniquement le travail original de Denœux [35, 149] qui présentent l'approche distance de la modélisation des fonctions de croyance comme la méthode des plus proches voisins. Par la suite, Denœux a effectué une mise en garde concernant les différences fondamentales entre la méthode originale des plus proches voisins et son approche par les fonctions de croyance. Denœux préfère désormais l'appellation de *classifieur évidentiel basé sur la distance*[1].

Considérons un nouvel individu de vecteur forme x connu et de vecteur d'appartenance u inconnu. Si un élément de vecteur forme $x^{(i)}$ et d'étiquette $u_n^i = 1$ de l'ensemble d'apprentissage $\mathcal{L}$ est proche de x dans l'espace des caractéristiques, alors une partie de la croyance sera affectée à H_n et le reste à l'ensemble des hypothèses du cadre de discernement. Ainsi, nous obtenons alors à partir de l'élément i une masse de croyance m_i. Jusqu'à présent, nous n'avons considéré pour l'appartenance de x qu'un seul élément de $\mathcal{L}$. Si la même opération est répétée pour l'ensemble des I exemples d'apprentissage, on obtient alors I fonctions de croyance qui peuvent être combinées

[1] En anglais « EDC » pour *Evidential Distance-based Classifier*

à l'aide de l'opérateur de Dempster. En pratique, les éléments éloignés de x ont peu d'influence et peuvent être négligés. Deux techniques peuvent alors être mises en œuvre. Dans la première approche, on ne prend en compte que les k plus proches voisins de x [149]. La seconde approche repose sur la caractérisation des données d'apprentissage à l'aide de prototypes [37]. Chacun des prototypes i, initialisés par un algorithme de type C-means, permet la construction d'une fonction de croyance ayant l'expression suivante :

$$\begin{cases} m_i(\{H_n\}) & = \alpha_i \phi_i(d_i) \\ m_i(\Theta) & = 1 - \alpha_i \phi_i(d_i) \end{cases} \tag{3.1}$$

où $0 < \alpha_i < 1$ est une constante, $\phi_i(.)$ est une fonction décroissante monotone vérifiant $\phi_i(0) = 1$ et $\lim_{d \to \infty} \phi_i(d) = 0$, d_i est la distance euclidienne entre le vecteur x et le $i^{\text{ème}}$ prototype. La fonction ϕ_i peut être une fonction exponentielle de la forme :

$$\phi_i(d_i) = exp^{-\gamma_i (d_i)^2} \tag{3.2}$$

où γ_i est un paramètre associé au $i^{\text{ème}}$ prototype. Le paramètre α_i empêche l'affectation de toute la masse de croyance à l'hypothèse H_n lorsque x et le $i^{\text{ème}}$ prototype sont égaux. Il traduit l'incertitude relative à la caractérisation du prototype i. En outre, la contrainte $\alpha_i < 1$ garantit la possibilité de combiner m_i avec n'importe quelle autre fonction de croyance puisque quelque soit d_i, on aura toujours $m_i(\Theta) > 0$ (la certitude de H_n pourrait entraîner un conflit total avec une autre source de croyance incompatible). Le paramètre γ_i, quant à lui, permet de spécifier la vitesse de décroissance de la masse avec la distance selon le prototype.

Les fonctions de croyance m_i, obtenues pour chacun chacun des prototypes, sont ensuite fusionnées avec la règle de combinaison de Dempster.

Dans [148], il est montré que cette approche par la théorie de l'évidence de la règle des plus proches voisins permet d'améliorer les performances de l'algorithme original. De la même manière, les performances obtenues par les méthodes d'ordonnancement des plus proches voisins [9, 10, 11] ont été aussi améliorées en utilisant la théorie de l'évidence [97].

Nous pouvons remarquer que cette méthode repose sur l'estimation de plusieurs paramètres :

- nombre de voisins k ou nombre de prototypes,
- position des prototypes,
- valeurs de γ_i,
- valeurs de α_i.

Au cours d'une première étude [148], il a été montré que l'approche présentée ici est très peu sensible au choix du paramètre k, et donc *a priori* au nombre de prototypes.

Pour le paramètre γ_i, une première estimation proposée, pour l'approche utilisant les plus proche voisins, a été l'inverse de la moyenne des distances intraclasses [35]. Bien que cette estimation donne de bons résultats en moyenne par rapport à l'approche classique des k-ppv, les performances en terme de classification varient fortement en fonction de la valeur de γ_i. C'est pourquoi une méthode d'optimisation automatique des paramètres (γ_i, α_i et la position des prototypes), fondée sur l'utilisation de l'information contenue dans l'ensemble d'apprentissage $\mathcal{L}$ a été introduite [149]. Cette optimisation est basée sur la minimisation d'un critère d'erreur quadratique moyenne E_{MS} entre les probabilités pignistiques et les vecteurs d'appartenance aux classes :

$$E_{MS} = \sum_{i=1}^{I} \sum_{n=1}^{N} [BetP^{(i)}(H_n) - u_n^i]^2 \tag{3.3}$$

où $BetP^{(i)}$ représente la probabilité pignistique d'un vecteur $x^{(i)}$ de la base d'apprentissage. Les valeurs optimales des paramétres sont recherchées itérativement par une méthode de descente de gradient sur E_{MS}. Une étude de cet estimateur a été faite par Zouhal et Denœux [149] et montre sa pertinence dans de nombreuses situations. Cependant, une étude a montré la variabilité de cet estimateur, pour l'approche des k plus proches voisins, dans le cadre d'une application où le nombre d'exemples d'apprentissage est petit et très bruité [61].

La méthode présentée ici est adaptée à un étiquetage précis puis elle a été généralisée aux étiquettes imprécises par Denœux [39].

3.1.2 Approche fondée sur la vraisemblance

Les approches fondées sur le calcul de vraisemblance pour l'obtention des fonctions de croyance considèrent généralement ces fonctions comme les bornes inférieure et supérieure d'une probabilité. Elles peuvent être classées en deux catégories : les méthodes globales et les méthodes séparables. Ces deux méthodes supposent connues $f(x|H_n)$ la densité de probabilité conditionnellement aux classes. A partir d'un vecteur forme x, la fonction de vraisemblance $L(H_n|x)$ est une fonction de Θ dans $[0, +\infty[$ définie par $L(H_n|x) = f(x|H_n)$ pour tout $n \in \{1, \ldots, N\}$.

Méthodes globales

Les méthodes globales sont ainsi nommées car la construction de la fonction de croyance nécessite l'ensemble des hypothèses H_n. La première approche globale basée sur le calcul de vraisemblance a été développée par Shafer [109]. Cette approche conduit à une fonction de croyance consonante. Afin d'établir un lien entre la fonction de plausibilité et une distribution de probabilité, Shafer impose deux axiomes :

- la plausibilité d'un singleton H_n doit être proportionnelle à sa vraisemblance. Si on pose c le coefficient de normalisation, la plausibilité peut être exprimée par la relation :

$$Pl(\{H_n\}) = cL(H_n|x) \quad \forall\, H_n \in \Theta \tag{3.4}$$

  avec x le vecteur forme à classer.
- la fonction de plausibilité Pl devant être consonante, elle doit donc vérifier la condition :

$$Pl(A \cap B) = \max[Pl(A), Pl(B)] \quad \forall\, A, B \subseteq \Theta. \tag{3.5}$$

On peut alors estimer la plausibilité d'un sous-ensemble A de la manière suivante :

$$Pl(A) = c \max_{H_n \in A} L(H_n|x). \tag{3.6}$$

La plausibilité du cadre de discernement Θ est toujours égale à 1, ainsi on peut déterminer c :

$$Pl(A) = \frac{\max\limits_{H_n \in A} L(H_n|x)}{\max\limits_{H_n \in \Theta} L(H_n|x)} \quad \forall\, A \subseteq \Theta. \tag{3.7}$$

Cette fonction de plausibilité Pl peut être transformée en un jeu de masses m consonant. Cette approche offre une interprétation correcte des connaissances probabilistes par une fonction de croyance consonante. Malheureusement, un système consonant interprète toujours les connaissances vers l'hypothèse ayant la plus grande probabilité. Ainsi d'autres approches globales par vraisemblance, dites *partiellement consonante*, ont vu le jour [135, 137].

Méthodes séparables

Au contraire des méthodes globales, les méthodes séparables vont construire une fonction de croyance pour chacune des hypothèses H_n du cadre de discernement Θ. Ainsi, pour chacune des hypothèses H_n appartenant au cadre de discernement Θ, nous allons définir un jeu de masses de croyance notée m_n. Ce type d'approches a été proposé pour la première fois par Smets [113] et repris ensuite par Appriou [5, 7].

Appriou a conduit une recherche exhaustive de l'ensemble des modèles susceptibles de satisfaire deux axiomes fondamentaux :

- **Axiome 1** : la cohérence avec l'approche bayésienne dans le cas où les vraisemblances $L(H_n|x)$ sont parfaitement représentatives des densités réellement rencontrées, et où les probabilités *a priori* sont connues.
- **Axiome 2** : la séparabilité de l'évaluation des hypothèses H_n ; chaque vraisemblance $L(H_n|x)$ doit être considérée comme une source d'information distincte donnant lieu à une fonction de croyance m_n, qui doit être susceptible d'intégrer son facteur de fiabilité en terme d'affaiblissement.

La recherche des modèles, menée par Appriou, satisfaisant ces axiomes permet alors de réduire le nombre de modèles possibles. Ainsi, deux modèles répondent aux deux propriétés énoncées précédemment. Le modèle 1 proposé par Appriou [7], qui est identique à celui obtenu par Smets [118], est particularisé par :

$$\begin{cases} m_n(\{H_n\}) & = 0 \\ m_n(\overline{H_n}) & = \alpha_n.\{1 - R.L(H_n|x)\} \\ m_n(\Theta) & = 1 - \alpha_n + \alpha_n.R.L(H_n|x) \end{cases} \tag{3.8}$$

et le modèle 2 :

$$\begin{cases} m_n(\{H_n\}) & = \alpha_n.R.L(H_n|x)/\{1 + R.L(H_n|x)\} \\ m_n(\overline{H_n}) & = \alpha_n/\{1 + R.L(H_n|x)\} \\ m_n(\Theta) & = 1 - \alpha_n \end{cases} \tag{3.9}$$

où R est un facteur de normalisation contraint par :

$$R \in [0, (\max_{n\in[1,N]} \{L(H_n|x)\})^{-1}]. \tag{3.10}$$

Les vraisemblances $L(H_n|x)$, du vecteur forme x sous les différentes hypothèses H_n, pourront être déterminées à partir de la connaissance *a priori* de la distribution de probabilité ou à partir d'estimateurs de densité parmi lesquels on trouve les méthodes paramétriques basées sur un modèle gaussien ou les méthodes de noyau non paramétriques. Celles-ci sont donc plus ou moins représentatives des densités réellement rencontrées. Les coefficients α_n caractérisent ici le degré de représentativité. Lorsque les densités sont parfaitement représentatives de l'apprentissage alors les coefficients α_n sont égaux à 1 et, dans ce cas, les masses de croyance ne sont pas affaiblies. A l'inverse, lorsque la distribution de probabilité est totalement méconnue, ce qui est caractérisé par un coefficient α_n égal à 0, alors les jeux de masses deviennent élément neutre de l'opérateur de combinaison de Dempster. Appriou [5] fixe ces coefficients de fiabilité à 1 lorsque la confiance accordée aux sources est élevée et à 0.9 dans le cas contraire. Une autre approche [88] détermine la valeur de ces coefficients à l'aide d'une fonction linéaire de l'écart-type des données d'apprentissage de chaque hypothèse selon chaque source à valeur dans $[0.9, 1]$. Pour un écart-type minimum, le coefficient est fixé à 1 alors que dans le cas d'un écart-type maximal le coefficient est fixé à 0.9.

En ce qui concerne le facteur de normalisation R, son domaine de définition garantit l'obtention d'une fonction de croyance normalisée. Une étude sur l'influence de ce facteur sur les distributions de croyance a été présentée dans [64]. La valeur de ce facteur est souvent fixée à la plus grande valeur autorisée afin d'avoir la fonction de croyance la moins spécifique possible. Toutefois, aucune étude théorique ne vient justifier ce choix.

En pratique, les performances des deux modèles proposés par Appriou semblent être quasi-équivalentes [58]. Toutefois, Appriou préconise l'utilisation du modèle 1 qui peut être obtenu à partir du *théorème de Bayes généralisé* proposé par Smets [118]. C'est ce modèle, dans sa forme la moins spécifique, c'est-à-dire avec la plus grande valeur autorisée pour le facteur R, que nous utilisons dans nos différentes simulations.

3.1.3 Remarque

Dans la description des modèles proposés par Appriou et Denœux, nous avons considéré l'espace des caractéristiques $\mathcal{X}$ dans sa globalité. Une autre stratégie consiste à modéliser les informations selon chaque caractéristique x_j (avec $j \in \{1, \ldots, J\}$) du vecteur x à classer.

Dans ce cas, Appriou impose un troisième axiome à ses modèles :

- **Axiome 3** : la cohérence avec l'association probabiliste des sources ; si l'on considère les sources d'information indépendantes et les densités de probabilités $f(x|H_n)$ parfaitement représentatives de la réalité alors les modélisations retenues doivent mener au même résultat si l'on effectue la somme orthogonale ou si l'on modélise directement les probabilités conjointes.

Les équations (3.8) et (3.9) deviennent alors respectivement pour le modèle 1 :

$$\begin{cases} m_{nj}(\{H_n\}) & = 0 \\ m_{nj}(\overline{H_n}) & = \alpha_{nj}.\{1 - R_j.L(H_n|x_j)\} \\ m_{nj}(\Theta) & = 1 - \alpha_{nj} + \alpha_{nj}.R_j.L(H_n|x_j) \end{cases} \tag{3.11}$$

et pour le modèle 2 :

$$\begin{cases} m_{nj}(\{H_n\}) & = \alpha_{nj}.R_j.L(H_n|x_j)/\{1 + R_j.L(H_n|x_j)\} \\ m_{nj}(\overline{H_n}) & = \alpha_{nj}/\{1 + R_j.L(H_n|x_j)\} \\ m_{nj}(\Theta) & = 1 - \alpha_{nj} \end{cases} \tag{3.12}$$

avec R_j est un facteur de normalisation contraint par :

$$R_j \in [0, (\max_{n \in [1,N]} \{L(H_n|x_j)\})^{-1}]. \tag{3.13}$$

Nous pouvons alors utiliser la combinaison de Dempster afin de fusionner les N fonctions de croyance obtenues. On note alors m_j les fonctions de croyance résultantes :

$$m_j = \bigoplus_{n \in [1,\ldots,N]} m_{nj}. \tag{3.14}$$

La fonction de croyance m unique est obtenue en suivant le même principe :

$$m = \bigoplus_{j \in [1,\ldots,J]} m_j. \tag{3.15}$$

Dans le cas de la démarche proposée par Denœux, le modèle présenté à l'équation (3.1) devient alors :

$$\begin{cases} m_{ij}(\{H_n\}) & = \alpha_{ij}\phi_{ij}(d_{ij}) \\ m_{ij}(\Theta) & = 1 - \alpha_{ij}\phi_{ij}(d_{ij}) \end{cases} \tag{3.16}$$

où d_{ij} représente la distance entre la composante x_j du vecteur x et la $j^{\text{ème}}$ composante du prototype i et où la fonction ϕ_{ij} s'exprime de la manière suivante :

$$\phi_{ij}(d) = exp^{-\gamma_{ij}d^2}. \tag{3.17}$$

De la même manière qu'avec les modèles proposés par Appriou, nous obtenons la fonction de croyance m_i à l'aide de la combinaison de Dempster :

$$m_i = \bigoplus_{j\in[1,\ldots,J]} m_{ij}. \tag{3.18}$$

et la même démarche est entreprise afin de fusionner les I fonctions de croyance issues des I prototypes pour obtenir la fonction de croyance unique m :

$$m = \bigoplus_{i\in[1,\ldots,I]} m_i. \tag{3.19}$$

Dans le cadre de l'approche développée par Appriou, les données sont modélisées à l'aide de distribution de probabilité. Les fonctions de croyance sont obtenues ensuite à partir de ces distributions en associant de la masse de croyance sur les éléments H_n, $\overline{H_n}$ et Θ ou sur $\overline{H_n}$ et Θ pour le modèle le moins spécifique.

Pour l'approche distance, proposée par Denœux, les données sont, soient gardées dans leur globalité, soient représentées par des prototypes. Les fonctions de croyance sont alors déterminées à l'aide d'une fonction de la distance entre les données caractéristiques et la nouvelle donnée. La masse de chacune des fonctions de croyance est répartie entre H_n, l'hypothèse d'appartenance de la donnée caractéristique à partir de laquelle la fonction est construite, et Θ.

Ainsi, ces modèles sont différents par leurs représentations des connaissances (distribution de probabilités ou prototypes) et par les ensembles où la masse de croyance est répartie. Toutefois, l'approche séparable et l'approche distance associent à chaque fonction de croyance un coefficient de fiabilité.

Dans la section suivante, nous nous attacherons à présenter deux méthodes pour l'obtention de ces coefficients de confiance pour les modèles présentés dans la section 3.1.3. La première de ces méthodes repose sur l'utilisation de critères d'information. La seconde utilise le critère d'erreur présenté dans la section 3.1.1 (équation (3.3)). Ces deux méthodes peuvent être employées pour les différentes modélisations.

3.2 Détermination des coefficients d'affaiblissement

3.2.1 Affaiblissement des fonctions de croyance à l'aide de critères d'information

Afin de clarifier la démarche entreprise, nous présentons l'obtention des coefficients de fiabilité dans le cadre de la modélisation d'Appriou, bien que cette démarche s'adapte très bien à la modélisation fondée sur les distances.

Ainsi, de manière à qualifier la bonne représentativité de l'apprentissage, nous proposons d'utiliser une mesure d'information qui est calculée pour chaque couple (S_j, H_n) où S_j est une source d'information et H_n une hypothèse. Le but de la démarche est de permettre d'évaluer la fiabilité de l'apprentissage de la source S_j relativement à l'hypothèse H_n. Pour satisfaire ce principe, on calcule la vraisemblance $L(H_n|x_j)$ à partir de l'estimation de la densité $f(x_j|H_n)$ sur la base d'apprentissage et une base de validation nous permet d'ajuster le coefficient α_{nj}. Le calcul du coefficient α_{nj} repose donc sur la ressemblance (ou la dissemblance) entre les lois de probabilité de chaque hypothèse issues de la base d'apprentissage et de la base de validation. Si ces lois sont ressemblantes, alors l'apprentissage de cette densité de probabilité sera considéré comme représentatif, et dans ce cas α_{nj} sera proche de 1. Au contraire, si les lois sont dissemblables, alors l'estimation de la densité $f(x_j|H_n)$ (et donc le calcul de la vraisemblance) sera considérée comme peu fiable et le coefficient α_{nj} sera alors proche de 0. Parmi les tests statistiques classiques paramétriques ou non paramétriques, tels que le test du (χ^2) ou le test de Kolmogorov, permettant de définir l'adéquation ou la ressemblance statistique entre des lois de probabilités, nous avons arrêté notre choix sur une méthode non paramétrique. Notre démarche est la suivante [82].

Nous approchons les lois de probabilité de chaque hypothèse par des histogrammes construits à l'aide d'un critère d'information que nous noterons IC [26, 28]. Ces histogrammes, optimaux au sens du maximum de vraisemblance, résument l'information contenue dans chaque source S_j selon chaque hypothèse H_n et permettent d'obtenir une estimation optimale de la loi au sens du critère choisi et ce de manière non paramétrique[2].

Nous utilisons ensuite une mesure de dissemblance entre lois de probabilité. Cette mesure, la distance de Hellinger [13], a été modifiée pour être applicable dans le cas des histogrammes [26] entre chacune des approximations de loi. L'emploi de cette mesure nous permet alors de définir le coefficient α_{nj}.

Soit $\mathcal{C}_{j,H_n}$ l'histogramme optimal à C_{j,H_n} classes des données de la source S_j sous l'hypothèse

[2]Une présentation détaillée de l'obtention des histogrammes optimaux est donnée en annexe.

H_n. Cet histogramme est construit en tenant compte de la base d'apprentissage et de la base de validation. Cet histogramme obtenu, on en déduit $\hat{\lambda}^a_{C_{j,H_n}}$ l'estimation de la loi sur $\mathcal{C}_{j,H_n}$ de la base d'apprentissage et $\hat{\lambda}^v_{C_{j,H_n}}$ l'estimation de la loi sur $\mathcal{C}_{j,H_n}$ de la base de validation. La distance de Hellinger entre ces deux estimations est donnée par [26] :

$$Hell(\hat{\lambda}^a_{C_{j,H_n}}, \hat{\lambda}^v_{C_{j,H_n}}) \triangleq 1 - \sum_{i=1}^{C_{j,H_n}} \sqrt{\hat{\lambda}^a_{C_{j,H_n}}(B_i) \times \hat{\lambda}^v_{C_{j,H_n}}(B_i)}. \tag{3.20}$$

Cette mesure entre lois de probabilité offre l'avantage, au contraire de mesures telles que l'information de Kullback, la divergence de Kullback ou la distance de Bhattacharyya [13], de varier entre 0 et 1. Lorsque les estimations de lois sont proches, alors la distance tend vers 0. Au contraire, si les lois sont fortement dissemblables, alors la distance est proche de 1. Le sens de variation de cette distance est l'inverse de celui du coefficient de confiance α_{nj}. Nous prendrons alors le coefficient α_{nj} égal à :

$$\alpha_{nj} = 1 - Hell(\hat{\lambda}^a_{C_{j,H_n}}, \hat{\lambda}^v_{C_{j,H_n}}). \tag{3.21}$$

Une présentation plus détaillée de la démarche se trouve en Annexe A.

3.2.2 Erreur quadratique moyenne

Une autre technique pour l'évaluation des coefficients d'affaiblissement α_{nj} consiste à utiliser la même approche que celle introduite par Denœux [37], en minimisant le critère d'erreur présenté dans l'équation (3.3). La minimisation du critère d'erreur revient alors à maximiser le taux de reconnaissance. Ainsi, au contraire de la méthode présentée précédemment, les coefficients de fiabilité obtenus ne sont pas forcément représentatifs de la fiabilité d'une source d'information mais plutôt de son pouvoir discriminant. Les coefficients de fiabilité, obtenus par cette méthode, peuvent ne pas refléter la confiance que l'on peut avoir en la modélisation des connaissances, mais ils permettent d'obtenir un taux de reconnaissance maximum vis-à-vis du critère choisi. Il est a noté que l'expression analytique des coefficients de fiabilité obtenus par minimisation de l'erreur quadratique se trouve dans [57].

3.3 Simulations

Les deux méthodes que nous avons présentées précédemment ont été testées sur des données synthétiques afin d'évaluer l'apport, en terme de discrimination, des coefficients de fiabilité des sources d'information relativement à une hypothèse. Nous étudions l'efficacité de ces coefficients de fiabilité dans le cas de la fusion d'une donnée incertaine avec une donnée sûre en situation

de contexte évolutif (Section 3.3.1). Pour ce test, nous utilisons la méthode de modélisation proposée par Appriou (équation (3.11)) et l'approche développée par Denœux (équation (3.16)). Les coefficients de fiabilité sont déterminés à partir de critères d'information et à partir de l'erreur quadratique moyenne. Nous rappelons ici que le but de ce test n'est pas de comparer les différentes modélisations, mais de voir l'apport de coefficients de fiabilité sur chacune d'entre-elles.

3.3.1 Constitution du jeu de données

Nous considérons un problème de discrimination à deux classes H_1 et H_2. Le but de ce test est de montrer l'efficacité des coefficients de fiabilité afin de pallier l'incertitude d'une donnée de bonne qualité en lui associant une donnée de moins bonne qualité mais sûre. La classe H_1 est constituée de deux distributions gaussiennes avec respectivement les moyennes μ_{11} et μ_{12} ainsi que les variances Σ_{11} et Σ_{12} :

$$\begin{array}{cc} \mu_{11} = (0,-3)^t & \mu_{12} = (0,3)^t \\ \Sigma_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \Sigma_{12} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{array} \tag{3.22}$$

L'apprentissage disponible pour la classe H_2 est donné sous forme d'une distribution normale de moyenne μ_2 et de variance Σ_2 :

$$\mu_2 = (3,0)^t \quad \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.23}$$

Les données de test générées pour la classe H_1 suivent des distributions gaussiennes de moyennes μ'_{11} et μ'_{12} et de variances Σ'_{11} et Σ'_{12} définies de la manière suivante :

$$\begin{array}{cc} \mu'_{11} = (0,-3)^t & \mu'_{12} = (0,3)^t \\ \Sigma'_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \Sigma'_{12} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \tag{3.24}$$

alors que l'hypothèse H_2 suit la distribution gaussienne de moyenne μ'_2 et de variance Σ'_2 suivantes :

$$\mu'_2 = (3,S)^t \quad \Sigma'_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.25}$$

La capacité de la seconde variable à détecter l'hypothèse H_2 est dépendante d'un signal S, représentant une évolution du contexte sous l'hypothèse H_2. Ce signal prend ces valeurs dans l'intervalle $[-6,6]$. La base de validation, nous permettant de définir les valeurs de coefficients de fiabilité est définie sur le même modèle que la base de test. Les bases d'apprentissage, de validation et de test sont constituées de 1500 vecteurs par classe. La figure FIG. 3.1 représente la répartition des données de la base d'apprentissage dans l'espace des caractéristiques.

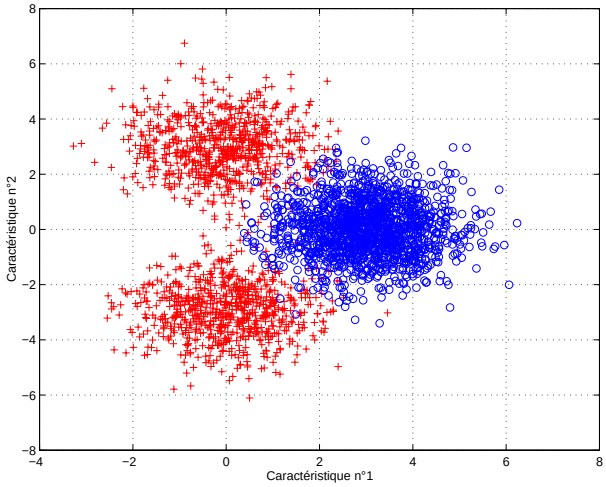

Fig. 3.1 : – Représentation des données simulées de la base d'apprentissage (H_1=+ ; H_2=o).

3.3.2 Modélisation d'Appriou

Pour les tests qui vont suivre, nous avons utilisé le modèle n°1 proposé par Appriou (équation (3.11)). Les coefficients R_j, pour $j \in \{1,2\}$, ont été fixés à la plus grande valeur autorisée afin d'obtenir les fonctions de croyance les moins spécifiques possibles. L'estimation des fonctions de densité associées à chaque classe selon chaque hypothèse peut être réalisée par plusieurs techniques [111] telles que les méthodes paramétriques (exemple : modèle gaussien) ou les méthodes de noyau non paramétriques. Dans les tests qui suivent, l'estimation des densités est obtenue en utilisant un modèle de mélange gaussien associé à une technique d'estimation des paramètres par l'algorithme EM [80]. Le nombre de modes pour l'estimation des différentes densités est fixé à deux pour la classe H_1 et à un pour la classe H_2.

En ce qui concerne la détermination des coefficients de fiabilité, utilisés dans les différentes modélisations, plusieurs stratégies peuvent être employées :

- stratégie n°1 : les coefficients de fiabilité α_{nj} sont fixés à 1, on considère ainsi que les estimations de densité ou les prototypes sont représentatifs de l'apprentissage,
- stratégie n°2 : un expert spécialiste du système sait que la seconde caractéristique est moins fiable que la première pour la reconnaissance de l'hypothèse H_2, on fixe alors $\alpha_{nj} = 1 \quad \forall\, n, j \neq 2$ et $\alpha_{22} = 0.9$,
- stratégie n°3 : les coefficients de fiabilité sont déterminés à l'aide de critères d'information,
- stratégie n°4 : les coefficients de fiabilité sont déterminés en minimisant l'erreur quadratique moyenne.

Nous allons évaluer le comportement de ces différentes stratégies pour les deux modélisations présentées précédemment.

Les résultats en terme de taux de classification, en prenant comme critère de décision le maximum de probabilité pignistique, sont représentés sur la figure FIG. 3.2. On constate que les stratégies impliquant un coefficient de confiance α_{22} différent de 1 permettent d'obtenir de meilleurs taux de classification par rapport à une stratégie sans affaiblissement dans le cas d'un apprentissage non représentatif. De plus, d'après cette figure, nous pouvons constater que les démarches de calcul des coefficients de fiabilité présentées ici permettent d'obtenir un gain de classification notable par rapport à une valeur de coefficient fixée à 0.9 lorsque la base d'apprentissage n'est plus représentative des données de la base de test. Il est à noter que la courbe obtenue pour $\alpha_{nj} = 1$ correspond à l'approche probabiliste classique qui suppose une représentativité parfaite des distributions apprises. Nous pouvons aussi constater que les performances obtenues en terme de classification dans le cas de la détermination des coefficients à l'aide de critères d'information sont semblables à celles obtenues en utilisant l'erreur quadratique moyenne. Enfin, lorsque l'apprentissage représente correctement les données de test ($S \approx 0$), les performances sont alors identiques dans les quatre cas.

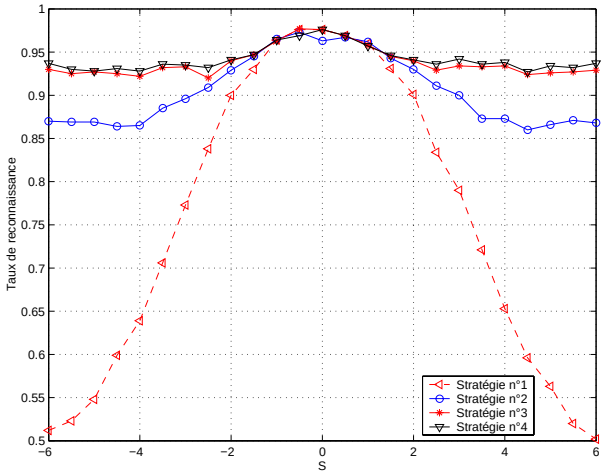

FIG. 3.2 : – Evolution du taux de classification en fonction du signal S ; pour $\alpha_{nj} = 1$ ($\triangleleft$–), pour $\alpha_{22} = 0.9$ et $\alpha_{nj} = 1\ \forall n, j \neq 2$ ($\circ$-), pour α_{nj} obtenu par la distance de Hellinger (*-) et pour α_{nj} obtenu par l'erreur quadratique moyenne (∇-).

Les coefficients de fiabilité obtenus par l'intermédiaire de critères d'information nécessitent, préalablement, la construction d'un histogramme initial. Une illustration de l'histogramme initial des données d'apprentissage et de validation pour l'hypothèse H_2 selon la seconde variable est présentée sur la figure FIG. 3.3 (gauche) pour $S = -6$. Le résultat de l'optimisation de cet histogramme à l'aide de critères d'information est représenté sur la figure FIG. 3.3 (droite).

A partir de cet histogramme, la répartition des données d'apprentissage et de validation (FIG. 3.3) permet alors de déterminer la valeur du coefficient de fiabilité à l'aide de la distance

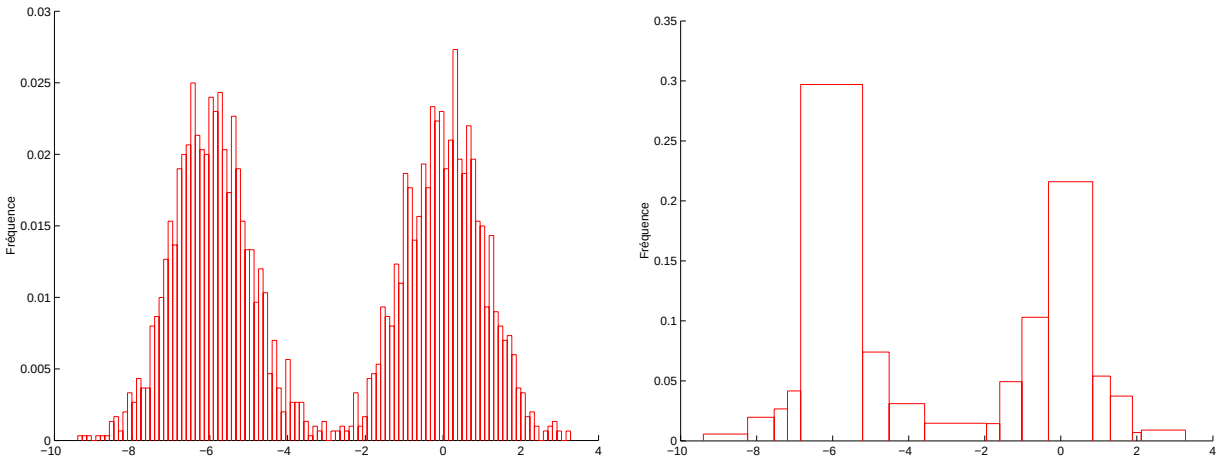

FIG. 3.3 : – Histogramme initial (à gauche) et optimal (à droite) des données d'apprentissage et de validation de l'hypothèse H_2 selon la seconde variable pour $S = -6$.

de Hellinger.

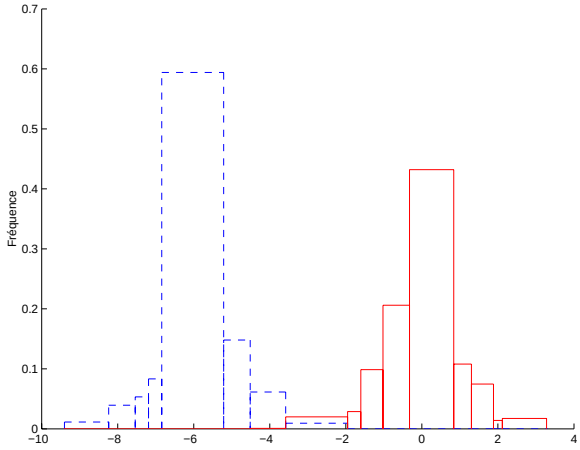

FIG. 3.4 : – Répartition des données d'apprentissage (trait pointillé) et de validation (trait plein) selon l'histogramme optimal pour l'hypothèse H_2 selon la seconde variable pour $S = -6$.

La figure FIG. 3.5 (gauche) représente l'évolution des coefficients de fiabilité déterminés à l'aide de critères d'information. Nous constatons que le coefficient α_{22} accordé à la caractéristique n°2 sous l'hypothèse H_2 varie en fonction du signal S. Ce coefficient est proche de 1, lorsque l'apprentissage est fiable $S \approx 0$ et proche de 0 dans le cas contraire. Les autres coefficients de fiabilité sont constants en fonction du signal S et proche de 1 car l'apprentissage est représentatif du contexte. La figure FIG. 3.5 (droite) représente l'évolution des coefficients de fiabilité déterminés à l'aide de l'erreur quadratique moyenne. On constate que l'évolution du coefficient α_{22} est proche de celle obtenue avec les critères d'information. On peut remarquer que l'évolution du coefficient α_{12} est semblable à celle du coefficient α_{22}. En effet, lorsque les données de H_1 et de H_2 sont confondues ($|S| > 2$) alors la seconde variable ne permet plus de discriminer ces deux

classes. Cette variable est alors considérée comme non fiable et la décision n'est prise qu'à l'aide de la première variable.

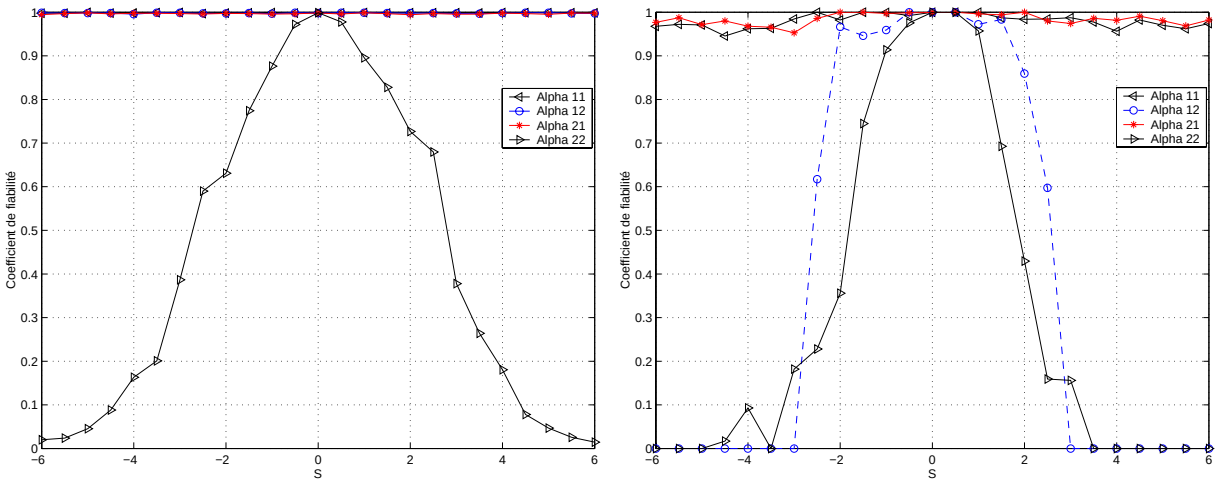

FIG. 3.5 : – Evolution des coefficients de fiabilité déterminés à l'aide de la distance de Hellinger (gauche) et à l'aide de l'erreur quadratique moyenne (droite) en fonction du signal S ; α_{11} (◁-), α_{12} (○-), α_{21} (*-) et α_{22} (▷-).

Les figures FIG. 3.6 et FIG. 3.7 représentent respectivement l'évolution des jeux de masses pour S égale -6 et -3. Pour chacune de ces figures, nous avons représenté selon chaque variable (une colonne correspond à une variable) les masses $m(H_1)$, $m(H_2)$ et $m(\Theta)$ pour les trois dernières stratégies étudiées (c'est-à-dire affaiblissement fixé par expertise $\alpha_{22} = 0.9$, coefficients de fiabilité déterminés à partir des critères d'information et coefficients optimisés en minimisant l'erreur quadratique moyenne). Pour chaque caractéristique, les données d'apprentissage ainsi que les données de test sont représentées. De plus, les positions des modes, utilisés pour l'obtention des densités, sont représentées en dessous des données d'apprentissage. On peut constater sur chacune de ces figures, que quelque soit la stratégie employée les fonctions de croyance obtenues selon la première caractéristique sont similaires. Ceci s'explique par le fait que les coefficients de fiabilité associés à cette caractéristique sont proches de 1 ($\alpha_{n1} \approx 1 \; \forall \; n \in \{1, 2\}$) pour chacune des stratégies. De la même manière, pour la stratégie optant pour des poids fixes (stratégie n°2), les distributions de masses de croyance sont identiques pour la seconde caractéristique quelque soit la valeur de S. Au contraire, on peut constater que dans le cas d'un apprentissage des coefficients de fiabilité effectués à partir de critères d'information (stratégie n°3) ou par minimisation de l'erreur quadratique moyenne (stratégie n°4), les fonctions de croyance évoluent en fonction du signal pour la seconde variable. Ainsi pour la stratégie n°3, lorsque la caractéristique n°2 est la moins fiable ($S = -6$), c'est-à-dire lorsque les données d'apprentissage sont les moins représentatives de la base de test, alors la masse accordée à Θ est proche de 1. De la même manière pour la stratégie

n°4, les coefficients de fiabilité associés aux hypothèses sont nuls ce qui implique une masse sur Θ égale à l'unité. Ainsi pour ces deux stratégies, dans le cas où la caractéristique est peu fiable pour la reconnaissance de l'hypothèse H_2, celle-ci n'a pas d'influence lors de la fusion avec la masse de croyance issue de l'autre caractéristique. La décision est alors prise uniquement sur la fonction de croyance issue de la première composante. En outre, plus la seconde caractéristique sera fiable ($S \approx 0$) et plus la masse de l'ensemble Θ diminuera au profit de la masse de H_1. Ainsi, elle sera de nouveau prise en compte lors de la fusion.

Globalement dans le cadre de la modélisation proposée par Appriou, les deux méthodes présentées permettent d'adapter des coefficients d'affaiblissement et d'atteindre, dans tous les cas, des taux de classification supérieurs à ceux obtenus avec un coefficient d'affaiblissement fixe.

3.3.3 Modélisation de Denœux

Pour l'utilisation du modèle de distance proposé par Denœux, nous avons utilisé deux prototypes pour représenter la classe H_1 et un seul pour la classe H_2. Les paramètres γ_{ij} ainsi que la position des prototypes ont été optimisés par minimisation de l'erreur présentée dans l'équation (3.3) sur la base d'apprentissage. Pour la suite, nous appelons le prototype associé aux données de la classe H_1 de moyenne $(3, 0)$ "prototype n°1" et le prototype associé aux données de moyenne $(-3, 0)$ "prototype n°2". Le prototype associé aux données de la classe H_2 est appelé "prototype n°3". Ainsi, on associe à chacun de ces prototypes et selon chaque caractéristique un coefficient de fiabilité. Pour l'obtention des coefficients de fiabilité nous avons mis en concurrence les différentes stratégies suivantes. Les deux premières stratégies fixent les valeurs de coefficients de fiabilité. Ces coefficients sont fixés à 1 pour la première stratégie ($\alpha_{ij} = 1 \; \forall \; i \in [1, \ldots, 3]$ et $\forall \; j \in [1, 2]$). Pour la seconde stratégie on pose l'ensemble des coefficients à 1 sauf le coefficient α_{32} qui est fixé à 0.9. Les coefficients de fiabilité sont obtenus à l'aide de critères d'information pour la troisième stratégie. Enfin, pour la stratégie n°4, les coefficients de fiabilité sont déterminés à partir de la minimisation de l'erreur quadratique moyenne sur la base de validation.

La figure FIG. 3.8 permet de visualiser les taux de reconnaissance, obtenus en prenant comme critère de décision le risque lié à la probabilité pignistique, en fonction du signal S pour les différentes stratégies mises en compétition. On peut constater que dans le cas où $S = -3$ (et de manière symétrique $S = 3$), c'est-à-dire lorsque les données de H_1 et de H_2 sont confondues aussi bien dans la base de validation que dans la base de test, alors les taux de reconnaissance obtenus avec les trois premières stratégies sont inférieurs à celui obtenu en déterminant les coefficients de fiabilité en minimisant l'erreur quadratique moyenne. Pour cette dernière stratégie, les taux de reconnaissance restent relativement constants en fonction de la valeur de S. D'autre part, on

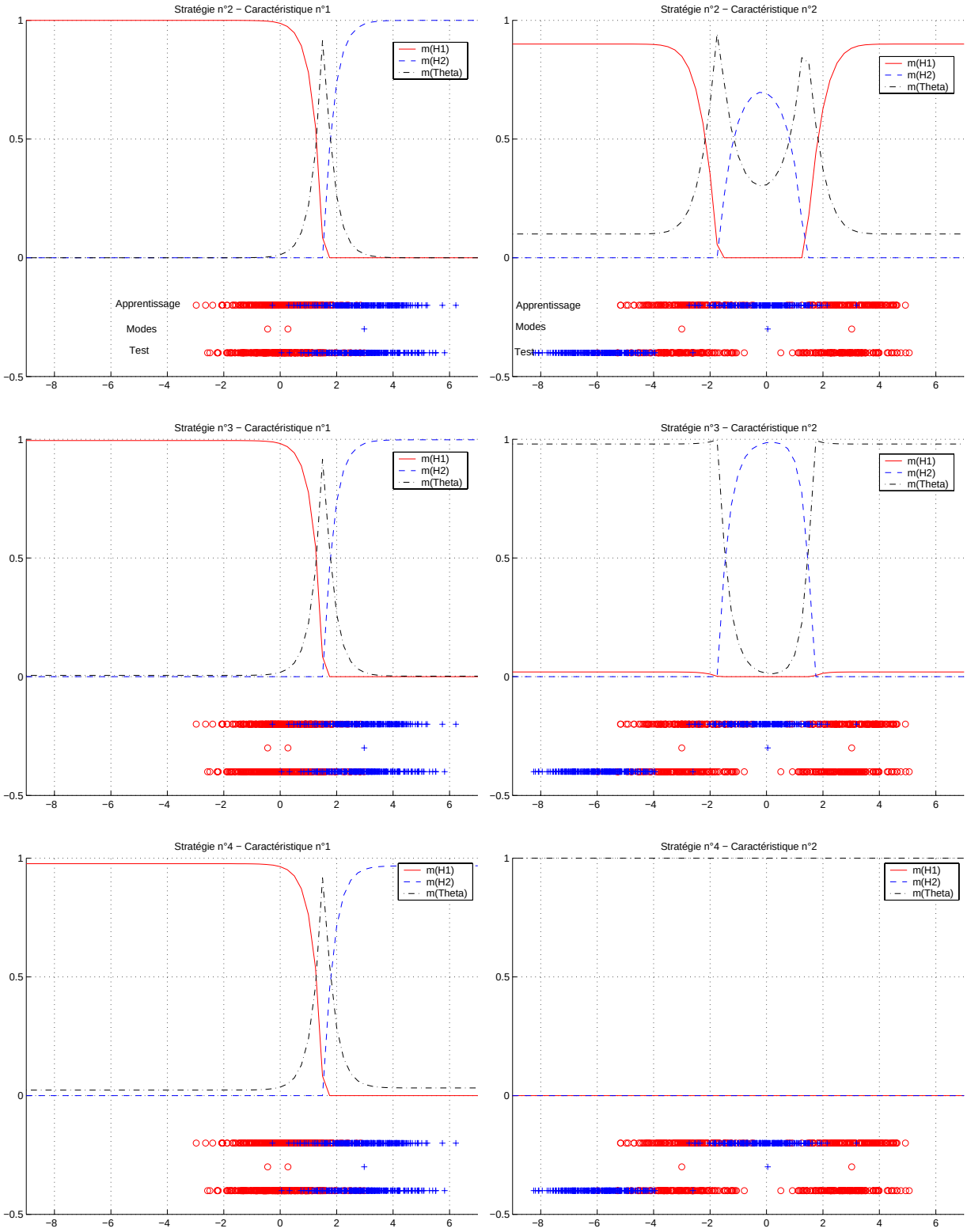

Fig. 3.6 : – Jeux de masses obtenus avec le modèle séparable pour $S = -6$ pour les trois stratégies d'affaiblissement envisagées pour chacune des caractéristiques ; $m(H_1)$ en trait plein, $m(H_2)$ en trait pointillé et $m(\Theta)$ en trait mixte ; $\circ$: hypothèse H_1, $+$: hypothèse H_2.

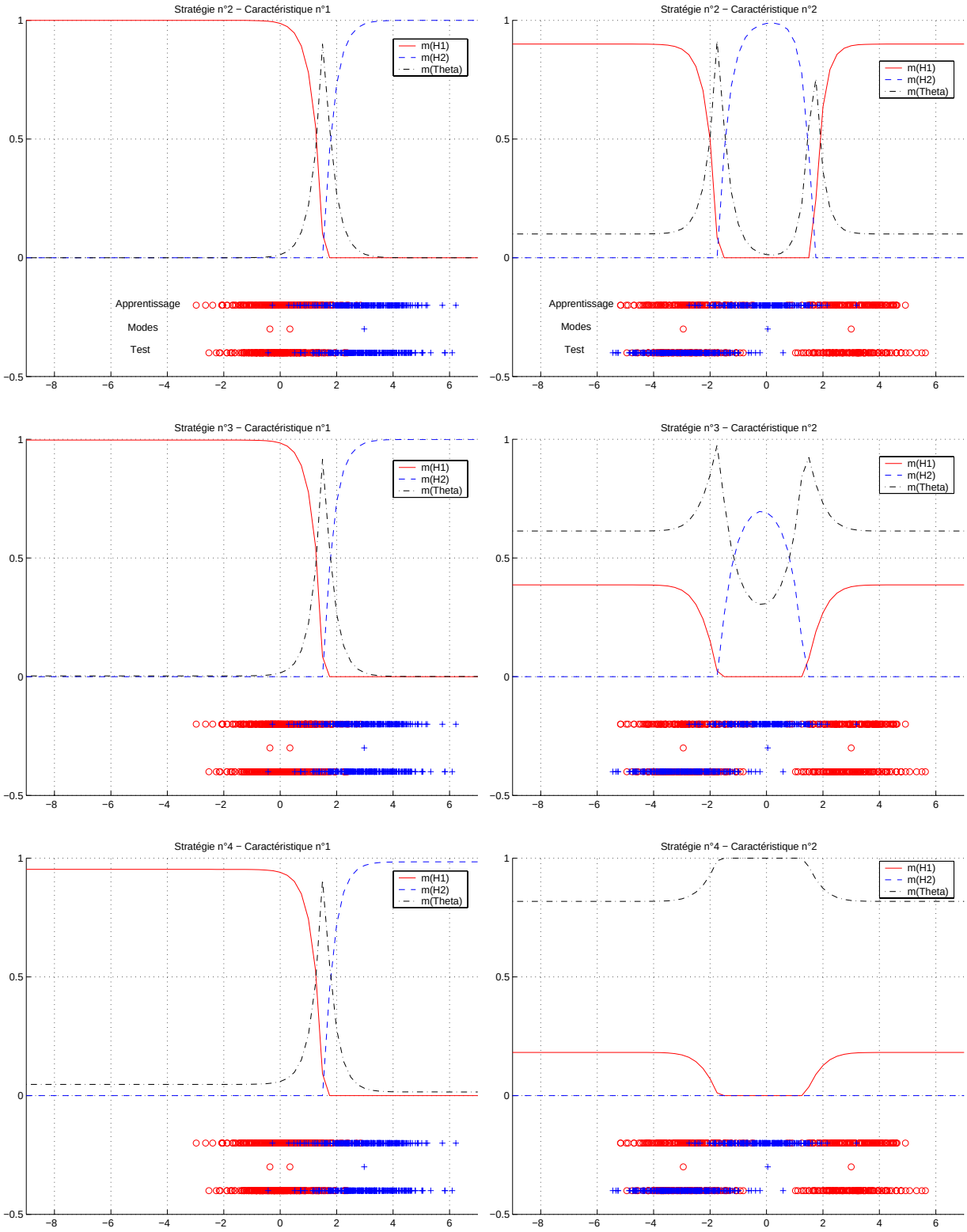

Fig. 3.7 : – Jeux de masses obtenus avec le modèle séparable pour $S = -3$ pour les trois stratégies d'affaiblissement envisagées pour chacune des caractéristiques ; $m(H_1)$ en trait plein, $m(H_2)$ en trait pointillé et $m(\Theta)$ en trait mixte ; $\circ$: hypothèse H_1, $+$: hypothèse H_2.

peut noter que l'approche sans affaiblissement (stratégie n°1) donne le taux de reconnaissance le plus faible lorsque la valeur absolue de S est supérieure à 2. Dans les mêmes conditions, la stratégie reposant sur l'estimation des coefficients de fiabilité à l'aide de critères d'information permet d'obtenir un gain en terme de classification par rapport aux stratégies de coefficients fixes. Enfin, dans le cas d'un apprentissage correct ($|S| \leq 1.5$), les différentes stratégies donnent des taux de reconnaissance semblables.

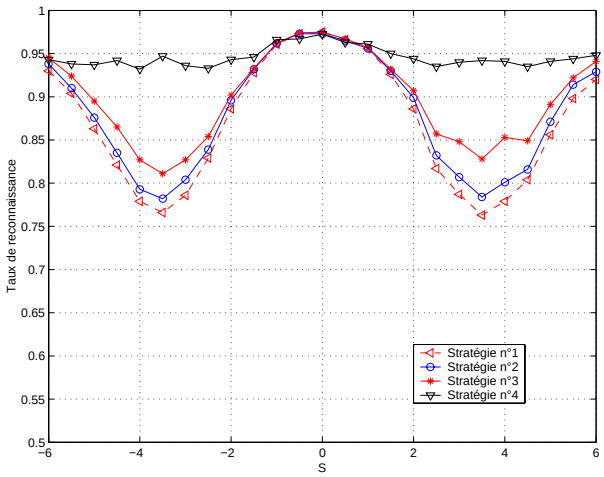

FIG. 3.8 : – Evolution du taux de classification en fonction du signal S ; pour $\alpha_{nj} = 1$ ($\triangleleft$–), pour $\alpha_{32} = 0.9$ et $\alpha_{ij} = 1\ \forall j \neq 2 \text{et} i \neq 3$ ($\circ$-), pour α_{ij} obtenu par la distance de Hellinger (*-) et pour α_{ij} obtenu par l'erreur quadratique moyenne (∇-).

Dans le cadre de cette modélisation, les coefficients de fiabilité déterminés à partir des critères d'information sont obtenus en construisant l'histogramme des distances entre un prototype et les données de la base d'apprentissage et de validation pour chaque caractéristique. Ainsi, la figure FIG. 3.9 (gauche) représente l'histogramme initial des distances entre le prototype n°3 et les données d'apprentissage et de validation selon la seconde variable. Dans les mêmes conditions, l'histogramme optimal obtenu à l'aide de critère d'information est représenté sur la figure FIG. 3.9 (gauche).

Les coefficients de fiabilité sont obtenus à l'aide de la distance de Hellinger entre l'histogramme optimal des distances entre les prototypes et les données d'apprentissage et l'histogramme optimal des distances entre les prototypes et l'ensemble des données de la base de validation. (FIG. 3.10).

La figure FIG. 3.11 représente les coefficients de fiabilité déterminés à partir des critères d'information. On constate sur cette figure que le coefficient α_{32} évolue en fonction du signal S. Lorsque les prototypes sont représentatifs de la base d'apprentissage ($S \approx 0$) alors ce coefficient est proche de 1, sinon il est proche de 0. Les autres coefficients de fiabilité sont constants et proches de 1. Le comportement de ces coefficients est identique à celui constaté avec la modélisation

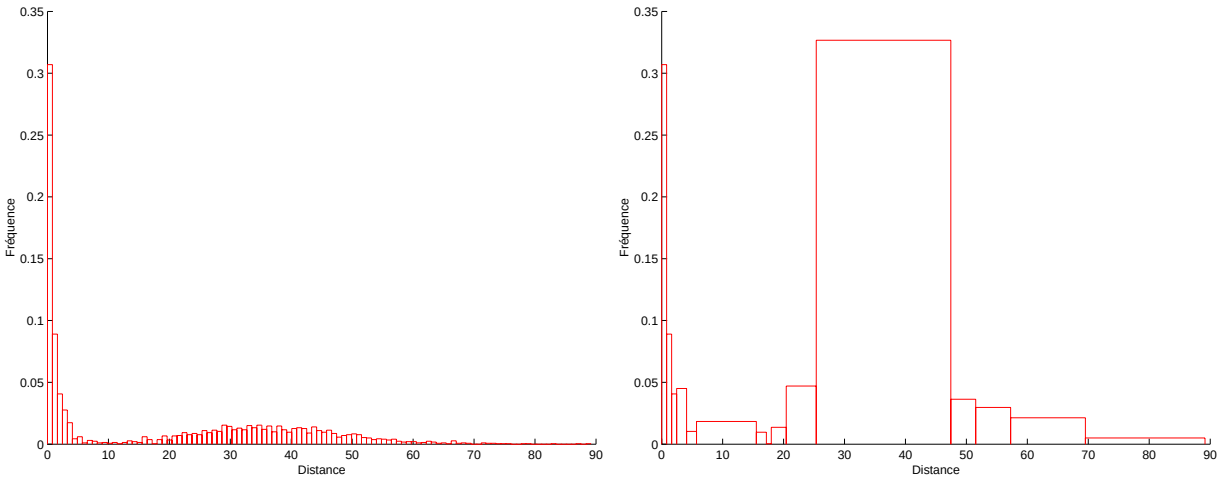

FIG. 3.9 : – Histogramme initial (à gauche) et optimal (à droite) des distances entre le prototype n°3 et les données d'apprentissage et de validation selon la seconde variable pour $S = -6$.

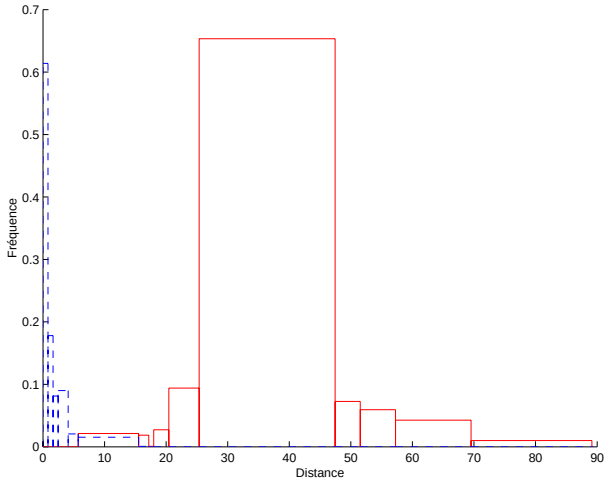

FIG. 3.10 : – Répartition selon l'histogramme optimal des distances entre le prototype n°3 et les données d'apprentissage (trait pointillé) et des distances entre le prototype n°3 et les données de validation (trait plein) pour la seconde caractéristique.

d'Appriou.

De la même manière que les coefficients de fiabilité déterminés à partir des critères d'information, les coefficients optimisés à l'aide de l'erreur quadratique moyenne sont représentés sur la figure FIG. 3.12. La figure FIG. 3.12 (gauche) illustre les coefficients de fiabilité des trois prototypes associés à la caractéristique n°1. On constate que les valeurs des coefficients associés aux prototypes représentant la classe H_1 (prototype n°1 et prototype n°2) suivent la même allure avec des amplitudes différentes. En effet, l'amplitude de variation du coefficient α_{21} associé au prototype n°2 est plus importante que celle de α_{11}. Toutefois, sur cette variable les informations apportées par ces deux prototypes sont redondantes car ils sont proches l'un de l'autre. Le coefficient α_{31} associé au prototype n°3 est proche de 1. On constate ainsi que cette caractéristique

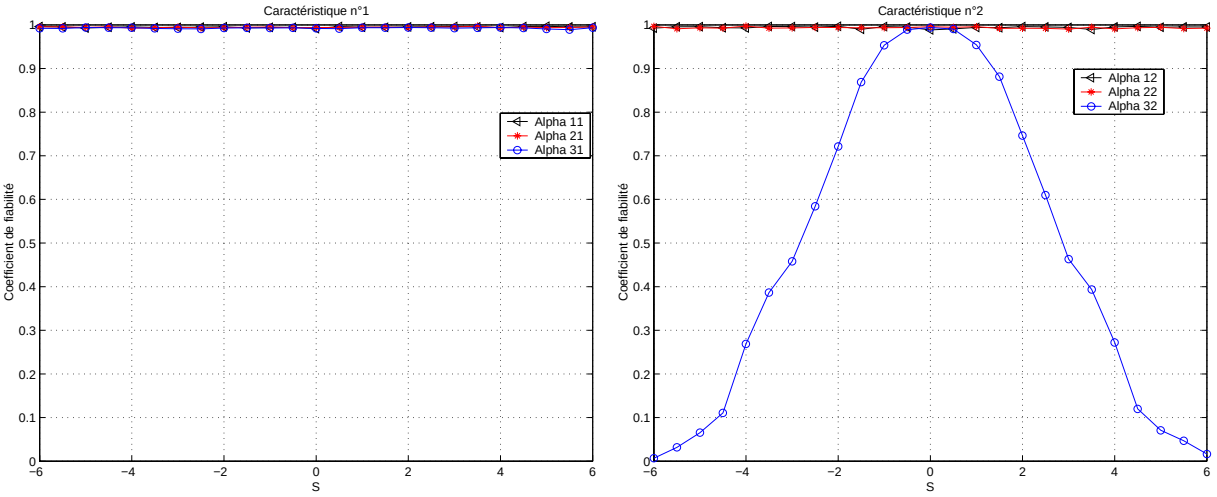

FIG. 3.11 : – Evolution des coefficients de fiabilité déterminés à l'aide des critères d'information en fonction du signal S pour la caractéristique n°1 (à gauche) et pour la caractéristique n°2 (à droite).

est fiable. La figure de droite représente les coefficients attribués à la seconde caractéristique. On remarque que le coefficient α_{32} évolue en fonction du signal S. Lorsque $S \approx 0$, alors le coefficient α_{32} est proche de 1, alors que dans le cas d'un prototype non représentatif des données de validation alors ce coefficient est proche de 0.5. De plus, on constate que les valeurs des coefficients de fiabilité associés aux prototypes de la classe H_1 ont des comportements symétriques. Ainsi, dans le cas où les données de H_2 sont confondues avec celle de H_1 (lorsque $S = -3$), le coefficient α_{22} associé au prototype n°2 est nul alors que le coefficient α_{12} vaut 1. A l'opposé, lorsque $S = 3$ le coefficient α_{22} vaut 1 et $\alpha_{12} = 0$. Ceci permet, à chaque fois qu'il y a confusion, d'avoir une masse de croyance importante sur l'ensemble Θ élément non informatif dans le cadre de la combinaison ce qui revient à décider uniquement à l'aide de la première caractéristique. Le taux de reconnaissance est alors plus élevé qu'avec les autres stratégies (Cf. FIG. 3.8).

Les figures FIG. 3.13 et FIG. 3.14 représentent respectivement les distributions de masses de croyance pour les valeurs de $S = -6$ et $S = -3$. Chacune de ces figures représente l'évolution des masses pour chacune des caractéristiques et pour la stratégie de poids fixes (stratégie n°2) et les stratégies où les coefficients de fiabilité sont déterminés à l'aide de critères d'information (stratégie n°3) ou en minimisant l'erreur quadratique moyenne (stratégie n°4). De plus, sur ces différentes figures les données d'apprentissage, de test ainsi que les prototypes de chaque hypothèse selon chaque caractéristique sont mentionnés.

On peut remarquer, dans un premier temps, que dans le cas de la stratégie à coefficients fixes (stratégie n°2) les distributions sont identiques quelque soit les valeurs de S. En outre, on constate sur ces figures que, quelque soit la stratégie utilisée, la masse de croyance accordée à

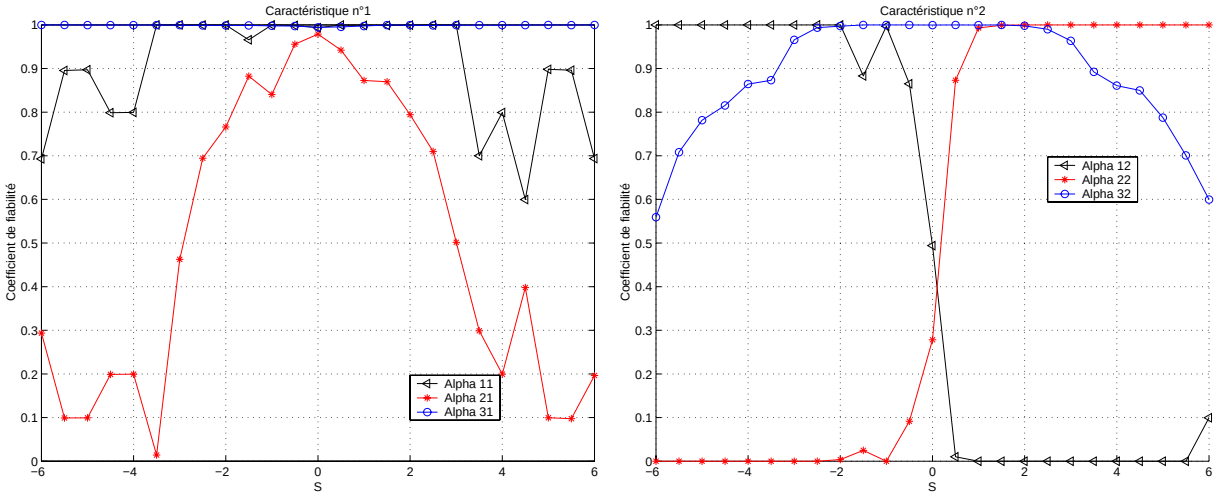

FIG. 3.12 : – Evolution des coefficients de fiabilité déterminés à l'aide de l'erreur quadratique moyenne en fonction du signal S pour la caractéristique n°1 (à gauche) et pour la caractéristique n°2 (à droite).

l'ensemble Θ croît lorsque l'on s'éloigne des données d'apprentissage. Ainsi, dans le cas de mesures aberrantes, ce modèle basé sur la distance, affecte la majorité de la masse à Θ. Dans [129, 130], il a été montré que cette modélisation permet alors une plus grande robustesse de décision vis-à-vis de mesures aberrantes par rapport aux approches basées sur les vraisemblances. Ceci entraîne un taux de classification supérieur, dans le cas sans affaiblissement, à celui obtenu avec le modèle développé par Appriou (Cf. figures FIG. 3.2 et FIG. 3.8).

De plus, l'affectation de masse de croyance à l'ensemble Θ, lorsqu'on s'écarte des données d'apprentissage, explique l'augmentation des taux de reconnaissance constatée pour les stratégies n°1, n°2 et n°3 sur la figure FIG. 3.8 lorsque $|S| > 4$. De la même manière, dans le cas où $|S| > 4$, la stratégie mettant en œuvre les critères d'information pour déterminer les coefficients permet d'obtenir des taux de reconnaissance supérieurs à ceux obtenus avec les stratégies n°1 et n°2 car la masse de croyance accordée à l'ensemble Θ est plus important ($\alpha_{32} \approx 0$). Pour les mêmes raisons, la stratégie n°2 où l'on fixe la valeur de α_{32} à 0.9 permet d'atteindre de meilleurs taux de reconnaissance qu'avec la stratégie sans affaiblissement.

Enfin, si l'on regarde plus précisément le cas où $S = -3$ sur la figure FIG. 3.14, c'est-à-dire lorsque la seconde caractéristique confond de manière importante H_1 et H_2, on remarque que l'approche à coefficients fixes (stratégie n°2) et l'approche avec les coefficients déterminés à l'aide de critères d'information (stratégie n°3) accordent une masse de croyance très importante pour l'hypothèse H_1. Cette affectation engendre alors une mauvaise affectation de la majorité des données étiquetées H_2 (Cf. FIG. 3.8). Dans l'approche des coefficients de fiabilité optimisés (stratégie n°4), on constate que la majorité de la masse est accordée à l'ensemble Θ. La seconde

caractéristique n'influence alors pas la première ce qui permet d'avoir un taux de reconnaissance supérieur à ceux obtenus dans le cas de coefficients fixes.

Ainsi, dans le cadre de la modélisation proposée par Denœux, les deux méthodes d'obtention des coefficients de fiabilité présentées ici permettent d'atteindre, dans tous les cas, des taux de classification supérieurs à ceux obtenus avec un coefficient d'affaiblissement fixe. La méthode d'adaptation reposant sur la minimisation de l'erreur quadratique moyenne permet d'atteindre le taux de reconnaissance optimal.

3.4 Conclusion

Dans ce chapitre, nous avons présenté différentes méthodes de modélisation des informations sous forme de fonctions de croyance. Ces approches peuvent être distinguées en deux catégories.

La première correspond aux approches basées sur la détermination de vraisemblances. Parmi ces approches deux méthodes ont été développées : la méthode globale introduite par Shafer [109] et la méthode séparable présentée par Appriou [5]. La première méthode offre une modélisation des informations sous forme de fonctions de croyance consonantes. La dernière méthode a permis, grâce à une recherche axiomatique, d'établir deux modèles qui offrent l'avantage d'être cohérents avec l'approche bayésienne dans le cas des connaissances parfaites.

La seconde catégorie a été introduite par Denœux [35] puis améliorée par Zouhal et Denœux [149]. Cette approche repose sur le calcul d'une distance entre un voisinage et un vecteur à classer. Chaque exemple du voisinage sera alors considéré comme une source d'information modélisée sous forme de fonctions de croyance. Cette méthode offre une plus grande robustesse de décision en présence de mesures aberrantes que les approches basées sur les vraisemblances [129, 130].

L'approche distance et l'approche séparable des vraisemblances associent à chaque fonction de croyance un coefficient de fiabilité. Ces coefficients permettent de prendre en compte la représentativité vis-à-vis de l'apprentissage. Cela peut correspondre, dans le cadre des vraisemblances, à savoir si les distributions apprises sont réellement celles rencontrées ou dans le cas de l'approche distance à prendre en compte la qualité de la représentativité du voisinage choisi.

Un coefficient de fiabilité est aussi révélateur du pouvoir discriminant de la source à laquelle il est associé. Ainsi, ces coefficients vont permettre d'affaiblir les fonctions de croyance issues de sources d'information peu discriminantes.

Dans ce chapitre, notre contribution repose dans la proposition de deux méthodes de détermination des coefficients de fiabilité. La première méthode repose sur la construction d'une mesure de dissemblance entre lois de probabilité pour chacune des hypothèses du cadre de dis-

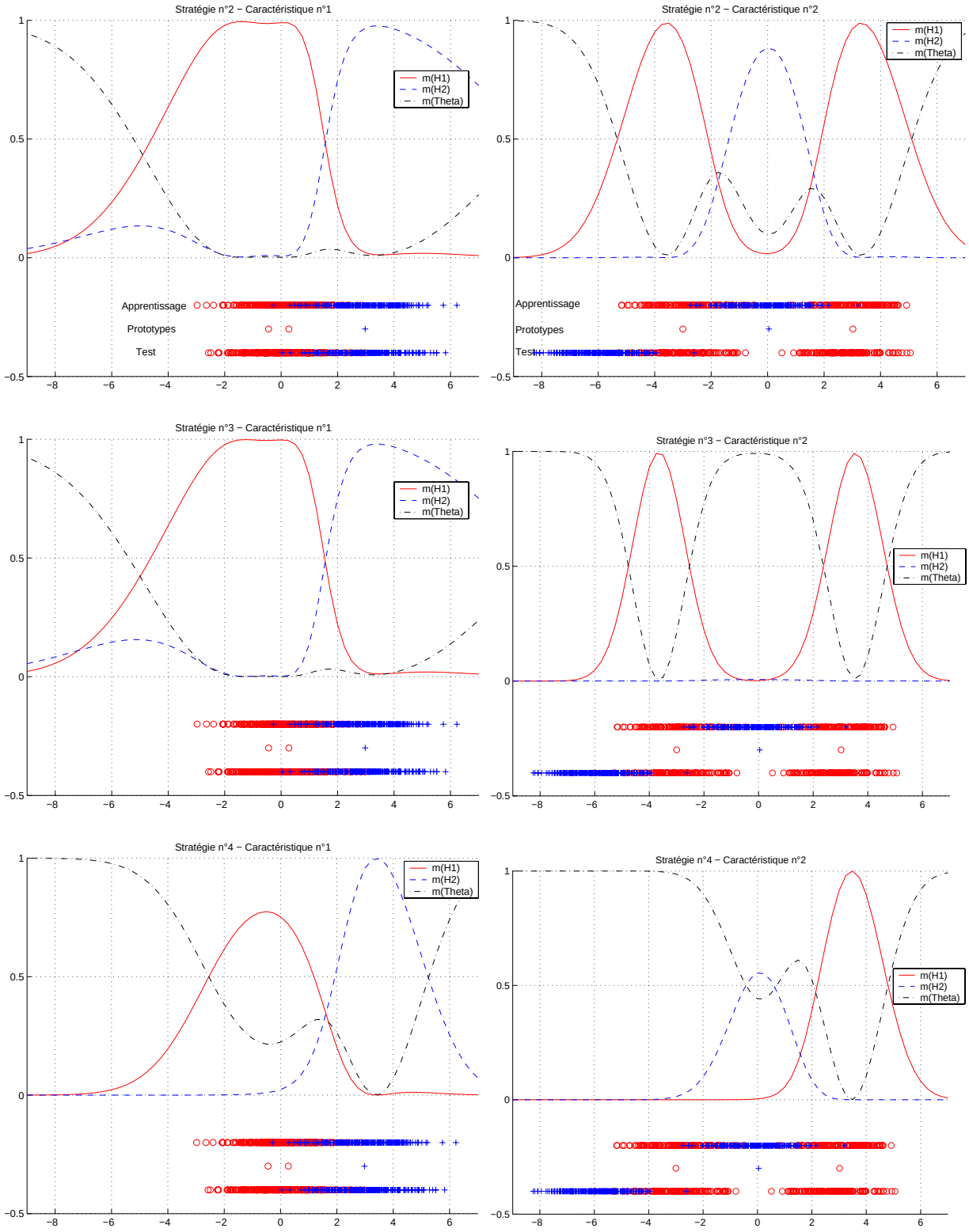

Fig. 3.13 : – Jeux de masses obtenus avec le modèle de distance pour $S = -6$ selon les trois stratégies envisagées pour chacune des caractéristiques ; $m(H_1)$ en trait plein, $m(H_2)$ en trait pointillé et $m(\Theta)$ en trait mixte ; $\circ$: hypothèse H_1, $+$: hypothèse H_2.

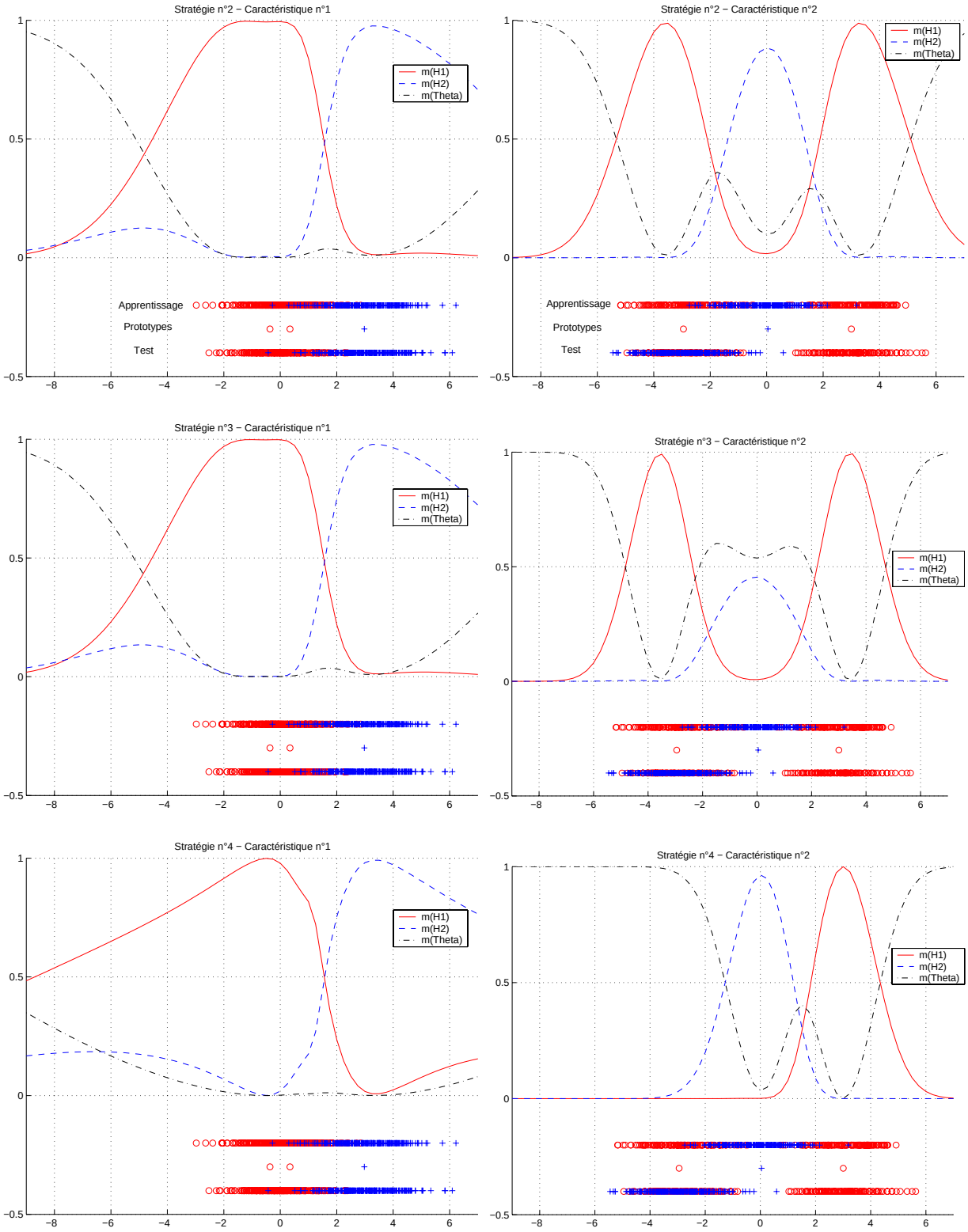

Fig. 3.14 : – Jeux de masses obtenus avec le modèle de distance pour $S = -3$ selon les trois stratégies envisagées pour chacune des caractéristiques ; $m(H_1)$ en trait plein, $m(H_2)$ en trait pointillé et $m(\Theta)$ en trait mixte ; $\circ$: hypothèse H_1, $+$: hypothèse H_2.

cernement. La seconde méthode reprend la même approche que celle utilisée par Denœux [37] pour l'estimation de ces paramètres, qui consiste à minimiser un critère d'erreur de classification.

Ces différentes méthodes d'optimisation des coefficients de confiance ont été testées sur des données synthétiques représentant une évolution de contexte. Ces tests montrent, quelque soit la méthode d'obtention des coefficients, l'apport de l'affaiblissement pour les différents modèles étudiés.

Chapitre 4

Gestion du conflit dans le cadre de la théorie de l'évidence

La fusion de données a depuis quelques années suscité un intérêt certain dans la communauté scientifique [1, 2, 16, 17]. Elle est utilisée dans de nombreux domaines tels que la fusion multi-capteurs [5] et le traitement d'images [18, 19]. De par ce formalisme, la fusion de données permet la prise en compte de données hétérogènes (numériques ou symboliques) bien souvent imparfaites (imprécises, incertaines et incomplètes) modélisées sous forme de sources qu'il s'agit de combiner, agréger, fusionner. Dans le cadre de la théorie de l'évidence, la fusion d'informations repose sur l'utilisation d'un opérateur permettant de combiner les fonctions de croyance pour les différentes propositions, ou hypothèses en compétition. L'opérateur de fusion de base dans la théorie de l'évidence est l'opérateur de Dempster (somme orthogonale). Lors de la fusion avec cet opérateur, une étape de normalisation est nécessaire afin de préserver les propriétés des jeux de masses. Dans [147], Zadeh a montré que cette étape de normalisation conduit à des comportements contre-intuitifs. Afin de remédier à ce problème, Yager [141], Dubois [51] et Smets [114] ont proposé de nouveaux opérateurs. Cependant, ces opérateurs ont des comportements plus ou moins satisfaisants. En particulier, les opérateurs de Dubois et Yager ont tendance à répartir la masse conflictuelle (liée à l'inconsistance des sources fusionnées) de manière globale. Smets, quant à lui, soutient que l'existence d'une masse conflictuelle réside dans le fait que le cadre de discernement retenu n'est pas exhaustif. Nous proposons une autre approche.

Ce chapitre sera organisé de la manière suivante. Dans un premier temps, nous rappelons dans la section 4.1, les propriétés de la combinaison de Dempster. Dans la section 4.2, nous présentons un cadre générique qui nous permet d'unifier les opérateurs classiques de combinaison qui ont été développés dans le cadre de la théorie de l'évidence, mais aussi de proposer une famille d'opérateurs adaptatifs.

4.1 Règle de combinaison et gestion du conflit

Comme nous l'avons vue dans le Chapitre 1, dans le cas de données imparfaites (incertaines, imprécises et incomplètes), la fusion de données est une solution intéressante pour l'obtention d'informations plus pertinentes. La théorie de l'évidence offre des outils appropriés de fusion. A partir des jeux de masses notés m_j obtenus sur chacune des sources d'information S_j, il est possible de mettre en œuvre une règle de combinaison permettant de fournir un jeu de masses combiné m, synthétisant la connaissance des différentes sources. Ce jeu de masses peut alors être utilisé par un module de décision en bénéficiant de toute la connaissance contenue dans les jeux de masses issus de chacune des sources. Historiquement, l'opérateur de Dempster est le premier opérateur de combinaison défini dans le cadre de la théorie de l'évidence. Son utilisation impose de respecter la condition d'indépendance des sources d'information à combiner. L'opérateur de combinaison de Dempster, appelé également somme orthogonale, vérifie les propriétés de commutativité et d'associativité. La masse résultant de la combinaison de J sources d'information S_j est notée $m_\oplus$ et est donnée par l'équation (2.37). Cette règle de combinaison vérifie certaines propriétés intéressantes et son utilisation a été justifiée de manière théorique par plusieurs auteurs [47, 76, 134]. Toutefois, dans certaines situations, cet opérateur ne peut pas être utilisé. En effet, lorsque :

- les sources d'information ne sont pas indépendantes [99, 143, 144], la combinaison n'étant pas idempotente, son emploi renforcerait abusivement les propositions soutenues,
- les sources ne sont pas parfaitement fiables et dans le cas où la construction des fonctions de masses est imprécise, un conflit K est engendré. Nous reviendrons en détail sur les origines de ce conflit dans la section 4.1.2. Le facteur de normalisation, qui dépend de ce conflit, rend l'opérateur sensible aux petites imprécisions des jeux de masses, comme l'a montré Zadeh [147] et comme nous l'illustrons dans la section suivante.

4.1.1 Sensibilité de l'opérateur de Dempster

Soit le cadre de discernement $\Theta = \{H_1, H_2, H_3\}$, et deux sources d'information S_1 et S_2 produisant respectivement deux jeux de masses de croyance m_1 et m_2 définis comme suit :

$$\begin{array}{lllllll} m_1(\{H_1\}) & = & \epsilon & & m_2(\{H_1\}) & = & 1-k-\epsilon \\ m_1(\{H_2\}) & = & k & & m_2(\{H_2\}) & = & k \\ m_1(\{H_3\}) & = & 1-k-\epsilon & & m_2(\{H_3\}) & = & \epsilon \end{array} \tag{4.1}$$

avec $0 \leq k \leq 1$. Dans le cas général, l'application de l'opérateur de Dempster donne le résultat suivant :

$$m_\oplus(\{H_1\}) = m_\oplus(\{H_3\}) = \frac{\epsilon(1-k-\epsilon)}{k^2 + 2\epsilon(1-k-\epsilon)}, \tag{4.2}$$

et :

$$m_{\oplus}(\{H_2\}) = \frac{k^2}{k^2 + 2\epsilon(1-k-\epsilon)}. \tag{4.3}$$

Ainsi, en prenant $k = 0.1$ et $\epsilon = 0.01$, on obtient le jeu de masses suivant :

$$m_{\oplus}(\{H_1\}) = m_{\oplus}(\{H_3\}) = 0.32 \qquad m_{\oplus}(\{H_2\}) = 0.36 \tag{4.4}$$

alors que pour $k = 0.1$ et $\epsilon = 0.001$, on obtient :

$$m_{\oplus}(\{H_1\}) = m_{\oplus}(\{H_3\}) = 0.08 \qquad m_{\oplus}(\{H_2\}) = 0.84. \tag{4.5}$$

Nous constatons, par cet exemple, que la règle de combinaison de Dempster est très sensible aux variations de ϵ. Pour $\epsilon = 0.01$, la masse est répartie de manière équi-crédible sur les trois hypothèses tandis que pour $\epsilon = 0.001$, la fusion a tendance à soutenir l'hypothèse H_2. Cette sensibilité est due aux fortes variations du coefficient de normalisation $\frac{1}{1-K}$. La figure FIG. 4.1, montre les variations de ce coefficient de normalisation en fonction du conflit K. On peut constater qu'au voisinage de $K = 1$, une faible variation de K entraîne une forte variation du coefficient de normalisation.

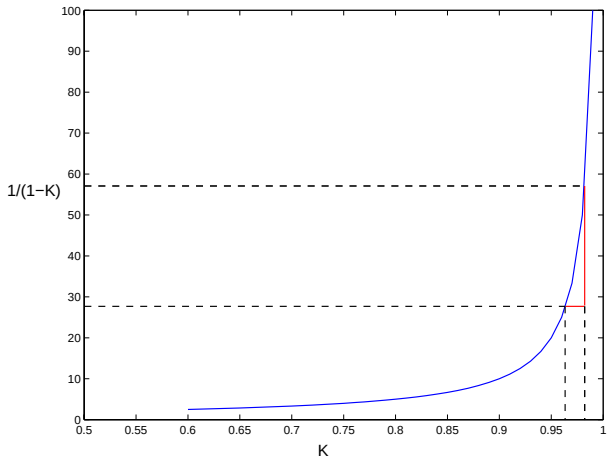

FIG. 4.1 : – Coefficient de normalisation en fonction de la masse conflictuelle K.

4.1.2 Origines et solutions au conflit

Dans cette section, nous évoquons les différentes origines possibles du conflit (Section 4.1.2) ainsi que les solutions envisageables pour la répartition de la masse conflictuelle en fonction de son origine (Section 4.1.2).

Origines du conflit

Les causes possibles du conflit peuvent être diverses. Nous pouvons toutefois regrouper ces origines en trois catégories.

La première catégorie regroupe les origines du conflit reposant sur une mesure aberrante issue d'un capteur. En effet, une mesure, située hors du domaine de fonctionnement obtenu lors de l'apprentissage, peut engendrer des conflits lors de la combinaison. Cette mesure aberrante peut être due :

- soit à un mauvais fonctionnement du capteur lors de son acquisition,
- soit à une mauvaise évaluation de la plage de fonctionnement du capteur lors de l'apprentissage. Dans le cas où le capteur fonctionne correctement, cette situation pourrait correspondre à la non prise en compte d'une hypothèse dans le cadre de discernement (classe inconnue).

La seconde catégorie repose sur la modélisation imprécise des fonctions de croyance. En effet, les principaux modèles de détermination des jeux de masses reposent soit sur l'étude d'un voisinage soit sur l'apprentissage de probabilités comme nous l'avons vu au chapitre précédent. Un mauvais choix de distance dans l'approche voisinage ou une mauvaise estimation des vraisemblances dans l'approche apprentissage de probabilités peut engendrer des variations au niveau des fonctions de croyance et ainsi favoriser l'apparition d'un conflit.

Enfin, lorsque le nombre de sources impliquées dans le processus de fusion est relativement important, un conflit apparaît. Considérons, par exemple, un ensemble de jeux de masses identiques répartissant la croyance de la manière suivante :

$$m(\{H_1\}) = 0.80 \qquad m(\{H_2\}) = 0.15 \qquad m(\Theta) = 0.05. \tag{4.6}$$

Prise isolément, chacune des sources soutient de manière importante l'hypothèse H_1. La figure FIG. 4.2, qui représente l'évolution du conflit en fonction du nombre de sources impliquées dans la combinaison, montre que lors de la combinaison de deux sources, 24% de la masse de croyance est conflictuelle et que celle-ci dépasse les 80% lors de la combinaison de dix sources.

Ces trois causes, qui interviennent généralement de manière simultanée dans la plupart des applications, soutiennent l'idée d'une redistribution adaptée de la masse conflictuelle.

Solutions au conflit

Plusieurs règles de combinaisons ont été proposées pour résoudre le problème du conflit. Les différentes solutions proposées dans la littérature peuvent être distinguées en deux familles, correspondant à deux philosophies de fusion des informations. La première regroupe des règles de combinaison reposant sur le postulat de fiabilité des sources à fusionner. Cela induit la définition d'opérateurs conjonctifs (Cf. Dempster [32] et Smets [114]). La seconde suppose qu'au moins une des sources est fiable mais en ignorant laquelle. Les opérateurs appartenant à cette famille

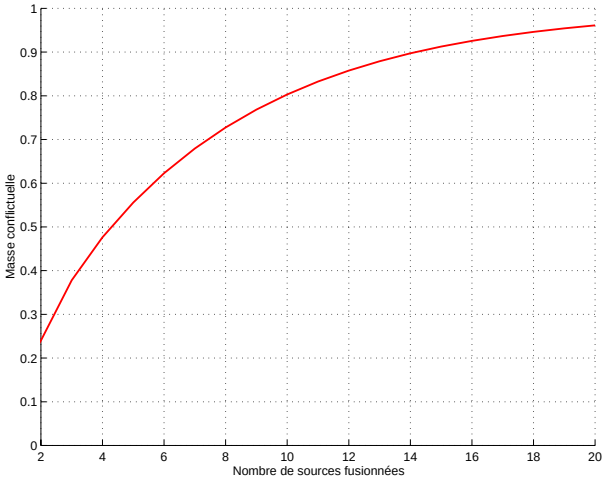

FIG. 4.2 : – Evolution de la masse conflictuelle en fonction du nombre de sources à combiner.

procèdent en une combinaison conjonctive et disjonctive (Cf. Yager [141] et Dubois-Prade [51]).

Combinaison de sources fiables : Dans la même idée que pour l'opérateur de Dempster, Smets considère que les sources à fusionner sont fiables. En partant de ce postulat, le conflit ne peut alors provenir que d'un problème mal posé, c'est-à-dire de la non prise en compte d'une ou de plusieurs hypothèses dans le cadre de discernement. Smets préconise alors de ne pas redistribuer la masse conflictuelle K sur l'ensemble des propositions mais uniquement sur l'ensemble vide $\emptyset$, ceci afin de constater les problèmes de représentativité du référentiel (non-exhaustivité de ce référentiel). La combinaison proposée par Smets est alors définie de la manière suivante :

$$\left\{ \begin{array}{lcll} m_S(A) & = & m_\cap(A) & \forall A \subseteq \Theta,\ A \neq \emptyset \\ m_S(\emptyset) & = & K. & \end{array} \right. \tag{4.7}$$

Notons qu'une approche similaire proposée par Yager [140] repose sur l'introduction d'une nouvelle hypothèse dans le cadre de discernement. Cette hypothèse va supporter toute la masse conflictuelle. En outre, on peut remarquer que les opérateurs supposant que les sources sont fiables reposent principalement sur une combinaison conjonctive.

Combinaison de sources non fiables : Le conflit peut être généré par un défaut de fiabilité d'une partie des sources d'information. Cet argument a été repris dans le cadre des opérateurs présentés par Yager [141] et par Dubois et Prade [51].

Dans le cas de l'opérateur de Yager [141], on suppose que l'une des sources intervenant dans la combinaison est fiable. Ainsi, la solution est obligatoirement dans le référentiel. Mais ne sachant quelle source donne la vraie solution, Yager propose d'attribuer la masse conflictuelle K à l'ensemble Θ. La masse résultante, m_Y, de cette combinaison, pour deux sources d'information

$\{S_1, S_2\}$, est obtenue de la manière suivante :

$$\begin{cases} m_Y(A) &= m_\cap(A) \qquad \forall A \subset \Theta,\ A \neq \emptyset \\ m_Y(\Theta) &= m_\cap(\Theta) + K \\ m_Y(\emptyset) &= 0. \end{cases} \tag{4.8}$$

La combinaison proposée par Dubois et Prade [51], dans le cadre de la fusion de deux sources d'information $\{S_1, S_2\}$, peut s'expliquer de la manière suivante. Soit S_1 une source soutenant la proposition B avec une masse de croyance $m_1(B)$ et soit une source S_2 soutenant la proposition C avec une masse de croyance $m_2(C)$. Lorsque les propositions soutenues par ces deux sources sont contradictoires et ne sachant pas quelle source est fiable, le *principe de minimum de spécificité* [48] impose de redistribuer la masse associée à cette contradiction, soit $m_1(B).m_2(C)$, sur l'union des propositions c'est-à-dire $(B \cup C)$. L'opérateur de Dubois et Prade est alors défini de la manière suivante pour deux sources d'information :

$$\begin{cases} m_D(A) &= m_\cap(A) + \sum_{\substack{B \cup C = A \\ B \cap C = \emptyset}} m_1(B).m_2(C) \qquad \forall A \subseteq \Theta,\ A \neq \emptyset \\ m_D(\emptyset) &= 0. \end{cases} \tag{4.9}$$

Une autre méthode, reposant sur la définition de coefficients de fiabilité, permet de gérer la combinaison de sources non fiables. Soit une fonction de croyance m_j fournie par une source S_j et un coefficient α_j qui représente la degré de confiance que l'on accorde à la source S_j. On obtient alors le formalisme suivant :

- $\alpha_j = 0$ signifie une remise en cause totale de la fiabilité de S_j,
- $\alpha_j = 1$ signifie une confiance absolue en la source S_j.

On note alors $m_{\alpha_j,j}$ la fonction de croyance m_j affaiblie par un coefficient $(1-\alpha_j)$. Cette fonction est définie ainsi :

$$\begin{cases} m_{\alpha_j,j}(A) &= \alpha_j m_j(A) \qquad \forall A \subset \Theta \\ m_{\alpha_j,j}(\Theta) &= 1 - \alpha_j + \alpha_j m_j(\Theta). \end{cases} \tag{4.10}$$

De cette manière, lorsque nous avons une confiance totale en la fiabilité de la source S_j, l'information apportée par cette source ne devrait pas engendrer de conflit lors de la combinaison. Le coefficient α_j est dans ce cas égal à 1 et la fonction de croyance n'est alors pas modifiée. Au contraire, si l'on suppose qu'une source S_j est non fiable, lors de la combinaison avec d'autres sources celle-ci peut alors produire une information conflictuelle. En introduisant un coefficient $\alpha_j = 0$, la fonction de croyance m_j associée à la source S_j devient alors une fonction de croyance d'ignorance totale ($m_{\alpha_j,j}(\Theta) = 1$) et donc élément neutre pour la combinaison de Dempster. Ainsi l'intérêt de l'affaiblissement est de maîtriser l'influence des sources d'information selon leur fiabilité avant de les combiner. Plusieurs méthodes ont été développées afin de définir les coefficients de fiabilité (Cf. Chapitre 3).

4.2 Cadre générique

Dans la section précédente, nous avons vu que différents auteurs [51, 114, 141] ont proposé un certain nombre de solutions pour l'interprétation du conflit. Le travail effectué dans cette thèse a consisté à proposer un cadre générique pour unifier ces différents opérateurs de combinaison. En outre, ce cadre permet de définir d'autres opérateurs destinés à permettre une redistribution de la masse conflictuelle de manière locale, adaptée ou répondant à des objectifs précis.

4.2.1 Présentation

Le but des opérateurs de combinaison proposés, est de redistribuer la masse conflictuelle K sur un ensemble de propositions. Une partie de la masse K sera affectée à chaque proposition A selon un poids noté w. Ce poids pourra être fonction de la proposition considérée et des masses engendrant le conflit. Ainsi, la masse totale après fusion pour une proposition A est la somme des deux masses et s'écrit :

$$\begin{cases} m(A) &= m_{\cap}(A) + m^c(A) \qquad \forall A \subseteq \Theta,\ A \neq \emptyset \\ m(\emptyset) &= 0. \end{cases} \tag{4.11}$$

Dans l'équation (4.11), le premier terme, $m_{\cap}(A)$, correspond à la règle de combinaison conjonctive. Le second, noté $m^c(A)$, est la partie de la masse de conflit affectée à la proposition A. Cette valeur peut s'écrire :

$$m^c(A) = w(A, m_1, \ldots, m_j, \ldots, m_J).K \qquad \forall\ A \subseteq \Theta \tag{4.12}$$

avec comme contrainte :

$$\sum_{A \subseteq \Theta} w(A, m_1, \ldots, m_j, \ldots, m_J) = 1 \tag{4.13}$$

afin de garantir que la somme des masses d'une structure de croyance soit égale à l'unité (cf. équation (2.4)). Le terme $w(A, m_1, \ldots, m_J)$ représente le facteur de poids, associé à la proposition A, participant à la redistribution de la masse conflictuelle K. Ce cadre générique permet de réécrire les opérateurs proposés par Smets [114], Yager [141] et Dubois et Prade [51]. Il suffit pour chacun de ces opérateurs de définir les poids $w(A, m_1, \ldots, m_J)$ associés à chacune des propositions $A \subseteq \Theta$, comme nous le présentons dans les sections suivantes.

4.2.2 Opérateur de combinaison de Smets

Au sein du cadre générique proposé, l'opérateur de combinaison de Smets [123], présenté dans la section 4.1.2, est défini de la manière suivante. L'ensemble, sur lequel la masse conflictuelle

sera redistribuée, est constitué uniquement de l'ensemble vide $\emptyset$. Nous obtenons alors les poids suivants :

$$\begin{cases} w(\emptyset, m_1, \ldots, m_j, \ldots, m_J) = 1 \\ w(A, m_1, \ldots, m_j, \ldots, m_J) = 0 \quad \forall\, A \subseteq \Theta,\ A \neq \emptyset. \end{cases} \tag{4.14}$$

En discrimination, l'ensemble vide $\emptyset$ peut être interprété comme une classe de rejet. Une variante de cette démarche [105], nécessitant une modélisation adéquate, est fondée sur l'introduction d'une nouvelle hypothèse dans le cadre de discernement. Cette approche permet de différencier la masse conflictuelle et la masse associée au rejet. L'opérateur de fusion proposé par Smets vérifie les propriétés de commutativité et d'associativité. Enfin notons, que dans [121], Smets définit les α-jonctions comme un cadre fédérateur pour la règle purement conjonctive, la règle de combinaison disjonctive, ainsi que leur négation.

4.2.3 Opérateur de combinaison de Yager

Selon le cadre générique proposé précédemment (Section 4.2), l'opérateur de combinaison de Yager peut s'exprimer de la manière suivante. La masse conflictuelle est placée sur Θ. Les poids prennent alors la forme suivante :

$$\begin{cases} w(\Theta, m_1, \ldots, m_j, \ldots, m_J) = 1 \\ w(A, m_1, \ldots, m_j, \ldots, m_J) = 0 \quad \forall A \subset \Theta. \end{cases} \tag{4.15}$$

Cette méthode a pour effet de séparer la totalité de la masse conflictuelle et donc de ne pas la faire intervenir dans le processus de discernement des hypothèses. Cette règle de combinaison est commutative. Malheureusement, elle n'est pas associative. Il est donc nécessaire de définir un ordre de fusion des sources.

4.2.4 Opérateur de combinaison de Dubois et Prade

De même que pour l'opérateur de combinaison de Yager, l'opérateur de combinaison de Dubois et Prade [51] repose sur l'hypothèse qu'au moins une des sources intervenant dans le processus de fusion est fiable. La répartition de la masse conflictuelle proposée par Dubois et Prade est plus spécifique que l'approche de Yager. Afin de décrire cette combinaison, nous introduisons la notion de masse conflictuelle partielle.

Chaque source d'information S_j donne une masse de croyance à chacun des éléments focaux appartenant à $\mathcal{F}(m_j)$. Quand les propositions soutenues par chacune des sources sont compatibles, c'est-à-dire lorsque les intersections entre ces propositions sont non vides, le produit des masses affectées à ces ensembles est attribué à leur intersection. Si les propositions sont incompatibles, c'est-à-dire lorsque leur intersection est égale à l'ensemble vide, nous sommes en

présence d'un conflit partiel auquel correspond une masse de croyance notée m^* qui s'exprime de la manière suivante :

$$m^* = m_1(A_1) \times m_2(A_2) \times m_3(A_3) \times \ldots \times m_J(A_J) \qquad \text{avec} \qquad A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_J = \emptyset. \quad (4.16)$$

La masse de conflit total K est la somme des masses de conflits partiels et peut s'exprimer par :

$$K = \sum\nolimits^* m^* \quad (4.17)$$

où $\sum^*$ est une somme dénombrable d'éléments. Ainsi, à partir de ce formalisme, nous pouvons décrire le principe de combinaison pour deux sources d'information de la manière suivante. Soit S_1 une source soutenant la proposition A_1 avec une masse de croyance $m_1(A_1)$ et soit une source S_2 soutenant la proposition A_2 avec une masse de croyance $m_2(A_2)$. Si les propositions A_1 et A_2 sont en contradiction, c'est-à-dire si $A_1 \cap A_2 = \emptyset$, alors on ne sait pas quelle source dit la vérité et l'on doit considérer que la solution est l'une des deux propositions. La masse conflictuelle partielle $m_1(A_1).m_2(A_2)$ sera affectée à la proposition $A_1 \cup A_2$. Dans la situation générale de ce type de combinaison, les masses de conflits partiels sont redistribuées sur une proposition A, définie de la manière suivante :

$$\forall A \subseteq \Theta \mid \exists\ A_1 \in \mathcal{F}(m_1),\ \exists\ A_2 \in \mathcal{F}(m_2),\ A = A_1 \cup A_2 \text{ et } A_1 \cap A_2 = \emptyset \quad (4.18)$$

avec pour poids :

$$\begin{cases} w(A, m_1, m_2) = \frac{1}{K} \sum\limits_{\substack{A_1, A_2 : A_1 \cup A_2 = A \\ A_1 \cap A_2 = \emptyset}} m_1(A_1) m_2(A_2) \\ w(B, m_1, m_2) = 0 \quad \text{ailleurs.} \end{cases} \quad (4.19)$$

Pour le cas de J fonctions de croyance $m_1, \ldots, m_J$, on peut alors écrire :

$$w(A, m_1, \ldots, m_J) = \frac{1}{K} \sum_{\substack{A_1, \ldots, A_J : \bigcup\limits_{j=1}^{J} A_j = A \\ \bigcap\limits_{j=1}^{J} A_j = \emptyset}} \prod_{j=1}^{J} m_j(A_j). \quad (4.20)$$

On peut remarquer que le calcul des poids ne dépend plus exclusivement des propositions auxquelles ils sont associés, mais aussi des masses de croyance à l'origine des conflits partiels. Les masses de croyance engendrant le conflit permettent ainsi de calculer la redistribution de la masse conflictuelle. Cette règle de combinaison est plus précise dans la redistribution de la masse de conflit et donc plus fine et adaptée que la règle de Yager. En outre, dans l'étape de décision, la masse conflictuelle à redistribuer interviendra dans le discernement des hypothèses en compétition. On peut noter que cette règle de combinaison utilise une approche conjonctive quand les sources sont en accord et une approche disjonctive en cas de conflit. Comme pour la

règle de combinaison de Yager, l'opérateur de fusion proposé par Dubois et Prade est commutatif mais n'est pas associatif. Une stratégie permettant de combiner les sources dans un ordre précis doit être définie.

4.2.5 Relation avec l'affaiblissement

Dans la section précédente (cf. Section 4.1.2), nous avons vu que l'on pouvait gérer le conflit à l'aide de coefficients de fiabilité. De la même manière que pour les opérateurs classiques (Smets, Yager, Dubois et Prade), nous pouvons, dans le cadre du formalisme que nous proposons, reproduire les résultats obtenus via cet affaiblissement. En effet, la relation liant les poids aux coefficients de fiabilité (le raisonnement pour l'obtention de cette relation est décrit en Annexe B) démontre que la gestion du conflit à l'aide d'un affaiblissement n'est qu'un cas particulier dans le cadre générique présenté.

Nous mettons, ainsi, en évidence le fait que l'opérateur introduit ici permet de prendre en compte la plupart des stratégies de gestion de conflit selon les poids accordés à chacune des propositions concernées par la redistribution de la masse conflictuelle.

4.2.6 Méthodes de calcul des poids

A partir de la définition des poids $w(A, m_1, \ldots, m_j, \ldots, m_J)$ associés à chaque sous-ensemble $A \subseteq \Theta$, il est ainsi possible de décliner différents opérateurs. Dans [83], nous avons présenté deux opérateurs particuliers de cette famille. Le but des opérateurs de combinaison ainsi proposés, est de redistribuer la masse conflictuelle parmi les sous-ensembles qui ont produit le conflit. Ainsi, le conflit partiel m^* est redistribué proportionnellement à un poids parmi les sous-ensembles qu'ils l'ont engendrés. Les poids sont calculés à l'aide des masses de chacun des sous-ensembles impliqués dans le conflit partiel. Pour le premier opérateur, ce conflit est redistribué uniquement sur les sous-ensembles l'occasionnant. Le second opérateur de combinaison permet de répartir le conflit sur les sous-ensembles mais aussi sur leur disjonction. Mais d'autres stratégies peuvent être mises en place. Dans cette section, nous présentons différentes méthodes afin de déterminer les poids $w(A, m_1, \ldots, m_j, \ldots, m_J)$ pour chaque sous-ensemble $A \subseteq \Theta$.

Répartition du conflit par une expertise

Le formalisme de combinaison ainsi introduit peut s'avérer utile dans le cas d'une connaissance supplémentaire afin de résoudre le conflit. Un expert spécialiste de l'application à traiter peut fournir cette connaissance. En effet, dans des domaines tels que les applications médicales, les applications de détection de cibles ou d'obstacles, les non-détections ont des conséquences plus

importantes dans la prise de décision. Dans ces domaines, la masse conflictuelle sera attribuée à l'hypothèse la plus prudente. Comme exemple, considérons un système de détection d'obstacles à l'avant d'un véhicule muni de deux capteurs de distance. Lors d'une mesure de distance à un obstacle, les informations issues des deux capteurs peuvent être conflictuelles (1 mètre pour l'un et 10 mètres pour l'autre). Il convient dans ce cas de privilégier l'information qui donne la distance la plus faible pour ne pas mettre en danger la vie du conducteur. Lorsqu'aucune connaissance supplémentaire ne peut être fournie par un expert, nous pouvons adopter une stratégie prudente en répartissant la masse conflictuelle sur les éléments focaux les plus grands (c'est-à-dire en accordant un poids plus important à ces éléments) ou en les apprenant de manière automatique comme nous le proposons dans la section suivante.

Apprentissage automatique à l'aide d'une fonction de coût

On se propose ici de tenir compte de l'imprécision de l'étiquetage lors de la détermination des poids. Pour cela nous étudions dans un premier le cas classique des étiquettes précises et nous présentons ensuite une extension à l'étiquetage imprécis.

Etiquetage certain - Nous proposons un apprentissage des poids à partir des données par minimisation d'une fonction d'erreur. Cette fonction d'erreur sera définie comme l'erreur quadratique moyenne entre la probabilité pignistique $BetP$ calculée à l'aide de l'équation (2.48) et l'indicateur d'appartenance à chaque hypothèse. L'erreur quadratique moyenne E_{MS} de l'ensemble des vecteurs de la base d'apprentissage est alors définie par l'équation suivante :

$$E_{MS}(w) = \sum_{i=1}^{I} \sum_{n=1}^{N} [BetP^{(i)}(H_n) - u_n^i]^2 \tag{4.21}$$

où $BetP^{(i)}$ représente la probabilité pignistique d'un vecteur $x^{(i)}$ de la base d'apprentissage. Ce critère a déjà été utilisé pour l'optimisation de paramètres [37, 149]. De la même manière, nous déterminons l'ensemble des poids $w(A, m_1, \ldots, m_j, \ldots, m_J)$ pour $A \subseteq \Theta$ en minimisant la fonction décrite par l'équation (4.21).

Etiquetage imprécis - Le critère utilisé précédemment est défini uniquement dans le cas d'étiquettes précises. Mais il arrive parfois que cet étiquetage revête un caractère imprécis. C'est le cas notamment lorsque l'étiquetage est réalisé par plusieurs experts qui ne sont pas en accord. Ainsi, le problème de discrimination peut se reformuler en présence d'étiquette imprécise en ces termes. On suppose que l'ensemble d'apprentissage $\mathcal{L}$ est composé de I pairs $(x^{(i)}, m_{\mathcal{L}}^{(i)})$ où $m_{\mathcal{L}}^{(i)}$ est une fonction de $2^{\Theta} \rightarrow [0, 1]$ qui vérifie :

$$\begin{cases} \sum\limits_{A \subseteq \Theta} m_{\mathcal{L}}^{(i)}(A) & = 1 \\ m_{\mathcal{L}}^{(i)}(\emptyset) & = 0. \end{cases} \tag{4.22}$$

Par ce biais, l'étiquetage prend alors la forme d'une fonction de croyance [109]. Le critère présenté dans l'équation (4.21) peut très bien être adapté aux étiquettes données sous forme de fonctions de croyance. Ainsi, on peut utiliser le critère introduit dans [39] qui est défini par :

$$C(w) = I - \sum_{i=1}^{I} \sum_{A \subseteq \Theta} BetP^{(i)}(A).m_{\mathcal{L}}^{(i)}(A) \tag{4.23}$$

où $BetP^{(i)}$ représente la probabilité pignistique estimée pour le $i^{\text{ème}}$ élément de la base d'apprentissage $\mathcal{L}$. Ainsi, la minimisation de ce critère peut se faire part le biais des poids w. De plus, il est intéressant de remarquer que le critère précédemment cité répond à quelques propriétés qui sont à satisfaire dans le cadre d'étiquettes imprécises. La première propriété repose sur l'adéquation du critère présenté avec les coûts $\{0, 1\}$ utilisés dans le cas d'étiquettes précises. La seconde propriété concerne les vecteurs qui ne seraient pas étiquetés dans l'ensemble d'apprentissage. En effet, on peut remarquer que l'ajout de vecteurs dont l'étiquetage correspond à $m_{\mathcal{L}}(\Theta) = 1$ ne modifie pas la valeur de C. Il est envisageable d'utiliser d'autres critères qui pourraient répondre à ces conditions [73].

4.3 Résultats

Nous présentons différents résultats qui nous permettent de décrire les comportements des différentes stratégies de calcul des poids pour la répartition de la masse conflictuelle. Dans un premier temps, nous reprendrons le test présenté par Zadeh [147] afin de comparer le comportement en fonction du conflit de l'opérateur de Dempster et de la stratégie que nous proposons (Section 4.3.1). Nous présentons ensuite une comparaison entre l'opérateur de Dempster et notre stratégie de redistribution de conflit en terme d'interprétation de la masse résultante (Section 4.3.2). De plus, nous étudierons l'évolution des frontières de décision sur des données synthétiques (Section 4.3.3). Un test sur une base de données médicales permettra de mettre en évidence les bénéfices issus de la connaissance d'un expert pour la répartition du conflit (Section 4.3.4). Enfin, une illustration des performances en terme de classification avec un apprentissage des poids sera présentée dans la section 4.3.5.

Afin de modéliser les masses initiales, plusieurs méthodes existent (cf. Chapitre 3). Pour les tests qui vont suivre, nous avons obtenu les fonctions de croyance à partir des techniques suivantes. Nous avons utilisé la méthode des distances associées aux prototypes [37] pour les tests réalisés dans les sections 4.3.1 à 4.3.4. L'approche par les k plus proches voisins [35] a été utilisée pour les tests élaborés dans la section 4.3.5.

4.3.1 Sensibilité de l'opérateur de Dempster

La description des fonctions de croyance utilisées pour ce test est présentée dans la section 4.1.1. Avec ce test, nous allons comparer les résultats de la combinaison de Dempster avec ceux obtenus avec deux répartitions de conflit différentes. La première combinaison considérée répartit la masse conflictuelle de manière uniforme sur l'ensemble des hypothèses singletons :

$$w_1(\{H_n\}, m_1, m_2) = 1/3 \quad \forall \quad n \in \{1, 2, 3\} \tag{4.24}$$

et la seconde :

$$w_2(\{H_1\}, m_1, m_2) = w_2(\{H_3\}, m_1, m_2) = 0.1 \quad w_2(\{H_2\}, m_1, m_2) = 0.8. \tag{4.25}$$

Les masses résultantes de ces combinaisons sont représentées en fonction du conflit sur les figures FIG. 4.3 et FIG. 4.4. A l'aide de ces figures, nous pouvons constater que les masses combinées

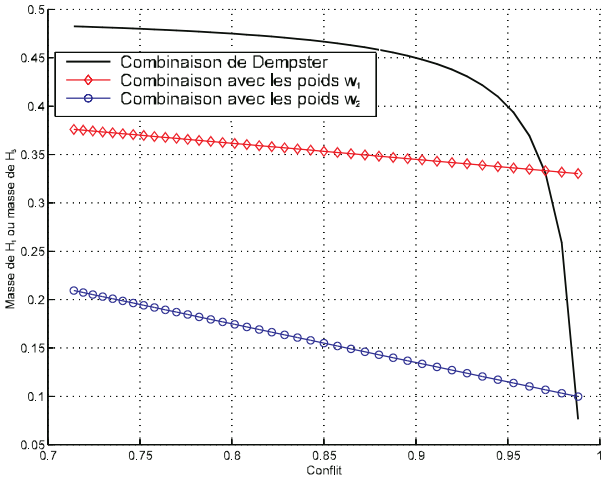

FIG. 4.3 : – Evolution de la masse de H_1 en fonction du conflit K.

avec les stratégies proposées aux équations (4.24) et (4.25) varient linéairement en fonction du conflit alors que les masses obtenues avec la combinaison de Dempster varient fortement en cas de conflit important. Ceci permet de s'affranchir de l'extrême sensibilité de l'opérateur de Dempster en cas de fort conflit et donc d'éviter des changements brutaux de comportements qui seraient contre-intuitifs.

4.3.2 Répartition de la masse résultante

Dans ce paragraphe, nous allons voir comment on peut interpréter le jeu de masses fusionné à l'aide de l'opérateur que nous proposons et l'interprétation qui peut être faite sur le jeu de masses issu de la combinaison de Dempster. Les fonctions de croyance sont construites à l'aide

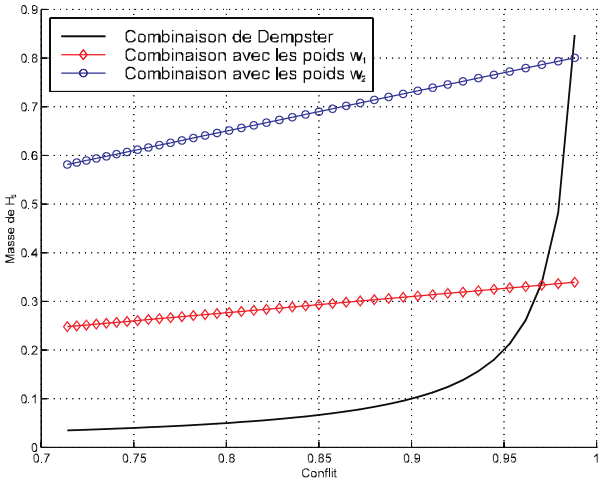

FIG. 4.4 : – Evolution de la masse de H_2 en fonction du conflit K.

de la méthode proposée par Denœux [37]. Pour ce test, nous considérons la base de données IRIS créée par Fischer en 1936 [59, 60], qui est l'une des bases les plus souvent utilisée et citée dans le cadre de la reconnaissance des formes [53]. Cette base de données est constituée de trois classes de 50 exemples chacune, où chaque classe correspond à une variété d'iris : Virginica que l'on note H_1, Versicolor que l'on note H_2 et Setosa que l'on note H_3. A chaque exemple de cette base est associé un vecteur de quatre caractéristiques (largeur et longueur des sépales et des pétales). Dans cette étude, nous considérons uniquement la largeur ainsi que la longueur des pétales afin d'avoir un représentation dans un espace à deux dimensions. Les fonctions de croyance sont construites en utilisant trois prototypes par classe. Nous avons ainsi trois fonctions de croyance par classe et donc neuf fonctions $(m_1, \ldots, m_9)$ à combiner. Sur la figure FIG. 4.5, nous avons représenté les exemples d'apprentissage constituant ces trois classes dans l'espace des deux caractéristiques retenues. De plus, sur cette figure, nous pouvons visualiser la localisation du conflit ainsi que son amplitude. Nous remarquons que les zones conflictuelles ainsi représentées reflètent l'ambiguïté d'appartenance des exemples à l'une ou à l'autre des classes (ici majoritairement entre H_2 et H_3). Sur la figure FIG. 4.6, nous avons représenté l'évolution, dans l'espace des caractéristiques, de la valeur maximale de la masse de croyance après fusion avec l'opérateur de Dempster et avec l'opérateur que nous proposons. Sur ces figures, plus les zones sont claires plus la valeur maximale de la masse de croyance est importante. Les valeurs de poids utilisés pour la redistribution du conflit sont les suivantes :

$$w(\{H_n\}, m_1, \ldots, m_9) = 1/3 \qquad \forall \quad n \in \{1, 2, 3\}. \tag{4.26}$$

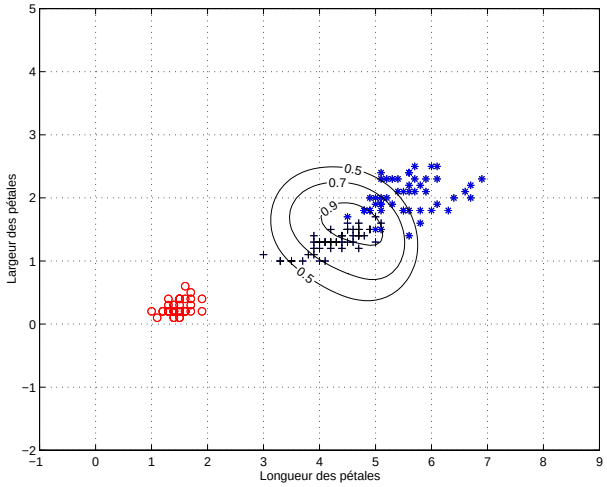

FIG. 4.5 : – Contour du conflit dans l'espace des caractéristiques (o=H_1, +=H_2 et *=H_3).

Avec ce choix de poids, nous ne privilégions aucune hypothèse lors de la redistribution du conflit. Nous pouvons constater sur la figure FIG. 4.6 (gauche) qu'avec l'opérateur de Dempster les transitions entre les valeurs maximales des masses se font de manière brutale. Sur la figure FIG. 4.6

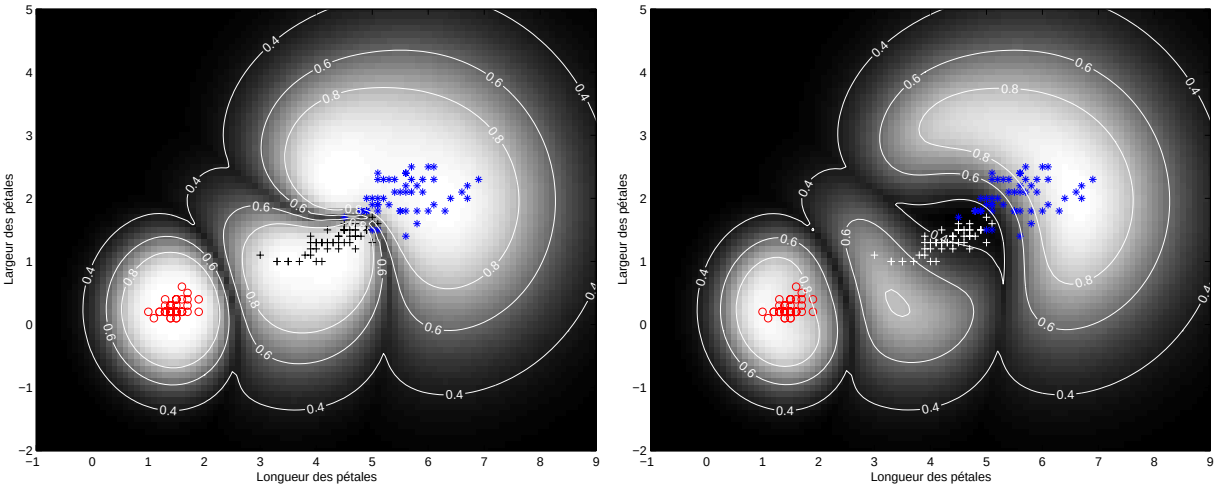

FIG. 4.6 : – Maxima des masses de croyance obtenus avec la combinaison de Dempster (gauche) et avec notre approche (droite) (o=H_1, +=H_2 et *=H_3).

(droite), la variation des masses de croyance maximales est moins brutale. En effet, le passage entre les différents maxima de la masse de croyance s'effectue en passant par un plateau caractérisant la zone où les sources d'information sont en conflit. La valeur de la masse de croyance pour les trois hypothèses dans cette zone est approximativement 1/3, reflétant ainsi le conflit et se traduisant par une absence de prise de décision (pas d'hypothèse favorisée). Ainsi, le jeu de masses résultant de la fusion avec notre approche permet de conserver l'information conflit pour la prise de décision au contraire de l'opérateur de Dempster.

4.3.3 Evolution des frontières de décision

Nous étudions, maintenant, l'évolution des frontières de décision en fonction des poids pour la redistribution de la masse conflictuelle. Pour cela, nous considérons un problème à deux hypothèses. Pour ce jeu de données, la première hypothèse H_1 suit une distribution normale de moyenne μ et de variance Σ avec :

$$\mu = (0,0)^t \qquad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.27}$$

La seconde classe suit un arc de parabole d'équation :

$$y = 0.4x^2 + 5 \quad \text{avec } x \in [-4;4]. \tag{4.28}$$

Ces données sont ensuite bruitées sur chacune des composantes à l'aide d'un bruit gaussien de moyenne μ_b et de variance Σ_b :

$$\mu_b = (0,0)^t \qquad \Sigma_b = 0.35. \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.29}$$

Chaque classe est constituée de 500 exemples. Les fonctions de masse, de la même manière que précédemment, sont obtenues en utilisant trois prototypes par classe. Nous avons ainsi six fonctions de croyance à combiner. Les données des deux classes sont représentées sur la figure FIG. 4.7. Les exemples appartenant à l'hypothèse H_1 sont symbolisés par des cercles et les exemples de l'hypothèse H_2 par des croix. Les frontières de décision caractérisées par les contours de la probabilité pignistique sont représentées sur les figures FIG. 4.7 et FIG. 4.8 pour différentes distributions de poids. Sur ces figures, les prototypes utilisés pour la modélisation sont représentés par une croix noire pour l'hypothèse H_1 et par un cercle noir pour l'hypothèse H_2. Sur la figure FIG. 4.7, les poids sont répartis de la manière suivante : $w(\{H_1\}, m_1, \ldots, m_6) = 0.4$ et $w(\{H_2\}, m_1, \ldots, m_6) = 0.6$. Dans ce cas, la masse conflictuelle sera répartie de manière plus importante sur l'hypothèse H_2 au détriment de l'hypothèse H_1. Nous pouvons constater que la frontière de décision, représentée dans ce cas par une probabilité pignistique égale à 0.5, est décalée vers le centre de l'hypothèse H_1. Ainsi les points appartenant à la zone d'ambiguïté seront affectés majoritairement à l'hypothèse H_2. De la même manière sur la figure FIG. 4.8, où les poids sont répartis ainsi : $w(\{H_1\}, m_1, \ldots, m_6) = 0.6$ et $w(\{H_2\}, m_1, \ldots, m_6) = 0.4$, la masse conflictuelle sera redistribuée plus sur l'hypothèse H_1 que sur l'hypothèse H_2 et ainsi les exemples où le conflit est important, seront affectés à l'hypothèse H_1. Enfin, sur la même base de données, nous avons simulé les frontières de décision obtenues à l'aide de la probabilité pignistique. Nous supposons deux cas.

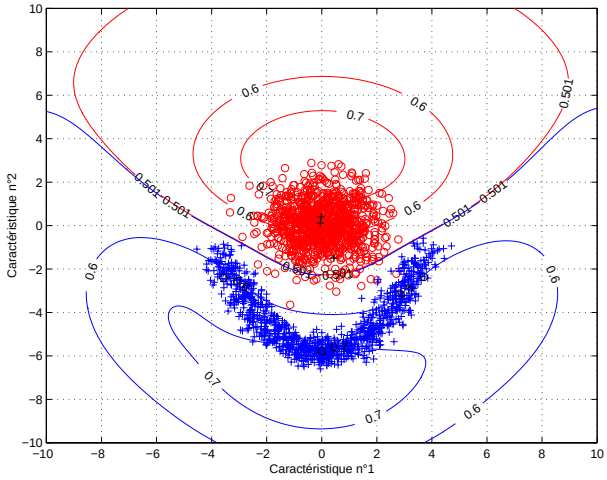

FIG. 4.7 : – Contour de la probabilité pignistique dans l'espace des caractéristiques avec $w(\{H_1\}, m_1, \ldots, m_6) = 0.4$ et $w(\{H_2\}, m_1, \ldots, m_6) = 0.6$ (o=$H1$, +=H_2).

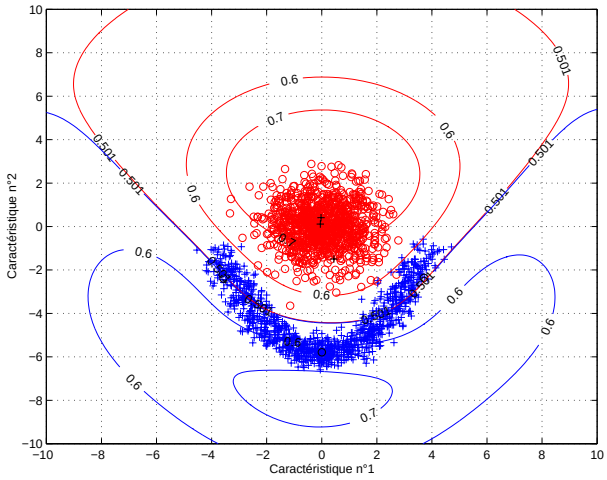

FIG. 4.8 : – Contour de la probabilité pignistique dans l'espace des caractéristiques avec $w(\{H_1\}, m_1, \ldots, m_6) = 0.6$ et $w(\{H_2\}, m_1, \ldots, m_6) = 0.4$ (o=$H1$, +=H_2).

Affaiblissement : Dans un premier temps, nous considérons le cas où les fonctions de croyance sont affaiblies et combinées ensuite avec l'opérateur de Dempster. Nous pouvons alors considérer plusieurs stratégies. Soit, nous considérons que les sources associées à l'hypothèse H_2 (c'est-à-dire les prototypes dans le cas de la modélisation choisie) sont peu fiables et sont donc affaiblies par rapport à celles associées à H_1. Cette situation est appelée α_1. La situation inverse, c'est-à-dire lorsque l'on affaiblit uniquement les sources associées à H_1, est notée α_3. Enfin, la situation où les sources sont affaiblies de la même manière est notée α_2. Pour ce test, nous avons associé un coefficient d'affaiblissement de 0.75 pour une source dite non fiable. Les fonctions de croyance issues de sources fiables ne sont pas modifiées. Les frontières de décision obtenues sont représentées sur la figure FIG. 4.9.

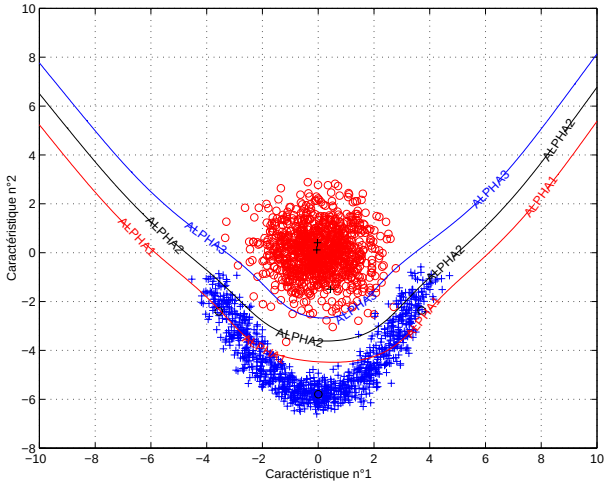

FIG. 4.9 : – Frontières de décision selon les différentes situations d'affaiblissement α_1, α_2 et α_3 (o=H_1, +=H_2).

Redistribution de la masse conflictuelle à l'aide de poids : Dans le second cas, les fonctions de croyance ne sont pas affaiblies et elles sont fusionnées avec l'opérateur proposé. On peut alors :

- soit privilégier l'hypothèse H_1 par rapport à H_2, on accorde alors un poids plus important à l'hypothèse H_1 qu'à l'hypothèse H_2, comme par exemple :

$$w_1(\{H_1\}, m_1, \ldots, m_6) = 0.6 \quad \text{et} \quad w_1(\{H_2\}, m_1, \ldots, m_6) = 0.4,$$

- soit au contraire associer un poids plus important à l'hypothèse H_2 par rapport à H_1, comme par exemple :

$$w_3(\{H_2\}, m_1, \ldots, m_6) = 0.6 \quad \text{et} \quad w_3(\{H_1\}, m_1, \ldots, m_6) = 0.4,$$

- soit ne privilégier aucune des deux hypothèses en leurs accordant un poids identique :

$$w_2(\{H_1\}, m_1, \ldots, m_6) = w_2(\{H_2\}, m_1, \ldots, m_6) = 0.5.$$

Les frontières de décision obtenues sont représentées sur la figure FIG. 4.10. On peut constater sur la figure FIG. 4.9, que les frontières de décision, obtenues avec différentes stratégies d'affaiblissement, ont uniquement subi une translation. Ainsi l'affaiblissement de sources d'information modifie la totalité de la frontière de décision. Sur la figure FIG. 4.10, on peut remarquer qu'en fixant les poids associés à chaque hypothèse pour la combinaison, les frontières de décision ne sont modifiées que sur la zone la plus conflictuelle (donc localement). Dans les zones où le conflit est faible, les frontières de décision convergent. Toutefois, on peut remarquer que, dans le cas où les poids dépendent des fonctions de masses, les résultats obtenus avec les coefficients de fiabilité auraient pu être retrouvés (Cf. Annexe B).

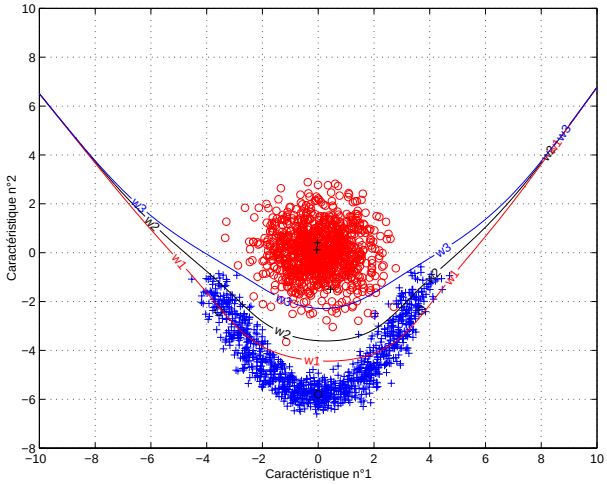

FIG. 4.10 : – Frontières de décision obtenues avec la méthode proposée pour différentes stratégies de redistribution de la masse conflictuelle w_1, w_2 et w_3 (o=H_1, +=H_2).

4.3.4 Connaissance experte

Dans la section 4.2.6, nous avons vu que les poids peuvent être déterminés à l'aide d'une connaissance supplémentaire issue d'un expert. Le but de ce test consiste à montrer l'utilité du cadre que nous proposons dans cette configuration. De manière à mettre en évidence les avantages de ce cadre, nous avons choisi de réaliser un test sur une base de données médicales. La pathologie à diagnostiquer est le diabète [89]. Le jeu de données est composé de 768 individus répartis en deux classes : patients sains (hypothèse H_1) et pathologiques (hypothèse H_2). Les huit variables observées sont extraites soit de tests cliniques réalisés sur les patients ou de caractéristiques propres aux patients (antécédents, âge, ...). La classe des patients sains est représentée par 500 individus tandis que la classe des patients atteints de diabète compte 268 patients dans la base. Le critère de décision employé, pour les résultats présentés ici, est le maximum de probabilité pignistique. On peut alors supposer que plusieurs experts ayant des objectifs différents interviennent afin de définir les valeurs de poids pour la redistribution du conflit. Un premier expert (Expert n°1) peut, par exemple, vouloir minimiser l'erreur de classification et par ce biais optimiser les valeurs des poids. Un autre expert (Expert n°2) souhaite détecter l'ensemble (ou le maximum) de cas pathologiques en s'autorisant des fausses alarmes c'est-à-dire des patients sains qu'il diagnostiquerait diabétiques. Cet avis consiste donc à répartir le conflit sur l'hypothèse pathologique en s'imposant les poids $w(\{H_1\}, m_1, \ldots, m_6) = 0$ et $w(\{H_2\}, m_1, \ldots, m_6) = 1$. Enfin, un troisième expert (Expert n°3) préfère considérer que l'existence d'un conflit est révélatrice d'une autre maladie (connue ou inconnue) que le diabète et ainsi rejeter la prise de décision. Pour cela, la masse conflictuelle sera conservée sur l'ensemble vide ($w(\emptyset, m_1, \ldots, m_2)$). Les différents

résultats sont présentés dans le tableau TAB. 4.1. Sur ce tableau, on peut constater que pour

	Non détection	**Fausse alarme**	**Bonne classification**	**Rejet**
Expert n°1 : poids optimisés $w(\{H_1\}, m_1, \ldots, m_6) = 0.57$ $w(\{H_2\}, m_1, \ldots, m_6) = 0.43$	0.552	0.03	**0.765**	0
Expert n°2 : poids fixés $w(\{H_2\}, m_1, \ldots, m_6) = 1$	**0.044**	0.850	0.431	0
Expert n°3 : poids fixés $w(\emptyset, m_1, \ldots, m_6) = 1$	0.132	0.280	0.807	**0.817**

TAB. 4.1 : – Résultats de bonne classification selon la combinaison employée.

l'expert n°1 les valeurs de poids permettent d'obtenir un taux de bonne classification de l'ordre de 0.765. De la même manière, la connaissance apportée par l'expert n°2 conduisent à un taux d'erreur de l'ordre de 0.569. Même si ce taux est important, il est à noter que seulement 12 patients pathologiques ont été classés sains minimisant ainsi le nombre de non-détection. Enfin, les connaissances de l'expert n°3, qui considère que le cadre de discernement n'est pas exhaustif et associe l'ensemble vide à une hypothèse de rejet, permettent d'obtenir un taux de bonne classification 0.807 en rejetant près de 82% des patients. Notons que dans le cas particulier d'un problème de discrimination, ces résultats aurait pu être obtenus de manière classique en modifiant les coûts de décision. Toutefois, dans certaines applications, où il ne s'agit pas, après la fusion, de prendre une décision sur la fonction de probabilité pignistique mais de dégager une tendance générale ou de déterminer une préférence entre plusieurs possibilités [43], il est nécessaire et utile de fournir une information (exemple : une fonction de croyance) capable de rendre compte de l'incertitude sur le résultat. Le choix de poids particuliers nous semble plus approprié. En effet, dans ce cas, l'utilisation de coûts de décision, qui ne modifient pas les masses de croyance, n'est pas adéquate.

4.3.5 Apprentissage des poids

Etiquetage certain

Pour ce test, nous considérons un problème à deux hypothèses et à deux caractéristiques. L'ensemble d'apprentissage disponible est issu de distributions normales de moyenne μ_n et de matrice de variance Σ_n :

$$\begin{array}{cc} \mu_1 = (2,2)^t & \mu_2 = (4,4)^t \\ \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \tag{4.30}$$

alors que les mesures simulées, constituant la base de test, suivent les distributions de moyenne μ'_n et de variance Σ'_n suivantes :

$$\begin{array}{ll} \mu'_1 = (2+S, 2+S)^t & \mu'_2 = (4+S, 4+S)^t \\ \Sigma'_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \Sigma'_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{array} \tag{4.31}$$

Ce test représente le cas classique de dérive de capteurs. En effet, dans l'exemple présenté, les données varient linéairement en fonction d'un signal S (FIG. 4.11 et FIG. 4.12). Les ensembles

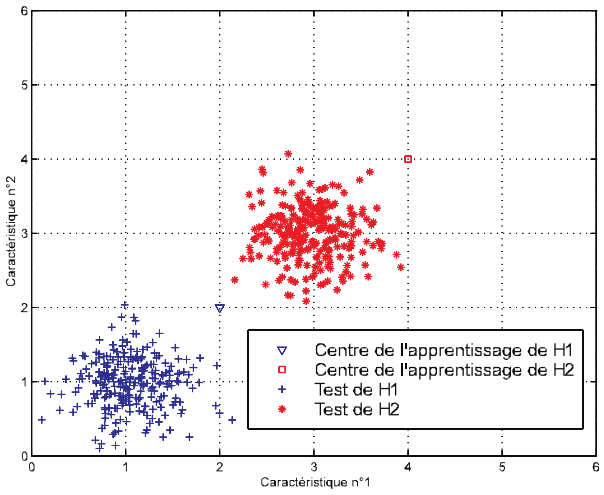

FIG. 4.11 : – Représentation des données simulées pour $S = -1$.

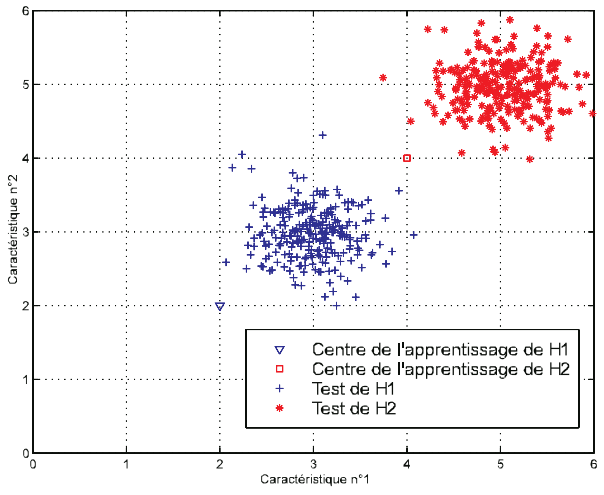

FIG. 4.12 : – Représentation des données simulées pour $S = 1$.

d'apprentissage et de test sont constitués de 250 vecteurs par hypothèse. Les fonctions de croyance sont obtenues à l'aide de la méthode des k-ppv crédibilistes proposée par Denœux [35, 149]. Pour ce test, nous prendrons un nombre de voisins égal à 5. On obtient ainsi cinq fonctions de croyance, notées $(m_1, \ldots, m_5)$, à fusionner. La décision est prise sur le critère du maximum de probabilité

pignistique. Les données constituant la base de validation seront simulées de la même manière que les données de la base de test avec les distributions présentées dans l'équation (4.31). Ces données permettent de définir les valeurs des poids $w(A, m_1, \ldots, m_5)$, avec $A \subseteq \Theta$, à l'aide de l'erreur quadratique moyenne définie par l'équation (4.21). Les résultats en terme de taux de classification, obtenus en réalisant une moyenne sur 10 tirages, sont représentés sur la figure FIG. 4.13. D'après cette figure, nous pouvons constater que la démarche de calcul des poids pour

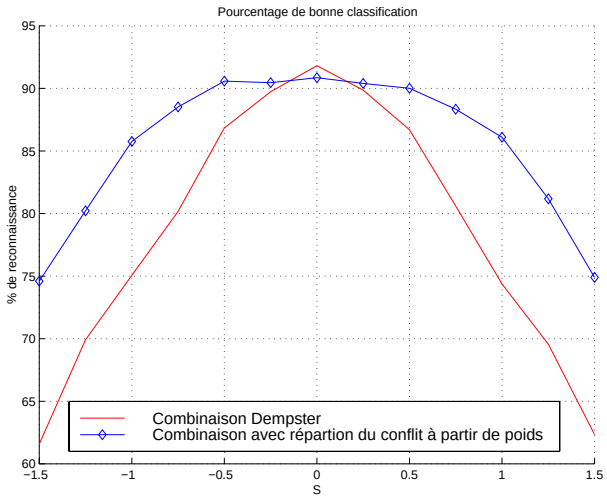

FIG. 4.13 : – Evolution du pourcentage de bonne classification en fonction de S.

la redistribution de la masse conflictuelle présentée ici permet d'obtenir un gain de classification notable par rapport à l'emploi de la combinaison classique de Dempster. En effet, lorsque la dérive des données est importante, c'est-à-dire lorsque $|S| \in [0.5, 1.5]$, le pourcentage de bonne classification obtenu par la répartition du conflit à l'aide de poids est de l'ordre de 10% supérieur à celui obtenu par la combinaison de Dempster. Lorsque la derive des données est faible, c'est-à-dire $S \approx 0$, les résultats entre les deux combinaisons sont similaires. La figure FIG. 4.14 représente l'évolution des poids de répartition du conflit en fonction de S. Nous pouvons constater que dans le cas d'un apprentissage correct, c'est-à-dire $S \approx 0$, la répartition du conflit se fera équitablement entre les deux hypothèses ($w(\{H_1\}, m_1, \ldots, m_5) = w(\{H_2\}, m_1, \ldots, m_5) = 0.5$). Par contre, dans le cas où les données de test de l'hypothèse H_1 se situe dans la zone conflictuelle (c'est-à-dire lorsque $S > 0$), le poids accordé à l'hypothèse H_1 sera supérieur au poids associé à l'hypothèse H_2 ($w(\{H_1\}, m_1, \ldots, m_5) > w(\{H_2\}, m_1, \ldots, m_5)$). De la même manière, lorsque $S < 0$ le poids associé à l'hypothèse H_2 pour la répartition de la masse conflictuelle est plus important que celui attribué à H_1.

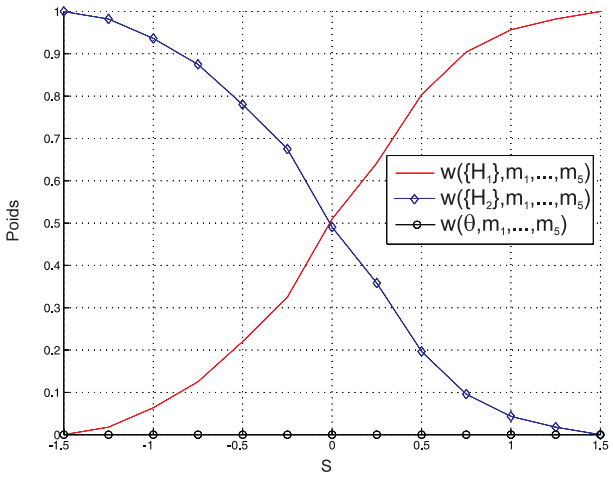

FIG. 4.14 : – Evolution des poids en fonction de S.

Etiquetage imprécis

Pour illustrer le comportement de la méthode en fonction de l'opérateur de combinaison choisi, dans le cadre de l'étiquetage imprécis, on génère un ensemble d'apprentissage $\mathcal{L}$ composé de 3 classes gaussiennes de 100 vecteurs chacune, issues des distributions $\mathcal{N}(\mu_q, \Sigma_q)$ suivantes :

$$\mu_1 = (0,0)^t,\ \mu_2 = (2,2)^t,\ \mu_3 = (10,10)^t \tag{4.32}$$

$$\Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},\ \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},\ \Sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.33}$$

A partir de ces distributions, on réalise un étiquetage imprécis des vecteurs de l'ensemble d'apprentissage de deux façons permettant ainsi d'illustrer deux situations différentes. L'un correspond à un étiquetage ensembliste au sens strict du terme tandis que l'autre est un étiquetage imprécis définie de la même manière qu'une fonction de croyance.

Etiquetage ensembliste - Pour le vecteur $x^{(i)}$ de classe précise H_n dans l'ensemble d'apprentissage, l'étiquetage imprécis est généré de la manière suivante. On tire aléatoirement un sous-ensemble A tel que $H_n \subset A \subseteq \Theta$. La totalité de la masse de croyance est affectée à ce sous-ensemble :

$$m^i_{\mathcal{L}}(A) = 1. \tag{4.34}$$

L'ensemble de test $\mathcal{T}$ est composé de 250 vecteurs par classe. Les paramètres de la méthode sont fixés de la manière suivante : $k = 15$, $\alpha = 0.99$ et le paramètre γ est pris comme préconisé dans [35] égal à la moyenne des distances aux k voisins dans l'ensemble d'apprentissage.

Les figures FIG. 4.15, FIG. 4.16 et FIG. 4.17 représentent, respectivement, les probabilités pignistiques pour les trois règles de combinaison envisagées (la règle de combinaison de Dempster, l'opérateur de Yager et l'opérateur pondéré avec optimisation des poids). Dans cet exemple à

3 classes, il est possible de représenter les probabilités pignistiques dans un triangle équilatéral dont chacune des hauteurs représente la probabilité pignistique de chacune des classes. De cette manière, les vecteurs de l'ensemble de test sont représentés dans le triangle ($\square = H_1$, $\circ = H_2$, $+ = H_3$) ainsi que la zone de rejet possible pour deux valeurs de coût ($\lambda_0 = 0.25$ et $\lambda_0 = 0.5$) en traits pleins. La prise de décision sans rejet est également représentée en trait pointillé dans ce même triangle.

En terme de classification, on a représenté sur les figures FIG. 4.18 et FIG. 4.19, le taux d'erreur en classification en fonction du taux de rejet pour les deux règles de décision qui correspondent à la minimisation du risque pignistique (cf. équation (2.50)) et du risque inférieur (cf. équation (2.51)). On peut remarquer que la règle de combinaison de Dempster obtient de moins bonnes performances que les deux autres opérateurs. Pour l'opérateur pondéré, la répartition des poids est la suivante : $w(\{H_1, H_2\}, m_1, \ldots, m_{15}) = 0.41$, $w(\{H_1, H_3\}, m_1, \ldots, m_{15}) = 0.1$, $w(\{H_2, H_3\}, m_1, \ldots, m_{15}) = 0.13$ et $w(\Theta, m_1, \ldots, m_{15}) = 0.36$.

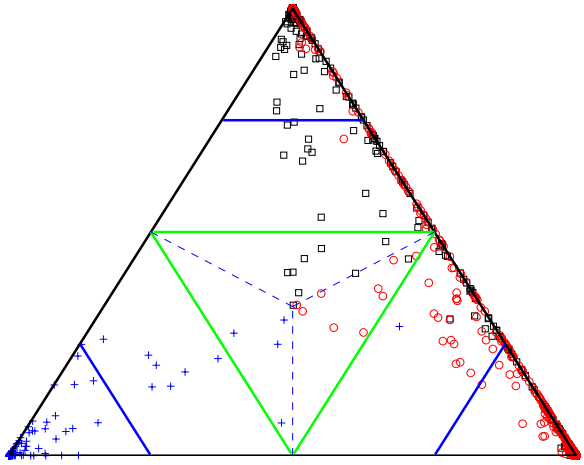

Fig. 4.15 : – Probabilité pignistique avec l'opérateur de Dempster, $\square = H_1$, $\circ = H_2$, $+ = H_3$

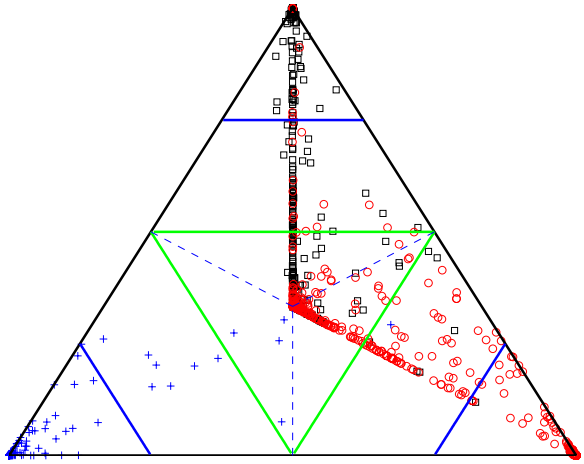

Fig. 4.16 : – Probabilité pignistique avec l'opérateur de Yager, $\square = H_1$, $\circ = H_2$, $+ = H_3$

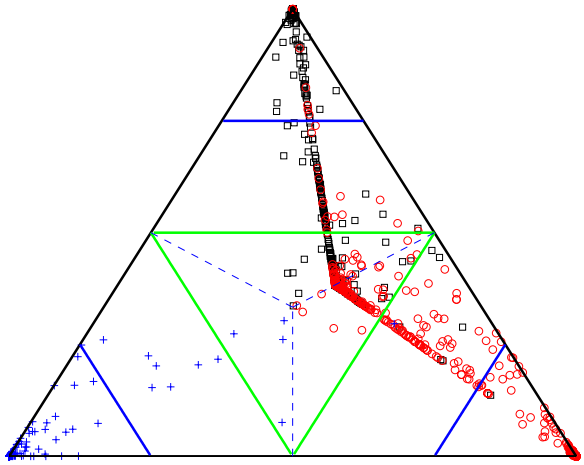

Fig. 4.17 : – Probabilité pignistique avec l'opérateur pondéré, $\square = H_1$, $\circ = H_2$, $+ = H_3$

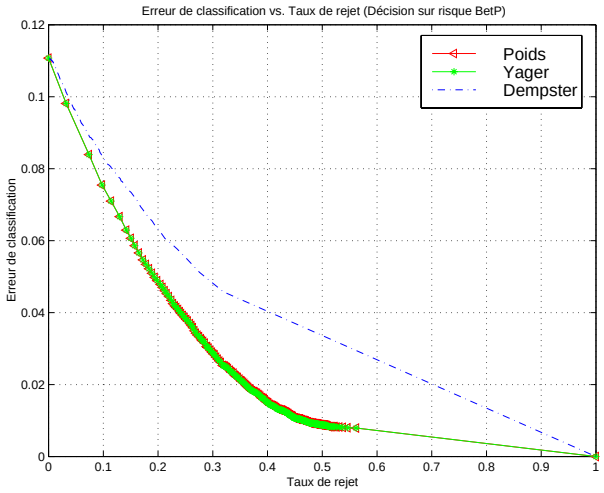

FIG. 4.18 : – Taux d'erreur vs. taux de rejet pour le risque pignistique, $\cdot\!-$ = Dempster, $*$ = Yager, $\triangleleft$ = Poids

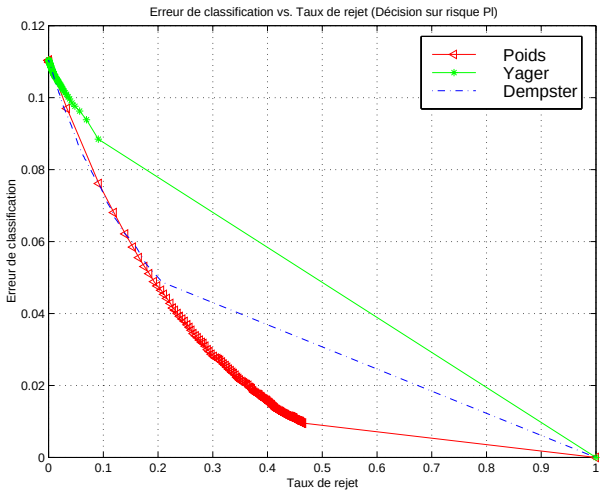

FIG. 4.19 : – Taux d'erreur vs. taux de rejet pour le risque inférieur, $\cdot\!-$ = Dempster, $*$ = Yager, $\triangleleft$ = Poids

Etiquetage imprécis -Dans cette seconde simulation, on souhaite un étiquetage imprécis pour chacun des vecteurs $x^{(i)}$. On tire aléatoirement δ qui suit une distribution de probabilité uniforme sur $[0, 1]$. On tire, de la même manière, u indépendamment du premier tirage. Si $\delta > u$, on tire aléatoirement un sous-ensemble A tel que $H_n \subseteq A \subseteq \Theta$. Sinon, on tire un sous-ensemble quelconque $A \subseteq \Theta$. Dans ce cas, A ne contient pas forcément H_n. On répartit la masse de croyance de la manière suivante :

$$m^i_{\mathcal{L}}(A) = \delta \tag{4.35}$$

$$m^i_{\mathcal{L}}(\Theta) = 1 - \delta. \tag{4.36}$$

De cette façon, la fonction de croyance $m^i_{\mathcal{L}}$ représente la confiance dans la vraie classe H_n à

laquelle on ajoute le degré de fiabilité δ de l'expert qui réalise l'étiquetage. De la même façon que précédemment, on a représenté les probabilités pignistiques pour les trois opérateurs (FIG. 4.20, FIG. 4.21 et FIG. 4.22), ainsi que les performances en classification pour les deux règles de décision (FIG. 4.23 et FIG. 4.24). Pour cette simulation, le paramètres de modélisation sont fixés de la même manière que lors du test sur l'étiquetage ensembliste. De plus, les poids $w(A, m_1, \ldots, m_{15})$ ont été fixés en répartissant l'unité de manière uniforme ($w(A, m_1, \ldots, m_{15}) = 1/(2^{|\Theta|}-1) \ \forall \ A \subseteq \Theta$). On constate donc que l'opérateur de Yager et l'opérateur proposé ont des comportements similaires au niveau des probabilités pignistiques. Ce constat n'est plus vrai pour les plausibilités et dans cette simulation, c'est l'opérateur pondéré qui obtient les meilleures performances.

Commentaires - Dans les deux simulations réalisées sur l'étiquetage imprécis, on peut remarquer que dans le cadre d'une prise de décision basée sur la probabilité pignistique, l'opérateur de Yager et celui proposé ont le même comportement. En effet, on constate que les courbes du taux d'erreur en classification en fonction du taux de rejet sont similaires (Cf. FIG. 4.18 et FIG. 4.23). Ceci s'explique par une répartition de la masse conflictuelle de manière uniforme sur les différentes hypothèses ce qui mène à des probabilités pignistiques identiques. (Cf. FIG. 4.16 et FIG. 4.17 puis FIG. 4.21 et FIG. 4.22).

En ce qui concerne le prise de la décision fondée sur le risque inférieur, cette remarque n'est plus vraie. En effet, l'opérateur proposé permet d'obtenir de meilleurs résultats par rapport aux opérateurs de Dempster et de Yager. Pour ce dernier, la masse conflictuelle étant redistribuée sur Θ, les plausibilités de chaque hypothèse H_n sont proches de 1 ce qui ne permet pas de rejeter certains exemples qui l'étaient dans le cas de la prise de décision avec la probabilité pignistique.

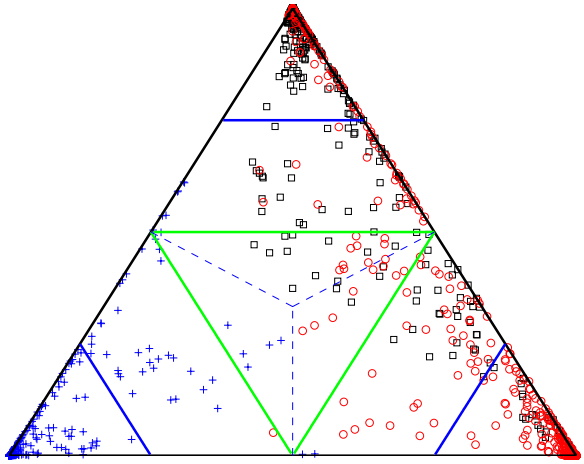

FIG. 4.20 : – Probabilité pignistique avec l'opérateur de Dempster, $\square = H_1$, $\circ = H_2$, $+ = H_3$

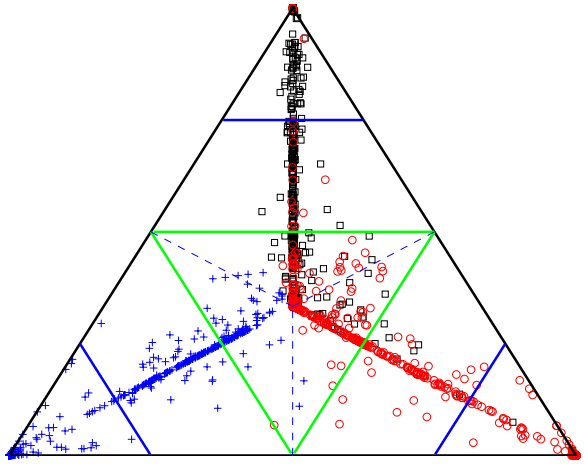

FIG. 4.21 : – Probabilité pignistique avec l'opérateur de Yager, $\square = H_1$, $\circ = H_2$, $+ = H_3$

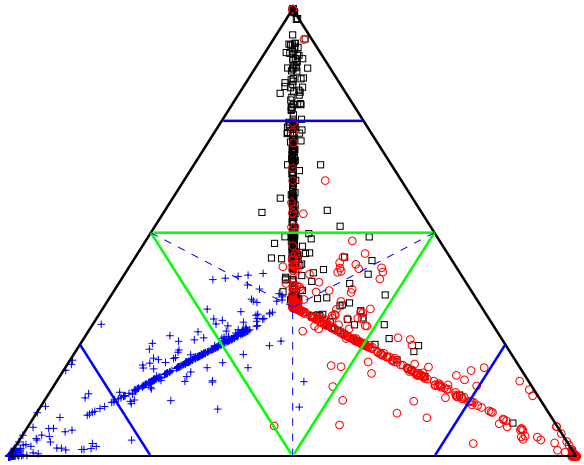

FIG. 4.22 : – Probabilité pignistique avec l'opérateur pondéré, $\square = H_1$, $\circ = H_2$, $+ = H_3$

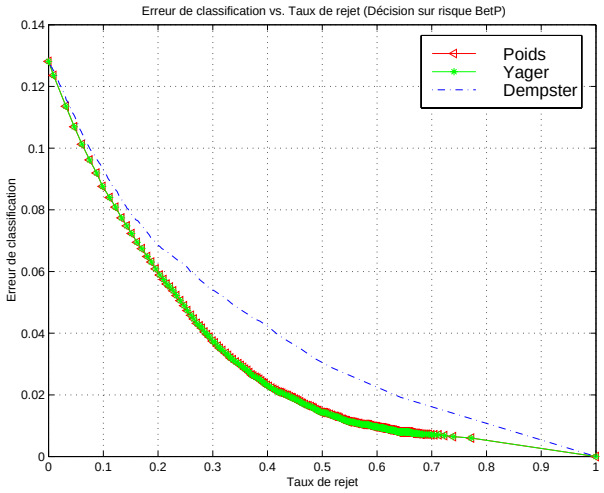

Fig. 4.23 : – Taux d'erreur vs. taux de rejet pour le risque pignistique, $\cdot\!-$ = Dempster, $*$ = Yager, $\triangleleft$ = Poids

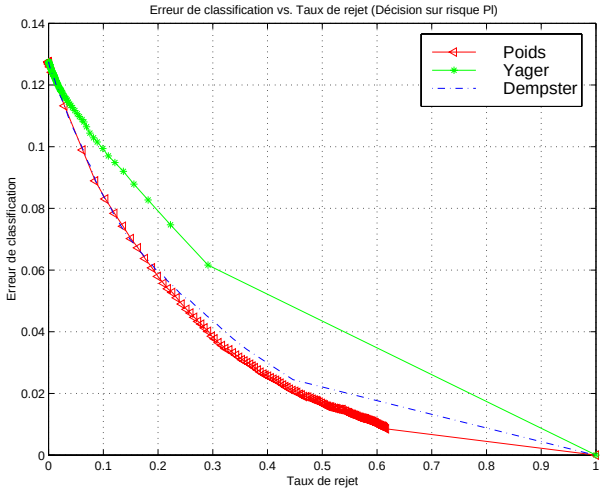

Fig. 4.24 : – Taux d'erreur vs. taux de rejet pour le risque inférieur, $\cdot\!-$ = Dempster, $*$ = Yager, $\triangleleft$ = Poids

4.4 Conclusion

Dans ce chapitre, nous avons présenté un cadre générique pour la fusion de sources d'information modélisées à l'aide de fonctions de croyance. A partir de ce cadre de travail, nous avons retrouvé les opérateurs classiques de combinaison utilisés au sein de la théorie de l'évidence. De fait, ce cadre générique permet de définir une famille d'opérateurs de combinaison. En effet, il est possible de décliner différents opérateurs à partir de poids, notés $w(A, m_1, \ldots, m_j, \ldots, m_J)$, associés à chacune des propositions $A \subseteq \Theta$. Nous avons proposé plusieurs méthodes d'obtention des poids. L'une de ces méthodes est fondée sur l'intégration d'une connaissance supplémentaire que pourrait fournir un expert afin de résoudre le conflit. La seconde méthode repose sur la dé-

termination des poids en minimisant l'erreur quadratique moyenne. Cette seconde technique est adaptée aux problèmes de discrimination auquel nous nous sommes intéressés dans ce chapitre. Ces deux méthodes ont été testées et comparées aux combinaisons utilisées classiquement dans la théorie de l'évidence. Ainsi, les simulations réalisées ont permis de mettre en évidence qu'un choix judicieux de la règle de combinaison peut permettre d'obtenir de meilleures performances en terme de classification. De cette manière, l'utilisation de l'opérateur proposé est une solution élégante qui permet non seulement de retrouver l'ensemble des opérateurs introduits dans la littérature mais aussi de créer une famille beaucoup plus générique pour la gestion du conflit. L'optimisation des poids relatifs à cet opérateurs permet en outre d'obtenir de meilleures performances quelque soit la stratégie choisie pour la prise de décision. Dans le chapitre suivant, nous allons voir comment utiliser les techniques proposée dans le cadre d'une application d'aide au diagnostic médical.

Chapitre 5

Application au diagnostic de mélanomes malins

Le mélanome malin est, parmi les cancers cutanés, l'une des formes les plus graves, bien qu'il ne représente que 10% de cette forme de cancer à l'heure actuelle. L'augmentation du nombre de mélanomes diagnostiqués ces dernières années pose un réel problème de santé publique. Dans ce type de cancer, le pronostic favorable de survie du patient est fortement dépendant à la fois de la qualité du diagnostic posé et du fait qu'il intervienne dans les plus courts délais après le début de la maladie. C'est pourquoi, les acteurs du monde médical tendent à développer et mettre en œuvre une politique de dépistage systématique. Il suffit, pour s'en convaincre, de constater les campagnes annuelles de dépistage gratuit auprès de la population et les reportages et enquêtes de plus en plus fréquents que mènent les média à l'orée des périodes estivales. A tel point que ce sujet en deviendrait presque ce que les professionnels appellent un "marronnier" ! Alors que le nombre de cas de mélanome est croissant, cette forme de cancer se prête *a priori* pourtant bien au diagnostic, puisqu'il atteint la couche cutanée visible à l'œil nu. Cependant, bien des difficultés sont rencontrées sur le plan clinique, lorsqu'il s'agit de poser un tel diagnostic.

5.1 Introduction

Les lésions næviques se distinguent en général de la peau par une différence de couleur ou de texture. Elles peuvent avoir des dimensions de l'ordre de quelques millimètres à un ou deux centimètres. Il existe plusieurs types de lésions dont la forme et les couleurs sont très variables. A titre d'indication, la figure Fig. 5.1 montre différents types de lésions que le dermatologue est amené à rencontrer dans la pratique quotidienne. On peut constater sur cette figure que la reconnaissance des nævi peut poser problème visuellement car certains d'entre-eux dits atypiques, mais bénins, peuvent être associés à une lésion maligne à tort. Ces images montrent une grande diversité de forme et de couleurs, qui entraîne une certaine imprécision, un manque de reproductibilité et donc une variabilité des interprétations.

Fig. 5.1 : – Exemples de lésions.

Le diagnostic clinique repose sur la règle ABCDE [62] qui a été conçue pour faciliter la reconnaissance clinique de mélanomes malins. Chaque lettre de cette règle correspond à une caractéristique visuelle permettant de détecter les mélanomes malins :

- le A signifie asymétrie de la forme,
- le B signifie bordure irrégulière,
- le C signifie couleurs différentes,
- le D signifie diamètre supérieur à 5mm,
- et enfin le E correspond à l'élévation de la lésion.

La fiabilité du diagnostic posé à l'aide de ces différentes caractéristiques visuelles n'excède pas 65%, même lorsque celui-ci est posé par un dermatologue chevronné. En moyenne, seulement 10% des exérèses effectuées révèlent, après analyse histologique, une pathologie cancéreuse. Ceci démontre, si tant est que cela soit nécessaire, la difficulté de mettre en place un système d'aide au diagnostic.

La méthodologie qui a été mise en œuvre afin de résoudre ce problème repose sur quatre grandes étapes :

1. L'acquisition des images de lésions cutanées issues de la numérisation de diapositives dont les conditions d'acquisition ne sont pas contrôlées,
2. La segmentation des images, c'est-à-dire la séparation de la lésion et de la peau saine qui l'environne,
3. L'extraction de caractéristiques pertinentes, c'est-à-dire informatives quant à la nature de la lésion,
4. La mise en œuvre d'un processus de discrimination des lésions en tenant compte de la fiabilité et l'inconsistance des caractéristiques extraites afin d'élaborer une stratégie de décision pour détecter les mélanomes malins.

Les trois premiers points ont été développés dans la thèse de Taouil [125]. Les travaux effectués dans le cadre de notre étude s'articule plus particulièrement autour du quatrième point.

Ce chapitre se compose de deux parties. La première partie (Section 5.2) résume les résultats des travaux obtenus dans le cadre de la thèse de Taouil [125], ce qui nous permet de décrire le début de la chaîne de traitement des images de lésions en allant de l'acquisition de l'image à l'extraction des caractéristiques des lésions. Nous terminons ce chapitre par l'application des travaux présentés dans les chapitres 3 et 4 sur les données ainsi extraites des lésions (Cf. Section 5.3).

5.2 Etudes préalables

Au sein de cette section, nous détaillons le procédé mis en place dans l'étude préliminaire [125] afin d'expliquer la méthodologie employée pour obtenir les caractéristiques qui seront utilisées pour la discrimination des lésions malignes. Trois points ont été abordés au cours de cette étude. Le premier concerne le protocole d'acquisition des images de lésions (Section 5.2.1). La suite de l'étude (Section 5.2.2) a porté sur la méthode de segmentation nécessaire à l'extraction de la zone d'intérêt de l'image, c'est-à-dire la lésion. Enfin, nous abordons la définition des caractéristiques numériques dans la section 5.2.3.

5.2.1 Acquisition des images

L'acquisition des lésions cutanées repose sur la numérisation de diapositives prises par les dermatologues. Cette méthode d'acquisition indirecte[1] offre l'avantage de ne pas changer les habitudes des dermatologues et par conséquence de faciliter l'étape de collecte des images. Cependant, la qualité de prise de vue reste à surveiller. En effet, les procédés d'utilisation, des matériels photographiques ainsi que les modifications des conditions d'éclairage et les effects des bains chimiques pour obtenir les diapositives impliquent une grande diversité des images.

Les diapositives sont recueillies auprès des dermatologues par la clinique dermatologique du CHU de Rouen. Elles sont ensuite numérisées afin d'obtenir des images au format de 150×150 pixels. Chacun de ces pixels est codé par trois octets représentant respectivement les trois plans couleurs rouge, vert et bleu.

5.2.2 Segmentation des images

Pour réaliser une classification des lésions dermatologiques, et en particulier pour discriminer les mélanomes malins des lésions bénignes, une étape préalable permettant d'isoler la lésion

[1] Au contraire du terme direct qui consisterait à obtenir directement une image numérique par l'intermédiaire d'un appareil photo numérique ou d'une caméra CCD.

dans chaque image couleur, est nécessaire. En substance, il s'agit de séparer la lésion de la peau saine qui l'environne, dans le but ultérieur d'extraire des informations de forme et de couleur caractérisant cette lésion, informations utilisables dans un processus de discrimination. Pour ce faire, il est nécessaire de mettre en place une méthodologie de segmentation, robuste vis-à-vis des conditions d'acquisition des images et permettant de localiser le plus "finement" possible, au regard de l'expertise médicale, les contours de la lésion[2].

Dans notre application, la principale zone d'intérêt est la lésion. C'est l'image binarisée issue de la segmentation qui définit la forme géométrique à caractériser. Le masque binaire, ainsi produit, permet d'éliminer ou de retenir la valeur d'un pixel donné de l'image lorsqu'on caractérise la lésion. Une erreur lors de l'étape de segmentation entraîne des erreurs sur les paramètres caractéristiques des formes et des couleurs de la lésion.

Une première méthode de segmentation fondée sur la coopération d'une approche région et d'une approche frontière a été étudiée [25, 24, 29]. A l'aide d'une approche région fondée sur la maximisation d'une entropie, une première phase consiste à déterminer une zone certaine de la lésion et une zone certaine de la peau saine. Ainsi, on obtient une première segmentation, dite "grossière", qui permet de distinguer dans les images une première région de pixels qui appartiennent sûrement à la peau saine, une seconde région de pixels qui appartiennent sûrement à la lésion et une couronne constituée de pixels ne pouvant être attribués en première analyse à la peau saine ou à la lésion ; on parlera alors de zone "floue", ou plutôt de zone ou région d'incertitude. C'est dans cette zone que se trouve la frontière qui sera déterminée de manière précise par une approche contour. Afin d'illustrer la méthode proposée, nous donnons à la figure FIG. 5.2 quelques résultats de contours obtenus sur des images de nævi et de mélanomes. Toutefois, dans certains cas, les contours obtenus ne sont pas satisfaisants. Les situations pour

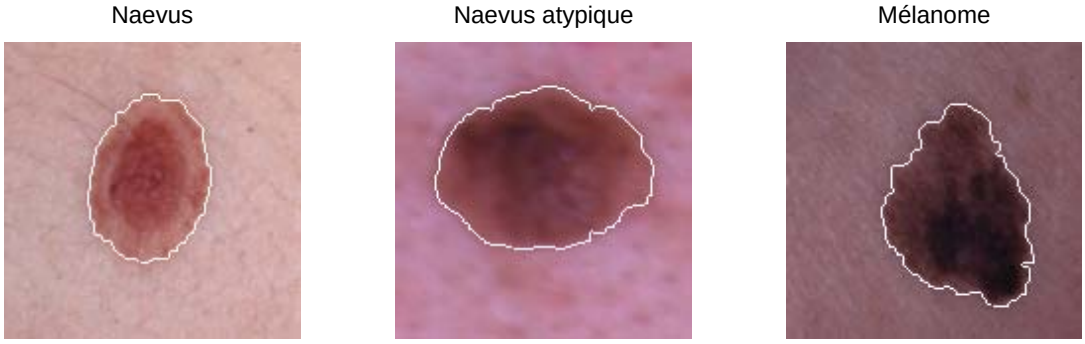

FIG. 5.2 : – Lésions avec contours détectés en surimpression.

[2]Les informations de forme des lésions sont des indicateurs utilisés par les dermatologues pour qualifier la malignité possible ou non d'une lésion.

lesquelles on rencontre le plus fréquemment des contours mal définis sont :

- la présence sur la peau saine de zones sombres (métastases, rougeurs, boutons, ...) qui ont été interprétées comme appartenant à la lésion,
- la présence de points brillants (surexposition locale ou étendue),
- la présence de poils.

Ces différentes situations sont représentées sur la figure FIG. 5.3. Bien qu'un certain nombre de

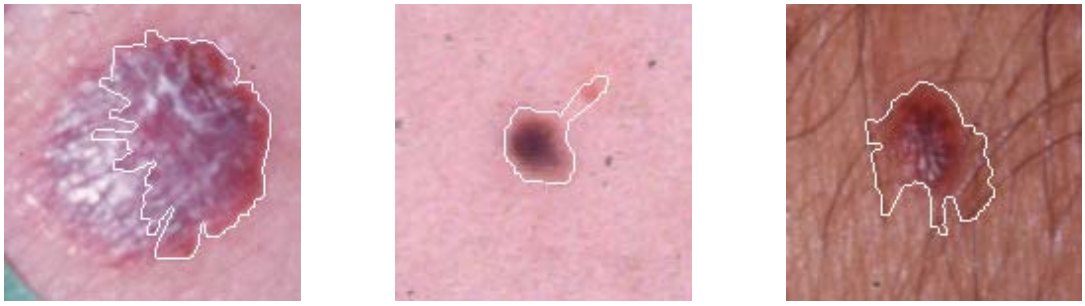

FIG. 5.3 : – Exemples d'images avec contours mal détectés.

ces erreurs soient imputables en grande partie aux conditions d'acquisition des images, une autre approche est à l'étude. Cette approche utilise la fusion des informations provenant de différents plans couleurs à l'aide de la théorie de l'évidence [84, 128]. L'utilisation de cette technique, dans le cadre proposé, permet d'employer plusieurs plans couleurs, de quantifier l'incertitude en ce qui concerne l'affectation d'un pixel à une zone et de rejeter des pixels n'appartenant ni à la lésion ni à la peau saine (exemple : poils, point brillant, ...) pour le calcul des paramètres. Toutefois, nous n'exposons pas d'avantage cette approche qui est en cours de développement et qui nécessite d'être approfondie notamment par l'utilisation des informations spatiales issues du voisinage de chaque pixel. Une fois la segmentation obtenue, les grandeurs numériques permettant de caractériser la forme sont définies.

5.2.3 Extraction des caractéristiques de la lésion

Le choix de caractéristiques convenables a nécessité la prise en compte de plusieurs contraintes. Ces contraintes sont principalement dues à l'effet de l'échantillonnage et aux prises de vue non contrôlées. Ces deux effets sont partiellement réduits en choisissant, lorsque cela est possible, des caractéristiques géométriques invariantes par similitude, sans grandeurs ou issues de calculs de moyennes.

La construction des caractéristiques permettant de discriminer les lésions bénignes des lésions malignes repose :

- sur l'étude des caractéristiques "exprimées" par les dermatologues et qu'ils exploitent en routine clinique,
- sur des caractéristiques classiques d'étude de forme (exemple : le rapport de finesse, l'allongement de la forme, ...),
- sur des caractéristiques issues de l'analyse des couleurs de l'image et plus particulièrement de la répartition spatiale de ces couleurs.

Les différentes caractéristiques ont été regroupées en deux catégories : les caractéristiques géométriques et les caractéristiques photométriques. Il existe cinq caractéristiques géométriques et quatorze caractéristiques photométriques. La liste des paramètres est la suivante :

Caractéristiques géométriques :

- *Rapport de finesse* : cette caractéristique correspond au rapport entre la surface de la lésion et celle d'un cercle dont le centre de gravité correspond au centre de la lésion et de rayon la distance moyenne des points de la lésion à ce centre. Ainsi, cette caractéristique vaut 1 si la lésion à la forme d'un disque, et est inférieure à 1 dans le cas d'une autre forme.
- *Etendue* : elle correspond à la différence entre la plus grande et la plus petite des distances d'un point du contour de la lésion au centre de gravité.
- *Ecart moyen au centre de la forme* : cette grandeur correspond à la distance moyenne entre un pixel de la lésion et le centre de gravité de cette lésion.
- *Régularité du contour* : un contour est régulier s'il ne présente aucune distortion ou encoche. Cette caractéristique est égale à la somme des distances séparant chaque point du contour à un cercle idéal. Une normalisation par la distance moyenne et par la longueur du périmètre du cercle est effectuée pour que cette caractéristique ne soit pas liée aux dimensions apparentes de la lésion.
- *Allongement de la forme* : cette caractéristique correspond au rapport entre la plus petite et la plus grande longueur suivant les axes d'inertie de la lésion.

Caractéristiques photométriques :

La première caractéristique photométrique est la *quantification de surface floue*. Cette caractéristique est égale au rapport entre la surface de zone d'incertitude et la surface totale de la lésion. Cette zone d'incertitude, qui correspond à la portion de l'image à l'apparence floue, est déterminée après la segmentation grossière. Elle correspond à la zone de peau séparant la peau saine de la lésion.

D'autres caractéristiques sont obtenues en déterminant l'homogénéité d'une couleur. Pour

calculer *l'homogénéité d'une couleur*, nous partons de l'histogramme de la couleur des pixels correspondant à la lésion afin de définir des niveaux. A chaque niveau est associée une fausse couleur comme suit :

- un premier tiers constitué des pixels de niveaux les plus bas (appelée zone "sombre"),
- un second tiers constitué des pixels de niveaux les plus élevés (appelée zone "claire"),
- un dernier tiers constitué des pixels de niveaux intermédiaires.

L'homogénéité de la couleur est définie comme la somme des transitions d'une zone à une autre. Cette caractéristique essaie de quantifier l'apport de flux sanguin au sein de la lésion.

Pour illustrer cette caractéristique, nous avons représenté sur la figure FIG. 5.4 la répartition de la couleur rouge pour une lésion nævique et pour une lésion maligne. La zone dite sombre est représentée en rouge, la zone intermédiaire en blanc et la zone claire en vert. On peut noter que la lésion bénigne a une répartition en "cible", alors que le mélanome a une distribution beaucoup plus "chaotique".

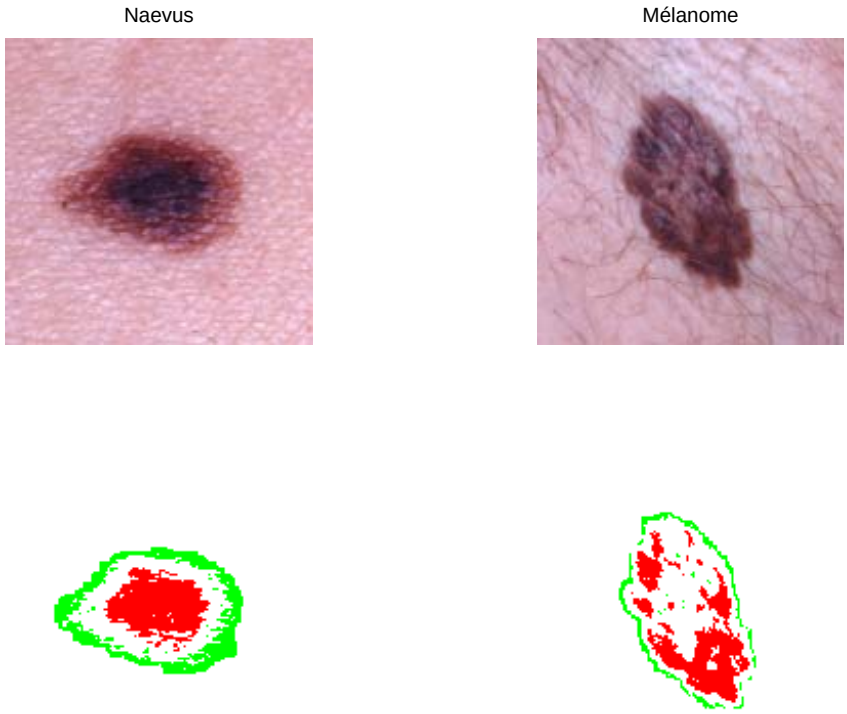

FIG. 5.4 : – Images originales et illustrations de l'homogénéité de la couleur rouge pour une lésions nævique et une lésion maligne.

En partant de l'histogramme couleur défini précédemment, nous déterminons *la symétrie d'une couleur* en calculant la différence entre la moyenne des distances au centre des pixels les moins intenses et celle des plus intenses. Cette différence est normalisée par la distance moyenne

au centre de la lésion. La symétrie est calculée sur le rouge, le vert, le bleu, la teinte et la saturation. Nous obtenons ainsi cinq caractéristiques quantifiant la symétrie de la lésion.

Enfin, *l'écart d'une couleur* se détermine en réalisant la différence entre la moyenne des pixels de la peau saine et la moyenne des pixels de la lésion sur la couleur choisie. Cet écart est calculé sur les plans couleurs rouge, vert et bleu ce qui permet d'avoir trois caractéristiques supplémentaires.

Les caractéristiques, qui viennent d'être décrites, ont été mises au point lors d'une première étude dans le cadre de la thèse de Taouil [125]. Malheureusement, cette élaboration des caractéristiques a été menée sur un petit nombre de diapositives. En passant à un éventail plus large d'images de lésions cutanées, il s'est avéré que certaines caractéristiques sont peu pertinentes. Ainsi, nous avons pris le parti d'être le plus proche du raisonnement du dermatologue en sélectionnant les caractéristiques les plus pertinentes qui correspondent à la règle ABCDE. Ainsi, parmi les 19 primitives présentées, nous n'en retenons que 5 :

- l'*allongement de la lésion* qui permet de quantifier l'**A**symétrie de la forme,
- la *quantification de surface floue* qui donne une information sur la **B**ordure,
- l'*écart sur le rouge*, l'*écart sur le vert* et l'*écart sur le bleu* qui symbolisent la différence de **C**ouleur.

Malheureusement, les acquisitions d'images étant non contrôlées, aucune caractéristique reflétant le **D**iamètre n'a pu être calculé. De la même manière, l'image ne permet pas d'obtenir des informations sur l'**E**paisseur de la lésion. Nous obtenons ainsi un ensemble de 5 grandeurs numériques susceptibles de discriminer les lésions bénignes des lésions malignes. De par la nature même de ces caractéristiques géométriques et photométriques extraites des images segmentées de lésions, nous sommes en présence d'informations incertaines et imprécises. L'utilisation de la théorie de l'évidence est donc, à notre sens, parfaitement justifiée dans le cadre applicatif visé.

5.3 Discrimination des mélanomes malins

5.3.1 Constitution des bases de données

Cette application émane d'une collaboration avec les praticiens de la clinique dermatologique de l'hôpital Charles Nicolle de Rouen. Le problème nous a été posé de la manière suivante.

Dans un premier temps, les dermatologues nous ont fourni une base d'apprentissage constituée de 80 lésions pour effectuer l'apprentissage de notre système de diagnostic. Cette base d'apprentissage a été fournie sous deux versions. Dans la première, l'étiquetage des lésions est précis et la répartition est la suivante : 61 lésions bénignes (nævi) et 19 lésions malignes (mélanomes). Le nombre de mélanomes est assez faible, mais il répond à une réalité de terrain. Il y

a peu de mélanomes au regard du nombre de lésions bénignes suspectes aux yeux des dermatologues. Pour s'affranchir de la variabilité lors de la constitution de cette base, nous découpons cette base en deux parties. La première partie est constituée d'environ deux tiers des lésions (41 lésions bénignes et 9 mélanomes) et sert à l'apprentissage et le reste des lésions constitue la base de validation nécessaire à l'estimation des paramètres. Les différents résultats présentés dans le cadre de cette section sont des moyennes sur dix tirages effectués de manière aléatoire. Pour la seconde version de cette base, à chaque lésion sont associées des opinions émanant d'un panel d'experts concernant l'appartenance de la lésion au groupe nævi ou au groupe mélanome. Ainsi pour chaque lésion, les opinions nous permettent de construire un étiquetage possibiliste comme introduit dans [46]. Soit q le nombre d'experts et q_k^i le nombre d'experts classant l'exemple i de la base d'apprentissage dans la classe k. Alors, u_k^i le degré de possibilité de l'exemple i d'appartenir à la classe k est obtenu de la manière suivante :

$$u_k^i = \frac{1}{q} \sum_{n=1}^{N} \min(q_k^i, q_n^i). \tag{5.1}$$

Ces degrés de possibilité peuvent être ensuite transformés en fonctions de croyance [44]. Nous obtenons, ainsi, un ensemble d'apprentissage $\mathcal{L}$ qui est composé de I paires $(x^{(i)}, m_{\mathcal{L}}^{(i)})$ où $m_{\mathcal{L}}^{(i)}$ est une fonction de croyance. Ces fonctions de croyance auraient pu être construites en associant à chaque expert un degré de fiabilité et en combinant l'ensembles des opinions. Mais, le choix des coefficients de fiabilité à associer à chaque expert reste délicat. Nous obtenons donc, dans le cadre de cette version de la base d'apprentissage, un étiquetage imprécis donné sous forme de fonctions de croyance. De la même manière que précédemment, afin de minimiser l'influence de la variabilité de la base d'apprentissage, les résultats présentés lors de l'utilisation de cette base sont des moyennes sur dix tirages effectués de manière aléatoire.

Enfin, pour valider les performances du système de diagnostic, les dermatologues ont fourni une base de test[3]. Cette base est constituée de 209 lésions (191 lésions bénignes et 18 mélanomes) qui ont été jugées exploitables parmi 300 lésions. Cette sélection a été faite au regard des résultats de segmentation obtenus. En effet, certaines lésions mal segmentées ne pouvaient être incluses dans la base de test et certaines diapositives étaient inexploitables au regard des conditions d'acquisition beaucoup trop mauvaises (peau saine environnant la lésion insuffisante ou absente de l'image, surexposition ou sous-exposition trop importante, diapositives floues, ...).

[3]Données obtenues dans le cadre d'un protocole prospectif en partenariat avec la clinique dermatologique de l'hôpital Charles Nicolle de Rouen.

5.3.2 Etude de l'influence des coefficients de fiabilité

Dans un premier temps, nous allons étudier l'influence des coefficients de fiabilité sur les performances en terme de discrimination pour les différentes méthodes de modélisation présentées dans le cadre du chapitre 3. La détermination des coefficients de fiabilité utilisés pour les différentes modélisations, repose sur différentes stratégies :

- stratégie n°2 : l'ensembles des coefficients de fiabilité intervenant dans la modélisation sont fixés à 0.9,
- stratégie n°3 : les coefficients de fiabilité sont déterminés à l'aide de critères d'information,
- stratégie n°4 : les coefficients de fiabilité sont obtenus en minimisant l'erreur quadratique moyenne.

Les comportements de ces différentes stratégies sont étudiés :

- soit en considérant les primitives de façon individuelle (modélisation mono-dimensionnelle),
- soit en considérant l'ensemble des primitives (modélisation multi-dimensionnelle).

Modélisation mono-dimensionnelle des fonctions de croyance

Modélisation d'Appriou - Dans le cadre la modélisation mono-dimensionnelle, nous avons, dans un premier temps, utilisé la modélisation proposée par Appriou. Ainsi pour les tests qui vont suivre, nous avons utilisé le modèle n°1 proposé par Appriou (équation 3.11). Les coefficients R_j, pour $j \in \{1, 2\}$, ont été fixés à la plus grande valeur autorisée afin d'obtenir les fonctions de croyance les moins spécifiques possibles. L'estimation des fonctions de densité est obtenue en utilisant un modèle de mélange gaussien associé à une technique d'estimation des paramètres par l'algorithme EM [80]. Le nombre de modes est obtenu par validation croisée sur chaque tirage de la base d'apprentissage. La figure FIG. 5.5 représente l'évolution du taux d'erreur en fonction du taux de rejet pour une décision fondée sur la minimisation du risque pignistique (équation (2.50)) à droite et pour une décision fondée sur la minimisation du risque inférieur (équation (2.51)) à gauche. On constate sur ces courbes que la stratégie n°4 qui consiste à déterminer les coefficients de fiabilité en minimisant l'erreur quadratique permet d'obtenir les meilleures performances. On peut aussi remarquer que la prise de décision sur la minimisation du risque pignistique permet de rejeter des vecteurs quelque soit la valeur du coût de rejet, alors que la prise de décision sur la minimisation du risque inférieur ne le permet pas. En effet, sur la figure FIG. 5.5 de droite, on constate qu'à partir d'un taux de rejet de 0.5 (pour la stratégie n°4) on ne rejette plus aucun vecteur car les plausibilités sont trop grandes. Ceci s'explique par le fait que nous avons choisi de fixer les valeurs de R_j, nous amenant au modèle le moins spécifique. Enfin, on remarque

que la stratégie n°3 mettant en œuvre les critères d'information obtient les performances les moins bonnes. Ainsi comme nous l'avons remarqué dans le chapitre 3, cette approche ne garantit pas l'obtention des meilleurs taux de classification mais quantifie à l'aide d'une distance les différences existant entre la base d'apprentissage et la base de validation. Dans le cadre de cette application, la base d'apprentissage est représentative de la base de test ainsi les coefficients obtenus par cette technique sont proches de 1. Ceci traduit le fait qu'il n'existe pas d'évolution de contexte. Toutefois, cette stratégie peut s'avérer utile lorsque la base d'apprentissage des lésions n'est pas représentative de la base de test. Nous pourrions voir un intérêt à cette méthode, par exemple, si la base d'apprentissage était constituée de lésions présentes sur des individus d'Europe du Nord (peau saine généralement claire) alors que la base de test serait constituée de lésions photographiées sur des individus du bassin Méditerranéen (peau saine généralement hâlée). Dans ce cas, les coefficients s'adapteraient suite à ce changement de contexte et seraient susceptibles de donner de meilleurs résultats.

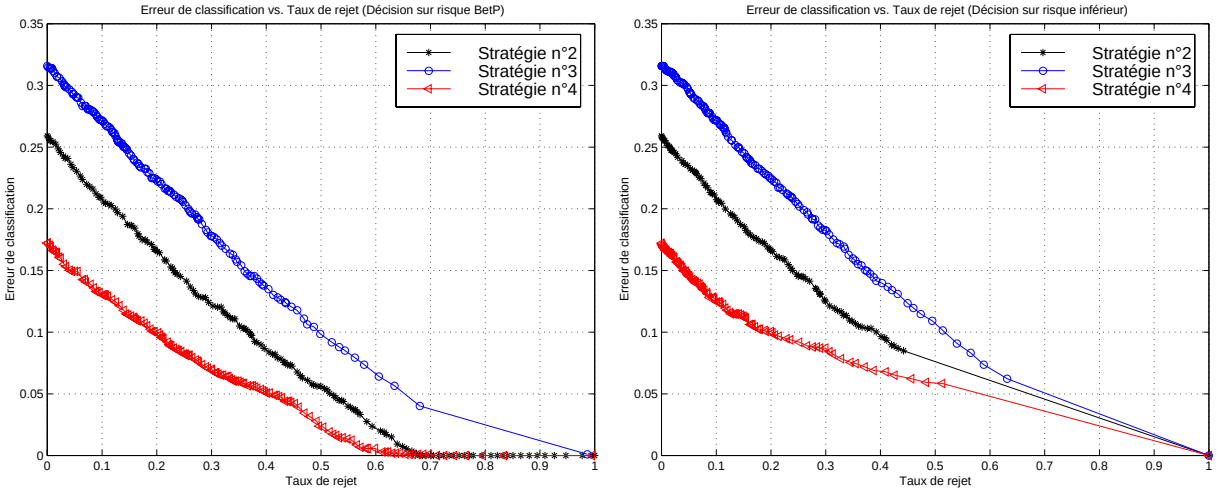

Fig. 5.5 : – Taux d'erreur vs. taux de rejet pour le risque pignistique R_{Bet} (à gauche) et le risque inférieur R_* (à droite) avec la modélisation d'Appriou

Modélisation de Denœux - Pour l'utilisation du modèle de distance proposé par Denœux, nous avons optimisé le nombre de prototypes par validation croisée pour chaque tirage de la base d'apprentissage. De la même manière, les paramètres γ_{ij} ainsi que la position des prototypes ont été optimisés par minimisation de l'erreur quadratique moyenne présentée à l'équation (3.3). La figure Fig. 5.6 représente l'évolution du taux d'erreur en fonction du taux de rejet pour une décision reposant sur la minimisation du risque pignistique (à droite) et sur la minimisation du risque inférieur (à gauche). On constate, même si cela est moins évident que dans la modélisation précédente, que les meilleurs taux de classification sont obtenus avec la stratégie n°4 où les

coefficients de fiabilité sont déterminés par minimisation de l'erreur quadratique moyenne.

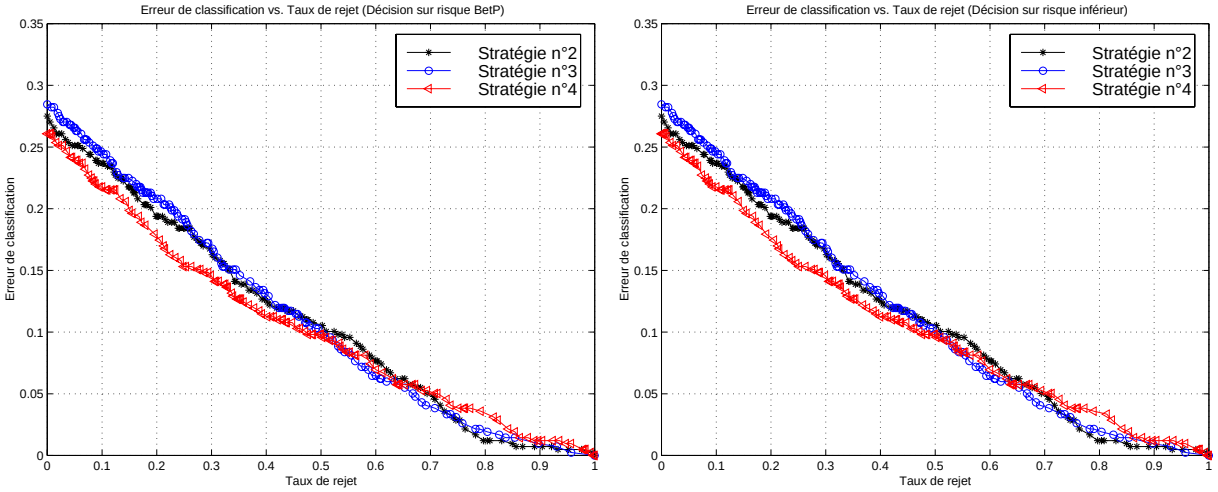

FIG. 5.6 : – Taux d'erreur vs. taux de rejet pour le risque pignistique R_{Bet} (à gauche) et le risque inférieur R_* (à droite) avec la modélisation de Denœux

Modélisation multi-dimensionnelle des fonctions de croyance

Dans le cadre de l'approche multi-dimensionnelle, on considère les caractéristiques dans leur ensemble. Pour la modélisation d'Appriou, nous utilisons l'équation (3.8). Dans ce cadre, les fonctions de densité sont estimées à l'aide d'un modèle de mélange gaussien multi-dimensionnel où le nombre de modes par classe est obtenu par validation croisée et les paramètres sont estimés par un algorithme EM. Pour l'approche distance présentée par Denœux, nous employons l'équation (3.1). De la même manière que pour l'approche vraisemblance, le nombre de prototypes est obtenu par validation croisée et les paramètres employés dans le cadre de cette modélisation sont obtenus en minimisant l'équation (3.3).

Avant d'aborder les performances des différentes stratégies étudiées, nous présentons, pour les deux modélisations, le comportement des distributions de probabilité pignistique et de plausibilité ainsi que les frontières de décision obtenues pour des décisions fondées sur la minimisation du risque pignistique et sur la minimisation du risque inférieur. On représente ces différentes figures dans un espace limité à deux caractéristiques qui sont l'allongement de la forme et la quantification de la surface floue. La figure FIG. 5.7 représente le maximum de probabilité pignistique en niveau de gris avec des niveaux de gris clairs pour des probabilités pignistiques proches de un et des niveaux de gris foncés pour des probabilités pignistique proches de zéro pour la modélisation d'Appriou. De la même manière, la figure FIG. 5.8 illustre le maximum de probabilité pignistique (à gauche) et le maximum de plausibilité (à droite) pour la méthode des distances. Sur ces différentes figures, les modes ainsi que les prototypes (fixés à 2 par classe) sont indiqués

par des astérisques (∗).

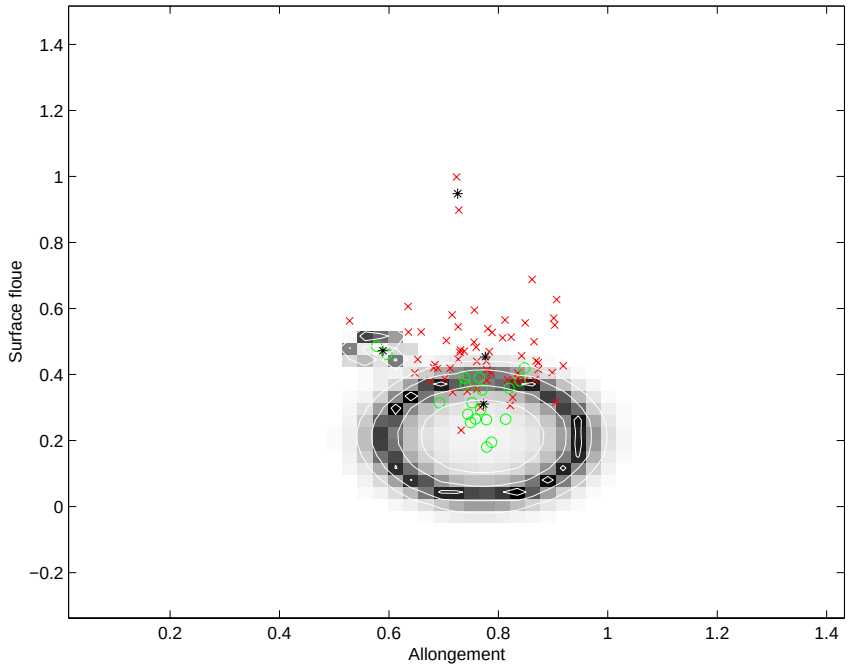

Fig. 5.7 : – Maximum de probabilité pignistique obtenu avec la modélisation d'Appriou avec les lignes de contour à 0.9, 0.75 et 0.6, (Nævus = x, Mélanome = o).

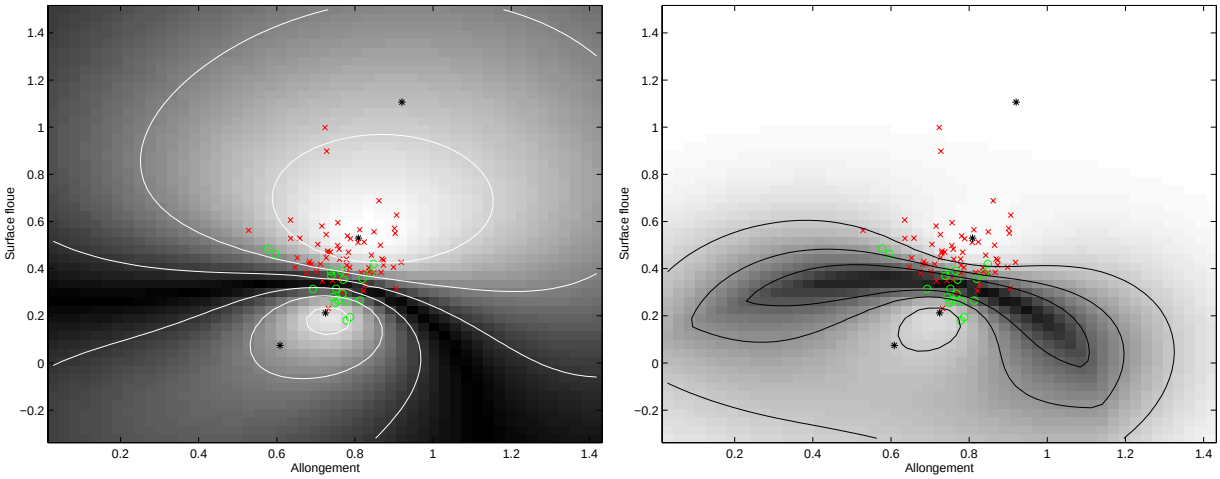

Fig. 5.8 : – Maximum de probabilité pignistique (à gauche) et maximum de plausibilité (à droite) obtenus avec la modélisation de Denœux avec les lignes de contours à 0.9, 0.75 et 0.6 (à gauche) et 0.9, 0.8 et 0.7 (à droite), (Nævus = x, Mélanome = o).

Sur les figures Fig. 5.9 et Fig. 5.10 sont représentées respectivement les frontières de décision obtenues à l'aide de l'approche distance et l'approche vraisemblance. Sur ces figures, les frontières de décision obtenues par minimisation du risque pignistique sont représentées à gauche et les frontières de décision obtenues par minimisation du risque inférieur sont elles représentées sur la droite. Pour ces figures, le coût de rejet λ_0 a été fixé à 0.4. On peut constater que dans le cadre

de la modélisation d'Appriou, la prise de décision sur le minimum de risque inférieur ne permet pas de rejet. En effet, cette modélisation impose d'avoir le maximum de plausibilité égale à 1 ($\max_n Pl(\{H_n\}) = 1$). Ainsi, dans le cas multi-dimensionnel, seule la modélisation reposant sur l'approche distance permet de faire du rejet en utilisant la règle de décision sur le maximum de plausibilité.

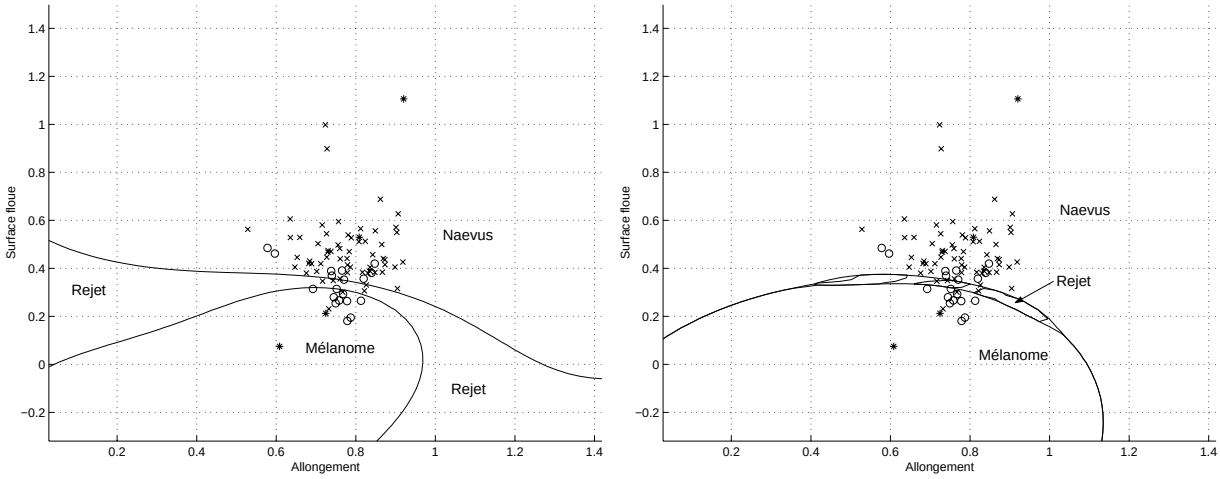

FIG. 5.9 : – Régions de décision avec la modélisation de Denœux pour le risque pignistique R_{Bet} (à gauche) et le risque inférieur R_* (à droite) avec un coût de rejet $\lambda_0 = 0.4$, (Naevus = x, Mélanome = o).

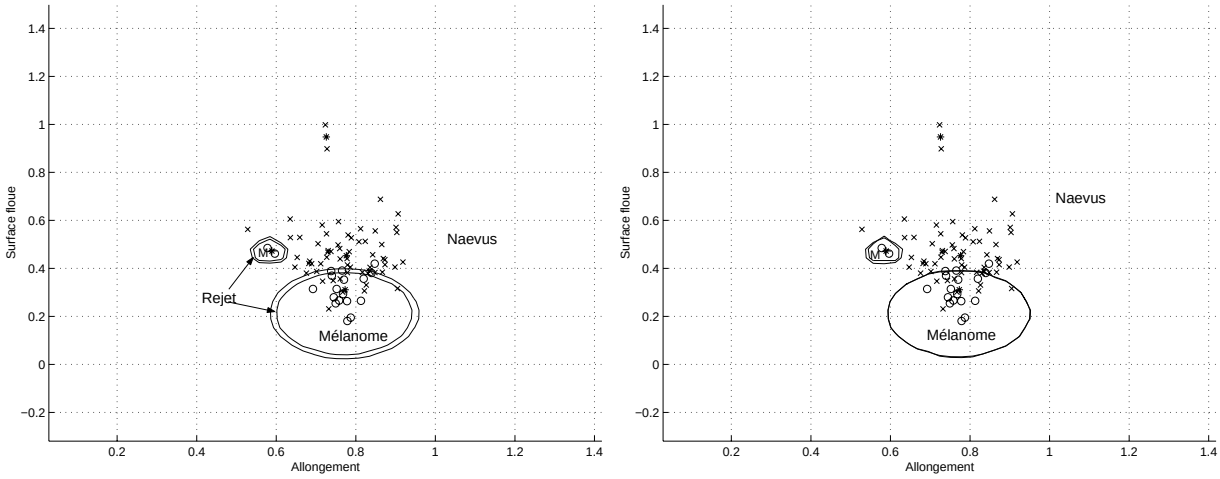

FIG. 5.10 : – Régions de décision avec la modélisation d'Appriou pour le risque pignistique R_{Bet} (à gauche) et le risque inférieur R_* (à droite) avec un coût de rejet $\lambda_0 = 0.4$, (Naevus = x, Mélanome = o).

Afin d'étudier les performances en terme de classification pour les différentes stratégies employées dans le cadre de la modélisation à l'aide de vraisemblance, nous présentons sur la figure FIG. 5.11 l'évolution du taux de classification en fonction du taux de rejet pour une décision fondée sur la minimisation du risque pignistique. Sur cette figure, nous constatons que l'utilisa-

tion de la stratégie n°4 pour la détermination des coefficients de fiabilité permet d'atteindre les taux d'erreur les plus faibles. Toutefois, ces performances sont moins bonnes que dans le cas de la modélisation mono-dimensionnelle. Deux raisons peuvent être avancées pour expliquer ce phénomène. Dans un premier temps, l'estimation des densités de probabilité peut s'avérer être plus précise pour la modélisation mono-dimensionnelle que pour la modélisation multi-dimensionelle. La seconde justification repose sur l'utilisation du modèle. Si l'on regarde de façon plus précise, le modèle n°1 présenté à l'équation (3.8), on constate que, pour notre cas à deux hypothèses dans le cadre de discernement, l'une des fonctions de croyance est égale à l'élément neutre. Ainsi, la décision, dans ce cas, est prise uniquement sur une seule fonction de croyance ce qui tend à minimiser les performances.

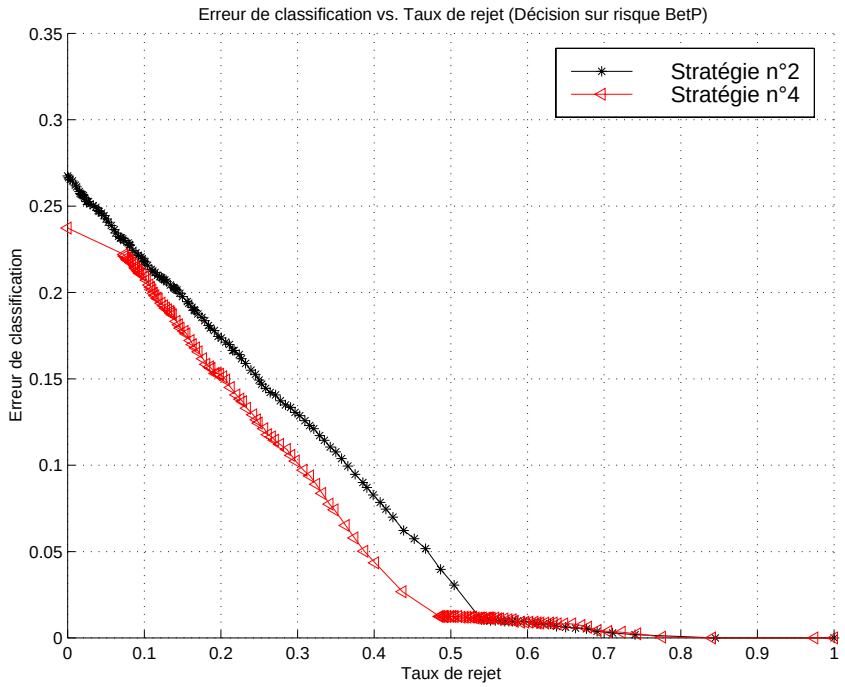

Fig. 5.11 : – Taux d'erreur vs. taux de rejet pour le risque pignistique pour la modélisation d'Appriou selon les deux stratégies étudiées.

Nous avons représenté sur la figure Fig. 5.12, l'évolution du taux d'erreur en fonction du taux de rejet obtenu avec la modélisation de Denœux pour une décision basée sur la minimisation du risque pignistique. Là encore, la stratégie n°4 permet d'atteindre les meilleures performances.

Enfin, nous avons représenté sur la figure Fig. 5.13 l'évolution du taux de non détection en fonction du taux de fausse alarme pour la stratégie n°4. Deux situations sont considérées : une première sans rejet et une seconde pour laquelle on s'accorde un taux de rejet de l'ordre de 20% pour des coûts $\{0, 1\}$. Pour la situation avec rejet, les taux de non détection et de fausse alarme sont calculés sur les vecteurs non rejetés. On remarque, sur la figure Fig. 5.13, que pour les deux

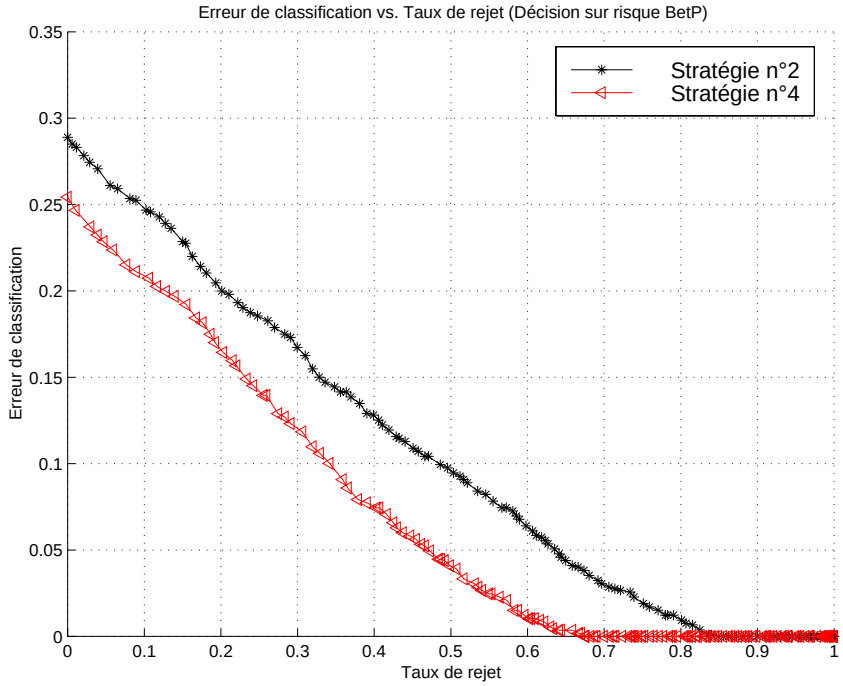

Fig. 5.12 : – Taux d'erreur vs. taux de rejet pour le risque pignistique pour la modélisation de Denœux selon les deux stratégies étudiées.

modélisations mises en œuvre (modélisation d'Appriou à gauche et modélisation de Denœux à droite) les performances s'améliorent dans le cas de rejet. De plus, le rejet agit aussi bien sur la fausse alarme que sur la non détection. Enfin, on constate que les taux de bonne classification ne se font pas au détriment du taux de non détection, ce qui est essentiel dans le cadre de cette application.

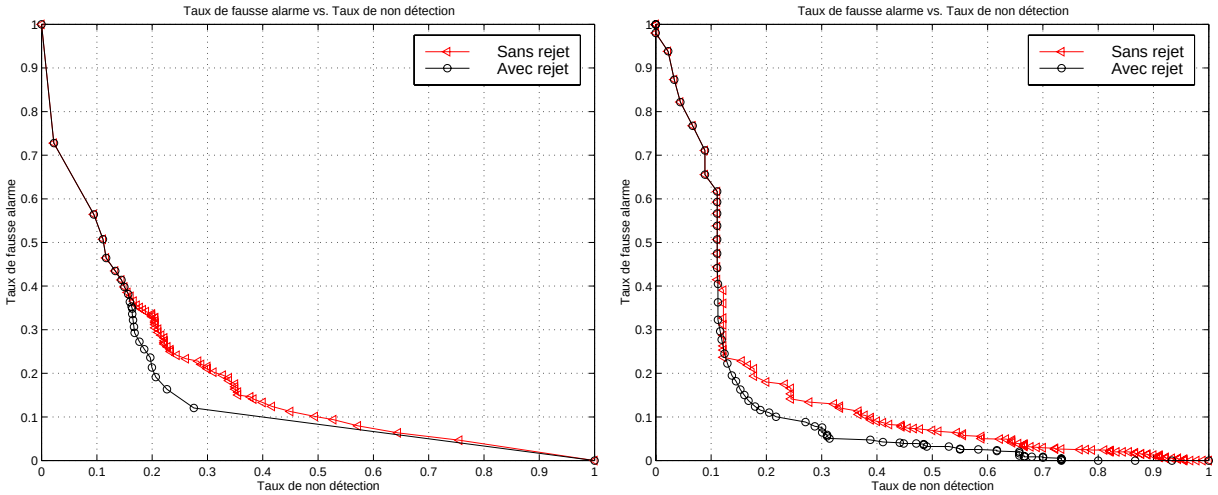

Fig. 5.13 : – Taux de non détection vs. taux de fausse alarme pour le risque pignistique pour la modélisation d'Appriou (à gauche) et la modélisation de Denœux (à droite).

Ainsi, nous avons présenté l'apport de l'adaptation des coefficients de fiabilité, dans le cadre

de la discrimination des mélanomes malins. Ces coefficients permettent de prendre en compte la fiabilité des sources, d'où sont issues les fonctions de croyance, et ainsi de remédier à l'une des origines possibles du conflit.

5.3.3 Etude de la répartition de la masse conflictuelle

D'autres causes peuvent être à l'origine de conflit (Cf. chapitre 4). Afin de gérer au mieux l'ensemble des origines possibles du conflit, nous pouvons opter pour la redistribution de la masse conflictuelle à l'aide de poids. Pour mettre en évidence ce conflit, sur la figure FIG. 5.14, nous avons représenté la localisation ainsi que l'amplitude du conflit obtenu dans le cas d'une modélisation fondée sur les distances avec deux prototypes par classe (représentés par des astérisques sur la figure) et avec des coefficients de fiabilité fixés à 0.99.

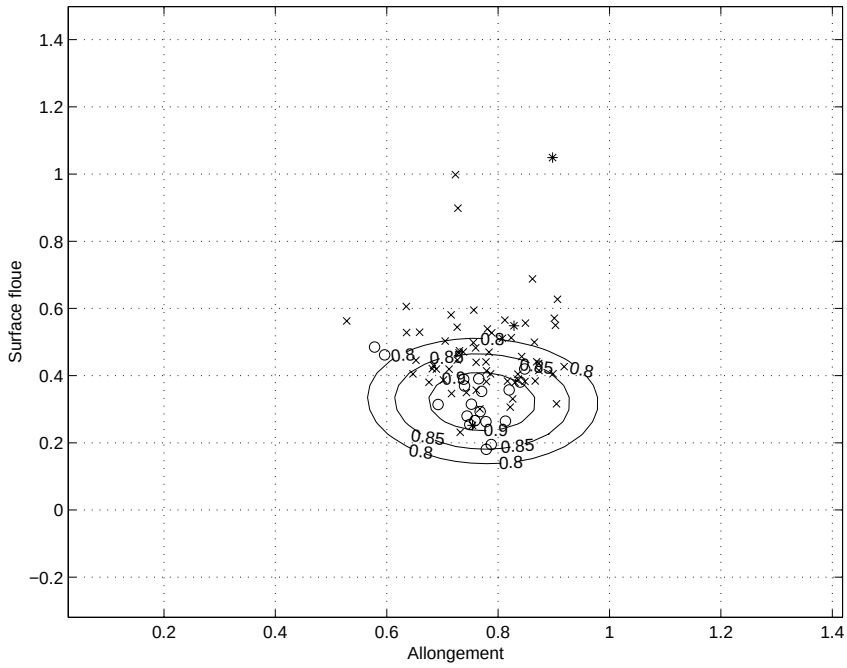

FIG. 5.14 : – Contour du conflit dans l'espace des caractéristiques (Naevus = × ; Mélanome = ○).

Dans cette section, nous étudions l'apport de la redistribution de la masse conflictuelle à l'aide de poids. Cette étude se fera en deux temps. Nous présentons, dans un premier temps, le comportement de cette technique dans le cadre d'une base d'apprentissage d'étiquettes précises. Nous abordons ensuite une approche où la base d'apprentissage est constituée d'étiquettes imprécises.

Pour la première étude, nous utilisons la modélisation des fonctions de croyance à partir de la méthode des distances associées aux prototypes [37]. Pour la seconde approche, nous utilisons une

modélisation reposant sur une approche des k plus proches voisins mais adaptée à un étiquetage imprécis [39]. Pour cette modélisation, on suppose que les k-ppv de x dans l'ensemble d'apprentissage $\mathcal{L}$ sont autant de sources d'informations indépendantes précisant le degré de croyance de x. Chacune des k fonctions de croyance m_i est construite par affaiblissement en fonction de la distance d_i entre x et $x^{(i)}$. Elle est définie $\forall A \in 2^\Theta$ par :

$$m_i(A) = \alpha\phi(d_i) \tag{5.2}$$

$$m_i(\Theta) = 1 - \alpha\phi(d_i). \tag{5.3}$$

Dans ces équations, α est un paramètre et $\phi(.)$ est une fonction décroissante monotone qui vérifie $\phi(0) = 1$ et $\lim_{d\to\infty} \phi(d) = 0$. Elle peut s'exprimer par :

$$\phi(d_i) = \exp^{-\gamma(d_i)^2} \tag{5.4}$$

avec γ un paramètre positif.

Etiquettes précises

Comme nous l'avons vu dans le cadre du chapitre 4, les poids permettant la redistribution de la masse conflictuelle, peuvent être déterminés, soit par un expert, soit de façon automatique en minimisant l'erreur quadratique moyenne (équation 3.3) afin d'adapter la redistribution selon l'objectif visé (exemple : maximisation du taux de reconnaissance, minimisation du taux de non détection,...).

La modélisation des fonctions de croyance est obtenue à partir de la méthode des distances associées aux prototypes [37]. Le nombre de prototype p est optimisé par validation croisée sur la base d'apprentissage pour chaque tirage. Le critère de décision employé est le maximum de probabilité pignistique. Comme nous l'avons fait dans le chapitre 4, nous pouvons supposer que plusieurs experts ayant des objectifs différents interviennent afin de définir les valeurs des poids pour la redistribution de la masse conflictuelle. Un premier expert (Expert n°1) peut, par exemple, vouloir minimiser l'erreur de classification. Un autre expert (Expert n°2) peut souhaiter une approche prudente dans son diagnostic et ainsi obtenir le minimum de non détection en s'autorisant un taux de fausse alarme convenable (vrai nævus classé mélanome). Enfin, on peut aussi supposer que toutes les pathologies cutanées n'ont pas été apprises dans le cadre de l'apprentissage et que la présence de conflit révèle la présence d'un autre type de lésion et ainsi rejeter la prise de décision en conservant le conflit sur l'ensemble vide $w(\emptyset, m_1, \ldots, m_p) = 1$. Les différents résultats sont présentés sur le tableau Tab. 5.1.

	Non détection	**Fausse alarme**	**Bonne classification**	**Rejet**
Expert n°1 : poids optimisés $w(\{H_1\}, m_1, \ldots, m_p) = 0.523$ $w(\{H_2\}, m_1, \ldots, m_p) = 0.477$	0.447	0.238	**0.78**	0
Expert n°2 : poids optimisés $w(\{H_1\}, m_1, \ldots, m_p) = 0.456$ $w(\{H_2\}, m_1, \ldots, m_p) = 0.544$	**0**	0.80	0.268	0
Expert n°3 : poids fixés $w(\emptyset, m_1, \ldots, m_p) = 1$	0.123	0.363	0.663	**0.820**

Tab. 5.1 : – Résultats de bonne classification selon la combinaison employée sur la base de test des lésions cutanées (H_1 : nævus ; H_2 : mélanome).

On constate sur ce tableau que le choix de l'expert n°1 permet d'obtenir un taux de classification de 0.780, mais en atteignant un taux de non détection de l'ordre de 0.45, ce qui engendre des conséquences relativement graves dans le cas de cette application. En ce qui concerne l'expert n°2, sa stratégie l'amène à un taux d'erreur de l'ordre de 0.73, mais il ne commet alors aucune non détection, c'est-à-dire que toutes les lésions malignes ont été diagnostiquées en tant que telles. Enfin, la stratégie définie par l'expert n°3 permet d'atteindre un taux de bonne classification de l'ordre de 0.66 en rejetant 82% des lésions. Cependant, cette dernière stratégie permet de n'avoir que 12% de non détection.

Etiquettes imprécises

En ce qui concerne la base avec étiquetage imprécis, nous avons ajusté les paramètres de la méthode de modélisation employée [39] de la manière suivante. Le nombre de voisins k est optimisé par validation croisée pour chaque tirage de la base d'apprentissage. Le coefficient de fiabilité α est fixé à 0.99 et le paramètre γ est pris comme préconisé dans [35] égal à la moyenne des distances aux k voisins dans l'ensemble d'apprentissage. Pour l'opérateur pondéré, les poids sont obtenus par minimisation de l'erreur quadratique moyenne (équation (3.3)).

Nous nous intéressons, dans un premier temps, au cas où les étiquettes de la base d'apprentissage sont données sous forme de fonctions de croyance. On considère dans ce cas que l'on ne possède plus la vérité terrain qui est issue de l'analyse histologique mais uniquement de la synthèse, donnée sous forme d'une fonction de croyance, des opinions émanant de dermatologues à propos de la lésion étudiée. L'optimisation des poids par minimisation de l'erreur quadratique moyenne, nous donne alors la répartition suivante : $w(\{H_1\}, m_1, \ldots, m_k) = 0.515$ et $w(\{H_2\}, m_1, \ldots, m_k) = 0.485$. En terme de classification, on a représenté sur les figures FIG. 5.15 et FIG. 5.16, le taux d'erreur en fonction du taux de rejet respectivement pour les deux règles

de décision qui correspondent à la minimisation du risque pignistique et du risque inférieur. On peut remarquer sur ces deux figures que la répartition de la masse conflictuelle que nous proposons amène à de meilleures performances que les opérateurs de Dempster et Yager et ce dans les deux principes de décisions étudiés. En ce qui concerne, la prise de décision sur la minimisation du risque pignistique l'opérateur de Dempster donne de moins bons résultats que l'opérateur de Yager. Cette remarque n'est plus vraie lorsque l'on s'intéresse à la prise de décision sur la minimisation du risque inférieur. Pour l'opérateur de Yager, la masse conflictuelle étant redistribuée sur Θ, les plausibilités sont alors proches de 1. Ceci ne permet pas de rejeter d'exemples ($\max_n Pl(\{H_n\}) < 1 - \lambda_0$) ce qui n'est pas le cas pour les deux autres opérateurs testés.

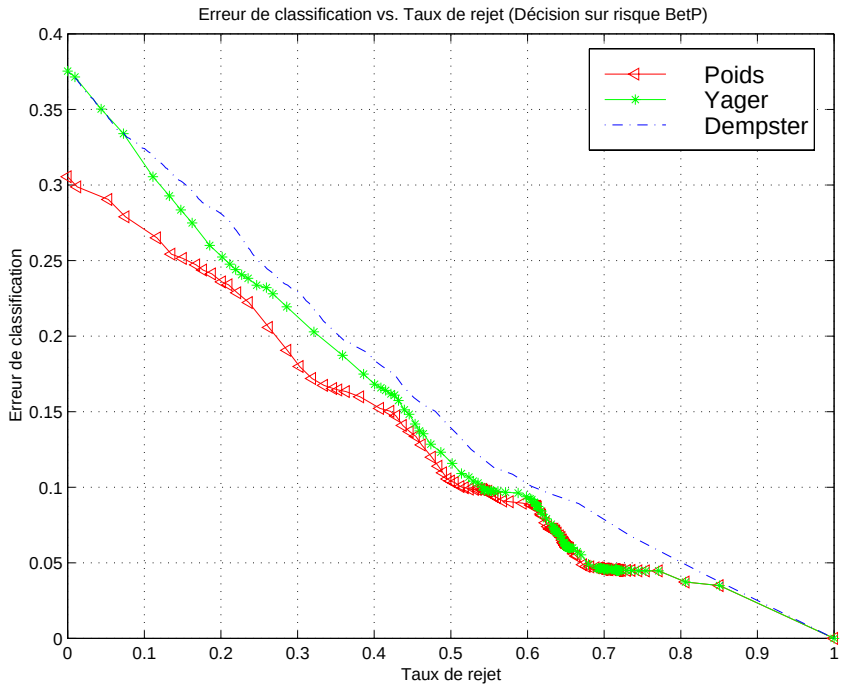

FIG. 5.15 : – Taux d'erreur vs. taux de rejet pour le risque pignistique, $\cdot-$ =Dempster, $*-$ =Yager, $\triangleleft$=Poids.

Enfin, nous avons étudié l'influence de la proportion de vecteur d'étiquette inconnue dans la base d'apprentissage sur le taux de classification (FIG. 5.17). Pour cela, nous partons de la base d'apprentissage étiquetée de manière précise et nous modifions cette base en donnant une étiquette inconnue ($m_{\mathcal{L}}^{(i)}(\Theta) = 1$) à des lésions tirées de manière aléatoire. La présence de lésions d'étiquette inconnue dans la base d'apprentissage peut survenir lorsque la fiche patient, qui contient les informations propres au patient, a été perdue ou lorsque l'histologie d'une lésion n'a pas été effectuée. On constate que la méthode de combinaison avec les poids permet d'obtenir un taux de reconnaissance supérieur à celui obtenu avec la combinaison de Dempster et cela quelque soit la proportion de vecteurs inconnus dans la base d'apprentissage. Enfin, on peut constater

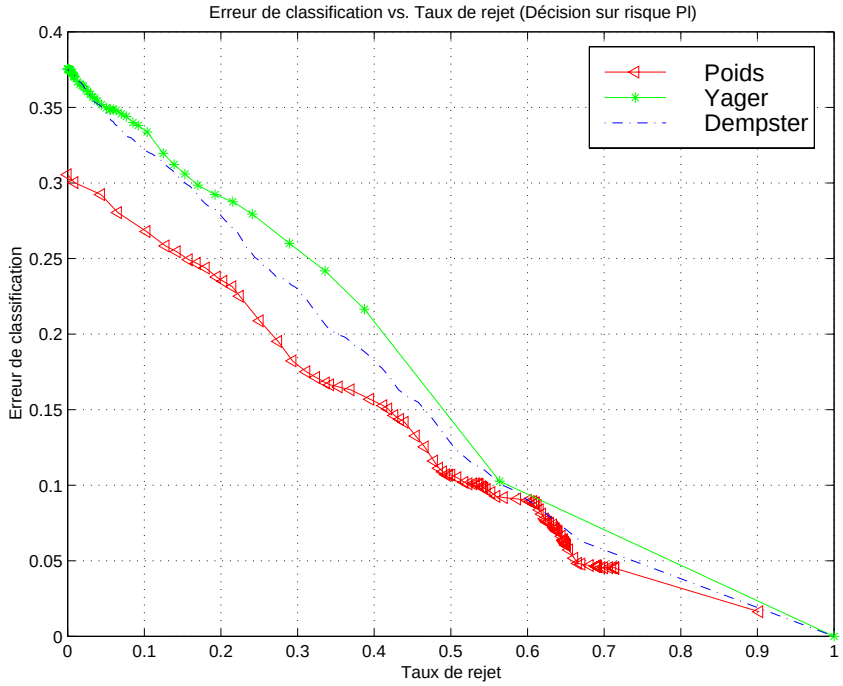

Fig. 5.16 : – Taux d'erreur vs. taux de rejet pour le risque inférieur, $\cdot-$ =Dempster, $*-$ =Yager, $\triangleleft$=Poids.

que, lorsque le nombre d'exemples étiquetés inconnus dans la base d'apprentissage est important les courbes tendent à se rejoindre. En effet, lorsque l'on introduit des vecteurs de classe inconnue le conflit diminue. Ainsi, les deux opérateurs tendent vers l'opérateur conjonctif et donnent des résultats similaires.

5.4 Conclusion

Dans ce chapitre, nous avons évoqué le problème du diagnostic du mélanome malin. Dans le cadre de l'"automatisation" de ce diagnostic, nous avons présenté les différentes procédures qui vont de l'acquisition non contrôlée des images de lésions cutanées à la discrimination des mélanomes, en passant par des étapes de segmentation et d'extraction de caractéristiques. Rappelons ici que le travail effectué dans le cadre de cette thèse se situe uniquement au niveau de la dernière étape c'est-à-dire au niveau de la discrimination des lésions.

Nous avons mis en œuvre les différentes contributions de cette thèse, présentées dans les chapitres 3 et 4, au sein de cette application de diagnostic médical.

Cette application nous a permis, dans un premiers temps, de mettre en évidence l'intérêt de l'affaiblissement des fonctions de croyance à l'aide de coefficients de fiabilité optimisés par minimisation de l'erreur quadratique. Cette optimisation des coefficients de fiabilité permet d'obtenir de meilleures performances en terme de classification par rapport à des stratégies à coefficients

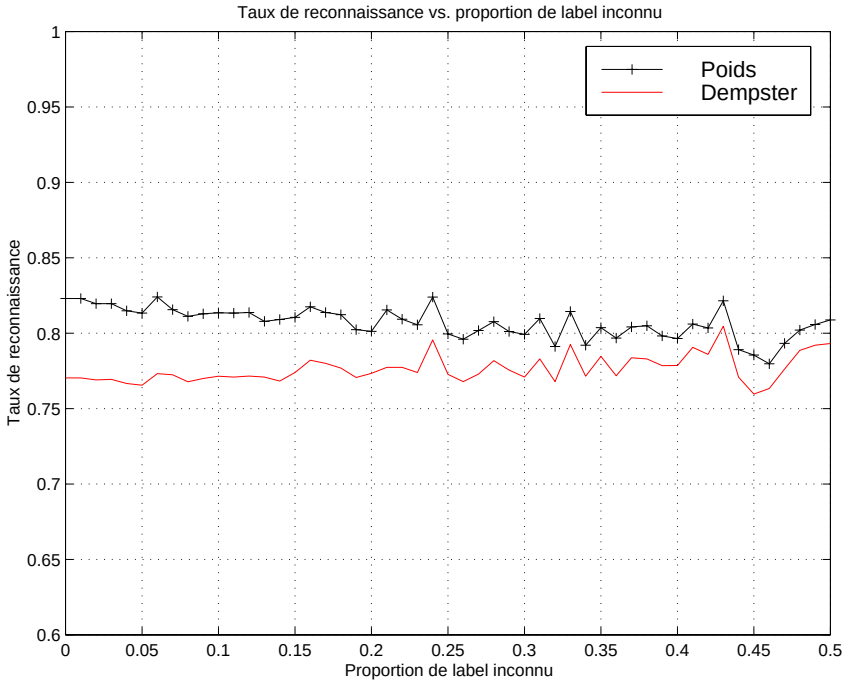

Fig. 5.17 : – Evolution du taux de reconnaissance en fonction du taux de vecteurs de label inconnu dans la base d'apprentissage.

fixes ou à coefficients calculés à partir de critères d'information. En effet, cette dernière méthode ne permet pas de déterminer les coefficients de fiabilité afin d'obtenir le taux de classification optimal, mais simplement de quantifier la dissemblance ou la ressemblance entre les données issues de différentes bases (par exemple : base d'apprentissage et base de validation). Ainsi, cette stratégie peut être utilisée dans le cas d'une évolution de contexte (différence de couleur de peau entre les individus constituant les deux bases, ...). Après analyse des performances en termes de classification, il apparaît judicieux d'employer, lors d'un apprentissage précis pour cette application donnée, la modélisation mono-dimensionnelle proposée par Appriou en optimisant les coefficients de confiance par minimisation de l'erreur quadratique moyenne. Par ailleurs, la modélisation mono-dimensionnelle correspond bien à la démarche employée par les dermatologues avec l'utilisation de la règle ABCDE.

D'autre part, une étude sur la redistribution de la masse conflictuelle à l'aide de poids adaptés a montré l'efficacité et l'intérêt de cette approche dans le cadre de cette application. En effet, la détermination de poids par un expert permet de répondre à ses propres objectifs (par exemple : minimisation du taux de non détection, minimisation du taux d'erreur, ...). De la même manière, pour un apprentissage étiqueté de manière imprécise, la version optimisée de la redistribution de la masse conflictuelle à l'aide de poids permet d'obtenir des taux d'erreur moins importants que ceux observés dans le cas de la combinaison de Demspter, Yager ou Dubois. Enfin, une étude

sur l'introduction dans l'ensemble d'apprentissage de vecteurs d'étiquette inconnue montre la robustesse de cette approche. Ainsi, dans le cas d'un apprentissage imprécis, il paraît judicieux d'employer la modélisation développée par Denœux permettant de gérer un apprentissage imprécis. En renfort de cette modélisation, la répartition de la masse conflictuelle à l'aide de poids semble relativement bien adaptée au problème étudié puisqu'elle permet au praticien de répartir la masse conflictuelle selon la stratégie qu'il désire.

Ainsi, les différentes contributions de cette thèse permettent d'obtenir des résultats, en terme de discrimination, similaires à ceux observables chez un dermatologue. En effet, il faut rappeler que le taux de reconnaissance des mélanomes malins est de l'ordre de 65% même dans le cas d'un dermatologue expérimenté. Ainsi, les différentes approches proposées permettent d'atteindre de tels résultats. Notons aussi, que l'ensemble des lésions constituant les bases d'apprentissage et de test peuvent être considérées comme des lésions relativement délicates à diagnostiquer. En effet, si ces lésions constituent les différentes bases c'est parce qu'elles ont subi une exérèse ; le dermatologue ayant diagnostiqué une suspicion de mélanomes. Ainsi, l'utilisation la chaîne de reconnaissance des mélanomes permettrait déjà de réduire le nombre de fausses alarmes (ablation de lésions non cancéreuses), ce qui diminuerait l'impact psychologique de l'intervention chirurgicale et le surcoût financier lié à ces exérèses. Toutefois, les résultats sur cette application sont perfectibles. Il semble souhaitable de revenir sur deux points importants, qui n'ont pas été pris en compte dans le cadre de cette thèse puisque situés en amont dans la chaîne de traitement des lésions cutanées.

Dans un premiers temps, il semble qu'un effort devrait être porté sur le contrôle de l'acquisition des images. Il serait souhaitable de développer un système d'acquisition dont l'éclairage serait contrôlé. Un tel dispositif doit pouvoir être obtenu, sans doute en partenariat avec des équipes compétentes et disposant d'un savoir-faire dans ce domaine.

Ensuite, une étude sur l'introduction de nouvelles caractéristiques pertinentes apparaît judicieuse. Au travers des réunions avec les dermatologues, il semble qu'un certain nombre d'informations essentielles à la discrimination des mélanomes n'ont pas été prises en compte. En effet, des informations telles que les antécédents de mélanomes, la localisation de la lésion, les modifications récentes de la lésion, le sexe du patient sont autant d'informations que la procédure de diagnostic automatique ne prend pas actuellement en compte.

Conclusion et perspectives

Le diagnostic peut-être abordé par différentes approches. Vis-à-vis des contraintes propres à l'application médicale, l'approche du diagnostic que nous proposons dans ce manuscrit est fondée sur la reconnaissance de formes, afin d'éviter la représentation du système à diagnostiquer à l'aide de modèles. Nous avons plus particulièrement adopté les méthodes de reconnaissance de formes qui utilisent la théorie de l'évidence. Cette théorie permet de représenter de manière naturelle l'incertitude ainsi que les imprécisions.

Nous nous sommes attachés, dans un premier temps (Cf. Chapitre 3), à la modélisation des connaissances, dans le cadre de la théorie de l'évidence. Nous avons, plus particulièrement, introduit la notion de fiabilité des informations et la prise en compte de cette fiabilité. Ainsi, deux méthodes permettant la détermination de coefficients de fiabilité ont été introduites. La première méthode repose sur le calcul, obtenu à l'aide de critères d'information, de la dissemblance entre des informations de même nature. La seconde approche proposée permet, par minimisation de l'erreur quadratique moyenne, d'obtenir les coefficients de fiabilité qui optimisent le taux de reconnaissance. Ces deux approches ont été testées dans le cadre d'une évolution de contexte et dans le cadre de l'application médicale visée. Il apparaît, dans le cas d'une évolution de contexte, que les deux approches arrivent à des taux de reconnaissance similaires et meilleurs que ceux obtenus avec des coefficients de fiabilité fixés. Dans le cadre de l'application médicale, l'approche par minimisation de l'erreur quadratique permet d'améliorer les taux de reconnaissance et d'atteindre ainsi les résultats obtenus par les dermatologues. La méthode fondée sur les critères d'information n'obtient pas, elle, les résultats escomptés, ce qui s'explique par la ressemblance des données de la base d'apprentissage. Toutefois, cette approche s'avérerait utile dans le cas où le système de diagnostic doit s'adapter à des populations différentes.

Nous nous sommes, ensuite, intéressés à la combinaison des fonctions de croyance (Cf. Chapitre 4). Cette combinaison est nécessaire afin de synthétiser les connaissances lorsque plusieurs fonctions de croyance sont disponibles. Cependant, elle soulève le problème de la gestion du conflit. En effet, lorsque les sources d'information sont peu fiables ou lorsque la modélisation des données est peu précise, alors un conflit peut apparaître. Nous avons proposé un cadre générique pour la fusion de sources d'information modélisées dans le cadre de la théorie de l'évidence. A partir de ce cadre de travail, nous avons montré que les différents opérateurs classiques de la théorie des fonctions de croyance pouvaient être retrouvés. Ainsi, à l'aide de poids permettant la

distribution du conflit, il est possible de décliner les opérateurs usuels de cette théorie. Nous avons décrit deux méthodes d'obtention des poids. La première de ces méthodes est fondée sur l'intégration d'une connaissance supplémentaire que pourrait introduire un expert afin de résoudre le conflit. La seconde méthode repose sur l'obtention des poids par minimisation de l'erreur quadratique. Les différentes simulations réalisées ainsi que la mise en œuvre de cette technique dans le cadre de la détection des mélanomes malins ont montré qu'un choix judicieux de l'opérateur peut permettre d'obtenir de meilleures performances en terme de taux de reconnaissance quelque soit la stratégie choisie pour la prise de décision. La solution proposée est une approche élégante qui permet de retrouver les différentes solutions existantes à la gestion du conflit (les différents opérateurs et l'affaiblissement), et ainsi de créer une famille plus générique pour la gestion du conflit.

Les différents travaux que nous avons réalisés dans le cadre de cette thèse, semblent offrir de nombreuses perspectives aussi bien dans le domaine du diagnostic médical que dans le cadre du diagnostic d'autres systèmes.

Une nouvelle voie consiste à montrer que l'opérateur de redistribution de la masse conflictuelle est suffisamment générique pour englober des opérateurs tels que les t-normes, t-conormes, les opérateurs disjonctifs, ...

Enfin, les travaux entrepris dans cette thèse pourront être employés dans différentes applications. Ainsi, pour la segmentation d'images couleurs, après avoir réalisé une première segmentation dans l'espace des couleurs, on peut envisager une fusion des informations issues du voisinage d'un pixel de l'image. L'interprétation de la masse conflictuelle peut alors nous renseigner sur la position des frontières entre les différentes régions de l'image. Cet aspect est à l'heure actuelle à l'étude et à donner lieu à de premiers développements. Enfin, dans le cadre de la combinaison de classifieurs, les travaux sur l'obtention de coefficients de confiance permettraient de réduire l'influence, en terme de taux de reconnaissance, d'un classifieur peu fiable ou sujet à des performances moindres dans le cas d'une évolution de contexte.

Annexes

Annexe A

Obtention des coefficients d'affaiblissement à partir de critères d'information

De manière à qualifier la bonne représentativité de l'apprentissage, nous proposons d'utiliser une mesure d'information qui est calculée pour chaque couple (S_j, H_n). Le but de la démarche est de permettre d'évaluer la fiabilité de l'apprentissage de la source S_j relativement à l'hypothèse H_n. Pour satisfaire ce principe, on calcule la vraisemblance $L(H_n|x_j)$ à partir de la base d'apprentissage et une base de validation nous permet d'ajuster le coefficient α_{nj}. Le calcul du coefficient α_{nj} repose donc sur la ressemblance (ou la dissemblance) entre les lois de probabilité de chaque hypothèse issues de la base d'apprentissage et de la base de validation. Parmi les tests statistiques classiques paramétriques ou non paramétriques, tels que le test du Chi-deux (χ^2) ou le test de Kolmogorov permettant de définir l'adéquation ou la ressemblance statistique entre des lois de probabilités, nous avons arrêté notre choix sur une méthode non paramétrique. Notre démarche est la suivante.

Nous approcherons les lois de probabilité de chaque hypothèse par des histogrammes (Section A.1) construits à l'aide d'un critère d'information que nous noterons $IC(.)$ [26, 28], qui sera fonction du nombre de classes de l'histogramme (Section A.3). Ce dernier pourra être le critère d'Akaïke (AIC) [4], un critère synthétisant AIC et le critère de Rissanen (RIC) [103] ou le critère de Hannan et Quinn [65]. Ces histogrammes, optimaux au sens du maximum de vraisemblance, résument l'information contenue dans chaque source S_j et permettent d'obtenir une estimation optimale de la loi au sens du critère choisi et ce de manière non paramétrique. Nous utilisons ensuite une mesure de dissemblance (Section A.5) entre lois de probabilité (distance de Hellinger) [13] modifiée pour être applicable dans le cas des histogrammes [26] entre chacune des approximations de loi afin de définir le coefficient α_{nj}. Cette mesure entre lois de probabilité offre l'avantage, au contraire de mesures telles que la divergence de Kullback ou la distance de Bhattacharyya [13], de varier entre 0 et 1.

A.1 Approximation de loi de probabilité par des histogrammes

Disposant d'un échantillon $\epsilon_1 \dots \epsilon_T$ de taille T d'un processus aléatoire $\mathcal{E}$ de loi de probabilité λ inconnue, que l'on suppose continue par rapport à une loi ν *a priori* donnée, on souhaite effectuer une approximation de λ à l'aide d'un histogramme. Soit Ω l'ensemble des valeurs prises par $\mathcal{E}$. La densité de probabilité f de λ s'exprime à l'aide de la dérivée de Radon-Nicodym par :

$$\forall\, \epsilon \in \Omega \qquad f(\lambda, \epsilon) \triangleq \frac{d\lambda}{d\nu}(\epsilon). \tag{A.1}$$

La densité f sera approchée à partir du seul échantillon de taille T de $\mathcal{E}$ et d'un histogramme à C classes construit à l'aide de cet échantillon. Il s'agit, dans un premier temps, de déterminer cet histogramme à C classes défini sur une partition $\mathcal{C}$ de Ω.

A.2 Estimateur du maximum de vraisemblance pour une partition $\mathcal{C}$

Soit $\mathcal{C}$ une partition à C classes de Ω, soit $\epsilon_1 \dots \epsilon_T$ un T-échantillon d'observation et soit $\lambda_{\mathcal{C}}$ la restriction de λ à la partition $\mathcal{C}$. L'estimateur du maximum de vraisemblance $\hat{\lambda}_{\mathcal{C}}$ de $\lambda_{\mathcal{C}}$ est donné par l'équation suivante :

$$\forall\, p \in \{1, \dots,\ C\} \qquad \hat{\lambda}_{\mathcal{C}}(A_p) = \frac{\mid A_p \mid}{T} \tag{A.2}$$

où A_p est une classe de la partition $\mathcal{C}$ et où $\bigcup_{p\in\{1,\dots,C\}} A_p = \mathcal{C}$. Ce résultat provient de l'expression de la densité de $\hat{\lambda}_{\mathcal{C}}$:

$$\forall\, \epsilon \in \Omega \qquad f(\hat{\lambda}_{\mathcal{C}}, \epsilon) = \sum_{A\in\mathcal{C}} \frac{\hat{\lambda}_{\mathcal{C}}(A)}{\nu_{\mathcal{C}}(A)} 1_A(\epsilon) \tag{A.3}$$

avec $1_A(\epsilon) = 1$ si $\epsilon \in A$ et 0 sinon.

A.3 Sélection du nombre de classes d'un histogramme approchant une loi inconnue

L'obtention d'un histogramme optimal est fondé sur l'utilisation d'un critère d'information noté IC. Celui-ci est établi à partir d'une fonction coût de type contraste de Kullback ou distance de Hellinger [26]. En effet, on définit le coût de prendre $\hat{\lambda}_{\mathcal{C}}$ quand λ est la vraie densité de probabilité par :

$$W(\lambda, \hat{\lambda}_{\mathcal{C}}) \triangleq E_\lambda \left(\psi \left[\frac{f(\hat{\lambda}_{\mathcal{C}}, \epsilon)}{f(\lambda, \epsilon)} \right] \right) \tag{A.4}$$

et le risque moyen par :

$$\overline{W}(\lambda, \hat{\lambda}_{\mathcal{C}}) \triangleq E_\lambda \left(W(\lambda, \hat{\lambda}_{\mathcal{C}}) \right) \tag{A.5}$$

où E_λ est l'espérance mathématique par rapport à λ et ψ une fonction convexe. Selon l'expression de ψ, la fonction risque conduit à divers critères d'information pour établir l'histogramme à C classes, c'est-à-dire pour choisir $\hat{\lambda}_\mathcal{C}$ minimisant le risque. Ainsi, si ψ est de type distance de Hellinger [54], nous obtenons relativement à la partition $\mathcal{C}$ le critère défini par :

$$AIC(C) = \frac{2C-1}{T} - 2\sum_{B_i \in \mathcal{C}} \hat{\lambda}_\mathcal{C}(B_i) \ln \frac{\hat{\lambda}_\mathcal{C}(B_i)}{\nu_\mathcal{C}(B_i)}. \tag{A.6}$$

On peut voir que cette expression est similaire à la formulation classique du critère d'Akaïke [4], mais adapté ici au contexte de l'histogramme. Si la fonction de coût $W(\lambda, \hat{\lambda})$ est exprimée à l'aide du contraste de KullBack [54], cela conduit à deux nouveaux critères ϕ^* et AIC^* respectivement définis par :

$$\phi^*(C) = \frac{C(1+\ln(\ln T))}{T} - 2\sum_{B_i \in \mathcal{C}} \hat{\lambda}_\mathcal{C}(B_i) \ln \frac{\hat{\lambda}_\mathcal{C}(B_i)}{\nu_\mathcal{C}(B_i)} \tag{A.7}$$

et :

$$AIC^*(C) = \frac{C(1+\ln T)}{T} - 2\sum_{B_i \in \mathcal{C}} \hat{\lambda}_\mathcal{C}(B_i) \ln \frac{\hat{\lambda}_\mathcal{C}(B_i)}{\nu_\mathcal{C}(B_i)}. \tag{A.8}$$

On reconnaît dans l'équation (A.7) une formulation semblable à celle du critère d'Hannan et Quinn et dans l'équation (A.8) une formulation synthétisant les critère d'Akaïke et Rissanen. L'ensemble de ces trois critères peut être synthétisé sous la forme suivante :

$$IC(C) \triangleq g(C) - 2\sum_{B_i \in \mathcal{C}} \hat{\lambda}_\mathcal{C}(B_i) \ln(\frac{\hat{\lambda}_\mathcal{C}(B_i)}{\nu_\mathcal{C}(B_i)}) \tag{A.9}$$

où $g(C)$ est un terme de pénalité, qui diffère selon le critère d'information choisi, $\nu_\mathcal{C}$ une loi *a priori* qui sera la loi uniforme pour traduire l'absence de connaissance *a priori* sur la structure et C le nombre de classes de la partition $\mathcal{C}$. Ces critères[1] peuvent être utilisés pour sélectionner l'histogramme à C classes approchant la loi inconnue du T-échantillon $\epsilon_1 \ldots \epsilon_T$. Des démonstrations détaillées sont disponibles dans [26].

A.4 Construction de l'histogramme optimal

Initialement, un histogramme à $C_{init} =\mid \mathcal{C} \mid = 2 \times [\sqrt{T} - 1]$ classes de même pas (même largeur) est construit sur la partition $\mathcal{C}$, où $[\cdot]$ est la partie entière. Le choix du nombre initial de classes est conforme à ce qui est préconisé dans [107] et utilisé dans [26]. Ensuite, une partition à $(C_{init} - 1)$ classes est considérée. Pour chaque fusion parmi les $(C-1)$ fusions possibles de deux classes adjacentes de l'histogramme à C classes, le critère $IC(C-1)$ est calculé. Le choix de la meilleure fusion est guidé par la minimisation de la quantité $IC(C-1)$. Ceci fait, la recherche

[1] Pour des raisons de convergence, il est préférable d'utiliser soit le critère AIC^* soit le critère ϕ^*.

de la meilleure partition à $(C-2)$ classes est menée selon le même principe que précédemment. Finalement, l'histogramme à C classes tel que $IC(C)$ est minimum pour $C \in \{1, \ldots, C_{init}\}$ est retenu. On note $\mathcal{C}_{opt}$ cette sous-partition optimale à C_{opt} classes définissant la meilleure estimation $\hat{\lambda}_{C_{opt}}$ de la loi inconnue λ. La figure FIG. A.1 montre un histogramme initial construit à l'aide d'un échantillon de taille $T = 90$ généré aléatoirement selon une loi gaussienne. L'histogramme final obtenu respectivement par l'utilisation des critères AIC, AIC^* et ϕ^* est donné en figure FIG. A.3. La figure FIG. A.2 donne le comportement des trois critères en fonction du nombre de classes. On constate que les critères AIC^* et ϕ^* donnent le même histogramme final. Le critère AIC donne quant à lui un histogramme final disposant d'un nombre de classes plus élevé. Cette différence tient au type de convergence des différents critères d'information [26]. On notera que le critère AIC tend à induire une sur-paramétrisation, inconvénient déjà montré dans la littérature sur des problèmes de sélection de modèles par ce type de critère.

A.5 Calcul du coefficient de confiance

La dissemblance entre deux estimations de lois peut être calculée en introduisant la distance de Hellinger [26, 13]. La partition optimale des données de la source S_j sous l'hypothèse H_n, $\mathcal{C}^j_{opt}/H_n$ à C^j_{opt}/H_n classes est tout d'abord construite à partir de la base d'apprentissage et de la base de validation. Une fois obtenue, on en déduit $\hat{\lambda}^a_{C^j_{opt}/H_n}$ l'estimation de la loi sur $\mathcal{C}^j_{opt}/H_n$ de la base d'apprentissage et $\hat{\lambda}^v_{C^j_{opt}/H_n}$ l'estimation de la loi sur $\mathcal{C}^j_{opt}/H_n$ de la base de validation. La distance de Hellinger entre ces deux estimations de lois est donnée par [26] :

$$Hell(\hat{\lambda}^a_{C^j_{opt}/H_n}, \hat{\lambda}^v_{C^j_{opt}/H_n}) \triangleq 1 - \sum_{i=1}^{C^j_{opt}/H_n} \sqrt{\hat{\lambda}^a_{C^j_{opt}/H_n}(B_i) \times \hat{\lambda}^v_{C^j_{opt}/H_n}(B_i)}. \tag{A.10}$$

Lorsque les estimations de lois sont proches, alors la distance tend vers 0. Au contraire, si les lois sont fortement dissemblables, alors la distance est proche de 1. Le sens de variation de cette distance est l'inverse de celui du coefficient de confiance α_{nj} que nous avons défini lors de la présentation des approches séparables à la section 3.1.2. Nous prendrons alors le coefficient α_{nj} égal à :

$$\alpha_{nj} = 1 - Hell(\hat{\lambda}^a_{C^j_{opt}/H_n}, \hat{\lambda}^v_{C^j_{opt}/H_n}). \tag{A.11}$$

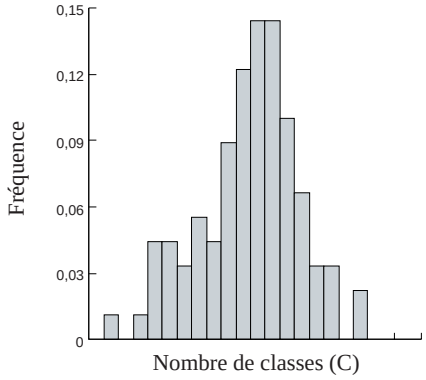

Fig. A.1 : – Histogramme initial.

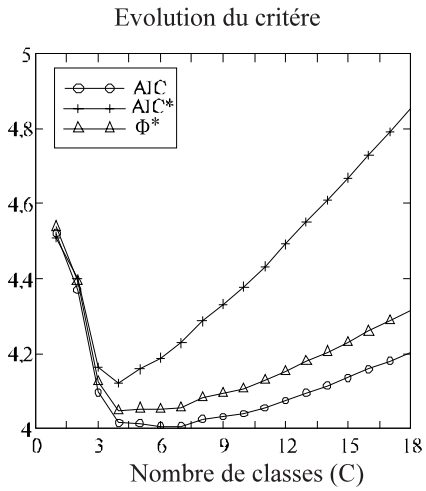

Fig. A.2 : – Evolution des critères d'information en fonction du nombre de classes.

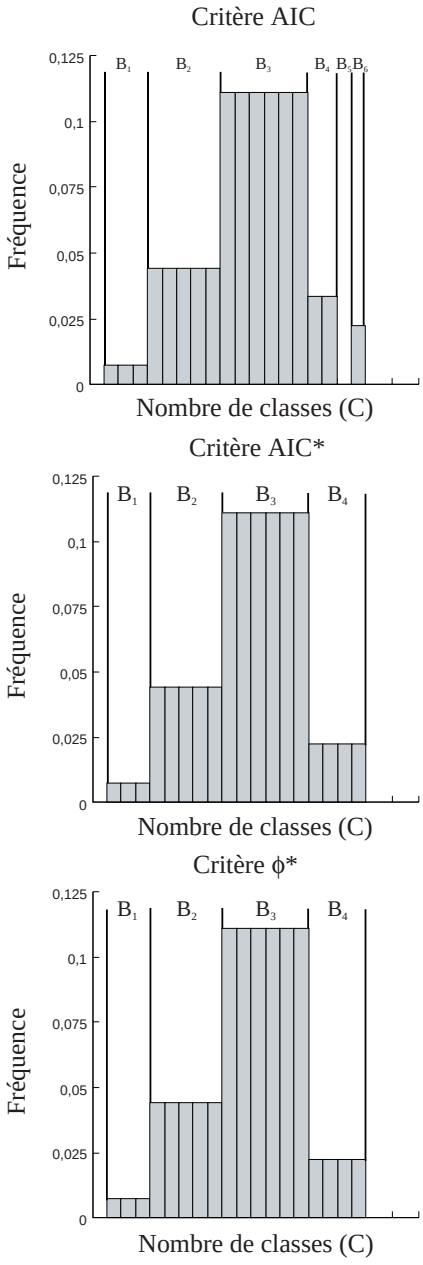

Fig. A.3 : – Histogrammes optimaux selon le critère d'information.

Annexe B

Affaiblissement vs. combinaison

Lors de la combinaison d'information, les sources impliquées dans cette combinaison ne soutenant pas obligatoirement des informations concordantes, un conflit peut se produire. Cette annexe détaille le calcul permettant de mettre en évidence le lien existant entre une gestion de conflit basée sur un affaiblissement (source non fiable) et une gestion basée sur une redistribution du conflit à l'aide de poids accordés à chaque proposition.

Soit une fonction de croyance m_j obtenue à partir d'une source d'information S_j. La fonction de communalité Q_j associée à la fonction m_j est définie de la manière suivante :

$$Q_j(A) = \sum_{A \subseteq B} m_j(B) \qquad \forall A \subseteq \Theta. \tag{B.1}$$

De plus, la transformée de Möbius inverse permet, à partir de la fonction de communalité Q_j, de retrouver la distribution de masse par la relation suivante :

$$m_j(A) = \sum_{A \subseteq B} (-1)^{|B-A|} Q_j(B) \qquad \forall A \subseteq \Theta. \tag{B.2}$$

B.1 Résultats de la combinaison de jeux de masses affaiblies

B.1.1 Expression de la fonction de communalité issue de jeux de masses affaiblies

Soit un ensemble de fonctions de croyance $\{m_1, \ldots, m_j, \ldots, m_J\}$. On note $m_{\alpha_j,j}$ la fonction de croyance m_j affaiblie par un coefficient α_j. Alors la fonction $m_{\alpha_j,j}$ peut s'écrire de la façon suivante :

$$\begin{cases} m_{\alpha_j,j}(A) &= \alpha_j m_j(A) \qquad \forall A \subset \Theta \\ m_{\alpha_j,j}(\Theta) &= 1 - \alpha_j + \alpha_j m_j(\Theta). \end{cases} \tag{B.3}$$

La fonction de communalité $Q_{\alpha_j,j}$, associée à $m_{\alpha_j,j}$, peut s'écrire :

$$\begin{array}{lll} \forall A \subseteq \Theta \qquad Q_{\alpha_j,j}(A) & = & \sum\limits_{A \subseteq B} m_{\alpha_j,j}(B) \\ & = & \sum\limits_{\substack{A \subseteq B \\ B \neq \Theta}} (\alpha_j m_j(B)) + 1 - \alpha_j + \alpha_j m_j(\Theta) \\ & = & \alpha_j \sum\limits_{A \subseteq B} m_j(B) + 1 - \alpha_j \\ & = & \alpha_j Q_j(A) + 1 - \alpha_j \\ Q_{\alpha_j,j}(A) & = & \alpha_j (Q_j(A) - 1) + 1 \end{array} \tag{B.4}$$

B.1.2 Fonction de croyance résultant de la combinaison

On peut alors écrire la combinaison des J sources d'information à l'aide des fonctions de communalité. Le résultat de cette fusion est noté Q_α et peut s'exprimer de la façon suivante :

$$\begin{array}{rcll} Q_\alpha(A) & = & K_\alpha \times Q_{\alpha_1,1}(A) \times \ldots \times Q_{\alpha_j,j}(A) \times \ldots Q_{\alpha_J,J}(A) & \forall A \subseteq \Theta \\ & = & K_\alpha \times \prod\limits_{j=1}^{J} Q_{\alpha_j,j}(A) & \forall A \subseteq \Theta \end{array} \tag{B.5}$$

où K_α représente le coefficient de normalisation de la combinaison. Ce coefficient s'exprime de la manière suivante (Cf. [109] page 42) :

$$K_\alpha = \frac{1}{- \sum\limits_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} (-1)^{|B|} Q_\alpha(B)}. \tag{B.6}$$

La masse résultante de la combinaison de Dempster (combinaison normalisée) peut alors s'écrire :

$$\begin{array}{lll} \forall A \subseteq \Theta \qquad m_\alpha(A) & = & \frac{1}{- \sum\limits_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} (-1)^{|B|} Q_\alpha(B)} \sum\limits_{A \subseteq B} (-1)^{|B-A|} Q_\alpha(B) \\ & = & \frac{1}{- \sum\limits_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} (-1)^{|B|} \prod\limits_{j=1}^{J} Q_{\alpha_j,j}(B)} \sum\limits_{A \subseteq B} (-1)^{|B-A|} \prod\limits_{j=1}^{J} Q_{\alpha_j,j}(B) \\ m_\alpha(A) & = & \frac{1}{- \sum\limits_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} (-1)^{|B|} \prod\limits_{j=1}^{J} [\alpha_j (Q_j(B)-1)+1]} \sum\limits_{A \subseteq B} (-1)^{|B-A|} \prod\limits_{j=1}^{J} [\alpha_j (Q_j(B) - 1) + 1]. \end{array} \tag{B.7}$$

B.2 Expression de la fonction de croyance résultant de la combinaison proposée

Soit m_c la fonction de croyance résultante de la combinaison proposée des J fonctions de croyance m_j. Cette fonction peut alors s'écrire de la manière suivante :

$$m_c(A) = m_\cap(A) + w(A, m_1, \ldots, m_J) K \qquad \forall A \subseteq \Theta \tag{B.8}$$

où $m_\cap(.)$ correspond à la masse issue de la combinaison conjonctive et où $w(A, m_1, \ldots, m_J)$ est le poids associé au sous-ensemble A lors de la redistribution du conflit K. Cette équation (B.8)

peut aussi s'écrire à l'aide de la fonction de communalité. En effet, le résultat de la combinaison conjonctive peut s'écrire :

$$m_{\cap}(A) = \sum_{A \subseteq B} (-1)^{|B-A|} \prod_{j=1}^{J} Q_j(B) \qquad \forall A \subseteq \Theta. \tag{B.9}$$

La masse conflictuelle engendrée par cette combinaison conjonctive peut s'écrire :

$$K = 1 + \sum_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} (-1)^{|B|} \prod_{j=1}^{J} Q_j(B). \tag{B.10}$$

En égalant la masse issue de la combinaison de masses affaiblies et l'équation B.8, nous obtenons :

$$w(A, m_1, \ldots, m_J) = \frac{(m_\alpha(A) - m_{\cap}(A))}{K}. \tag{B.11}$$

Ainsi, à partir de l'équation précédente et des équations (B.7), (B.9) et (B.10), nous pouvons alors déterminer les valeurs de poids de redistribution de la masse conflictuelle à partir des coefficients d'affaiblissement α_j et des jeux de masse m_j :

$$\forall\, A \subseteq \Theta \quad w(A, m_1, \ldots, m_J) = \zeta \left[\frac{\sum\limits_{A \subseteq B} (-1)^{|B-A|} \prod\limits_{j=1}^{J} [\alpha_j (Q_j(B) - 1) + 1]}{- \sum\limits_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} (-1)^{|B|} \prod\limits_{j=1}^{J} [\alpha_j (Q_j(B) - 1) + 1]} - \sum_{A \subseteq B} (-1)^{|B-A|} \prod_{j=1}^{J} Q_j(B) \right] \tag{B.12}$$

avec :

$$\zeta = \frac{1}{1 + \sum\limits_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} (-1)^{|B|} \prod\limits_{j=1}^{J} Q_j(B)}. \tag{B.13}$$

Bibliographie

[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 and S. Moral. Building classification trees using the total uncertainty criterion. In *2nd International Symposium on Imprecise Probabilities and their Applications (ISIPTA'01)*, New York, 2001. (Version electronique disponible à : http ://ipper.rug.ac.be/ isipta01/proceedings/060.html).

[4] H. Akaike. Information theory and an extension of the maximum likelihhod principle. In B.N Petrov and F. Csaki, editors, *Proceedings of the 2nd International Symposium of Information Theory*, pages 267–281, Budapest, 1973.

[5] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, 11 :27–40, 1991.

[6] A. Appriou. Perspectives liées à la fusion de données. *Science et Défense*, mai 1990.

[7] 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.

[8] B. Bouchon-Meunier. *La logique floue et ses applications.* Addison-Wesley, 1995.

[9] S. C. Bagui. *Nearest neighbor classification rules for multiple observations.* PhD thesis, University of Alberta, Edmonton, AB, Canada, 1989.

[10] S. C. Bagui. Classification using first stage rank nearest neighbor for multiple classes. *Pattern Recognition Letters*, 14 :537–544, 1993.

[11] S. C. Bagui and N. R. Pal. A multistage generalization of the rank nearest neighbor classification rule. *Pattern Recognition Letters*, 16 :601–614, 1995.

[12] A. Barr and E. Feigenbaum. *The Handbook of Artificial Intelligence.* William Kaufmann, Los Altos, CA, 1981.

[13] M. Basseville. Distance measures for signal processing and pattern recognition. Rapport de recherche interne 899, INRIA, Rennes, 1988.

[14] A. Belaïd and Y. Belaïd. *Reconnaissances des formes : Méthodes et applications.* Inter-Editions, Paris, 1992.

[15] J. C. Bezdek. Fuzziness vs. probability - the n-th round. *IEEE Transactions on Fuzzy Systems*, 2(1) :1–42, 1994.

[16] I. Bloch. Incertitude, imprécision et additivité en fusion de données : point de vue historique. *Revue Traitement du Signal*, 13(4) :267–288, 1996.

[17] I. Bloch. Information combination operators for data fusion : A comparative review with classification. *IEEE Transactions on Systems, Man and Cybernetics*, 26(1) :52–67, 1996.

[18] I. Bloch. Some aspects of Dempster-Shafer evidence theory for classification of multimodality medical images taking partial volume effect into account. *Pattern recognition Letters*, 17 :905–919, 1996.

[19] I. Bloch and H. Maître. Fusion de données en traitement d'images : Modèles d'informations et décisions. *Revue Traitement du Signal*, 11 (Numéro Spécial : Fusion de données)(6) :435–446, 1994.

[20] H. Bracker. *Utilisation de la théorie de Demspter-Shafer pour la classification d'images satellitaires à l'aide de données multi-sources et multi-temporelles.* PhD thesis, Université de Rennes I, 1996.

[21] W. F. Caselton and W. Luo. Decision making with imprecise probabilities : Dempster-Shafer theory and application. *Water Resources Research*, 28(12) :3071–3083, 1992.

[22] L. Cholvy. Applying Theory of Evidence in multisensor data fusion : a logical interpretation. In *Thrid International Conference on Information Fusion (FUSION'2000)*, pages TuB4 17–24, July 2000.

[23] C. K. Chow. An optimum character recongnition system using decision functions. *R. E. Transactions on Electronic Computers*, pages 247–254, 1957.

[24] O. Colot, R. Devinoy, A. Sombo, and D. de Brucq. Une méthode de traitement et d'analyse d'image pour la classification dermatologique. In *Proceeding of the Mediterranean Conference on Electronics and Automatic control (MCEA'98)*, pages 54–57, Marrakech, Maroc, 1998.

[25] O. Colot, R. Devinoy, A. Sombo, and D. de Brucq. *A Colour Image Processing Method for Melanoma Detection*, pages 562–569. Lecture Notes in Computer Science, 1496. Springer-Verlag, October 1998.

[26] O. Colot, C. Olivier, P. Courtellemont, A. El-Matouat, and D. de Brucq. Information criteria and abrupt changes in probability laws. In M. Holt, C. Cowan, P. Grant, and

W. Sandham, editors, *Signal Processing VII : Theories and Applications*, pages 1855–1858. EUSIPCO'94, September 1994.

[27] B. V. Dasarathy. Nosing around the neighborhood : A new system structure and classification rule for recognition in partially exposed environments. *IEEE Transactions Pattern Analysis Machine Intelligence*, PAMI-2(1) :67–71, 1980.

[28] D. de Brucq. Characterization of the optimal number of classes for a histogram with Hellinger's distance. In J.L. Lacoume, A. Chehikian, N. Martin, and J. Malbos, editors, *Signal Processing IV : Theories and Applications*, pages 1117–1120. EUSIPCO'88, September, 1988.

[29] D. de Brucq, K. Taouil, and O. Colot. Segmentations d'images et extractions de contours pour l'analyse de lésions dermatologiques. In *XVème Colloque GRETSI*, volume 3, pages 1205–1208, Septembre 1995.

[30] M. Degroot. *Optimal Statistical Decisions.* McGraw-Hill, New York, 1970.

[31] C. Delannoy. Un algorithme rapide de recherche de plus proche voisins. *RAIRO Informatique*, 14(3) :208–218, 1980.

[32] A. Dempster. Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics*, AMS-38 :325–339, 1967.

[33] A. Dempster. A generalization of bayesian inference. *Journal of Royal Statistical Society, Serie B*, 30 :205–247, 1968.

[34] A. Dempster. A class of random convex polytopes. *Annals of mathematical Statistics*, 43 :250–272, 1972.

[35] T. Denoeux. A k-nearest neighbour classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5) :804–813, 1995.

[36] T. Denoeux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7) :1095–1107, 1997.

[37] T. Denoeux. A neural network classifier based on Demspter-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics Part A*, 30(2) :131–150, 2000.

[38] T. Denoeux and M. Skartein-Bjanger. Induction of decision trees from partially classified data. In *International Conference IEEE System, Man, and Cybernetics (SMC'2000)*, pages 2923–2928, October 2000.

[39] T. Denoeux and L. M. Zouhal. Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3) :47–62, 2001.

[40] S. Deveughele. *Etude d'une méthode de combinaison adaptative d'informations incertaines dans un cadre possibiliste.* PhD thesis, Université de Technologie de Compiègne, 1993.

[41] P. A. Devijver and J. Kittler. *Pattern recognition, a statitical approach.* Prentice-Hall, Englewood Cliffs, 1982.

[42] A. Dromigny-Badin. *Fusion d'images par la thèorie de l'évidence en vue d'applications médicales et industrielles.* PhD thesis, INSA de Lyon, 1998.

[43] D. Dubois, M. Grabisch, H. Prade, and P. Smets. Using the transferable belief model and a qualitative possibility theory approach on an illustrative example : The assessment of the value of candidate. *International Journal of Intelligent Systems (To appear)*, 2001.

[44] D. Dubois and H. Prade. On several representations of an uncertainty body of evidence. In M.M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, New-York, 1982.

[45] D. Dubois and H. Prade. A note on measures of specificity for fuzzy sets. *International Journal of General Systems*, 10 :279–283, 1985.

[46] D. Dubois and H. Prade. Unfair coins and necessity measures : towards a possibilistic interpretation of histograms. *Fuzzy Sets and Systems*, 10(1) :15–20, 1985.

[47] D. Dubois and H. Prade. On the unicity of Dempster rule of combination. *International Journal of Intelligent Systems*, 1 :133–142, 1986.

[48] D. Dubois and H. Prade. *The Principle of Minimum Specificity as a Basic for Evidential Reasoning*, volume 50 of *Lecture Notes in Computer Science : Uncertainty in Knowledge-Based Systems*, pages 75–84. Springer-Verlag, Berlin, B. Bouchon and R. R. Yager edition, 1986.

[49] D. Dubois and H. Prade. Propertie of measures of information in evidence and possibility theories. *Fuzzy Sets and Systems*, 24 :161–182, 1987.

[50] D. Dubois and H. Prade. *Possibility Theory : An Approach to Computerized Processing of Uncertainty.* Plenum Press, New-York, 1988.

[51] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.*, 4 :244–264, 1988.

[52] B. Dubuisson. *Diagnostic et Reconnaissances de Formes.* Hermes, 1990.

[53] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley & Sons, New-York, 1973.

[54] A. El-Matouat and C. Olivier. Sélection du nombre de classe d'un histogramme et contraste de Kullback. In *25ème Journées Internationales de Statistiques*, pages 193–196. ASU, 1993.

[55] Z. Elouedi, K. Mellouli, and P. Smets. Decision trees using belief function theory. In *8th international conference Information Processing and Management of Uncertainty in knowledge-based systems (IPMU'2000)*, volume 1, pages 141–148, 2000.

[56] Z. Elouedi, K. Mellouli, and P. Smets. Induction of belief decision trees : a conjunctive approach. In G. Govaert, J. Janssen, and N. Limnios, editors, *Applied Stochastic Models and Data Analysis (ASMDA'2001)*, pages 404–409, Juin 2001.

[57] Z. Elouedi, K. Mellouli, and P. Smets. The evaluation of sensors'reliability and their tuning for multisensor data fusion within the transferable belief model. In *6éme European Conference on symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'2001)*, Septembre 2001.

[58] S. Fabre, A. Appriou, and X. Briottet. Presentation and description of two classidication methods using data fusion based on sensor management. *Information Fusion*, 2 :49–71, 2001.

[59] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annual Eugenies*, 7(Part II) :179–188, 1936.

[60] R. A. Fisher. *Contribution to Mathematical Statistics.* John Wiley, New-York, 1950.

[61] J. Francois. *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] R. J. Friedman. Early detection of melanoma : the role of physician examination and self examination of the skin. *CA*, 35 :130–151, 1985.

[63] T. George and N. Pal. Quantification of conflict in the Dempster-Shafer framework. *International Journal of General Systems*, 24(4) :407–423, 1994.

[64] 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.

[65] E. J. Hannan. The estimation of the order of an ARMA process. *Annals of Statistics*, 8(5) :1071–1081, 1980.

[66] D. Harmanec and G. Klir. Measuring totale uncertainty in Dempter-Shafer theory : A novel approach. *International Journal of General Systems*, 22(4) :405–419, 1994.

[67] P. E. Hart. The condensed Nearest Neighbor Rule. *IEEE Transactions Information Theory*, 14, 1968.

[68] R. V. L. Hartley. Transmission of information. *The Bell Systems Technical Journal*, 7(3) :535–563, 1928.

[69] M. E. Hellman. The nearest neighbor classification rule with a reject option. *IEEE Transactions on System, Man and Cybernetics*, 6(3) :179–185, 1970.

[70] U. Höhle. Entropy with respect to plausibility measures. In *12th IEEE International Symposium Multiple-Valued Logic*, pages 167–169, 1982.

[71] U. Höhle. Fuzzy Plausibility Measures. In E.P. Klement, editor, *Proceedings of 3rd International Seminar on Fuzzy Set Theory*, pages 7–30, Linz, September 1981.

[72] F. Janez. *Fusion de sources d'information définies sur des référentiels non exhaustifs différents.* PhD thesis, Université d'Angers, 1996.

[73] A. Jousselme, D. Grenier, and E. Bossé. A new distance between two bodies of evidence. *Information Fusion*, 2 :91–101, 2001.

[74] R. Kennes and P. Smets. *Uncertainty in Artificial Intelligence*, chapter Computational aspect of the Möbius transformation, pages 401–416. Elsevier Science Publishers, 1991.

[75] H. Kim and 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.

[76] F. Klawonn and E. Schwecke. On the axiomatic justification of Dempster's rule combination. *International Journal of Intelligent Systems*, 7 :469–478, 1992.

[77] G. Klir and A. Ramer. Uncertainty in the Dempster-Shafer theory : a critical reexamination. *International Journal of General Systems*, 18(2) :155–166, 1990.

[78] G. Klir and M. J. Wierman. *Uncertainty-Based information : elements of generalized information theory.* Physica-Verlag, New-York, 1998.

[79] J. Kohlas and P. A. Monney. *A mathematical theory of hints. An approach to Dempster-Shafer theory evidence.* Springer-Verlag, 1995.

[80] G. J. M. Lachlan and T. Krishnan. *The EM algorithm and extensions.* Wiley, New-York, 1997.

[81] M. T. Lamata and S. Moral. Measures of entropy in the theory of evidence. *International Journal of General Systems*, 14(4) :297–305, 1988.

[82] E. Lefevre, O. Colot, P. Vannoorenberghe, and D. de Brucq. Contribution des mesures d'information à la modélisation crédibiliste des connaissances. *Revue Traitement du Signal*, 17(2) :87–97, 2000.

[83] E. Lefevre, O. Colot, P. Vannoorenberghe, and D. de Brucq. A generic framework for resolving the conflict in the combination of belief structures. In *Third International Conference on Information Fusion (FUSION'2000)*, pages MOD4 11–18, July 2000.

[84] E. Lefevre, P. Vannoorenberghe, and O. Colot. About the use of Dempter-Shafer theory for color image segmentation. In *First International Conference on Color in Graphics and Image Processing (CGIP'2000)*, pages 164–169, October 2000.

[85] W. B. Luo and B. Caselton. Using Dempster-Shafer theory to represent climate change. *Journal of Environnemental Management*, 49 :73–93, 1997.

[86] Y. Maeda and H. Ichihashi. An uncertainty measure with monotonicity under the random set inclusion. *International Journal of General Systems*, 21(4) :379–392, 1993.

[87] E. Mandler and 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.

[88] S. Mathevet, L. Trassoudaine, P. Checchin, and J. Alizon. Application de la théorie de l'évidence à la combination de segmentations en régions. In *Dix-Septième Colloque GRETSI (Vannes)*, pages 635–638, Septembre 1999.

[89] D. Michie, D.J. Spiegelhalter, and C.C. Taylor (Eds.). *Machine Learning, Neural and Statistical Classification.* Ellis Horwood Series in Artificial Inteligence. Ellis Horwood, Chichester, U.K., 1994.

[90] L. Miclet. *Méthodes structurelles pour la reconnaissances de formes.* Eyrolles, Paris, 1984.

[91] N. Milisavljevic. *Analyse et fusion, par théorie des fonctions de croyances, de données multi-sensorielles pour la détection de mines humanitaires.* PhD thesis, Ecole Nationale Supérieure des Télécommunications, 2001.

[92] N. Milisavljevic and I. Bloch. A two-level approach for modeling and fusion of humanitarian mine detection sensors within the belief function framework. In *Applied Stochastic Models and Data Analysis (ASMDA'2001)*, pages 743–748, Juin 2001.

[93] G. Ng and H. Singh. Data equalisation with evidence combination for pattern recognition. *Pattern Recognition Letters*, 19 :227–235, 1998.

[94] G. Ng and H. Singh. Democracy in pattern classifications : Combinations of votes from various pattern classifiers. *Artificial Intelligence in Engineering*, 12 :189–204, 1998.

[95] A. Nifle. *Modélisation comportementale en fusion de données. Application à la détection et l'identification d'objets ou de situations.* PhD thesis, Université Paris XI Orsay, 1998.

[96] A. Nifle and R. Reynaud. Un argument pour le choix entre décision pignistique et maximum de plausibilité en théorie de l'évidence. In *Seizième Colloque GRETSI*, pages 1411–1414, Grenoble, 1997.

[97] N. R. Pal and S. Ghosh. Some classification algorithms integrating Dempster-Shafer theory of evidence with the rank nearest neighbor rules. *IEEE Transactions on Systems, Man and Cybernetics*, 31(1) :59–66, 2001.

[98] S. Petit-Renaud and T. Denoeux. Regression analysis using fuzzy evidence theory. In *8th International Conference on Fuzzy Systems (FUZZ'IEEE 99)*, pages 1229–1234, August 1999.

[99] P. Quinio and T. Matsuyama. Random closed sets : A unified approach to the representation of imprecision and uncertainty. In P. Siegel (Eds.) R. Kruse, editor, *Symbolic and Quantitative Approaches to Uncertainty (ECSQARU)*, pages 282–286, Marseille, France, 1991. Spinger Verlag.

[100] R. Kennes. Computational aspects of the Möbius tarnformation of graphs. *IEEE Transactions on Systems, Man and Cybernetics*, 22 :201–223, 1992.

[101] A. Ramer. Uniqueness of information measure in the theory of evidence. *Fuzzy Sets and Systems*, 24(2) :183–196, 1987.

[102] A. Ramer and G. Klir. Measures of discord in the Dempster-Shafer theory. *Informations Sciences*, 67(1 and 2) :35–50, 1993.

[103] J. Rissanen, T.P. Speed, and B. Yu. Density estimation by stochastic complexity. *IEEE Transactions on Information Theory*, 58(2) :315–323, 1992.

[104] G. Rogova. Combining the results of several neural networks classifiers. *Neural Networks*, 7(5) :777–781, 1994.

[105] C. Royère, D. Gruyer, and V. Cherfaoui. Identification d'objets par la combinaison d'experts à l'aide de la théorie de l'évidence. In *Rencontre Francophone sur la Logique Floue et Ses Applications LFA'2000*, pages 237–244, Octobre 2000.

[106] S. Petit-Renaud. *Application de la théorie des croyances et des systèmes flous à l'estimation*

fonctionnelle en présence d'informations incertaines ou imprécises. PhD thesis, Université de Technologie de Compiègne, 1999.

[107] Y. Sakamoto, M. Ishiguro, and G. Kitagawa. *Akaïke Information Criterion Statistics.* KTK Scientific Publishers, Tokyo, 1986.

[108] D. Salomon. Le mélanome malin, un défi diagnostique et thérapeutique. *Médecine et Hygiène*, 2108 :475–476, 1996.

[109] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, New Jersey, 1976.

[110] C. E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27(3) :379–423,623–656, 1948.

[111] B. W. Silverman. *Density estimation for statistics and data analysis.* Chapman and Hall, London, UK, 1986.

[112] P. Smets. Information content of an evidence. *International Journal of Man-Machine Studies*, 19 :33–43, 1983.

[113] P. Smets. Bayes'theorem generalized for belief functions. In *European Conference on Artificial Intelligence (ECAI'86)*, pages 169–171, 1986.

[114] P. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :447–458, 1990.

[115] P. Smets. Constructing the pignistic probability function in a context of uncertainty. In M. Henrion, R. D. Schachter, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*, pages 29–40, Amsterdam, 1990. North-Holland.

[116] P. Smets. The transferable belief model and random sets. *International Journal of Intelligent Systems*, 7 :37–46, 1990.

[117] P. Smets. Resolving misunderstandings about belief functions. *International Journal of Approximate Reasoning*, 6 :321–344, 1992.

[118] P. Smets. Belief functions : The disjonctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9 :1–35, 1993.

[119] P. Smets. Quantifying beliefs by beliefs functions : An axiomatic justification. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence, IJCAI'93*, pages 598–603. Morgan Kaufmann, 1993.

[120] P. Smets. Non standard probabilistic and non probabilistic representations of uncertainty. Technical Report TR/IRIDIA/95-2, IRIDIA, 1995.

[121] P. Smets. *Qualitative and Quantitative Practical Reasoning*, chapter The Alpha-Junctions : Combination Operators Applicable to Belief Functions., pages 131–153. Springer Verlag, Berlin, Kruse R. and Nonengart A. and Ohlbach H. J. edition, 1997.

[122] P. Smets. La théorie des possibilités quantitatives épistémiques vue comme un modèle de croyances transférables très prudent. In *Rencontres Francophones sur la Logique Floue et ses Applications LFA'2000*, pages 343–353, Octobre 2000.

[123] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2) :191–234, 1994.

[124] M. Sugeno. *Theory of fuzzy integral and its applications.* PhD thesis, Tokyo Institute of Technology, 1974.

[125] K. Taouil. *Faisabilité en détection des mélanomes.* PhD thesis, Université de Rouen, 1995.

[126] F. Tupin. *Reconnaissance de formes et analyse de scénes en imagerie radar.* PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1997.

[127] P. Vannoorenberghe and O. Colot. Fusion of color information for image segmentation based on Dempster-Shafer's theory. In *Second International Conference on Information Fusion (FUSION'99)*, pages 816–821, 1999.

[128] P. Vannoorenberghe and O. Colot. Color image segmentation using the Dempster-Shafer's theory. In *International Conference on Image Processing (ICIP'99)*, pages CD–ROM, Kobe, Japon, Octobre 1999.

[129] P. Vannoorenberghe and T. Denoeux. Likelihood-based vs. Distance-based Evidential Classifiers. In *10th international conference on Fuzzy Systems FUZZ-IEEE'2001*, 2001.

[130] P. Vannoorenberghe and T. Denoeux. Diagnostic de la pollution atmosphérique par une approche RDF utilisant les fonctions de croyance. In *Colloque Automatique et Environnement AE'2001*, Juillet 2001.

[131] P. Vannoorenberghe, E. Lefevre, and O. Colot. Application de la théorie des fonctions de croyance à la surveillance de l'environnement. In *Rencontres Francophones sur la Logique Floue et Ses Applications LFA'2000*, pages 229–236, Octobre 2000.

[132] J. Vejnarova and G. Klir. Measure of strife in Dempster-Shafer theory. *International Journal of General Systems*, 22(1) :25–42, 1993.

[133] A. Verikas, K. Malmqvist, and M. Bacauskiene. Combining Neural Networks, Fuzzy Sets, and Evidence Theory based approaches for analysing colour images. In *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, pages 297–302, Como, Italy, July 2000.

[134] F. Voorbraak. On the justification of Dempster's rule of combinations. *Artificial Intelligence*, 48 :171–197, 1991.

[135] P. Walley. Belief function representations of statistical evidence. *The Annals of Statistics*, 15(4) :1439–1465, 1987.

[136] P. Walley. *Statistical with imprecise probabilities.* Chapman and Hall, London, 1991.

[137] P. Walley and S. Moral. Upper probabilities based only on the likelihood function. *Journal of Royal Statistical Society, Serie B*, 61(Part 4) :831–847, 1999.

[138] L. Xu, A. Krzyzak, and 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.

[139] R. R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9 :249–260, 1983.

[140] R. R. Yager. Hedging in the combination of Evidence. *Journal of Information and Optimization Sciences*, 4(1) :73–81, 1983.

[141] R. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41 :93–138, 1987.

[142] R. R. Yager, M. Fedrizzi, and J. Kacprzyk. *Advances in the Dempster-Shafer theory of evidence.* John Wiley and sons, 1994.

[143] B. Yaghlane, P. Smets, and K. Mellouli. Independence Concepts for Belief Functions. In *8th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, volume 1, pages 357–364, 2000.

[144] B. Yaghlane, P. Smets, and K. Mellouli. On Conditional Belief Function Independence. In R. Scozzafava and B. Vantaggi, editors, *Workshop on Partial Knowledge and Uncertainty : Independence, Conditioning, Inference*, pages 4–5, Italy, 2000.

[145] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8 :338–353, 1965.

[146] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.

[147] L. A. Zadeh. *On the Validity of Dempster's Rule of Combination of Evidence.* University of California, Berkeley, 1979. ERL Memo M79/24.

[148] L. M. Zouhal. *Contribution à l'application de la théorie des fonctions de croyance en reconnaissance des formes.* PhD thesis, Université de Technologie de Compiègne, 1997.

[149] L. M. Zouhal and T. Denoeux. An evidence-theoretic k-NN rule with parameter optimization. *IEEE Transactions on Systems, Man and Cybernetics - Part C*, 28(2) :263–271, 1998.

Bibliographie de l'auteur

Publication dans des revues

1. **E. Lefevre**, O. Colot , P. Vannoorenberghe et D. de Brucq, "*Contribution des mesure d'information à la modélisation crédibiliste de connaissance.*", Revue Traitement du Signal, vol. 17, n°2, pp. 87-97, 2000.
2. **E. Lefevre**, O. Colot , P. Vannoorenberghe et D. de Brucq, "*Informations et combinaison : les liaisons conflictuelles.*", Revue Traitement du Signal (A paraître), 2001.
3. **E. Lefevre**, O. Colot et P. Vannoorenberghe , "*About the belief function combination and the conflict management problem.*", Revue Information Fusion (A paraître), 2001.

Congrès internationaux avec comité de lecture

1. **E. Lefevre**, O. Colot et P. Vannoorenberghe, "*A classification method based on the Dempster-Shafer's theory and information criteria.*", Second International Conference on Information Fusion, FUSION'99, pp. 1179-1184, 6-8 Juillet 1999, Silicon Valley-San Francisco, Californie, USA.
2. **E. Lefevre**, O. Colot et P. Vannoorenberghe, "*Using information criteria in Dempster-Shafer's basic belief assignment.*", 8th International Conference on Fuzzy Systems, FUZZ-IEEE'99, pp. 173-178, 22-25 Août 1999, Séoul, Corée.
3. **E. Lefevre**, O. Colot et P. Vannoorenberghe, "*Basic Belief Assignment in the framework of information theory.*", 7th European Congress on Intelligent Techniques and Soft Computing, EUFIT'99, 13-16 Septembre 1999, Aix-la-Chapelle, Allemagne.
4. P. Vannoorenberghe, **E. Lefevre**, O. Colot et D. de Brucq, "*Cadre générique pour la résolution du conflit lors de la combinaison de structure de croyance.*", Fusion 2000 : Workshop sur l'aggregation et la fusion de données, 6-8 Avril 2000, Douz, Tunisie.
5. **E. Lefevre**, O. Colot, P. Vannoorenberghe et D. de Brucq, "*A generic framework for resolving the conflict in the combination of belief structures.*", 3rd International Conference on Information Fusion, FUSION'2000, pp. MOD4 11-18, 22-25 Août 2000, Paris, France.
6. **E. Lefevre**, P. Vannoorenberghe et O. Colot, "*Evidence theory and color image segmentation.*", The first international conference on Color in Graphics and Image Processing, CGIP'2000,

pp. 164-169, 1-4 Octobre 2000, Saint-Etienne, France.

7. **E. Lefevre**, O. Colot, P. Vannoorenberghe et D. de Brucq, "*Knowledge modeling methods in the framework of belief theory.*", IEEE international Conference on Systems, Man and Cybernetics, SMC'2000, pp. 2806-2811, 8-11 Octobre 2000, Nashville, Tennessee, USA.
8. **E. Lefevre**, P. Vannoorenberghe et O. Colot, "*Combination rules for belief functions.*", 10th International Symposium on Applied Stochastic Models and Data Analysis, ASMDA'2001, pp. 678-683, 12-15 Juin 2001, Compiègne, France.

Congrès nationaux avec comité de lecture

1. P. Vannoorenberghe, **E. Lefevre** et O. Colot, "*Application de la théorie des fonctions de croyance à la surveillance de l'environnement*", Rencontres Francophone sur la logique floue et ses applications, LFA'2000, pp. 229-236, Octobre 2000, La Rochelle, France.
2. P. Vannoorenberghe, **E. Lefevre** et O. Colot, "*Fonctions de croyance appliquées à la surveillance de la pollution atmosphérique*", Congrès ATMO'2000, Bordeaux, France.
3. **E. Lefevre**, P. Vannoorenberghe, O. Colot et T. Denoeux "*Combinaison d'évidence pour la discrimination en présence d'étiquetage imprécis*", Rencontres Francophone sur la logique floue et ses applications, LFA'2001, pp. 65-72, Novembre 2001, Mons, Belgique.

Communications dans des conférences nationales et groupes de travails nationaux avec comité de lecture

1. P. Vannoorenberghe, O. Colot et **E. Lefevre**, "*Segmentation d'images couleur : Application à la détection de mélanomes malins en dermatologie*", Communication orale au GDR-PRC ISIS dans le cadre du GT3.2 Imagerie couleur, 8 Décembre 1998, ENST Paris.
2. **E. Lefevre**, P. Vannoorenberghe, O. Colot et D. de Brucq, "*Segmentation d'images multicomposantes par fusion d'informations imparfaites.*", Communication orale au GDR-PRC ISIS dans le cadre de la journée commune Imagerie Couleur du GT3.2 et Imagerie Multicomposantes GT3.3, 9 Mars 2000, ENST Paris.
3. **E. Lefevre**, P. Vannoorenberghe, O. Colot et D. de Brucq, "*Combiner en présence d'informations conflictuelles.*", Communication orale au GDR-PRC ISIS dans le cadre du GT6 Fusion d'informations, 18 Janvier 2001, ENST Paris.
4. P. Vannoorenberghe, Y. Grandvalet, C. Ambroise, T. Denoeux, **E. Lefevre** et O. Colot, "*Méthodes ensembliste pour la surveillance de l'environnement.*", Communication orale au GDR-PRC ISIS dans le cadre du GT6 Fusion d'informations, 18 Janvier 2001, ENST Paris.