

# La fusion d'informations

Polycopié de cours ENSIETA - Réf. : 1484

Arnaud MARTIN

Janvier 2005

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Pourquoi la fusion ? . . . . .	1
1.2	Position du problème . . . . .	2
1.3	Les méthodes . . . . .	2
1.4	Plan . . . . .	3
<b>2</b>	<b>Définition et architecture</b>	<b>5</b>
2.1	Définitions de la fusion d'informations . . . . .	5
2.2	Architecture de la fusion d'informations . . . . .	6
<b>3</b>	<b>Typologie de la fusion d'informations</b>	<b>11</b>
3.1	Niveaux de fusion . . . . .	11
3.2	Type d'information . . . . .	13
3.3	Imperfections de l'information . . . . .	14
3.4	Informations supplémentaires et connaissances externes . . . . .	18
<b>4</b>	<b>Méthodes du vote</b>	<b>21</b>
4.1	Propriétés . . . . .	22
4.2	Généralisation du principe de vote majoritaire absolu . . . . .	23
4.3	Autres combinaisons et règles de décision . . . . .	24
4.4	Applications . . . . .	25
4.5	Conclusion . . . . .	25
<b>5</b>	<b>L'approche bayésienne</b>	<b>27</b>
5.1	Modélisation . . . . .	27
5.2	Estimation . . . . .	27
5.3	Combinaison . . . . .	28
5.3.1	Combinaison bayésienne . . . . .	29
5.3.2	Combinaison des probabilités . . . . .	29
5.4	Décision . . . . .	30
5.5	Applications . . . . .	31
5.6	Réflexions . . . . .	31
5.6.1	Limites . . . . .	31
5.6.2	Intérêts . . . . .	32

<b>6</b>	<b>La théorie des croyances</b>	<b>33</b>
6.1	Modélisation . . . . .	33
6.1.1	Fonction de masse . . . . .	34
6.1.2	Fonction de croyance . . . . .	35
6.1.3	Fonction de plausibilité . . . . .	37
6.1.4	Fonction de communalité . . . . .	37
6.1.5	Fonction d'implicabilité . . . . .	37
6.1.6	Propriétés . . . . .	38
6.1.7	Affaiblissement . . . . .	39
6.1.8	Conditionnement . . . . .	40
6.1.9	Raffinement et grossissement . . . . .	40
6.2	Estimation . . . . .	41
6.2.1	Fonction à support simple . . . . .	41
6.2.2	Fonction de croyance complémentaire . . . . .	42
6.2.3	Fonction sur les singletons . . . . .	42
6.2.4	Modèles probabilistes . . . . .	43
6.2.5	Modèles de distances . . . . .	45
6.2.6	Autres approches . . . . .	47
6.3	Combinaison . . . . .	47
6.3.1	Combinaison conjonctive . . . . .	47
6.3.2	Combinaison disjonctive . . . . .	53
6.3.3	Combinaison mixte . . . . .	54
6.3.4	Autres Combinaisons . . . . .	55
6.4	Décision . . . . .	58
6.4.1	Maximum de plausibilité . . . . .	58
6.4.2	Maximum de croyance . . . . .	59
6.4.3	Maximum de croyance sans recouvrement des intervalles de confiance . . . . .	59
6.4.4	Maximum de croyance avec rejet . . . . .	59
6.4.5	Maximum de probabilité pignistique . . . . .	59
6.4.6	Règle de décision avec coût . . . . .	60
6.4.7	Autres règles de décision . . . . .	61
6.5	Applications . . . . .	61
6.6	Réflexions . . . . .	62
<b>7</b>	<b>La théorie des possibilités</b>	<b>65</b>
7.1	théorie des sous-ensembles flous et des possibilités . . . . .	65
7.1.1	Théorie des sous-ensembles flous . . . . .	66
7.1.2	Théorie des possibilités . . . . .	72
7.2	Fusion d'informations par la théorie des possibilités . . . . .	75
7.2.1	Modélisation . . . . .	76
7.2.2	Estimation . . . . .	76
7.2.3	Combinaison . . . . .	77
7.2.4	Décision . . . . .	89

<i>TABLE DES MATIÈRES</i>	iii
7.3 Applications . . . . .	90
<b>Glossaire</b>	<b>90</b>



# Liste des tableaux

6.1	Illustration de la règle orthogonale pour deux sources distinctes. . . . .	48
6.2	Exemple de deux fonctions de masse issues de deux sources en conflit. . . .	49
6.3	Fusion de deux fonctions de masse issues de deux sources en conflit. . . .	49
6.4	Exemple de deux fonctions de masses identiques. . . . .	51
6.5	Exemple de la fusion de deux fonctions de masses identiques. . . . .	51
6.6	Exemple de table d'intersection. . . . .	51
6.7	Exemple de table d'intersection tenant compte de l'ignorance initiale. . . .	52
6.8	Exemple de table d'union. . . . .	54
6.9	Exemple de fusion pour des sources en conflit. . . . .	56
6.10	Exemple de fusion pour des sources en conflit avec l'introduction de l'igno- rance. . . . .	56
7.1	Exemple d'opérateurs de moyenne bissymétriques, continus et strictement croissants pour une fonction $k$ de type $x^\alpha$ . . . . .	84



# Table des figures

2.1	Représentation de la fusion. . . . .	7
2.2	Représentation du bloc FUSION. . . . .	8
3.1	Les différents niveaux de fusion. . . . .	12
3.2	Processus ECD (Extraction de connaissances à partir de données) généralisé. . . . .	13
6.1	Représentation des fonctions de croyance et de plausibilité. . . . .	39
7.1	Fonction caractéristique de l'ensemble $A$ et fonction d'appartenance de l'ensemble $B$ . . . . .	66
7.2	Noyau, support et hauteur d'un sous-ensemble flou. . . . .	68
7.3	Exemple de nombre flou représentant <i>environ 170</i> . . . . .	69
7.4	Union de deux sous-ensembles flous $A$ et $B$ . . . . .	70
7.5	Intersection de deux sous-ensembles flous $A$ et $B$ . . . . .	71

# Chapitre 1

## Introduction

### 1.1 Pourquoi la fusion ?

La fusion d'informations est apparue afin de gérer des quantités très importantes de données multisources dans le domaine militaire [1, 141, 76, 77] et reste d'actualité dans ce domaine [131, 132]. Depuis quelques années des méthodes de fusion ont été adaptées et développées pour des applications en traitement du signal. Dans ce domaine la fusion d'informations doit tenir compte des spécificités des données à toutes les étapes du processus de fusion. Elle diffère des autres domaines, d'une part à cause de ces spécificités (données hétérogènes, quantité importante, données surtout issues de capteurs (donc principalement objectives), *etc.*), d'autre part des principales raisons de la fusion en traitement du signal et des images. En effet, il s'agit ici d'améliorer la connaissance du monde observé pour le décrire du mieux possible tel qu'il est. Nous ne cherchons donc pas ici à fusionner des informations pour que le monde devienne tel que nous le souhaitons, ou tel qu'il devrait être. Par exemple dans le cas d'élections il faut fusionner les votes pour que le monde devienne tel que la population le souhaite, il n'y a pas de vérité préexistante.

Les données à fusionner peuvent être nombreuses et issues de sources hétérogènes (capteurs, bases de données, expertises [36]). Les raisons de la fusion se justifient selon les applications, et elles sont nombreuses. La fusion s'applique en imagerie satellitaire et aérienne, imagerie médicale, sonar et radar, robotique, intelligence artificielle [133], imagerie pour des applications aérospatiales [37], *etc.* Nous reviendrons sur ces différents domaines et sur des applications concrètes par la suite.

Nous pouvons distinguer trois buts dans ces domaines d'applications. Tout d'abord la fusion d'informations est employée de plus en plus pour des applications de reconnaissance (détection, identification) et c'est surtout dans ce cadre que nous nous plaçons. Les méthodes présentées permettent également l'estimation d'un paramètre obtenu par la combinaison de valeurs issues de différentes sources, mais dans ce cas d'autres approches sont envisageables [78, 185]. Enfin pour l'association de données qui permet de trouver parmi différents signaux ceux qui sont émis par la même source, les approches employées sont soit déterministes soit probabilistes (principalement bayésien, la méthode la plus employée étant le filtre de Kalman). L'association de données est particulièrement

importante pour la poursuite de cibles, nous ne traiterons pas ces approches ici.

## 1.2 Position du problème

Nous considérerons dans la suite que nous disposons de  $m$  sources  $S_j$  avec  $j \in \{1, \dots, m\}$ . Chaque source  $S_j$  peut être caractérisée par l'information fournie par la  $j^{\text{ème}}$  source comme une fonction de l'observation  $x$  notée  $s_j(x)$  ce qui permet souvent de simplifier la connaissance sur la source. Ces sources doivent prendre une décision sur une observation  $x$  dans un ensemble de  $n$  décisions  $d_1, \dots, d_n$ . Par exemple la décision  $d_i$  correspond au fait que l'observation  $x$  vérifie une hypothèse  $H_i$ , ou encore dans le cadre de la classification l'observation  $x$  appartient à la classe  $C_i$ ,  $x \in C_i$ .

Chaque source  $S_j$  fournit une information sur la décision  $d_i$  pour l'observation  $x$  que nous représenterons par  $M_i^j(x)$ . Ainsi la décision finale de l'expert sur l'observation  $x$ ,  $E(x)$  sera prise à partir de la combinaison des informations contenues dans la matrice  $(M_i^j(x))$  donnée par (1.1).

$$\begin{array}{c}
 S_1 \\
 \vdots \\
 S_j \\
 \vdots \\
 S_m
 \end{array}
 \begin{bmatrix}
 d_1 & \dots & d_i & \dots & d_n \\
 M_1^1(x) & \dots & M_i^1(x) & \dots & M_n^1(x) \\
 \vdots & \ddots & \vdots & \ddots & \vdots \\
 M_1^j(x) & \dots & M_i^j(x) & \dots & M_n^j(x) \\
 \vdots & \ddots & \vdots & \ddots & \vdots \\
 M_1^m(x) & \dots & M_i^m(x) & \dots & M_n^m(x)
 \end{bmatrix}
 \quad (1.1)$$

## 1.3 Les méthodes

Les principales méthodes de fusion d'informations sont celles que nous présentons dans ce document; elles sont issues essentiellement de deux cadres de modélisation, les probabilités et les approches floues. Des probabilités sont issues les approches bayésiennes et les méthodes issues de la théorie des croyances, des approches floues reposant sur la théorie des sous-ensembles flous est issue la théorie des possibilités. Le principe du vote, sans aucun doute l'approche la plus simple permet de fusionner des données telles que des décisions, c'est à ce titre que nous l'étudions ici.

Bien d'autres approches peuvent être considérées comme des méthodes de fusions d'informations, telles que les filtres de Kalman [47, 62, 108] davantage étudiées pour l'association de données; les classifieurs peuvent être vus comme des approches de fusion: les machines à vecteurs de support (*SVM*) [22], les chaînes de Markov cachées [138], les réseaux de neurones [2, 181], ou encore des méthodes d'extraction d'informations telles que les algorithmes génétiques [23, 170, 126, 130], ou les réseaux bayésiens [79] que nous n'aborderons pas ici.

## 1.4 Plan

Avant de présenter les différentes méthodes de fusion d'informations, nous nous attachons à donner une définition de la fusion et à présenter les différentes architectures puis nous présentons la typologie de la fusion d'informations. Nous présentons ensuite quatre cadres théoriques pour fusionner qui sont les méthodes de vote, l'approche bayésienne, la théorie des croyances et la théorie des possibilités.



# Chapitre 2

## Définition et architecture

### 2.1 Définitions de la fusion d'informations

De nombreuses définitions ont été données à la fusion d'informations à l'issue de groupes de travail. Nous ne présentons ici que les principales définitions apportées par ces groupes.

Historiquement, un premier groupe de travail le *Joint Directors of Laboratories* (JDL) du département de défense américain s'est créé en 1986. Ce groupe s'est intéressé à la définition de la terminologie de la fusion de données [1].

**Définition 2.1.1** *La fusion est un processus multi-niveaux et à facettes multiples ayant pour objet la détection automatique, l'association, la corrélation, l'estimation et la combinaison d'informations de sources singulières et plurielles [1].*

A la différence de ce cadre fonctionnelle, L. Wald [166, 167, 168] définit la fusion de données dans un cadre formel.

**Définition 2.1.2** *La fusion de données constitue un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses. Elle vise à l'obtention d'information de plus grande qualité; la définition exacte de "plus grande qualité" dépendra de l'application [168].*

Le groupe de travail européen FUSION (de 1996 à 1999) a défini la fusion d'informations dans un cadre conceptuel très général.

**Définition 2.1.3** *La fusion d'informations consiste à regrouper des informations issues de plusieurs sources d'informations et à exploiter l'information regroupée.*

I. Bloch [13] propose une définition plus spécifique.

**Définition 2.1.4** *La fusion d'informations consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision.*

Dans toutes ces définitions plusieurs notions apparaissent autour de la fusion. Tout d'abord le terme de *donnée* est employé dans la définition 2.1.2 alors que le terme *information* est préféré dans les définitions 2.1.3 et 2.1.4. Nous choisissons ici d'employer le terme de *fusion d'informations* afin de désigner l'ensemble de ce qui peut être fusionner et de ne pas le confondre avec le terme *donnée* qui sera précisé dans le chapitre 3.

Nous constatons également que dans toutes ces définitions l'information est issue de plusieurs sources, ceci n'exclue pas l'obtention de l'information par une même source mais à des instants différents, ou d'un point de vue différent [17]. Dans cette article, I. Bloch définit : *La fusion d'images est l'utilisation conjointe d'images hétérogènes pour l'aide à la prise de décision.*

Enfin l'action de fusionner est défini dans ces définitions par les termes *regrouper* et *combinaison*. Il est important de bien préciser le but de la combinaison comme dans la définition 2.1.4. En effet, de nombreux auteurs issus de la communauté informatique [145, 109] ou autres [61, 28, 29] entendent la fusion d'informations comme une concaténation d'informations provenant de différentes sources. La difficulté porte alors souvent sur la suppression de l'information redondante et la gestion d'un gros volume de données. Nous entendons ici la fusion comme un opérateur de combinaison de l'information qui ne se réduit pas à l'union de l'information des différentes sources.

Le sens de la fusion est donc plus ou moins large selon les auteurs. La définition retenue par la suite est la définition 2.1.4.

## 2.2 Architecture de la fusion d'informations

Contrairement à une première approche qui consistait à caractériser l'architecture de la fusion d'informations par les données à fusionner [47, 35] ou par les applications recherchées [1, 77], il est plus intéressant de la caractériser par sa structure [168]. L. Wald propose la cellule de l'opération de fusion de manière synthétique. Nous reprenons cette vision en la modifiant quelque peu (*cf.* figure 2.1). Cette approche intègre les informations supplémentaires sur lesquelles nous reviendrons dans le chapitre 3. Il est intéressant d'obtenir à la sortie de la fusion un indice de qualité en plus du résultat issu d'une décision qui dépend de l'application. Cet indice de qualité permettra d'évaluer l'opération de fusion et de compléter les informations supplémentaires. Les connaissances externes au système permettent le choix de la modélisation adéquate pour l'opération de fusion. Les résultats sur les données de l'expert apportent de la connaissance à l'expert qui peut en tenir compte pour une autre application ou pour modifier l'existant.

Il apparaît ainsi deux étapes importantes de la case FUSION : la modélisation et la décision. A ces deux étapes doivent être ajoutées [13] deux autres étapes : l'estimation et la combinaison (*cf.* figure 2.2). Nous définissons ici plus en détails ces quatre étapes, que nous considérons dans la suite pour toutes les méthodes de fusion.

### – Modélisation

La modélisation est la première étape ; elle consiste au choix du formalisme (*i.e.* la représentation des informations à fusionner). Cette étape est donc particulièrement

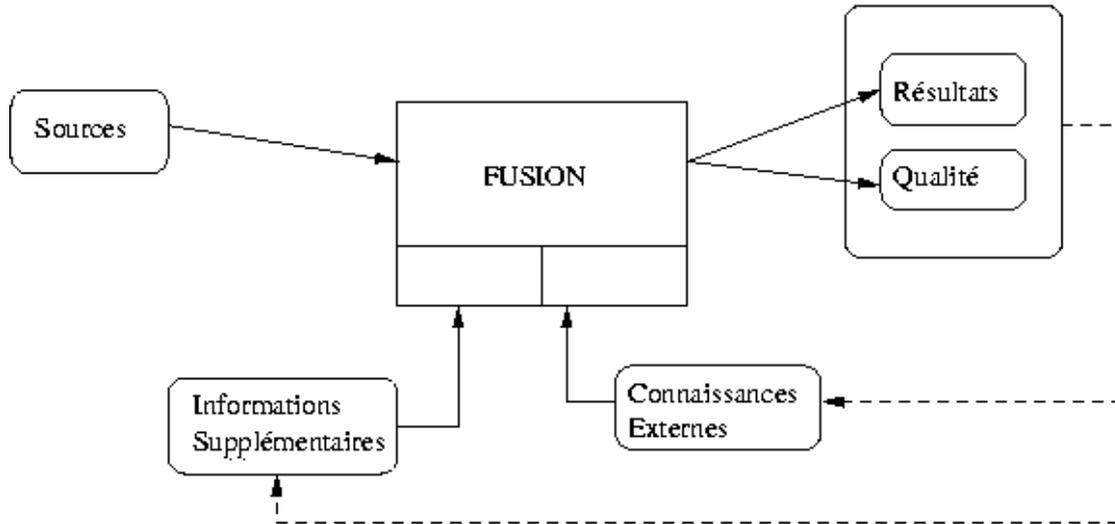


FIG. 2.1 – Représentation de la fusion.

déterminante. Elle repose essentiellement sur les connaissances de l'expert, ainsi que, dans une moindre mesure, sur les informations supplémentaires disponibles. En reprenant le formalisme de la section 1.2, chaque source  $S_j$  fournit une information représentée par  $M_i^j$  sur la décision  $d_i$ . La modélisation consiste à déterminer  $M_i^j$ , qui peut être une distribution, une fonction de coût, une formule, *etc.*

– **Estimation**

L'estimation est dépendante de la modélisation, elle n'est pas une étape systématique, mais souvent nécessaire pour la plupart des formalismes. Il s'agit par exemple de l'estimation des distributions, si c'est le formalisme choisi. Pour cette étape les informations supplémentaires peuvent aider.

– **Combinaison**

L'étape de combinaison est la phase de regroupement des informations. C'est le cœur de la fusion. Il s'agit de choisir un opérateur adapté au formalisme de la modélisation. Ici aussi des informations supplémentaires peuvent guider ce choix. Les opérateurs de combinaison fournissent un résultat de même nature que les informations combinées. L'information ainsi fournie s'interprète de la même façon que les informations de départ. Les propriétés intéressantes des opérateurs pour la fusion sont l'associativité, la commutativité, l'idempotence [10], ainsi que l'adaptabilité [107, 27, 38]. Contrairement aux probabilités, il est préférable d'avoir des opérateurs non additifs [10].

Une décision partielle prise avant la combinaison pourrait entraîner des contradictions. Le fait de conserver la nature de l'information permet de ne prendre aucune décision partielle avant la combinaison.

– **Décision**

La dernière étape de la fusion d'informations est la décision. En effet, d'après notre définition de la fusion, la combinaison est réalisée en vue d'améliorer la prise de décision. Le choix du critère se fait en fonction du formalisme (*i.e.* du choix de la

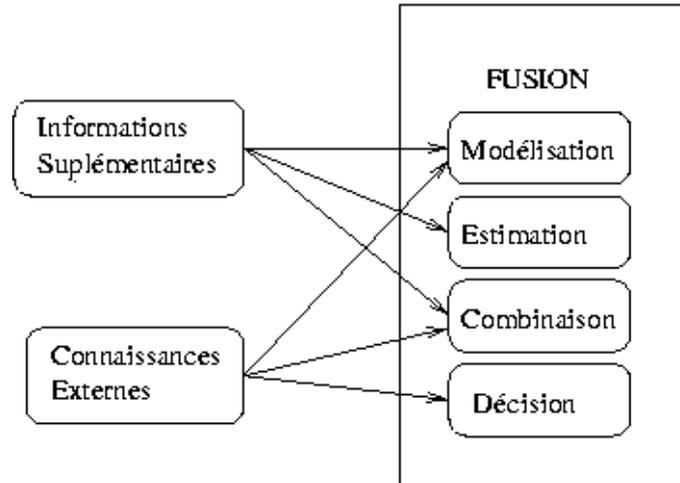


FIG. 2.2 – Représentation du bloc FUSION.

modélisation et de la combinaison). Les connaissances externes peuvent ici apporter une aide importante au choix du critère. Classiquement, il s'agit de la minimisation ou de la maximisation d'une fonction issue de la combinaison. C'est cette même fonction, ou une autre qui peut servir au calcul d'un indice de qualité. Cette étape doit donc fournir à l'expert la "meilleure" décision  $d_k$ .

Les façons de considérer ces quatre étapes définissent l'architecture de la fusion d'informations. Toute l'information de l'ensemble des sources est exprimée par la matrice (1.1) de la section 1.2. Idéalement la décision doit être prise en considérant toute la matrice. Ainsi nous tenons compte de toute l'information. C'est ce qui est appelée la fusion globale. Toutefois la complexité de cette architecture a poussé au développement d'autres approches.

Afin de simplifier la fusion globale, une première approche consiste à considérer une première étape de décision au niveau de chaque source. Il s'agit de prendre une décision locale pour chaque source prise indépendamment. Ces  $m$  décisions sont ensuite combinées pour finalement établir la décision finale. L'intérêt d'une telle approche réside dans le fait que l'information des sources n'est pas forcément disponible simultanément. Il est ainsi facile d'ajouter une source. Cette approche revient à effectuer la *fusion de décisions* (cf. Chapitre 3). Cette approche ne tient cependant pas compte des relations possibles (dépendances, corrélations) entre les sources. En outre les conflits apparaissent plus facilement (l'espace des décisions étant généralement plus petit que celui choisi pour  $M_i^j$  par la modélisation). Cette approche revient à résumer dans un premier temps la matrice (1.1) par un vecteur colonne de  $m$  paramètres (les décisions locales).

Une deuxième approche simplificatrice consiste à résumer la matrice (1.1) par un vecteur ligne de  $n$  paramètres. C'est-à-dire que nous cherchons dans une première étape à combiner les informations des  $m$  sources pour une décision  $d_i$  donnée. Il suffit dans une seconde étape de décision de prendre la décision sur le résultat de cette combinaison. Cette approche permet de conserver le formalisme choisi au départ pour  $M_i^j$ , diminuant ainsi les ambiguïtés et les conflits. Cependant, comme pour la fusion globale, toute l'information

doit être présente simultanément.

Il est également possible d'envisager un modèle hybride et tenant compte des spécificités du problème et des sources. Ce type d'architecture tente de reproduire l'expert humain en intégrant en particulier des connaissances symboliques.



# Chapitre 3

## Typologie de la fusion d'informations

### 3.1 Niveaux de fusion

Les différents niveaux de fusion sont également très discutés dans leur terminologie et dans leur nombre. B.V. Dasarathy a proposé trois niveaux de fusion [35] qui sont les données, les caractéristiques, et les décisions. La fusion de données (ou fusion de bas niveau) correspond à la fusion d'informations directement en sortie du capteur ou de la source. La donnée est donc par exemple un coefficient d'un signal à une fréquence donnée, ou encore un pixel si le signal a été reconstruit sous forme d'image. La fusion de caractéristiques (ou fusion de moyen niveau) est la fusion d'informations extraites des données issues directement de la source. Les caractéristiques sont donc par exemple, les caractéristiques phonétiques de la voix déterminées par un procédé quelconque à partir du signal, ou encore les paramètres de texture d'une image. Enfin la fusion de décisions (ou fusion de haut niveau) est relative à la fusion d'informations correspondant à la formulation d'hypothèses issues d'un expert ou d'un système (par exemple un classifieur). Ces trois niveaux ont été souvent repris [165].

A ces trois niveaux peut être ajoutée la fusion de modèles, qui est relative à la fusion d'informations décrivant des méthodes. Nous retiendrons ici ces quatre niveaux.

L. Wald [168] propose la fusion d'attributs, d'analyses et des représentations. Ces trois niveaux correspondent assez bien aux trois derniers niveaux de notre représentation : caractéristiques décisions et modèles. Dans [47] quatre niveaux sont donnés : le niveau symbolique, le niveau caractéristique, le niveau pixel et le niveau signal. A notre sens, il y a ici une confusion entre le type d'information et le niveau de la fusion. De même dans [64] une confusion dans la terminologie peut être faite entre le type d'information et le niveau de fusion. Trois niveaux sont décrits : la fusion bas niveau (au niveau du pixel) en accord avec notre premier niveau, la fusion numérique où les données sont de "type primitives ou attributs" en accord avec notre deuxième niveau et le premier niveau de L. Wald (le terme numérique se rapportant cependant pour nous au type d'informations *cf.* 3.2) et le niveau symbolique qui correspond aux informations issues d'hypothèses ou de propositions en accord avec notre troisième niveau (le terme symbolique se rapportant cependant pour nous au type d'informations *cf.* 3.2).

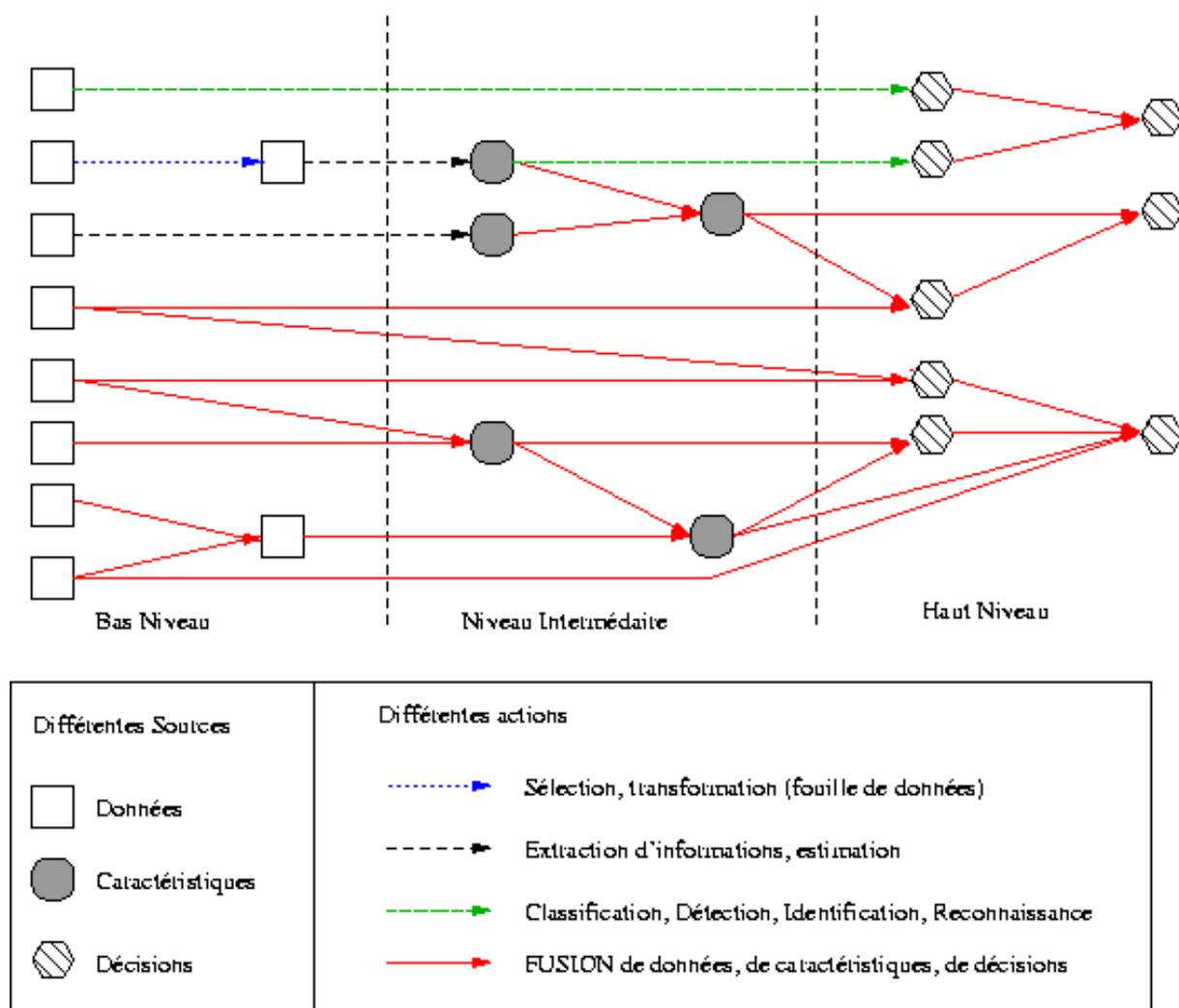


FIG. 3.1 – Les différents niveaux de fusion.

La figure 3.1 reprend les différentes entrées/sorties proposées dans [35], auxquelles nous avons ajouté la possibilité d'avoir des entrées de niveaux différents. Ceci en pratique, reviendra souvent - mais pas forcément - à homogénéiser sur le plus haut niveau. Cette figure est un exemple illustratif de l'ensemble des cas que nous pouvons rencontrer. La sélection ou transformation des données est une phase souvent nécessaire pour de nombreuses applications. Nous rappelons ici que la fusion peut jouer le rôle de sélection, transformation, extraction, et classification de l'information à partir de plusieurs sources. L'extraction d'informations à partir de données, peut être réalisée après une sélection ou transformation, ou directement à partir des données brutes conformément au processus ECD (Extraction de connaissances à partir de données) généralisé (*cf.* figure 3.2). Les approches de classifications ont été développées à partir des données, à partir de caractéristiques ou de décisions. Nous présentons donc la fusion à ces trois niveaux : données, caractéristiques et décisions. Nous pouvons fusionner des données pour obtenir des don-

nées ou des caractéristiques (dans le cadre de l'estimation) ou encore une décision (dans le cadre de la classification). La fusion de caractéristiques peut donner des caractéristiques (dans le cadre de l'estimation) ou une décision (dans le cadre de la classification). Et enfin la fusion de décisions permet de synthétiser plusieurs décisions. Cette figure illustre également les différentes possibilités de fusion sur plusieurs niveaux (données et caractéristiques pour donner des caractéristiques et des décisions, données et décisions, caractéristiques et décisions, ou données, caractéristiques et décisions pour donner des décisions).

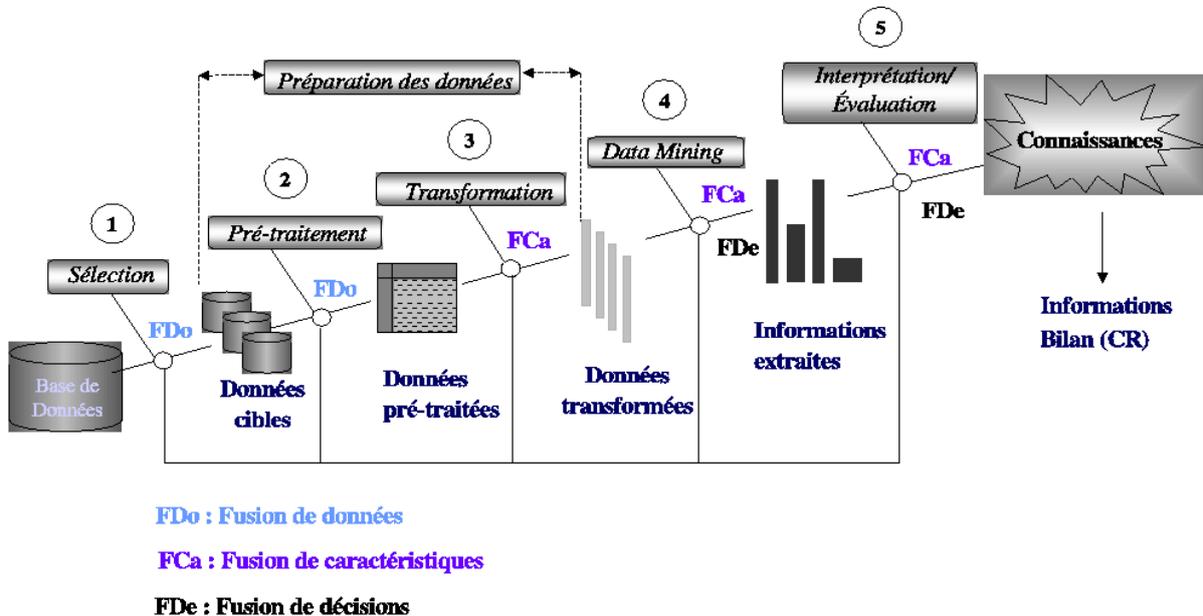


FIG. 3.2 – Processus ECD (Extraction de connaissances à partir de données) généralisé.

## 3.2 Type d'information

Il est possible de caractériser l'information en deux types [141] voire trois [13]. Le premier type concerne *l'information numérique* (*i.e.* qui s'exprime sous forme de nombre). C'est ce type d'information qui est le plus étudié, particulièrement pour des applications en traitement du signal. Il peut s'agir de l'intensité du signal, de niveaux de gris, de temps d'arrivées, *etc.*

Le deuxième type d'information est le *symbolique*. Nous appelons information symbolique toute information pouvant s'exprimer sous forme de symboles, de propositions (*ex.*

ce qui est grand n'est pas petit), de règles d'associations (*ex.* si c'est grand et que ça vole, c'est un avion). Les informations de type symbolique ont peu été étudiées dans le cadre de la fusion [25, 98], alors qu'elles peuvent être une source d'informations importante. La difficulté se trouve sur la formulation des connaissances de l'expert sur les données et les capteurs. Quelques études concernent cependant la fusion numérique symbolique [46, 143].

Enfin, I. Bloch [13] propose un type d'information hybride (*i.e.* une information symbolique codée par des nombres). Par exemple, le résultat de différentes classifications peut être modélisés par des nombres correspondant aux classes déterminées par chaque classifieur. Ces nombres peuvent ensuite être fusionnés pour améliorer la classification. Pour nous ce dernier type est discutable, car peut être considéré comme un type numérique. Nous restreindrons donc le nombre de type d'information à deux : le numérique et le symbolique.

Notons de plus que certains auteurs [47, 64] désignent par numériques les informations quantitatives et par symboliques les informations qualitatives (qui ne peuvent être codées par du numérique). Nous ne retenons pas cette définition.

### 3.3 Imperfections de l'information

Les approches de fusion d'informations reposent sur une bonne modélisation des imperfections de l'information pour pouvoir en tenir compte le mieux possible. En effet, à partir du moment où les informations sont imparfaites trois actions s'offrent à nous :

- soit nous cherchons à les supprimer,
- soit nous les tolérons et nous devons alors faire en sorte que les algorithmes mis en jeu soient robustes face à ces imperfections,
- soit nous cherchons à les modéliser.

C'est cette dernière qui est retenue par les approches de fusion d'informations. L'information est relative soit à la source soit aux informations à combiner - données, caractéristiques ou décisions selon le niveau de fusion.

L'information liée à la source peut être vue comme une information supplémentaire [13] venant s'ajouter aux informations à combiner. Elle est alors intégrée dans la modélisation. Deux qualificatifs des sources sont à considérer en priorité : l'*indépendance* et la *fiabilité*. L'indépendance est une hypothèse souvent faite, et parfois nécessaire selon le formalisme retenue. Il est rare de rencontrer une indépendance statistique des sources, car elles sont souvent de même nature, leur modélisation statistique est donc la même. Nous n'avons donc pas :

$$P(S_i \cap S_{i'}) = P(S_i).P(S_{i'}). \quad (3.1)$$

Par contre, l'indépendance cognitive est moins restrictive. L'indépendance cognitive des sources est réalisée si chaque source n'a aucune information sur les autres sources, c'est-à-dire que chaque source fournit son information sans consulter les autres sources. L'indépendance cognitive des sources est une hypothèse souvent suffisante pour les formalismes retenus, par exemple dans la théorie des croyances (*cf.* chapitre 6).

La fiabilité d'une source d'informations est souvent difficile à quantifier. Il est cependant souhaitable qu'elle soit intégrée dans la modélisation [10, 39, 63]. Il y a deux types de fiabilité :

- la fiabilité de la source à donner une information sûre,
- la fiabilité de la source à donner une information.

Ces deux types de fiabilité peuvent se quantifier à partir d'un degré de fiabilité qui est une mesure de certitude de l'information dans le premier cas et de la certitude de la source à donner une information dans le second. Ces deux notions sont donc liées à deux notions d'imperfections de l'information : l'*incertitude* et l'*incomplétude*, définies ci-après.

Les imperfections des informations à fusionner sont nombreuses. Les principales imperfections considérées sont l'incertitude et l'imprécision [58, 59, 10, 155]. A ces deux imperfections peuvent être ajoutées l'incomplétude [152, 11, 163, 40, 82, 83, 6], la redondance et la complémentarité [80, 13], le conflit [163, 39, 6, 85, 104], et enfin l'ambiguïté [13]. Nous décrivons ci-dessous ces notions, une brève description est rappelée dans le glossaire 7.3.

#### – Incertitude

L'incertitude caractérise le degré de conformité à la réalité d'une information, ou encore l'assurance d'une source en l'information fournie. Une information incertaine décrit donc une connaissance partielle de la réalité, alors que ce qui est certain entraîne nécessairement la connaissance de toute la réalité. Pour mesurer l'incertitude, la réalité doit donc être connue. Par exemple : *Cette lettre arrivera demain*. Le fait que la réalité (le futur) ne soit pas connue, ne nous permet pas d'affirmer que la lettre arrivera demain. Cette information est donc incertaine.

L'incertitude de l'information est l'imperfection la plus étudiée [156, 30]. Une des principales raisons de ces études provient du fait que les approches probabilistes permettent de modéliser l'incertitude parfaitement (*cf.* chapitre 5). Comme l'incertitude correspond à une ignorance (partielle ou totale) de la connaissance, l'incertitude peut être décrite à ce titre comme objective ou subjective [155, 25]. L'objectivité de l'incertitude est liée à la description du monde et à l'information tandis que la subjectivité de l'incertitude est liée à l'opinion de l'expert sur la vérité.

#### – Imprécision

L'imprécision d'une information est caractérisée par le contenu de l'information. Elle est relative à l'information ou à la source. Elle mesure donc un défaut quantitatif de connaissance. Nous distinguons [155] l'imprécision sans erreur (par exemple : *il pleut beaucoup*) et l'imprécision avec erreur (par exemple : *il pleut 10 mm<sup>3</sup>/h* alors qu'il y en a 15). Dans le premier cas l'imprécision est difficile à mesurer et elle reste relative à l'idée que l'expert s'en fait, alors que dans le second cas, elle se mesure par l'erreur de l'information fournie (défaut quantitatif de l'information). La réalité doit donc être connue ou estimée. L'imprécision est donc rencontrée dans de nombreux problèmes ce qui a entraîné de nombreux développements de modèles de l'imprécision [169]. L'étude de l'imprécision de l'information a vu le jour avec

la théorie des sous-ensembles flous, puis en fusion d'informations dans le cadre de la théorie des possibilités (*cf.* chapitre 7) [56] et de la théorie des croyances (*cf.* chapitre 6).

– **Lien entre incertitude et imprécision**

L'imprécision d'une information est souvent par erreur, confondue avec l'incertitude [13]. L'imprécision est donc un défaut quantitatif de la connaissance, alors que l'incertitude est un défaut qualitatif de la connaissance. De plus, la principale cause de l'incertitude d'une information provient de l'imprécision de l'information [155]. En effet, dans le cas d'informations quantitatives, l'imprécision d'une donnée entraîne une incertitude sur l'information véhiculée. De même, l'incertitude peut induire l'imprécision. Par exemple dire qu'il pleut  $10 \text{ mm}^3/\text{h}$  alors qu'il y en a 15, provoque une incertitude sur le temps qu'il fait. Par exemple, le doute sur l'arrivée de la lettre demain peut entraîner une imprécision sur l'estimation du nombre de lettres qui arriveront demain. Ces deux imperfections sont en outre souvent présentes simultanément.

– **Incomplétude**

L'incomplétude est le manque d'information apporté par la source. L'incomplétude de l'information peut être la cause de l'incertitude et de l'imprécision. Elle peut se mesurer par la différence de la quantité d'information réellement fournie par la source et de la quantité d'information que la source doit fournir (en cas de bon fonctionnement ou pour répondre à un problème posé). La quantité d'information est cependant difficile à quantifier. Une des premières mesure de la quantité d'information a été introduite par C.E. Shannon [148], connue sous le nom d'entropie de Shannon. Nous pouvons distinguer deux cas d'incomplétude de l'information : soit c'est un défaut d'information (par exemple une caractéristique ne fournit pas d'information alors qu'elle est sensée en donner (cas d'un défaut de transmission)), soit l'incomplétude est issue d'un défaut de modélisation de la source ou du problème (par exemple un radar ne permet pas de fournir une image des sous-marins immergés, l'information ne portant que sur la surface de l'eau).

– **Redondance**

La redondance d'informations ou de sources est le fait d'avoir plusieurs fois la même information. Cette imperfection, n'en est pas systématiquement une ; la fusion s'appuie sur la redondance des sources pour confirmer une information. Par exemple l'observation d'un même objet par différentes sources peut permettre de situer l'objet avec précision et de le représenter dans un espace de dimension supérieure. La redondance d'informations d'une source avec une autre peut être définie à partir d'une mesure  $H$  de la quantité d'informations fournie par une source. Ainsi la redondance d'informations d'une source  $S_1$  avec une source  $S_2$  peut s'écrire :

$$R(S_1, S_2) = H(S_1) + H(S_2) - H(S_1, S_2), \quad (3.2)$$

où  $H(S_1, S_2)$  désigne la quantité d'informations fournies par les deux sources  $S_1$  et  $S_2$ . Cette définition ne peut cependant pas être étendue sans perdre la positivité de la redondance. Nous pouvons cependant calculer la redondance d'une source ou de plusieurs sources avec une ou plusieurs autres sources en considérant la quantité d'information de ces sources :

$$R((S_1, \dots, S_j), (S_{j+1}, \dots, S_m)) = H(S_1, \dots, S_j) + H(S_{j+1}, \dots, S_m) - H((S_1, \dots, S_j), (S_{j+1}, \dots, S_m)). \quad (3.3)$$

#### – Complémentarité

Des sources ou des informations sont complémentaires si elles sont de nature différente et permettent de préciser la réponse du système. La complémentarité des informations d'une source par rapport à une autre peut se mesurer à partir d'une mesure de la quantité d'informations d'une source par exemple :

$$C(S_1|S_2) = H(S_1|S_2) = H(S_1, S_2) - H(S_2). \quad (3.4)$$

Cette définition permet d'écrire, de manière logique, que la quantité des informations d'une source est la somme de la redondance des informations de cette source avec une seconde source et de la complémentarité des informations de cette source par rapport à la seconde :

$$H(S_1) = R(S_1, S_2) + C(S_1|S_2). \quad (3.5)$$

Cette définition de la mesure de la complémentarité de l'information peut s'écrire pour une ou plusieurs sources par rapport à une ou plusieurs sources de la même façon que l'équation (3.3).

#### – Conflit

Deux informations ou plus sont en conflit si elles sont contradictoires. Le conflit peut aussi être rapporté à la source par extension. Dans le cadre de la théorie des croyances (*cf.* chapitre 6) une mesure de conflit entre l'information de plusieurs sources est définie, elle peut être nommée *inconsistance de la fusion* [80, 6]. Ce terme décrit cependant une notion plus large dans [155].

L'*inconsistance* représente à la fois le *conflit* (par exemple une source sonar détecte une mine à un endroit, alors que le sondeur décrit une matière rocheuse), l'*incohérence* (le conflit entraîne une incohérence dans la conclusion *existence d'une mine en matière rocheuse*), l'*incohérence temporelle* (par exemple un radar détecte une nappe de pétrole sur la surface de l'eau à 16h, alors qu'un autre radar décrit la surface sans pollution à 16h15) et la *confusion* qui est une incohérence plus faible (par exemple un radar annonce l'arrivée d'un avion à 16h15, alors qu'il arrive à 16h).

Les origines du conflit proviennent essentiellement de trois situations [6] :

- Les sources ne sont pas fiables. L'information est erronée et peut conduire à une ambiguïté.
  - Le cadre de discernement est non exhaustif. L'hypothèse d'un monde fermé est alors fausse (*cf.* section 3.4).
  - Les sources observent des phénomènes différents. Dans ce cas il ne faut pas combiner les informations.
- **Ambiguïté**

L'ambiguïté d'une information est le fait que celle-ci entraîne deux ou plus interprétations. L'ambiguïté peut provenir d'une autre imperfection de l'information (incertitude, imprécision, conflit, ...), mais pas nécessairement. Par exemple, la description d'une cible radar (par sa signature) ne permet pas toujours de distinguer cette cible d'une autre, lorsque les signatures sont trop proches, c'est le cas d'une signature imprécise. La fusion d'informations apportées par une autre source peut permettre de lever cette ambiguïté.

Nous avons vu que les mesures des imperfections de l'information peuvent passer par la *mesure de la quantité d'information*. L'entropie de Shannon est souvent employée pour cette mesure. L'entropie d'une source, ou encore l'apport d'informations d'une source pour un signal  $x$  est définie par :

$$H(S_j(x)) = -p(S_j(x)) \cdot \ln(p(S_j(x))) = -\sum_{i=1}^n M_i^j(x) \cdot \ln(M_i^j(x)). \quad (3.6)$$

L'entropie d'un ensemble de sources est alors donnée par :

$$H(S_1(x), \dots, S_m(x)) = -p(S_1(x), \dots, S_m(x)) \cdot \ln(p(S_1(x), \dots, S_m(x))). \quad (3.7)$$

D'autres formes d'entropie ont été étudiées, l'entropie floue de Termini, ou de Kirt Forger [25].

Outre les imperfections, d'autres aspects des données sont étudiées tels que l'aspect temporel des données [84] ou encore les caractéristiques géographiques des données indispensables pour le traitement d'images (pour des applications de recalage par exemple).

### 3.4 Informations supplémentaires et connaissances externes

Les informations supplémentaires sont des informations qui viennent compléter les informations issues des capteurs. Elles concernent les sources (leur dépendance, leur fiabilité), ou le domaine de l'application. Ces informations supplémentaires peuvent être issues des connaissances externes de l'expert si celui-ci formalise ces connaissances, elles peuvent également être issues de sources différentes telles que les résultats de la fusion. Ces informations supplémentaires peuvent aider à la modélisation, l'estimation, la combinaison

ou la décision de la fusion. Le principal problème pour l'intégration de ces informations et de la connaissance externe vient du fait qu'il est souvent difficile de les formaliser, particulièrement dans le même formalisme que les informations à combiner.

Une information supplémentaire particulièrement importante pour la modélisation et le choix du formalisme de la fusion est la connaissance sur le monde *ouvert* ou *fermé*.

**Définition 3.4.1** *Lorsque l'expert connaît toutes les possibilités de décision du problème, il est dans un monde fermé ou clos. Au contraire, s'il peut envisager d'autres réponses a priori non attendues, il doit admettre qu'il est dans un monde ouvert.*

Par exemple, un sédimentologue connaît parfaitement l'ensemble des sédiments qu'il peut rencontrer au large des côtes finistériennes. Il peut donc supposer être en monde fermé. Cependant, il n'est pas à l'abri d'une nouvelle épave que le classifieur devra reconnaître comme un sédiment. Dans ce cas, l'hypothèse d'un monde fermé est fausse. En fait, l'hypothèse d'un monde fermé est le plus souvent fausse, mais suffisante pour beaucoup d'applications.



# Chapitre 4

## Méthodes du vote

Le principe du vote est la méthode de fusion d'informations la plus simple à mettre en œuvre. Plus qu'une approche de fusion, le principe du vote est une méthode de combinaison particulièrement adaptée aux décisions. Notons  $S_j(x) = i$  le fait que la source  $S_j$  décide  $d_i$ , *i.e.* par exemple attribue la classe  $C_i$  à l'observation  $x$ . Nous supposons ici que les décisions  $d_i$  sont exclusives. A chaque source on associe la fonction indicatrice :

$$M_i^j(x) = \begin{cases} 1 & \text{si } S_j(x) = i, \\ 0 & \text{sinon.} \end{cases} \quad (4.1)$$

La combinaison des sources s'écrit par :

$$M_k^E(x) = \sum_{j=1}^m M_k^j(x), \quad (4.2)$$

pour tout  $k$ . L'opérateur de combinaison est donc associatif et commutatif. La règle du vote majoritaire consiste à choisir la décision prise par le maximum de sources, c'est-à-dire le maximum de  $M_k^E$ . Cependant cette règle simple n'admet pas toujours de solutions dans l'ensemble des décisions  $\{d_1, \dots, d_n\}$ . En effet, par exemple si le nombre de sources  $m$  est paire et que  $m/2$  sources décident  $d_{i_1}$  et  $m/2$  autres sources disent  $d_{i_2}$ , ou encore dans le cas où chaque source affecte à  $x$  une classe différente. Nous sommes donc obligé d'ajouter une classe  $d_{n+1}$  qui représente l'incertitude totale dans un monde fermé  $d_{n+1} = \{d_1, \dots, d_n\}$ . La décision finale de l'expert prise par cette règle s'écrit donc par :

$$E(x) = \begin{cases} k & \text{si } \max_k M_k^E(x), \\ n+1 & \text{sinon.} \end{cases} \quad (4.3)$$

Cette règle est cependant peu satisfaisante dans les cas où les sources donnent le maximum pour des classes différentes. La règle la plus employée est la règle du vote majoritaire absolu (souvent désignée simplement par la règle du vote majoritaire tellement cette dernière a peu d'intérêt en soi [97, 95]). La règle du vote majoritaire absolu s'écrit par :

$$E(x) = \begin{cases} k & \text{si } \max_k M_k^E(x) > \frac{m}{2}, \\ n+1 & \text{sinon.} \end{cases} \quad (4.4)$$

A partir de cette règle il a été démontré plusieurs résultats prouvant que la méthode du vote permet d'obtenir de meilleures performances que toutes les sources prises séparément. Ceci montre de manière théorique l'intérêt de la fusion d'informations.

## 4.1 Propriétés

En effet, dans [97] il est considéré le cas de sources indépendantes (statistiquement) et tel que chaque source a la même probabilité  $p$  de donner un résultat correct. Ainsi la performance est évaluée en terme de résultat correct ou non ( $p$  représente le taux de bonne décision, bonne classification, bonne reconnaissance, ...). Dans le cas où les performances sont évaluées en terme de matrice de confusion pour une classification, l'erreur de classer un objet d'une classe  $C_{i_1}$  donnée en une classe  $C_{i_2}$  donnée, peut être évaluée en terme de probabilité d'existence de cette erreur ou non. Ainsi chaque évènement peut être représenté par une variable de Bernoulli de paramètre  $p$ , et la probabilité de donner une réponse juste après combinaison suit une distribution binomiale telle que :

$$P_E(m) = \sum_{i=r}^m C_i^M p^i (1-p)^{m-i}, \quad (4.5)$$

où  $r$  est donné par :

$$r = \begin{cases} \frac{m}{2} & \text{si } m \text{ est pair,} \\ \frac{m+1}{2} & \text{si } m \text{ est impair.} \end{cases} \quad (4.6)$$

**Théorème 4.1.1** *Sous les deux hypothèses d'indépendance des sources et de même probabilité  $p$  de réussite et supposant  $m$  impair et  $m \geq 3$ , les propositions suivantes sont toujours vraies :*

- Si  $p > 0.5$ , alors  $P_E(m)$  est une fonction monotone croissante telle que :

$$\lim_{m \rightarrow \infty} P_E(m) = 1.$$

- Si  $p < 0.5$ , alors  $P_E(m)$  est une fonction monotone décroissante telle que :

$$\lim_{m \rightarrow \infty} P_E(m) = 0.$$

- Si  $p = 0.5$ , alors  $P_E(m) = 0.5$  pour tout  $m$ .

Ce théorème donne un résultat important : si le nombre de sources est impair et que chaque source a une probabilité de succès  $p > 0.5$  alors la fusion des résultats de chaque source par le principe du vote majoritaire absolu donne une meilleure probabilité de succès. Dans le cas où  $m$  est pair, ces résultats ne sont plus vrais. D'après le théorème ci-dessus, nous pouvons déduire que :

1 - pour tout  $m$  et  $p$

$$P_E(2m) < P_E(2m + 1) \quad \text{et} \quad P_E(2m) < P_E(2m - 1). \quad (4.7)$$

- 2 - S'il y a  $2m$  experts, et si  $p > \frac{m}{2m+1}$  alors  $P_E(2m)$  est une fonction monotone croissante telle que :

$$\lim_{m \rightarrow \infty} P_E(2m) = 1.$$

Ceci est vrai pour tout  $m$  et  $p \geq 1/2$ .

- 3 - S'il y a  $2m$  experts, et si  $p < \frac{m}{2m+1}$  alors  $P_E(2m)$  est une fonction monotone décroissante telle que :

$$\lim_{m \rightarrow \infty} P_E(2m) = 0.$$

Ceci est vrai pour tout  $m$  et  $p \leq 1/3$ .

- 4 - S'il y a  $2m$  experts, et si  $p < \frac{m}{2m+1}$  alors  $P_E(2m) = P_E(2m + 2)$ .

Les relations ci-dessus entraînent que pour garantir un meilleur résultat que chaque source prise séparément, dans le cas où  $m$  est pair il faut avoir un grand nombre de sources. Dans tous les cas il est préférable d'en prendre un nombre impair. [97] donne un tableau de valeurs de  $P_E$  pour différents  $m$  et  $p$ .

Ces résultats ont été étendus par [95] dans le cas des trois classifieurs en fonction de leur dépendance. La dépendance entre deux classifieurs est estimée par le coefficient de corrélation.

Une autre propriété importante de la règle du vote majoritaire absolu est le fait qu'aucune connaissance *a priori* n'est nécessaire.

Cette règle ne tient cependant pas compte des imperfections des données et n'est intéressante que dans le cas où les sources sont supposées de performances identiques.

## 4.2 Généralisation du principe de vote majoritaire absolu

Au lieu de combiner les réponses des sources par une somme simple comme dans l'équation (4.2), l'idée est d'employer une somme pondérée [172] :

$$M_k^E(x) = \sum_{j=1}^m \alpha_j M_k^j(x), \quad (4.8)$$

où  $\sum_{j=1}^m \alpha_j = 1$ . Les poids  $\alpha_j$  représentent le degré de fiabilité des sources. Si les poids sont tous égaux, nous nous ramenons au cas précédent du principe du vote majoritaire. L'estimation des poids et donc de la fiabilité, peut se faire selon plusieurs critères, par exemple le taux de réussite d'un classifieur. Contrairement à la règle précédente, celle-ci nécessite donc un apprentissage. Il est également possible de prendre en compte la fiabilité

d'une source  $S_j$  par rapport à une décision  $d_k$ . Dans ce cas la combinaison est donnée par :

$$M_k^E(x) = \sum_{j=1}^m \alpha_{jk} M_k^j(x), \quad (4.9)$$

où  $\sum_{j=1}^m \sum_{k=1}^n \alpha_{jk} = 1$ . Les poids  $\alpha_{jk}$  représente la fiabilité d'une source pour une décision donnée, et l'estimation de ces poids peut se faire comme précédemment. Les différentes règles de décision possibles peuvent se résumer par la formule suivante :

$$E(x) = \begin{cases} k & \text{si } M_k^E(x) = \max_i M_i^E(x) \geq cm + b(x), \\ n+1 & \text{sinon,} \end{cases} \quad (4.10)$$

où  $c$  est une constante,  $b(x)$  est une fonction de  $M_k^E(x)$ .

**Exemple 4.2.1** Selon les valeurs de  $c$  et de  $b(x)$ , nous retrouvons les règles précédentes.

- Si  $b(x) = 0$  et  $c = 0$ , l'observation  $x$  est attribuée à la décision majoritaire.
- Si  $b(x) = 0$  et  $c = 0.5 + \epsilon$ , avec  $\epsilon$  réel positif petit, l'observation  $x$  est attribuée à la décision de majorité absolue.
- Si  $b(x) = \max_{k', k' \neq k} M_{k'}^E(x)$  qui est le nombre de sources décidant de la deuxième meilleure décision, l'observation  $x$  est attribuée à la décision  $d_k$  s'il existe une différence significative (supérieure à  $cm$ ) entre la décision  $d_k$  et la deuxième décision possible.

### 4.3 Autres combinaisons et règles de décision

Dans [172] il est proposé d'autres combinaisons des réponses des sources, et par suite d'autres règles de décision. En introduisant les opérateurs logiques, nous pouvons écrire la règle la plus conservatrice du ET logique :

$$E(x) = \begin{cases} k & \text{si } \exists k, \bigcap_{j=1}^m \{M_k^j(x) > 0\}, \\ n+1 & \text{sinon.} \end{cases} \quad (4.11)$$

Pour adoucir cette règle, il est possible d'introduire le OU logique :

$$E(x) = \begin{cases} k & \text{si } \exists k, \bigcap_{j=1}^m \{M_k^j(x) > 0\} \cup \bigcup_{i=1}^n \{1 - M_i^j(x) > 0\}, \\ n+1 & \text{sinon.} \end{cases} \quad (4.12)$$

La décision prise sera  $d_k$  si des sources soutiennent  $d_k$  et aucune autre source ne soutient une autre décision. Cette règle peut être intéressante dans le cas d'un monde ouvert, *i.e.* si les sources peuvent décider  $d_{n+1}$  (par exemple  $x \in C_{n+1}$ ).

## 4.4 Applications

La méthode du vote est appliquée lorsque l'information s'exprime de façon symbolique sous forme d'hypothèses, c'est pourquoi elle est davantage adaptée à la fusion de décisions pour des applications de classifications (reconnaissance, détection, identification). Du fait que ce soit une méthode simple à mettre en œuvre, et que dans sa version du vote majoritaire absolu ou non il n'y a pas besoin de connaissance *a priori*, elle a souvent été comparée à d'autres approches de fusion d'informations. Ainsi, [172] la compare avec différentes approches bayésiennes et l'approche de Dempster-Shafer pour de la reconnaissance manuscrite. Sa conclusion montre que l'approche par vote reste moins bonne que les deux autres. La règle de décision (4.10) est meilleure lorsque  $b(x) = \max_{k', k' \neq k} M_{k'}^E(x)$  qu'avec  $b(x) = 0$ .

Cependant [135] montre que dans le cas de la reconnaissance de cibles sur des images, la fusion des résultats de quatre classifieurs (*Support Vecteur Machine*, deux réseaux de neurones et une classification fondée sur un vecteur de quantification) par la méthode du vote donne des résultats équivalents à l'approche bayésienne. En fait l'approche employée est la règle du vote majoritaire, lorsque la réponse est la classe  $C_{n+1}$  alors l'approche bayésienne est appliquée.

Dans le cadre de la classification, L.I. Kuncheva a comparé le vote majoritaire avec beaucoup d'autres méthodes telles que l'approche bayésienne, Dempster-Shafer, l'intégrale floue, les combinaisons par maximum, minimum, moyennes, médiane, oracle, *etc* [93, 92, 91]. Les données sur lesquelles ces méthodes ont été appliquées sont diverses, phonèmes, données médicales pour les maladies du cœur, cancer, diabète, ... Les résultats montrent que l'approche du vote majoritaire est moins bonne, mais reste compétitive selon les données. Notons également que Yager s'est penché dans [177] sur différents opérateurs de moyennes, et introduit la *moyenne puissante*.

Dans [89] il est montré que la règle du maximum de la somme des sorties des classifieurs est toujours meilleure que l'approche du vote majoritaire sous des hypothèses de même distribution gaussienne pour les sources indépendantes.

De nombreuses études ont comparées le vote majoritaire dans le cas de la classification ou de la reconnaissance. Même si cette règle reste simple, elle donne souvent des résultats comparables aux approches plus complexes.

## 4.5 Conclusion

Nous résumons ici les quatre étapes de la méthode du vote.

### - Modélisation

La modélisation est fondée sur la définition des fonctions indicatrices :

$$M_i^j(x) = \begin{cases} 1 & \text{si } S_j(x) = i, \\ 0 & \text{sinon,} \end{cases} \quad (4.13)$$

avec  $S_j(x) = i$  qui désigne le fait que la source  $S_j$  décide  $d_i$  (*i.e.* par exemple attribue la classe  $C_i$  à l'observation  $x$ ). Les décisions  $d_i$  sont exclusives.

- **Estimation**

A partir de cette modélisation, il n'y a pas d'étape d'estimation.

- **Combinaison**

La combinaison peut s'exprimer simplement par une somme pondérée des fonctions indicatrices :

$$M_k^E(x) = \sum_{j=1}^m \alpha_{jk} M_k^j(x), \quad (4.14)$$

où  $\sum_{j=1}^m \sum_{k=1}^n \alpha_{jk} = 1$ . Les poids  $\alpha_{jk}$  représentent la fiabilité d'une source pour une classe donnée et sont à estimer. Cette combinaison est associative et commutative, mais n'est pas idempotente. D'autres combinaisons telles que l'intersection et l'union peuvent être envisagées.

- **Décision**

L'étape de décision peut être décrite par plusieurs règles selon la combinaison choisie. Pour une combinaison s'exprimant sous la forme d'une somme pondérée des fonctions indicatrices, la règle de décision peut s'exprimer sous la forme :

$$E(x) = \begin{cases} k & \text{si } M_k^E(x) = \max_i M_i^E(x) \geq cm + b(x), \\ n + 1 & \text{sinon,} \end{cases} \quad (4.15)$$

où  $c$  est une constante,  $b(x)$  est une fonction de  $M_k^E(x)$ .

Les avantages de l'approche par vote sont double, elle est très simple et naturelle et elle ne nécessite pas de connaissances *a priori*. La règle de base peut cependant être modifiée de façon à intégrer les imperfections des données sous forme de fiabilité.

Du fait que cette méthode soit adaptée à la fusion de données symboliques, elle a été employée au niveau des décisions de classifieurs et d'approches de reconnaissance de formes.

Cette méthode a permis de montrer théoriquement que la fusion de classifieurs peut donner sous certaines hypothèses de meilleures performances que chaque classifieur pris individuellement. Nous avons également remarqué qu'il est plus intéressant pour cette approche de considérer un nombre impair de classifieurs qu'un nombre pair. En pratique, il donc plus intéressant de ne considérer qu'un nombre impair de sources pour accroître les performances.

# Chapitre 5

## L'approche bayésienne

L'approche bayésienne et plus généralement les approches probabilistes reposent sur un cadre mathématique rigoureux : la théorie des probabilités. En fait, l'approche bayésienne fait plus référence à l'étape de combinaison et en moindre mesure à l'étape d'estimation. Dans le cadre probabiliste, les imperfections de l'information et les informations elles-même sont modélisées à partir de distributions de probabilités ou de mesures statistiques à estimer telle que la quantité d'informations que nous avons vu à la section 2.2. De ce fait, cette approche ne permet de modéliser que l'incertitude de l'information.

Nous présentons ici les quatre étapes de l'approche bayésienne : la modélisation, l'estimation, la combinaison et la décision, à la suite desquelles nous exposons quelques applications de la littérature pour conclure sur les limites et intérêts de cette approche.

### 5.1 Modélisation

L'étape de modélisation est fondée sur les probabilités conditionnelles. Nous avons donc :

$$M_i^j(x) = p(d_i|S_j), \quad (5.1)$$

avec  $j \in \{1, \dots, m\}$  et  $i \in \{1, \dots, n\}$ . Dans cette approche probabiliste, nous sommes forcés d'avoir des décisions exclusives et exhaustives. Dans le cas de la classification, ceci signifie que si  $x \in C_i$  alors  $x \notin C_k$  pour tout  $k \neq i$  et que toutes les classes sont connues par avance (hypothèse de monde fermé).

### 5.2 Estimation

A partir du moment où l'étape de modélisation fait intervenir des distributions de probabilités, l'étape d'estimation est indispensable. En effet, les distributions  $p(d_i|S_j)$  sont rarement connues en traitement du signal et des images. Pour l'estimation de ces distributions, deux cas se présentent :

– **Cas discret**

En appliquant la formule de Bayes, nous estimons  $p(S_j|d_i)$  par dénombrement *i.e.* à l'aide des fréquences sur une base d'apprentissage.

– **Cas continu**

Dans ce cas, il faut chercher à approcher les distributions  $p(d_i|S_j)$  par des distributions connues. Les distributions gaussiennes sont très couramment employées et trouvent leur justification dans l'application du théorème de la limite centrale. Pour les distributions gaussiennes il s'agit donc d'estimer les moyennes et variances. Pour simplifier l'estimation des variances, une distribution laplacienne est parfois employée.

Une autre approche qui peut servir dans les deux cas, est fondée sur les  $k$  plus proches voisins [172]. Il s'agit de trouver les  $k$  plus proches voisins de  $s_j(x)$  (la représentation simplifiée de la réponse de la source  $S_j$ ) dans une base d'apprentissage avec :

$$k = \sum_{i=1}^n k_i, \quad (5.2)$$

où  $k_i$  est le nombre de points d'apprentissage les plus proches dans l'espace de  $s_j(x)$  pour lesquels il faut décider  $d_i$  (appartenant à la classe  $C_i$  dans le cas de la classification). Il est ainsi possible d'estimer  $p(d_i|S_j)$  :

$$p(d_i|S_j) = \frac{k_i}{k}. \quad (5.3)$$

[172] propose également une approche fondée sur une distance (euclidienne, Mahalanobis, ou pseudo distances). Une fonction  $p_i^j$  (qui n'est plus réellement la probabilité  $p(d_i|S_j)$ ) est estimée à partir des distances  $d_i^j(x)$  entre  $s_j(x)$  et les centres des classes  $C_i$  par une fonction  $f_j$  :

$$p_i^j = f_j(d_i^j(x)), \forall i \in \{1, \dots, n\}, \quad (5.4)$$

par exemple :

$$p_i^j = \frac{1/d_i^j(x)}{\sum_{i=1}^n 1/d_i^j(x)}. \quad (5.5)$$

### 5.3 Combinaison

Nous présentons ici deux types de combinaison [13]. La première s'appuie sur la règle de Bayes, la seconde cherche à combiner les probabilités par différents types d'opérateurs.

### 5.3.1 Combinaison bayésienne

La combinaison par l'approche bayésienne consiste à déterminer les probabilités  $p(d_i|S_1, \dots, S_m)$ . Ces probabilités peuvent être estimées directement au niveau de la modélisation par la règle de Bayes :

$$p(d_i|S_1, \dots, S_m) = \frac{p(S_1, \dots, S_m|d_i)p(d_i)}{p(S_1, \dots, S_m)}, \quad (5.6)$$

où les différentes probabilités sont alors à estimer dans l'étape d'estimation.

Ces probabilités peuvent également être calculées par la règle de Bayes de façon adaptative avec l'arrivée de l'information d'une nouvelle source :

$$p(d_i|S_1, \dots, S_m) = \frac{p(S_1|d_i)p(S_2|S_1, d_i) \dots p(S_m|S_1, \dots, S_{m-1}, d_i)p(d_i)}{p(S_1)p(S_2|S_1) \dots p(S_m|S_1, \dots, S_{m-1})}, \quad (5.7)$$

où les différentes probabilités sont également à estimer dans l'étape d'estimation. Ces deux manières de calculer les probabilités sont parfaitement équivalentes, mais la seconde permet d'intégrer les informations disponibles successivement alors que la première nécessite d'avoir les informations des sources simultanément.

En fait la difficulté d'estimer les différentes probabilités dans les équations (5.6) et (5.7) (car il faut un nombre de données d'apprentissage important) pousse à faire une hypothèse d'indépendance statistique des sources conditionnellement à une décision (qu'il est possible de vérifier par différents critères tels que le calcul des matrices de confusion). Nous obtenons ainsi :

$$p(d_i|S_1, \dots, S_m) = \frac{\prod_{j=1}^m p(S_j|d_i)p(d_i)}{\prod_{j=1}^m p(S_j)}, \quad (5.8)$$

où les différentes probabilités sont à estimer dans l'étape d'estimation. Cette équation montre que le type de combinaison s'exprime sous la forme d'un produit, c'est donc une combinaison conjonctive. Elle est de plus associative. Sous l'hypothèse d'indépendance, elle devient commutative et associative alors qu'elle ne l'est pas dans sa version (5.7). Lorsque l'hypothèse d'indépendance est faite, l'approche bayésienne est qualifiée de naïve.

### 5.3.2 Combinaison des probabilités

Une autre approche possible est de considérer la probabilité  $p_j(x) = (p(d_1|S_j), \dots, p(d_n|S_j))$  comme le résultat de chaque source, puis de combiner ces probabilités afin d'avoir la probabilité globale d'obtenir une décision  $d_i$ . Nous nous ramenons ainsi à un problème d'estimation, et nous pouvons utiliser des opérateurs de combinaison tels que le maximum, le minimum, la moyenne, la médiane, oracle, *etc*

[172, 93, 91, 91]. Cette combinaison se fait sur les vecteurs colonnes de la matrice (1.1), ainsi si l'opérateur est noté  $\phi$ , la décision sera prise sur le vecteur :

$$\mu(d_i|S_1, \dots, S_m) = \phi(p(S_1|d_i), \dots, p(S_m|d_i)), \quad (5.9)$$

où  $\mu(d_i|S_1, \dots, S_m)$  est un estimateur de  $p(d_i|S_1, \dots, S_m)$  dans le cas des opérateurs de moyenne et médiane. Dans le cas du maximum et minimum,  $\mu$  reste une probabilité.

## 5.4 Décision

Pour la dernière étape de nombreux critères de décisions sont envisageables. La règle la plus employée est celle du maximum *a posteriori*, nous choisirons  $d_k$  si :

$$p(d_k|S_1, \dots, S_m) = \max_{i \in \{1, \dots, n\}} p(d_i|S_1, \dots, S_m). \quad (5.10)$$

Il existe de nombreux critères statistiques, en voici quelques uns :

- L'espérance maximale, nous choisirons  $d_k$  si :

$$E[d_k|S_1, \dots, S_m] = \max_{i \in \{1, \dots, n\}} E[d_i|S_1, \dots, S_m]. \quad (5.11)$$

- Le maximum de vraisemblance (cas particulier de l'espérance maximale), nous choisirons  $d_k$  si  $d_k$  est solution de :

$$\frac{\partial L(S_1, \dots, S_m, d_i)}{\partial d_i} = 0, \quad (5.12)$$

avec

$$\frac{\partial^2 L(S_1, \dots, S_m, d_i)}{\partial d_i^2} < 0, \quad (5.13)$$

et la fonction de vraisemblance est donnée par :

$$L(S_1, \dots, S_m, d_i) = p(d_i|S_1, \dots, S_m). \quad (5.14)$$

- Le maximum d'entropie, nous choisirons  $d_k$  si :

$$\begin{aligned} H(d_k|S_1, \dots, S_m) &= \max_{i \in \{1, \dots, n\}} H(d_i|S_1, \dots, S_m) \\ &= \max_{i \in \{1, \dots, n\}} \{-p(d_i|S_1, \dots, S_m) \ln p(d_i|S_1, \dots, S_m)\}. \end{aligned} \quad (5.15)$$

- Le critère minimax, qui consiste à prendre la décision  $d_k$  qui donne la performance maximale dans le cas le plus défavorable.

Bien sûr le choix d'un critère sera guidé par le formalisme de la modélisation et de la combinaison.

## 5.5 Applications

Nous recensons dans cette partie quelques travaux de manière non exhaustive ayant employé l'approche bayésienne pour la fusion d'informations.

Citons tout d'abord les travaux de L.I. Kuncheva [91, 94], qui compare l'approche bayésienne naïve au maximum, minimum, moyenne et à la méthode du vote. Ces approches sont comparées dans le cadre de la fusion de classifieurs. La conclusion obtenue sur différentes bases de données montre les bonnes performances de l'approche naïve, l'approche de la moyenne donne des taux raisonnables et garde une bonne stabilité à la différence des autres approches.

Chauvin [26] étudie les performances du modèle bayésien de fusion pour la segmentation d'images satellitaires (en 3 classes) avec pour opérateur de combinaison le produit et pour règle de décision le maximum *a posteriori*. Il est montré que les performances sont fonction de l'apprentissage (choix de la densité de probabilité conditionnelle et des probabilités *a priori*) et de la qualité des images, mesurée par un critère fondé sur l'inertie intra-classes de Fisher.

[172] compare le classifieur bayésien à l'approche par vote et à la théorie de Dempster-Shafer pour une application en reconnaissance manuscrite. Il est montré que si les matrices de confusion sont bien apprises, l'approche bayésienne est la meilleure. Cependant cette approche est instable ; les performances sont vite dégradées si l'apprentissage est mauvais.

L'approche bayésienne est également employée dans la fusion d'images [53].

## 5.6 Réflexions

### 5.6.1 Limites

Le problème majeur des probabilités est qu'elles représentent essentiellement l'incertitude et très mal l'imprécision, ceci entraîne souvent une confusion des deux notions.

De plus, dans l'étape de modélisation probabiliste nous raisonnons sur des singletons qui représentent les différentes décisions. Ces décisions sont exhaustives et exclusives. Ceci entraîne que le monde doit être fermé, ce qui ne correspond pas toujours à la réalité. L'exclusivité provient du fait de l'hypothèse d'additivité des probabilités. Cette hypothèse peut être trop forte. Par exemple [13], dans le domaine du diagnostic médical, si un symptôme  $s$  est toujours présent quand un patient possède une pathologie  $A$  (*i.e.*  $P(s|A) = 1$ ) alors si ce symptôme  $s$  est présent, la probabilité pour que le sujet observé ait la pathologie  $A$  augmente. En effet  $p(A|s) = \frac{p(s|A)p(A)}{p(s)} = \frac{p(A)}{p(s)}$  et donc  $p(A|s) \geq p(A)$ . L'additivité des probabilités impose donc que la probabilité que le sujet n'ait pas  $A$  diminue ( $p(\bar{A}|s) = 1 - p(A|s)$  donc  $p(\bar{A}|s) \leq p(\bar{A})$ ) or rien ne justifie cela si ce symptôme peut être rencontré dans d'autres pathologies. L'hypothèse d'additivité sur les probabilités conditionnelles est ici trop forte. Il est également difficile de modéliser des connaissances qui ne se traduisent pas par des probabilités, ou l'absence de connaissances. Par exemple, l'ignorance sur l'existence d'une cible à détecter peut se modéliser par  $p(\text{cible}) = p(\text{non$

cible) = 0.5. Si le problème est posé de manière différente et que l'important est la détection de cibles ennemies, c'est-à-dire qu'il peut y avoir une cible amie, ennemie ou ne pas y avoir de cibles, l'ignorance peut alors être modéliser par  $p(\text{cible amie}) = p(\text{cible ennemie}) = p(\text{non cible}) = 1/3$  et donc  $p(\text{cible}) = 2/3$ . Il est ainsi possible par un raisonnement analogue d'exprimer cette probabilité par autant de valeurs que de façons d'aborder le problème.

L'étape d'estimation est donc également soumise à ces difficultés. De plus il est important d'avoir suffisamment de données pour estimer les probabilités conditionnelles avec une faible erreur, l'estimation pouvant se faire facilement à partir des fréquences d'occurrence. Par contre l'estimation des probabilités *a priori* est beaucoup plus délicate et il est souvent nécessaire de faire appel à une connaissance plus subjective sur l'application.

La combinaison est également contrainte par les axiomes des probabilités et son utilisation, en pratique elle nécessite souvent l'hypothèse d'indépendance.

### 5.6.2 Intérêts

Le principal intérêt de l'approche bayésienne ou plus généralement les approches probabilistes vient du fait qu'elle repose sur une base mathématique solide éprouvée par de nombreuses études depuis des années. La théorie des probabilités trouve sa justification dans le cas où la connaissance a été acquise au cours d'apprentissage. Elle est également justifiée car elle répond aux cinq postulats de Cox [9, 10] :

- 1- cohérence ou non-contradiction,
- 2- continuité de la méthode,
- 3- universalité ou complétude,
- 4- énoncés sans équivoque,
- 5- pas de refus d'information et prise en compte de la dépendance du contexte.

En particulier, la théorie des probabilités offre un grand nombre de méthodes pour les étapes de modélisation et d'estimation, spécialement dans le cadre d'apprentissage.

L'étape de combinaison repose également sur des bases mathématiques solides. La combinaison dans sa version de l'équation (5.7) permet la mise à jour si les informations sont successives et l'hypothèse d'indépendance (équation (5.8)) n'est pas obligatoire. De plus de nombreux opérateurs permettent la combinaison des informations issues des sources en vue de la classification mais aussi de l'estimation de paramètres.

Enfin, de nombreux critères statistiques ont été développés et peuvent être employés dans l'étape de décision. Si le choix d'un critère optimal n'est pas toujours aisé, il devra être fait en fonction des étapes de modélisation et combinaison.

Sous l'hypothèse forte et rarement vérifiée que l'information d'une source en fonction des différentes décisions possibles peut s'exprimer sous la forme de distributions de probabilités, la méthode bayésienne est optimale.

# Chapitre 6

## La théorie des croyances

La théorie des croyances est issue des travaux de Dempster en 1967 [41], repris par Shafer [146] sous le nom de *theory of evidence*, elle porte également le nom de théorie de Dempster-Shafer ou théorie de l'évidence (qui est une mauvaise traduction de *evidence*). Cette méthode repose sur la modélisation de la croyance en un évènement. Cette modélisation se réalise à partir de fonctions de masse permettant une bonne représentation des connaissances. Inspirée de la théorie des probabilités supérieures et inférieures [41, 146], elle est souvent présentée d'un point de vue probabiliste, mais peut être vue comme un modèle formel de degrés de confiance dans le modèle transférable des croyances de Smets [150, 160], qui a une approche plus subjective. Cette théorie permet particulièrement de bien modéliser l'incertitude, mais aussi l'imprécision. Elle tient compte également des ambiguïtés et des conflits entre les sources. Dezert et Smarandache [49], [50] ont proposé une extension coûteuse dans le cas où le conflit entre les sources est important.

Enfin du fait de sa modélisation, la théorie des croyances est plus adaptée à des problèmes de classification que des problèmes d'estimation, jamais aujourd'hui abordés dans le cadre de cette approche. En effet, pour ce faire il faudrait aborder cette théorie dans un cadre continu, ce qui est proposé dans [134] pour de la classification, et dans [159] dans un cadre plus générale.

Nous présentons ci-dessous, les quatre étapes de la fusion d'informations pour la théorie des croyances.

### 6.1 Modélisation

Le principe de la théorie des croyances repose sur la manipulation de fonctions définies sur des sous-ensembles et non sur des singletons comme dans la théorie des probabilités. Ces fonctions sont habituellement à valeur dans  $[0, 1]$ , et sont appelées *fonctions de masse* ou encore *masses élémentaires* ou *masses de croyance*.

### 6.1.1 Fonction de masse

Considérons l'ensemble appelé *cadre de discernement*  $D = \{d_1, d_2, \dots, d_n\}$ , ensemble de toutes les décisions possibles - par exemple l'ensemble de toutes les classes envisageables dans une problématique de classification. L'espace des fonctions de masse  $m$  est donné par l'ensemble de toutes les disjonctions possibles des décisions  $d_i$  noté

$2^D = \{\emptyset, \{d_1\}, \{d_2\}, \{d_1 \cup d_2\}, \{d_3\}, \{d_1 \cup d_3\}, \{d_2 \cup d_3\}, \{d_1 \cup d_2 \cup d_3\}, \dots, D\}$ , ou encore  $2^D = \{\emptyset, \{d_1\}, \{d_2\}, \{d_1, d_2\}, \{d_3\}, \{d_1, d_3\}, \{d_2, d_3\}, \{d_1, d_2, d_3\}, \dots, D\}$ . Les décisions  $d_i$  doivent être exclusives, mais pas nécessairement exhaustives. La différence introduite par la généralisation de Dezert-Smarandache [49, 50] consiste à considérer les décisions  $d_i$  pas forcément disjonctives, il faut donc ajouter les toutes les intersections possibles entre deux éléments de  $2^D$ . Les fonctions de masse sont alors définies sur un ensemble  $D^D$ , tel que  $\forall A, B \in D^D, A \cup B \in D^D$  et  $A \cap B \in D^D$ .

Nous définissons ici une fonction de masse comme une fonction définie sur  $2^D$  à valeurs dans  $[0, 1]$ . Pour une source  $S_j$ , la fonction de masse  $m_j$  vérifie en général par construction :

$$\sum_{A \in 2^D} m_j(A) = 1, \quad (6.1)$$

ici, la différence avec les probabilités réside dans le fait que  $A$  peut être l'union de deux décisions  $d_1$  et  $d_2$ . C'est grâce à ce principe que la théorie des croyances permet de modéliser les imprécisions. Une hypothèse souvent faite par commodité est celle d'un *monde clos*, ou *monde fermé*, c'est-à-dire que toutes les décisions possibles sont représentées dans  $D$ , et dans ce cas :

$$m_j(\emptyset) = 0. \quad (6.2)$$

Au contraire, si nous pouvons accepter le fait qu'une décision hors de  $D$  est envisageable, alors :

$$m_j(\emptyset) > 0, \quad (6.3)$$

et dans ce cas nous supposons un *monde ouvert*, c'est-à-dire que les décisions ne sont pas exhaustives. C'est l'hypothèse préconisée par Smets [150, 160], retenue par exemple dans [81, 83].

La fonction de masse  $m_j(A)$  caractérise le degré de croyance en la proposition  $A$  donnée par la source  $S_j$ , et strictement en la proposition  $A$ . Si  $A$  est une disjonction,  $m_j(A)$  est complètement indépendante des sous-ensembles ou sur-ensembles de  $A$ .

**Définition 6.1.1** *Un élément focal est un élément  $A$  de  $2^D$  tel que  $m_j(A) > 0$ . La réunion des éléments focaux est appelé le **noyau**.*

### Cas particuliers

- Si les seuls éléments focaux sont les décisions  $d_i$ , la fonction de masse  $m_j$  est alors une probabilité si elle vérifie l'hypothèse (6.1). En effet, elle est alors définie uniquement sur des singletons.
- Si  $m_j(D) = 1$ , alors la source  $S_j$  est dans l'ignorance totale.
- Si  $m_j(A) = 1$ , alors la source  $S_j$  a une connaissance imprécise, elle ne croit qu'en  $A$  qui peut être une disjonction de décisions.
- Si  $m_j(d_i) = 1$ , alors la source  $S_j$  a une connaissance précise, elle croit pleinement en  $d_i$ .
- Si  $m_j(A) = s$  et  $m_j(D) = 1 - s$ , avec  $s \in [0, 1]$  (cas des fonctions à support simple), alors la source  $S_j$  a une connaissance incertaine et imprécise, elle croit en partie en  $A$ , mais rien de plus.

### 6.1.2 Fonction de croyance

La *fonction de croyance* ou *fonction de crédibilité* permet de modéliser l'incertitude. Lorsque nous sommes incertains d'une chose, nous pouvons toujours croire plus ou moins en la réalité de cette chose. Cette fonction de croyance  $Cr_j$  pour une source  $S_j$ , est une fonction de  $2^D$  dans  $[0, 1]$  qui vérifie  $Cr_j(\emptyset) = 0$  et pour tout  $A_1, A_2, \dots, A_k$  :

$$Cr_j\left(\bigcup_{i=1}^k A_i\right) \geq \sum_{I \subseteq \{1, \dots, k\}, I \neq \emptyset} (-1)^{|I|+1} Cr_j\left(\bigcap_{i \in I} A_i\right), \quad (6.4)$$

où  $|I|$  représente le cardinal de l'ensemble  $I$ . Par exemple, cette relation pour  $k = 3$  s'écrit :

$$\begin{aligned} Cr_j(A_1 \cup A_2 \cup A_3) &\geq Cr_j(A_1) + Cr_j(A_2) + Cr_j(A_3) \\ &\quad - Cr_j(A_1 \cap A_2) - Cr_j(A_1 \cap A_3) - Cr_j(A_2 \cap A_3) \\ &\quad + Cr_j(A_1 \cap A_2 \cap A_3). \end{aligned} \quad (6.5)$$

Cette relation est difficile à comprendre sous cette forme, illustrons la dans le cas où  $k = 2$  et  $A_1 \cap A_2 = \emptyset$  :

$$Cr_j(A_1 \cup A_2) \geq Cr_j(A_1) + Cr_j(A_2). \quad (6.6)$$

Dans ce cas, la croyance en l'union disjointe de deux éléments est supérieure ou égale à la somme des croyances. Nous croyons plus à une union d'hypothèses disjonctives qu'en ces hypothèses prises individuellement. Si nous avons l'égalité dans l'équation (6.4), alors la fonction de croyance pourrait être une probabilité vérifiant ainsi les axiomes de Kolmogorov [157] : soit  $(\Omega, \mathcal{F}, P)$  un espace probabilisé, la probabilité  $P$  est définie par :

- $0 \leq P(A) \leq 1, \forall A \in \mathcal{F}$ ,
- $P(\Omega) = 1$ ,
- $P(\cup_{i=1}^{+\infty} A_i) = \sum_{i=1}^{+\infty} P(A_i), \forall A_i \in \mathcal{F}$  deux à deux disjoints.

Cette fonction de croyance  $Cr$  peut aussi être définie simplement à partir des fonctions de masse sous l'hypothèse (6.1), ainsi pour tout  $A \in 2^D$  :

$$Cr_j(A) = \sum_{B \subseteq A, B \neq \emptyset} m_j(B), \quad (6.7)$$

où  $B \subseteq A$  équivaut à  $B \in 2^A$ , l'ensemble de toutes les disjonctions de  $A$ . L'ensemble vide ne doit pas intervenir à la fois pour  $A$  et son complémentaire  $A^c$ . Nous obtenons ainsi une bijection entre  $Cr_j$  vérifiant  $Cr_j(\emptyset) = 0$  et  $m_j$  vérifiant  $\sum_{A \in 2^D} m_j(A) = 1$ . Nous remarquons immédiatement que  $Cr_j(D) = 1 - m_j(\emptyset)$  qui vaut 1 en monde fermé. Les hypothèses  $m_j(\emptyset) = 0$  et  $Cr_j(D) = 1$  initialement imposées par Shafer [146] ne sont pas indispensables.

Cette fonction mesure toute la croyance sur  $A$  *i.e.* l'intensité que les informations fournies par la source  $S_j$  soutiennent la proposition  $A$ . Elle s'interprète comme la croyance minimum en  $A$ .

Il est possible d'avoir  $Cr_j(A) > 0$ , alors que  $m_j(A) = 0$ , ce qui ne signifie donc pas que  $A$  est impossible, mais que nous ne savons pas affecter précisément un degré à  $A$ . Ceci a son importance pour l'étape de modélisation. En effet, nous ne sommes pas obligés d'affecter un degré, si nous ne sommes pas capables de le faire, au lieu de forcer l'information là où nous n'en disposons pas [13].

Nous pouvons de même retrouver la fonction de masse correspondant à la fonction de croyance, pour tout  $A \in 2^D$  :

$$m_j(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Cr_j(A). \quad (6.8)$$

**Fonction de non-croyance** Une *fonction de non-croyance* ou *doute* a été introduite dans [146] et étudiée dans [85], elle est donnée pour tout  $A \in 2^D$  par :

$$dt_j(A) = \sum_{B \cap A = \emptyset} m_j(B) = Cr_j(A^c), \quad (6.9)$$

où  $A^c$  représente le complémentaire de  $A$ .

**Fonction de croyance Bayésienne** Shafer [146] définit les *fonctions de croyances bayésiennes* comme toute fonction de  $2^D$  dans  $[0, 1]$  vérifiant :

1.  $Cr_j(\emptyset) = 0$ ,
2.  $Cr_j(D) = 1$ ,
3.  $Cr_j(A \cup B) = Cr_j(A) + Cr_j(B)$  pour tout  $A, B \in 2^D$  tel que  $A \cap B = \emptyset$ .

Toute fonction de croyance bayésienne est alors une fonction de croyance.

Cette notion de croyance est une notion très étudiée, et d'autres approches plus probabilistes ont été proposées [125].

### 6.1.3 Fonction de plausibilité

La *fonction de plausibilité* fut introduite dans [146] sous une hypothèse de monde fermé comme une probabilité supérieure. Par extension elle est donnée pour tout  $A \in 2^D$  par :

$$Pl_j(A) = \sum_{B \cap A \neq \emptyset} m_j(B) = Cr_j(D) - Cr_j(A^c). \quad (6.10)$$

En monde fermé, nous retrouvons la définition de Shafer :  $Pl_j(A) = 1 - Cr_j(A^c) = 1 - dt_j(A)$ . Elle est également en bijection avec  $m_j$ . Cette fonction mesure l'intensité avec laquelle nous trouvons  $A$ , *i.e.* avec laquelle nous ne doutons pas en  $A$ . Elle s'interprète aussi comme la croyance maximum en  $A$ .

### 6.1.4 Fonction de communalité

La *fonction de communalité* introduite par Dempster [41] et nommée par Shafer [146] est une autre fonction en bijection avec la fonction de masse, qui permet un calcul simple dans l'étape de combinaison (et plus précisément pour la combinaison de Dempster). Elle n'apporte donc pas d'information supplémentaire. Elle a été introduite surtout à des fins calculatoires ; son interprétation reste peu aisée. Elle est donnée, pour une source  $S_j$ , pour tout  $A \in 2^D$  par :

$$Q_j(A) = \sum_{B \in 2^D, B \supseteq A} m_j(B). \quad (6.11)$$

C'est une fonction décroissante qui fournit une autre façon de calculer la fonction de croyance. Shafer [146] propose le théorème suivant :

**Théorème 6.1.2** *Soit  $Cr_j$  une fonction de croyance et  $Q_j$  sa fonction de communalité, alors pour tout  $A \in 2^D$ , nous avons :*

$$Cr_j(A) = \sum_{B \subseteq A^c} (-1)^{|B|} Q_j(B), \quad (6.12)$$

et

$$Q_j(A) = \sum_{B \subseteq A} (-1)^{|B|} Cr_j(B^c). \quad (6.13)$$

### 6.1.5 Fonction d'implicabilité

La *fonction d'implicabilité* est une fonction peu usitée, qui semble avoir été introduite par Smets [157] à des fins calculatoires pour l'étape de combinaison (plus particulièrement pour la combinaison disjonctive). Elle n'a de sens qu'en hypothèse de monde ouvert. Elle est également en parfaite bijection avec la fonction de masse. Pour tout  $A \in 2^D$ , elle est donnée par :

$$b_j(A) = Cr_j(A) + m_j(\emptyset) = \sum_{B \subseteq A} m_j(B). \quad (6.14)$$

### 6.1.6 Propriétés

Nous avons vu que toutes les fonctions précédentes sont toutes déductibles les unes des autres, puisqu'elles sont toutes définies à partir des fonctions de masse, et que les fonctions de masse peuvent être déduites de chacune des fonctions précédentes. Des algorithmes ont été établis afin d'obtenir une fonction à partir d'une autre [88, 158]. Nous donnons ici quelques propriétés intéressantes entre les fonctions de croyance et de plausibilité.

**Proposition 6.1.3** *Pour tout  $A \in 2^D$ , nous avons :*

$$Pl_j(A) \geq Cr_j(A). \quad (6.15)$$

*Sous l'hypothèse d'un monde fermé, nous avons :*

$$Cr_j(A) + Cr_j(A^c) \leq 1, \quad (6.16)$$

$$Pl_j(A) + Pl_j(A^c) \geq 1, \quad (6.17)$$

$$Cr_j(A) + Cr_j(A^c) = 1 \Leftrightarrow Cr_j(A) = Pl_j(A). \quad (6.18)$$

**Preuve** Pour montrer l'équation (6.15), il suffit de constater que :

$$Pl_j(A) \geq Cr_j(A) \Leftrightarrow \sum_{B \cap A \neq \emptyset} m_j(B) \geq \sum_{B \subseteq A, B \neq \emptyset} m_j(B), \quad (6.19)$$

or si  $B \subseteq A$  (*i.e.*  $B \in 2^A$ ) et  $B \neq \emptyset$ , alors  $B \cap A \neq \emptyset$ . Par exemple si  $A = \{C_1, \dots, C_k\}$ ,  $2^A = \{\emptyset, \{C_1\}, \{C_2\}, \{C_1, C_2\}, \dots, A\}$ , et si  $B \in 2^A$  tel que  $B \cap A \neq \emptyset$ , il est alors évident que  $B \cap A \neq \emptyset$ . Par contre la contropposé est fausse. Soit  $B = \{C_1, C_{k+1}\}$ , alors  $B \cap A = \{C_1\}$  et  $B \notin 2^A$ .

Pour l'équation (6.16), comme nous avons  $Pl_j(A) \geq Cr_j(A)$  pour tout  $A \in 2^D$ , cette inégalité est équivalente en appliquant les définitions à :

$$1 - Cr_j(A^c) \geq Cr_j(A) \Leftrightarrow Cr_j(A) + Cr_j(A^c) \leq 1. \quad (6.20)$$

Montrons l'équation (6.17) : d'après l'inégalité (6.16), nous avons :

$$1 - Cr_j(A^c) + 1 - Cr_j(A) \geq 1 \Leftrightarrow Pl_j(A^c) + Pl_j(A) \leq 1. \quad (6.21)$$

Enfin, pour l'équation (6.18), il est clair que :

$$Cr_j(A) + Cr_j(A^c) = 1 \Leftrightarrow Cr_j(A) = 1 - Cr_j(A^c) = Pl_j(A). \quad (6.22)$$

□

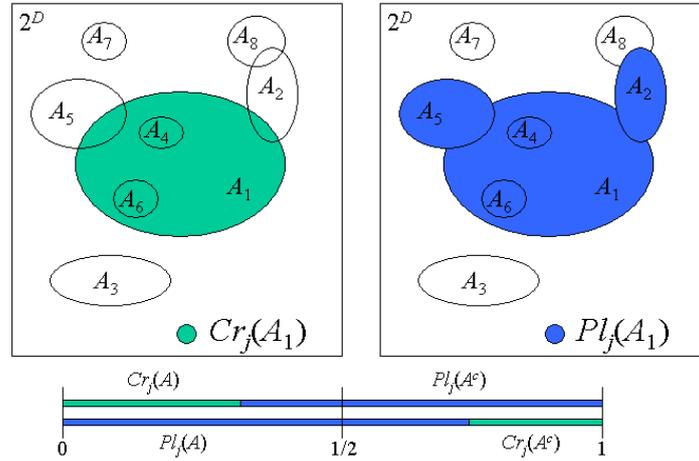


FIG. 6.1 – Représentation des fonctions de croyance et de plausibilité.

La figure 6.1 illustre les fonctions de croyance et de plausibilité. L'intervalle  $[Cr_j(A), Pl_j(A)]$  forme un intervalle de confiance pour la source  $S_j$ . L'ignorance en  $A$  est parfois mesurée par  $Pl_j(A) - Cr_j(A)$ , [82], il faut cependant faire attention à la variété d'ignorance que nous pouvons rencontrer [152].

La relation entre les probabilités et les fonctions de croyance et de plausibilité est fondée sur l'interprétation des probabilités supérieures et inférieures de Dempster. Pour un ensemble  $A$ , la probabilité de  $A$  pour une source  $S_j$ ,  $P_j(A)$  est donnée par [174] :

$$Cr_j(A) \leq P_j(A) \leq Pl_j(A). \quad (6.23)$$

Shafer [146] a montré que cette probabilité est complètement déterminée pour tous les sous-ensembles  $A \in 2^D$ , uniquement pour les fonctions de croyance Bayésienne.

### 6.1.7 Affaiblissement

L'affaiblissement des fonctions de masse, introduite dans [146] permet de modéliser la fiabilité des sources  $S_j$  en introduisant un coefficient  $\alpha_j$ . Ainsi, nous redéfinissons la fonction de masse pour tout  $A \in 2^D$ ,  $A \neq D$  par :

$$\begin{cases} m'_j(A) = \alpha_j m_j(A), \\ m'_j(D) = 1 - \alpha_j (1 - m_j(D)). \end{cases} \quad (6.24)$$

$\alpha_j \in [0, 1]$  est le *coefficient d'affaiblissement* de la source  $S_j$ .

Dans le cas où  $\alpha_j = 0$ , la source  $S_j$  n'est pas du tout fiable, et dans ce cas toute la masse est affectée à  $D$  :  $m_j(D) = 1$ , ce qui représente l'ignorance totale.

L'affaiblissement de la fonction de masse entraîne l'affaiblissement de la fonction de croyance et le renforcement de la fonction de plausibilité. Ainsi, nous augmentons les intervalles  $[Cr_j, Pl_j]$ , et nous réduisons le conflit lors de l'étape de combinaison [13, 6].

### 6.1.8 Conditionnement

Le conditionnement permet de tenir compte d'une information certaine. Par exemple, si nous sommes sûrs que le sous-ensemble  $A$  de  $2^D$  est vérifié, nous pouvons transformer une fonction de masse  $m_j(\cdot)$  quelconque en une fonction de masse  $m_j(\cdot|A)$  dont tous les éléments focaux sont inclus dans  $A$ . Nous définissons ainsi la fonction de masse [6], pour tout  $B \in 2^D$  par :

$$m_j(B|A) = \frac{\sum_{C \cup A = B} m_j(C)}{\sum_{C \cup A \neq \emptyset} m_j(C)}. \quad (6.25)$$

Ainsi  $m_j(B|A) = 0$  si  $A$  n'est pas inclus dans  $B$  et  $m_j(A|A) = 1$ . Ce conditionnement s'écrit également sous une forme plus pratique, à partir de la fonction de plausibilité, pour tout  $B \in 2^D$  par :

$$Pl_j(B|A) = \frac{Pl_j(B \cap A)}{Pl_j(A)}. \quad (6.26)$$

Le déconditionnement, *i.e.* l'opération inverse est plus délicate au niveau de la fonction de masse. Au niveau de la fonction de plausibilité, nous avons simplement la plausibilité de tous les sous-ensembles  $B$  de  $A$  :

$$Pl_j(B) = Pl_j(B|A)Pl_j(A). \quad (6.27)$$

En pratique,  $Pl_j(A)$  est inconnue, ce qui conduit à prendre :

$$\begin{aligned} Pl_j(B) &= Pl_j(B|A), \forall B \subset A, \\ Pl_j(B) &= 1, \forall B \not\subset A. \end{aligned} \quad (6.28)$$

### 6.1.9 Raffinement et grossissement

En pratique, il est possible d'avoir des sources d'information dont les cadres de discernement sont différents, mais cependant compatibles. Par exemple, si le cadre de discernement d'une des sources est une partition de celui d'une autre source. Reprenons l'exemple de [81], dans le cas de défense antiaérienne, un capteur peut être défini sur  $D_1 = \{\text{avion, hélicoptère, missile}\}$ , et un second sur  $D_2 = \{\text{Rafale, Mirage, ..., Puma, Gazelle, ..., Cro-tale, Mistral, ...}\}$ . Il convient donc d'établir des relations entre ces référentiels, fournies par deux opérations le *raffinement* et le *grossissement*.

Considérons deux sources  $S_1$  et  $S_2$ , ayant respectivement pour cadre de discernement  $D_1 = \{d_1^1, \dots, d_{n_1}^1\}$  et  $D_2 = \{d_1^2, \dots, d_{n_2}^2\}$  tel que  $\{R(d_1^1), \dots, R(d_{n_1}^1)\}$  constitue une partition

de  $D_2$ , où  $R$  est une fonction de raffinement. Ainsi, une fonction de masse  $m_1(\cdot)$  définie sur  $D_1$  fournit une fonction de masse  $m_2(\cdot)$  sur  $D_2$  pour tout  $A \in 2^{D_1}$  :

$$m_2(R(A)) = m_1(A). \quad (6.29)$$

Le grossissement est l'opération inverse  $R^{-1}$  d'un raffinement  $R$  [6]. Il faut donc regrouper les singletons de  $D_2$  en sous-ensembles exclusifs  $R(d_i^1)$ , associés aux singletons de  $D_1$ . Ainsi, une fonction de masse  $m_2$  définie sur  $D_2$  fournit une fonction de masse  $m_1$  de  $D_1$  pour tout  $A \in 2^{D_1}$  :

$$m_1(A) = \sum_{B \in 2^{D_2}, A = \{d_i^1 : R(d_i^1) \cap B \neq \emptyset\}} m_2(B). \quad (6.30)$$

Nous pouvons ainsi définir des fonctions de masse à tout niveau d'une hiérarchie, qui est une suite de partitions imbriquées. La fonction  $R$  reste à déterminer, souvent de manière empirique, et entraîne en outre une complexité algorithmique plus importante du fait de l'augmentation du cadre de discernement.

## 6.2 Estimation

Toute la difficulté de la théorie des croyances réside dans l'estimation des fonctions de masse et donc des fonctions de croyances. Plusieurs fonctions de masse ont été proposées dans la littérature et leur choix relève de la modélisation et doit être fait selon les données et l'application recherchée. Nous exposons ici différentes fonctions de masse et les solutions pour leur estimation.

### 6.2.1 Fonction à support simple

Une première fonction de masse très simple présentée dans [146] propose d'affecter toute la masse d'une source  $S_j$  à un sous-ensemble non-vide  $A$  de  $2^D$ . Ainsi, nous avons :

$$\begin{cases} m_j(A) = s \\ m_j(D) = 1 - s \\ m_j(B) = 0, \forall B \in 2^D, B \neq A, B \neq D, \end{cases} \quad (6.31)$$

où  $s \in [0, 1]$ , un réel. Si  $s = 0$ , alors  $m_j(D) = 1$ , ce qui représente l'ignorance totale. La fonction de croyance est alors donnée par :

$$\begin{cases} Cr_j(B) = s, \text{ si } A \subseteq B, \text{ et } B \neq D, \\ Cr_j(D) = 1, \\ Cr_j(B) = 0, \text{ sinon,} \end{cases} \quad (6.32)$$

et la fonction de plausibilité :

$$\begin{cases} Pl_j(B) = 1, \text{ si } A \cap B \neq \emptyset, \\ Pl_j(B) = 1 - s, \text{ si } A \cap B = \emptyset. \end{cases} \quad (6.33)$$

Nous avons donc  $P_j(A) \geq s$  et  $P_j(A^c) \leq 1 - s$  d'après la relation  $Cr_j(A) \leq P_j(A) \leq Pl_j(A)$ . Ces fonctions représentent donc la situation où toute la connaissance porte sur l'évènement  $A$  et sa probabilité est supérieure à  $s$ .

### 6.2.2 Fonction de croyance complémentaire

Yager [174] propose une fonction de masse définie sur un sous-ensemble  $A$  de  $2^D$ , pour une source  $S_j$  par :

$$\begin{cases} m_j(A) = s \\ m_j(A^c) = 1 - s \\ m_j(B) = 0, \forall B \in 2^D, B \neq A, B \neq A^c, \end{cases} \quad (6.34)$$

où  $s \in [0, 1]$ , un réel. Ces fonctions ont la particularité de vérifier  $P_j(A) = s$  et  $P_j(A^c) = 1 - s$ . En effet la fonction de croyance est donnée par :

$$\begin{cases} Cr_j(B) = s, \text{ si } A \subseteq B, \\ Cr_j(B) = 1 - s, \text{ si } A^c \subseteq B, \\ Cr_j(D) = 1, \\ Cr_j(B) = 0, \text{ sinon,} \end{cases} \quad (6.35)$$

et la fonction de plausibilité :

$$\begin{cases} Pl_j(\emptyset) = 0, \\ Pl_j(B) = s, \text{ si } B \subseteq A, \\ Pl_j(B) = 1 - s, \text{ si } B \subseteq A^c, \\ Pl_j(B) = 1, \text{ sinon.} \end{cases} \quad (6.36)$$

Les deux égalités découlent alors de la relation générale  $Cr_j(A) \leq P_j(A) \leq Pl_j(A)$ . Les fonctions de croyance complémentaires correspondent donc à la situation où toute la connaissance porte sur l'évènement  $A$  et sa probabilité de réalisation est  $s$ .

Bien sûr, ces fonctions trop simples ne permettent pas de modéliser la plupart des problèmes, car généralement une source a un avis sur plusieurs sous-ensembles, par exemple sur les différentes décisions envisageables (ou classes dans le cas de la classification).

### 6.2.3 Fonction sur les singletons

Dans le cas où la source  $S_j$  n'a d'opinion que sur les singletons (*i.e.* les décisions qu'elle peut prendre), la fonction de masse peut être définie uniquement sur les singletons. Nous avons donc :

$$\begin{cases} m_j(d_i) = M_i^j \\ m_j(A) = 0, \text{ si } A \notin \{\{d_1\}, \{d_2\}, \dots, \{d_n\}\} \end{cases} \quad (6.37)$$

Ainsi pour une observation  $x$ , la fonction de masse  $m_j(d_i)(x) = M_i^j(x)$ .  $M_i^j(x)$  peut ensuite être estimée par une probabilité conditionnelle  $p(f_j(x)|d_i)$ , comme par exemple dans [106]

et [3]. Nous retrouverons bien l'hypothèse de l'équation (6.1) :

$$\sum_{A \in 2^D} m_j(A) = 1, \text{ dans le cas des probabilités.}$$

Cette approche est cependant réductrice, et n'utilise pas la puissance des fonctions de croyance. En fait, nous nous ramenons à une approche de type bayésienne. Cependant, il peut être intéressant de partir de ces fonctions pour répartir ensuite la masse sur les hypothèses composées. Cette répartition se fait souvent de manière heuristique, mais peut être envisagée en se fondant sur différents modèles.

### 6.2.4 Modèles probabilistes

Les modèles probabilistes reposent sur l'estimation de  $p(f_j(x)|d_i)$ , que nous notons  $p(S_j|d_i)$ . Plusieurs modèles ont été proposés.

Appriou [4, 5, 6] propose deux modèles dont le principe est d'étendre le modèle sur les singletons à l'aide des fonctions simples, en intégrant un coefficient de fiabilité (affaiblissement) de la source pour une classe donnée  $\alpha_{ij}$ . Ces deux modèles reposent sur deux axiomes simples :

Axiome 1- Les  $n.m$  couples  $[M_i^j, \alpha_{ij}]$  sont des sources d'information distinctes dont les éléments focaux sont  $\{d_i\}$ ,  $\{d_i^c\}$  et  $D$ .

Axiome 2- Si  $M_i^j = 0$  et  $M_i^j$  est valide (*i.e.*  $\alpha_{ij} = 1$ ), alors il est certain que  $d_i$  n'est pas vérifiée.

L'axiome 1 implique le calcul de  $n * m$  fonctions de masse  $m_{ji}$  à partir des  $n * m$  couples  $[M_i^j, \alpha_{ij}]$ , au lieu des  $m$  fonctions de masse  $m_j$  initiales. Pour chacune d'elles la masse allouée aux seuls éléments focaux possibles  $\{d_i\}$ ,  $\{d_i^c\}$  et  $D$ , est définie à partir de la vraisemblance  $M_{ij}$  correspondante.  $M_{ij}$  représente donc la vraisemblance de  $d_i$  et de toute les décisions qui lui sont proches compte tenu de la source  $S_j$ , tandis que  $1 - M_{ij}$  représente la vraisemblance de  $d_i^c$ . L'axiome 2 limite les fonctions de masse à deux types, ces deux seuls modèles sont :

– Modèle 1 :

$$\begin{cases} m_{ji}(d_i)(x) = M_{ij} \\ m_{ji}(d_i^c)(x) = 1 - M_{ij} \\ m_{ij}(D)(x) = 0 \end{cases} \quad (6.38)$$

– Modèle 2 :

$$\begin{cases} m_{ij}(d_i)(x) = 0 \\ m_{ji}(d_i^c)(x) = 1 - M_{ij} \\ m_{ji}(D)(x) = M_{ij} \end{cases} \quad (6.39)$$

Si nous considérons le coefficient d'affaiblissement  $\alpha_{ij}$ , les deux modèles deviennent :

– Modèle 1 :

$$\begin{cases} m_{ji}(d_i)(x) = \alpha_{ij} M_{ij} \\ m_{ji}(d_i^c)(x) = \alpha_{ij} (1 - M_{ij}) \\ m_{ij}(D)(x) = 1 - \alpha_{ij} \end{cases} \quad (6.40)$$

– Modèle 2 :

$$\begin{cases} m_{ij}(d_i)(x) = 0 \\ m_{ji}(d_i^c)(x) = \alpha_{ij}(1 - M_{ij}) \\ m_{ji}(D)(x) = 1 - \alpha_{ij}(1 - M_{ij}) \end{cases} \quad (6.41)$$

Dans les 2 modèles si  $\alpha_{ij} = 0$ , nous retrouvons le cas de l'ignorance totale pour la source  $S_j$  pour la décision  $d_i$  ( $m_{ji}(D)(x) = 1$ ).

La difficulté réside dans l'estimation de  $M_{ij}$ . Pour ce faire Appriou [6] ajoute un troisième axiome qui repose sur la connaissance de l'étape de combinaison. Cet axiome garantit la conformité avec l'approche bayésienne, dans le cas où celle-ci est optimale, *i.e.* si la réalité est parfaitement connue.

Axiome 3- Si  $p(S_j|d_i)$  représente parfaitement la réalité ( $\alpha_{ij} = 1, \forall i, j$ ), et si toutes les probabilités *a priori*  $p(d_i)$  sont disponibles, alors la fonction de masse obtenue par la règle de combinaison de Dempster doit être la probabilité obtenue par l'approche bayésienne.

Si les sources  $S_j$  sont indépendantes, les résultats doivent être les mêmes si nous calculons individuellement  $p(S_j|d_i)$  ou directement la probabilité conjointe  $p(S_1, \dots, S_m|d_i) = \prod_{j=1}^m p(S_j|d_i)$ . Nous obtenons ainsi deux estimations, selon le modèle,

de  $M_{ij}$  :

– Modèle 1 :

$$M_{ij} = \frac{R_j p(S_j|d_i)}{1 + R_j p(S_j|d_i)}, \quad (6.42)$$

avec  $R_j$  un facteur de normalisation tel que  $R_j \geq 0$ .

– Modèle 2 :

$$M_{ij} = R_j p(S_j|d_i), \quad (6.43)$$

avec  $R_j \in [0, (\max_{S_j, i} p(S_j|d_i))^{-1}]$ .

Nous obtenons ainsi les fonctions de masse correspondant aux deux modèles :

– Modèle 1 :

$$\begin{cases} m_{ji}(d_i)(x) = \frac{\alpha_{ij} R_j p(S_j|d_i)}{1 + R_j p(S_j|d_i)} \\ m_{ji}(d_i^c)(x) = \frac{\alpha_{ij} R_j}{1 + R_j p(S_j|d_i)} \\ m_{ji}(D)(x) = 1 - \alpha_{ij} \end{cases} \quad (6.44)$$

avec  $R_j \geq 0$ .

– Modèle 2 :

$$\begin{cases} m_{ji}(d_i)(x) = 0 \\ m_{ji}(d_i^c)(x) = \alpha_{ij}(1 - R_j p(S_j|d_i)) \\ m_{ji}(D)(x) = 1 - \alpha_{ij}(1 - R_j p(S_j|d_i)) \end{cases} \quad (6.45)$$

avec  $R_j \in \left[0, (\max_{S_j, i} p(S_j|d_i))^{-1}\right]$ .

Appriou [6] propose différentes estimations de  $\alpha_{ij}$  et des probabilités  $p(S_j|d_i)$  (stochastiques ou floues) selon le type de mesure de  $f_j(x)$  (exacte, incertaine, imprécise). Cette approche est bien adaptée dans les cas où il est possible de considérer une décision contre toutes les autres.

Dans [52], les fonctions sur les singletons sont étendues en comparant les probabilités  $p(S_j|d_i)$ . Les disjonctions sont définies en fonction d'un critère de significativité des probabilités conditionnelles. Toute la difficulté porte donc sur les seuils de significativité. Si aucune probabilité n'est trouvée significative, la masse totale porte sur  $D$  ( $m_j(D) = 1$ , cas de l'ignorance totale). Dans le cas où une seule probabilité  $p(S_j|d_i)$  est trouvée significative, alors ce modèle est équivalent au modèle des singletons ce qui est équivalent à l'approche bayésienne. Sinon, il faut prendre en compte les disjonctions. Par exemple [13], si trois probabilités conditionnelles sont significatives, telles que  $p(S_j|d_{i1}) > p(S_j|d_{i2}) > p(S_j|d_{i3})$ , la fonction de masse est donnée par :

$$\begin{cases} m_j(d_{i1})(x) = p(S_j|d_{i1}) - p(S_j|d_{i2}) \\ m_j(\{d_{i1}, d_{i2}\})(x) = p(S_j|d_{i2}) - p(S_j|d_{i3}) \\ m_j(\{d_{i1}, d_{i2}, d_{i3}\})(x) = p(S_j|d_{i3}) \\ m_j(D)(x) = 1 - p(S_j|d_{i1}) \end{cases} \quad (6.46)$$

Ces masses sont ensuite normalisées. Il est aussi possible de les affaiblir afin de tenir compte de la fiabilité des sources  $S_j$ .

### 6.2.5 Modèles de distances

Dencœur a proposé une autre approche issue de la reconnaissance de formes [42, 43] qui ne fait plus appel aux probabilités. Les fonctions de masse et fonctions de croyance sont ici construites uniquement à partir de la représentation des vecteurs d'apprentissage. Dans cette approche, seuls  $\{d_i\}$  et  $D$  sont considérés comme éléments focaux.

#### Méthode des prototypes

Une première version consiste à considérer chaque groupe (classe)  $d_i$  par un prototype (*i.e.* son centre  $\mathbf{x}_i$  estimé sur les données d'apprentissage). Nous considérons ainsi l'ensemble  $\{(x^{(1)}, d^{(1)}), \dots, (x^{(Q)}, d^{(Q)})\}$  des  $Q$  vecteurs  $x^{(t)}$ ,  $t = 1, \dots, Q$  et la décision qui est associée  $d^{(t)}$  à valeur dans le cadre de discernement  $D = \{d_1, \dots, d_n\}$ . Nous considérons ici chaque paire  $(x^{(t)}, d^{(t)})$  comme source d'information pour le vecteur à classer  $x$ . Les fonctions de masse sont alors définies à partir de ces centroïdes :

$$\begin{cases} m_j(d_i|\mathbf{x}_i)(x) = \alpha_{ij}\varphi_i(d(x, \mathbf{x}_i)) \\ m_j(D|\mathbf{x}_i)(x) = 1 - \alpha_{ij}\varphi_i(d(x, \mathbf{x}_i)) \end{cases} \quad (6.47)$$

où  $\alpha_{ij} \in [0, 1]$  est un coefficient d'affaiblissement,  $d$  est une distance à déterminer en fonction des données (cf. [110]), et  $\varphi_i$  est une fonction décroissante vérifiant :

$$\begin{cases} \varphi_i(0) = 1, \\ \lim_{d \rightarrow +\infty} \varphi_i(d) = 0. \end{cases} \quad (6.48)$$

Il existe un grand nombre de fonctions  $\varphi_i$  vérifiant ces égalités, sans qu'il y ait une méthode pour le choix de ces fonctions. Dans le cas d'une distance euclidienne, Dencœur [43] propose la fonction :

$$\varphi_i(d) = \exp(-\gamma_i d^2), \quad (6.49)$$

où  $\gamma_i > 0$  est un paramètre associé à la décision  $d_i$ .  $\gamma_i$  peut être initialisé comme l'inverse de la distance moyenne entre les vecteurs d'apprentissage vérifiant  $d_i$  [42]. Une méthode d'optimisation de  $\gamma_i$  est proposée dans [187] et [44] obtenue par minimisation du critère d'erreur :

$$E = \sum_{n=1}^N \sum_{t=1}^Q (p^{(t)}(d_i) - u_i^{(t)})^2, \quad (6.50)$$

où  $u_i^{(t)}$  est l'indicateur d'appartenance du vecteur d'apprentissage  $x^{(t)}$  au groupe  $d_i$ , (i.e.  $u_i^{(t)} = 1$  si la décision de  $x^{(t)}$  est  $d_i$  et 0 sinon).  $p^{(t)}(d_i)$  est la probabilité pignistique introduite par Smets [151] présentée à la section 6.4, de  $d_i$  pour le vecteur  $x^{(t)}$ . En répétant  $Q$  fois l'opération, nous obtenons  $Q$  fonctions de masse.

Si un vecteur d'apprentissage  $x^{(t)}$  d'étiquette  $d^{(t)} = d_i$  est proche de  $x$  dans l'espace de représentation selon la métrique  $d$ , alors une grande partie de la croyance de  $x$  sera affectée à  $d_i$ , et le reste à l'ensemble des décisions du cadre de discernement. En revanche, si  $d(x, \mathbf{x}_i)$  est grande,  $x$  a peu de chance d'appartenir au groupe  $d_i$ , ce qui pousse vers une situation d'ignorance avec  $m_j(D)$  qui sera grande.

### Méthode des $k$ plus proches voisins

Un autre méthode consiste à non plus considérer les proximités aux centroïdes  $\mathbf{x}_i$ , mais à l'ensemble des vecteurs d'apprentissage. En pratique, il est possible de négliger l'influence des vecteurs  $x^{(t)}$  qui sont éloignés du vecteur  $x$  considéré, en ne considérant que les  $k$  vecteurs d'apprentissage  $x^{(t,k)}$  les plus proches de  $x$ . Les fonctions de masse sont alors données par :

$$\begin{cases} m_j(d_i | x^{(t,k)})(x) = \alpha_{ij} \varphi_i(d(x, x^{(t,k)})) \\ m_j(D | x^{(t,k)})(x) = 1 - \alpha_{ij} \varphi_i(d(x, x^{(t,k)})) \end{cases} \quad (6.51)$$

où  $d_i$  est la décision associée à  $x^{(t,k)}$  (i.e.  $d^{(t,k)} = d_i$ ).

### 6.2.6 Autres approches

D'autres approches ont été proposées pour l'estimation des fonctions de masse. Rombaut [136, 137] propose une approche pour les cas où  $x$  est dans un espace à une dimension, facilement représentable par un histogramme. Les masses sur les hypothèses composées sont définies dans les zones de recouvrement et chaque pic correspond à un singleton. Toute la difficulté réside dans la détermination des seuils de l'histogramme, mais cette approche conduit à des résultats intéressants en segmentation d'images médicales [162, 137] ou d'images sonar [115].

## 6.3 Combinaison

Plusieurs modes de combinaison ont été développés dans le cadre de la théorie des croyances. Il existe principalement deux types de combinaison : la combinaison conjonctive et la combinaison disjonctive, qui ont été déclinés en un grand nombre d'opérateurs de combinaison dont des combinaisons mixtes.

### 6.3.1 Combinaison conjonctive

L'approche initiale est celle introduite par Dempster [41] et reprise par Shafer [146], elle combine les fonctions de masse en considérant les intersections des éléments de  $2^D$ .

#### Règle orthogonale de Dempster-Shafer

La règle (ou somme) orthogonale de Dempster-Shafer permet de combiner deux fonctions de masse ou plus en une seule. Elle est donnée pour tout  $A \in 2^D$  par :

$$m(A) = (m_1 \oplus m_2)(A) = \sum_{B \cap C = A} m_1(B)m_2(C). \quad (6.52)$$

Pour les  $m$  fonctions de masse elle s'écrit :

$$m(A) = (m_1 \oplus m_2 \oplus \dots \oplus m_m)(A) = \sum_{B_1 \cap \dots \cap B_m = A} \prod_{j=1}^m m_j(B_j), \quad (6.53)$$

et dans le cas de  $m.n$  fonctions de masse  $m_{ji}$  :

$$m(A) = \sum_{B_1 \cap \dots \cap B_m = A} \prod_{i=1}^n \prod_{j=1}^m m_{ji}(B_j). \quad (6.54)$$

Cette règle permet ainsi de synthétiser toutes les fonctions de masse définies sur  $2^D$ . Elle s'écrit plus simplement à l'aide des fonctions de communalité par :

$$m(A) = \prod_{j=1}^m Q_j(A). \quad (6.55)$$

	$m_1(A_1)$	...	$m_1(A_i)$	...	$m_1(A_{m_1})$
$m_2(B_1)$					
$\vdots$					
$m_2(B_j)$			$m_1(A_i)m_2(B_j)$ assignée à $A_i \cap B_j$		
$\vdots$					
$m_2(B_{m_2})$					

TAB. 6.1 – Illustration de la règle orthogonale pour deux sources distinctes.

**Illustration** Prenons un exemple de deux sources distinctes  $S_1$  et  $S_2$  ayant pour fonction de masse  $m_1$  et  $m_2$  définies respectivement sur les éléments focaux  $A_1, \dots, A_{m_1}$  et  $B_1, \dots, B_{m_2}$  respectivement. La règle orthogonale de ces deux sources peut être illustrée par le tableau 6.1.

Sous cette forme de la règle de Dempster-Shafer, l'ensemble vide peut présenter une masse non nulle, ce qui en monde ouvert représente une solution non envisagée par le cadre de discernement, et ce qui en monde fermé est inacceptable.

Ainsi, si nous supposons être en monde fermé, il est important de forcer la masse de l'ensemble vide à 0 ce qui pousse à normaliser les autres masses. Nous obtenons ainsi la forme normalisée de la fonction de masse combinée, pour tout  $A \in 2^D$  :

$$\begin{cases} m(A) = \frac{1}{1-k} \sum_{B_1 \cap \dots \cap B_m = A} \prod_{j=1}^m m_j(B_j), & \text{si } A \neq \emptyset, \\ m(\emptyset) = 0, \end{cases} \quad (6.56)$$

où  $k = \sum_{B_1 \cap \dots \cap B_m = \emptyset} \prod_{j=1}^m m_j(B_j) < 1$ .  $k$  est une *mesure de conflit* entre les sources encore appelée *inconsistance de la fusion*. Elle dépend bien sûr de la modélisation retenue et de l'estimation des fonctions de masse.

Cette règle de combinaison est donc intéressante sous l'hypothèse de monde fermé et si il n'y a pas de conflit entre les sources. En effet, le fait de normaliser masque le conflit. Prenons l'exemple cité dans [13], où il n'y a que deux fonctions de masse définies sur  $D = \{d_1, d_2, d_3\}$  où seuls les singletons sont des éléments focaux. Les valeurs de ces fonctions de masse sur les éléments focaux sont données par le tableau 6.2 et représentent les valeurs de deux sources en conflit important, l'une prenant plutôt la décision  $d_1$  tandis que l'autre décide  $d_2$ .

Leur fusion normalisée et non normalisée simple à obtenir est donnée par le tableau 6.3.

Dans le cas de la forme normalisée de la règle de Dempster-Shafer toute la masse est focalisée sur la décision  $d_3$ , qui est la seule où les deux sources sont d'accord pour

	$d_1$	$d_2$	$d_3$
$m_1$	0.9	0.0	0.1
$m_2$	0.0	0.9	0.1

TAB. 6.2 – Exemple de deux fonctions de masse issues de deux sources en conflit.

	$d_1$	$d_2$	$d_3$	$\emptyset$
$m_1 \oplus m_2$ non normalisée	0.0	0.0	0.01	0.99
$m_1 \oplus m_2$ normalisée	0.0	0.0	1	0.0

TAB. 6.3 – Fusion de deux fonctions de masse issues de deux sources en conflit.

qu'elle ne soit que faiblement plausible. La forme non normalisée permet d'affecter la plus grande partie de la masse à l'ensemble vide. Ainsi l'hypothèse d'un monde ouvert permet d'affecter la masse de conflit à l'ensemble vide. Nous avons vu précédemment au chapitre 3 page 17, que le conflit peut provenir de l'hypothèse d'un monde ouvert ou d'une faible fiabilité d'au moins une des deux sources.

Ainsi, pour la forme normalisée, la fonction de masse obtenue par combinaison a tendance à renforcer la croyance sur les décisions pour lesquelles les sources sont concordantes, et à l'atténuer en cas de conflit.

### Règle de Smets

Historiquement la forme normalisée est celle proposée par Dempster [41] et Shafer [146], tandis que la forme non normalisée a été proposée plus tard par Smets [150], notamment pour aborder les mondes ouverts. Cette dernière est aujourd'hui la plus utilisée, pour les raisons présentées précédemment. Ainsi dans son modèle des fonctions de croyances transférables (*Transferable Belief Functions*), Smets propose donc la combinaison :

$$\begin{cases} m(A) = \sum_{B_1 \cap \dots \cap B_m = A \neq \emptyset} \prod_{j=1}^m m_j(B_j), \\ m(\emptyset) = \sum_{B_1 \cap \dots \cap B_m = \emptyset} \prod_{j=1}^m m_j(B_j). \end{cases} \quad (6.57)$$

Notons que dans ce cas la règle s'écrit simplement si les fonctions de masse sont modélisées par les modèles de distance de Dencœur [42]. Nous avons ainsi :

$$\begin{cases} m(d_i | \mathbf{x}_i)(x) = \frac{1}{L} \left( 1 - \prod_{t \in I_{k,i}} (1 - \alpha_{ij} \varphi_i(d(x, \mathbf{x}_i))) \right) \prod_{r \neq i} \prod_{t \in I_{k,r}} (1 - \alpha_{rj} \varphi_r(d(x, \mathbf{x}_r))) \\ m(D | \mathbf{x})(x) = \frac{1}{L} \prod_{r=1}^m \prod_{t \in I_{k,r}} (1 - \alpha_{rj} \varphi_r(d(x, \mathbf{x}_r))) \end{cases} \quad (6.58)$$

où  $L$  est une constante de normalisation, et  $I_{k,i}$  l'ensemble des  $k$  voisins de  $x$  appartenant à la décision  $d_i$ .

D'autres méthodes de normalisation ont cependant été proposées.

### Règle de Yager

Dans [174] Yager propose un modèle en monde fermé où la mesure de conflit est affectée au cadre de discernement totale  $D$ . Le conflit est ainsi transformé en ignorance. Nous obtenons pour tout  $A \in 2^D$  tel que  $A \neq \emptyset$  :

$$m(A) = \sum_{B_1 \cap \dots \cap B_m = A \neq \emptyset} \prod_{j=1}^m m_j(B_j), \quad (6.59)$$

ainsi la masse de  $D$  est donnée par :

$$m(D) = 1 - \sum_{A \in 2^D} \sum_{B_1 \cap \dots \cap B_m = A \neq \emptyset} \prod_{j=1}^m m_j(B_j). \quad (6.60)$$

Comme la masse sur l'ensemble vide est forcée à 0 ( $m_j(\emptyset) = 0$ ), pour conserver l'hypothèse de l'équation (6.1) sur les fonctions de masse, la fonction de masse sur  $D$  devient :

$$m(D) = m(D) + \sum_{B_1 \cap \dots \cap B_m = \emptyset} \prod_{j=1}^m m_j(B_j). \quad (6.61)$$

### Technique du *hedging*

La technique du *hedging* présentée dans [6] est une autre approche toujours dans le cas d'un monde fermé (*i.e.*  $m_j(\emptyset) = 0$ ). Cette technique consiste à affecter la mesure de conflit (*i.e.* la masse de l'ensemble vide) à un nouvel élément  $e$  inconnu qui est donc ajouté au cadre de discernement. Nous obtenons ainsi :

$$\begin{cases} m(A \cup e) = \sum_{B_1 \cap \dots \cap B_m = A \neq \emptyset} \prod_{j=1}^m m_j(B_j), \\ m(e) = \sum_{B_1 \cap \dots \cap B_m = \emptyset} \prod_{j=1}^m m_j(B_j). \end{cases} \quad (6.62)$$

### Propriétés

Notons tout d'abord que l'opérateur de combinaison (normalisée ou non) est commutatif et associatif du fait de la commutativité et de l'associativité de l'addition et de la multiplication. L'opérateur de Yager n'est lui que commutatif.

L'élément neutre de la combinaison est donné par la fonction de masse :

$$\begin{cases} m_0(D) = 1, \\ m_0(A) = 0, \forall A \in 2^D, A \neq D. \end{cases} \quad (6.63)$$

	$d_1$	$d_2$	$d_3$
$m_1$	0.7	0.2	0.1
$m_2$	0.7	0.2	0.1

TAB. 6.4 – Exemple de deux fonctions de masses identiques.

	$d_1$	$d_2$	$d_3$	$\emptyset$
$m_1 \oplus m_2$ non normalisée	0.49	0.04	0.01	0.46
$m_1 \oplus m_2$ normalisée	0.91	0.07	0.02	0.0

TAB. 6.5 – Exemple de la fusion de deux fonctions de masses identiques.

La fonction de masse élément neutre représente donc l'ignorance complète d'une source. Ainsi, si une source n'apporte aucune information, elle ne joue aucun rôle dans la combinaison.

La loi  $\oplus$  n'est pas idempotente (*i.e.*  $m_1 \oplus m_1 \neq m_1$ ). En effet, reprenons l'exemple de [13], avec deux fonctions de masse  $m_1$  et  $m_2$  identiques données par le tableau 6.4. Leur fusion normalisée et non normalisée est donnée par le tableau 6.5. Par cet exemple, nous constatons que l'opérateur de combinaison n'est pas idempotent. De plus, les masses fortes sont renforcées, tandis que les masses faibles sont diminuées. Notons de plus que dans le cas non normalisé  $m_1 \oplus m_2(\emptyset) = 0.46$ , ce qui signifie que le conflit de deux sources identiques est non nul, et qu'il est d'autant plus fort que la masse est répartie sur les décisions (les singletons).

Initialement, cette règle a été énoncée sous l'hypothèse d'indépendance des sources. Il a été montré que cette règle ne nécessite pas cette hypothèse trop forte. Il suffit que les sources vérifient l'hypothèse d'indépendance cognitive (*cf.* section 3.3 page 14).

Notons encore que cette règle permet de réduire l'imprécision et l'ambiguïté, ce qui est, entre autre, initialement recherché par la fusion d'informations. En effet, considérons l'exemple de deux fonctions de masse  $m_1$  et  $m_2$  dont les éléments focaux sont  $d_1 \cup d_2$  et  $d_3$  pour  $m_1$  et  $d_1$  et  $d_2 \cup d_3$  pour  $m_2$ . Les éléments focaux de  $m_1 \oplus m_2$  sont donnés par la table 6.6 d'intersection. Nous remarquons que les éléments focaux de  $m_1 \oplus m_2$  ne sont plus que des singletons et l'ensemble vide. Par cet exemple, nous montrons comment la combinaison conjonctive réduit l'imprécision et réduit l'ambiguïté de chaque source (ici disparue). Du fait que les masses portent sur des ensembles plus petits (elles sont focalisées) que les

	$d_1 \cup d_2$	$d_3$
$d_1$	$d_1$	$\emptyset$
$d_2 \cup d_3$	$d_2$	$d_3$

TAB. 6.6 – Exemple de table d'intersection.

	$d_1 \cup d_2$	$d_3$	$D$
$d_1$	$d_1$	$\emptyset$	$d_1$
$d_2 \cup d_3$	$d_2$	$d_3$	$d_2 \cup d_3$
$D$	$d_1 \cup d_2$	$d_3$	$D$

TAB. 6.7 – Exemple de table d’intersection tenant compte de l’ignorance initiale.

masses initiales, la combinaison conjonctive permet de réduire les intervalles  $[Cr, Pl]$ . Cependant la mesure de conflit (la masse sur l’ensemble vide) augmente.

Reprenons cet exemple en tenant compte de l’ignorance des deux sources (*i.e.* qu’une masse non nulle est affectée à  $D$ ). La table 6.7 d’intersection montre alors que l’ambiguïté n’est que partiellement réduit (il reste des éléments focaux qui ne sont pas des singletons). Le conflit (*i.e.* la masse sur l’ensemble vide) va cependant diminuer, ce qui est une conséquence logique : l’introduction d’une masse sur  $D$  diminue l’ensemble des masses sur les éléments focaux.

Dans [31], nous pouvons trouver une interprétation géométrique des fonctions de croyance et de la règle de Dempster. Les fonctions de masse sont considérées comme des points d’un espace  $[0, 1]^{|2^D|-1}$ , et la combinaison peut être vue en tant que construction géométrique. L’interprétation géométrique des fonctions de croyance donnée dans [87] permet de définir une distance sur l’ensemble des éléments focaux.

Lorsque les seuls éléments focaux sont les décisions (*i.e.* les singletons), les fonctions de masse, croyance et plausibilité sont alors des probabilités. La règle orthogonale de Dempster-Shafer est parfaitement cohérente avec les lois des probabilités, ce qui place la théorie des probabilités comme limite de la théorie des croyances lorsqu’il n’y a ni ambiguïté, ni imprécision et que seul l’incertitude est à considérer.

## Conditionnement

Nous avons présenté à la page 40, l’intérêt du conditionnement pour prendre en compte une information certaine. Le conditionnement s’écrit simplement comme la combinaison de  $m_j$  et d’une fonction de masse  $m_A$  exprimant la certitude en  $A$  (*i.e.*  $m_A(A) = 1$ ) [6], ainsi pour tout  $B \in 2^D$  :

$$m_j(B|A) = m_j \oplus m_A(B) = \sum_{B=A \cap C} m_j(C), \quad (6.64)$$

qui s’écrit aussi [13], pour tout  $B \in 2^D$  :

$$\begin{cases} m_j \oplus m_A(B) = 0, \forall B \notin 2^A, \\ m_j \oplus m_A(B) = \sum_{X \in A^c} m_j(X \cup A). \end{cases} \quad (6.65)$$

### Fonctions à support séparable

Introduites par [146] comme une extension des fonctions à support simple, les *fonctions à support séparable* sont un cas particulier des fonctions de masse.

Si nous considérons deux fonctions à support simple  $m_1$  et  $m_2$  de même support  $A$ , avec des poids  $s_1$  et  $s_2$  de  $[0, 1]$ , *i.e.* :

$$\begin{cases} m_1(A) = s_1 \\ m_1(D) = 1 - s_1 \\ m_1(B) = 0, \forall B \in 2^D, B \neq A, B \neq D, \end{cases} \quad (6.66)$$

et

$$\begin{cases} m_2(A) = s_2 \\ m_2(D) = 1 - s_2 \\ m_2(B) = 0, \forall B \in 2^D, B \neq A, B \neq D. \end{cases} \quad (6.67)$$

Dans ce cas la combinaison des fonctions de masse donne :

$$\begin{cases} m_1 \oplus m_2(A) = s_1 + s_2 - s_1 s_2 \\ m_1 \oplus m_2(D) = 1 - s_1 + s_2 - s_1 s_2 \\ m_1 \oplus m_2(B) = 0, \forall B \in 2^D, B \neq A, B \neq D, \end{cases} \quad (6.68)$$

ce qui correspond à la combinaison de Bernoulli [146]. Ainsi la combinaison de fonctions de masse à support simple identique est une fonction de masse à support simple de même support. Les fonctions de masse à support simple identique ne sont donc jamais conflictuelles.

Considérons à présent deux fonctions à support simple  $m_1$  et  $m_2$  de support  $A_1$  et  $A_2$  respectivement, avec des poids  $s_1$  et  $s_2$  de  $[0, 1]$ . La combinaison de  $m_1$  et  $m_2$  donne :

$$\begin{cases} m_1 \oplus m_2(A_1 \cap A_2) = s_1 s_2, \\ m_1 \oplus m_2(A_1) = s_1 (1 - s_2), \\ m_1 \oplus m_2(A_2) = s_2 (1 - s_1), \\ m_1 \oplus m_2(D) = (1 - s_1) (1 - s_2), \\ m_1 \oplus m_2(B) = 0, \forall B \in 2^D, B \neq A_1, B \neq A_2, B \neq A_1 \cap A_2, B \neq D. \end{cases} \quad (6.69)$$

Si  $A_1 \cap A_2 = \emptyset$ , alors les deux fonctions sont conflictuelles, et une masse non nulle ( $s_1 s_2$ ) est affectée à l'ensemble vide. Lorsque  $s_1 s_2 < 1$ , la fonction de masse issue de la combinaison est appelée *fonction à support séparable*. De telles fonctions sont appropriées lorsque la croyance peut être décomposée en composantes homogènes selon les singletons.

### 6.3.2 Combinaison disjonctive

D'autres approches de combinaison ont été proposée telle que la combinaison disjonctive [153]. La combinaison disjonctive est donnée non plus en considérant les intersections, mais les unions. Les éléments focaux de la combinaison s'obtiennent alors par des tables

	$d_1 \cup d_2$	$d_3$
$d_1$	$d_1 \cup d_2$	$d_1 \cup d_3$
$d_2 \cup d_3$	$D$	$d_2 \cup d_3$

TAB. 6.8 – Exemple de table d’union.

d’union. La combinaison de  $m$  fonctions de masse  $m_j$  est donnée donc pour tout  $A \in 2^D$  par :

$$m(A) = \sum_{B_1 \cup B_2 \cup \dots \cup B_m = A} \prod_{j=1}^m m_j(B_j). \quad (6.70)$$

Cette combinaison s’écrit simplement avec les fonctions d’implicabilité. En effet, pour tout  $A \in 2^D$  nous avons :

$$m(A) = \prod_{j=1}^m b_j(A). \quad (6.71)$$

Il est clair qu’alors  $m(\emptyset) = 0$ , par cette combinaison le conflit ne peut donc pas apparaître. En contrepartie, les éléments focaux de la fonction de masse résultante sont élargis ce qui pousse à une perte de spécificité. Si nous reprenons l’exemple précédent, la table d’union est donnée par le tableau 6.8. Cette combinaison est peu ou pas employée car ce qui est recherché dans la plupart des applications est une fonction de masse plus focalisée ou au moins autant que les fonctions de masse fournies par les sources. Une prise de décision sur les singletons ne sera possible que si les sources sont en accord sur ces singletons. Cette approche est cependant intéressante si nous ne savons pas modéliser les fiabilités des sources, leurs ambiguïtés et imprécisions.

### 6.3.3 Combinaison mixte

Pour chercher à conserver les avantages de la combinaison conjonctive et disjonctive, Dubois et Prade [57] ont proposé un compromis : une combinaison mixte. Cette combinaison est donnée pour tout  $A \in 2^D$ ,  $A \neq \emptyset$ , par :

$$m(A) = \sum_{B_1 \cap B_2 \cap \dots \cap B_m = A} \prod_{j=1}^m m_j(B_j) + \sum_{\substack{B_1 \cup B_2 \cup \dots \cup B_m = A \\ B_1 \cap B_2 \cap \dots \cap B_m = \emptyset}} \prod_{j=1}^m m_j(B_j). \quad (6.72)$$

Au même titre que la combinaison de Yager, ce modèle suppose que le conflit provient de la non fiabilité des sources et peut donc être modélisé par la fiabilité. Cette combinaison est donc un compromis raisonnable entre la précision et la fiabilité.

De nombreuses approches mixtes ont été proposées, telles que les  $\alpha$ -jonctions de Smets [154].

### 6.3.4 Autres Combinaisons

Bien d'autres approches de combinaisons des fonctions de masse ou fonctions de croyance ont été étudiées. Citons l'approche de [85] qui a introduit un opérateur de consensus sur l'ensemble des fonctions de croyance, de non-croyance, d'incertitude et sur le réel de relative atomicité. La fonction d'incertitude est définie pour tout  $A \in 2^D$  par :

$$u_j(A) = \sum_{\substack{B \cap A \neq \emptyset \\ B \in 2^D, B \notin 2^A}} m_j(B), \quad (6.73)$$

ce qui donne  $Cr_j(A) + dt_j(A) + u_j(A) = 1, \forall A \in 2^D, A \neq \emptyset$ . D'autres mesures d'incertitude ont été proposées par la littérature dans le cadre de la théorie des croyances, par exemple dans [90]. Le réel de relative atomicité de  $A$  par rapport à  $B$  ( $A, B \in 2^D$ ) est donné par :

$$a_j(A|B) = \frac{|A \cap B|}{|B|}, \quad (6.74)$$

où  $|B|$  représente le cardinal de  $B$ . Nous notons  $a_j(A) = a_j(A|D)$ . L'opérateur de consensus est définie pour deux sources  $S_1$  et  $S_2$  décrites par les quadruplets  $(Cr_1, dt_1, u_1, a_1)$  et  $(Cr_2, dt_2, u_2, a_2)$  par les relations :

$$\left\{ \begin{array}{l} Cr_{1,2}(A, B) = \frac{1}{k}(Cr_1(A)u_2(B) + Cr_2(B)u_1(A)), \\ dt_{1,2}(A, B) = \frac{1}{k}(dt_1(A)u_2(B) + dt_2(B)u_1(A)), \\ u_{1,2}(A, B) = \frac{1}{k}u_1(A)u_2(B), \\ a_{1,2}(A, B) = \frac{a_1(A)u_2(B) + a_2(B)u_1(A) - (a_1(A)a_2(B)u_1(A)u_2(B))}{u_1(A) + u_2(B) - 2u_1(A)u_2(B)}, \end{array} \right. \quad (6.75)$$

où  $k = u_1(A) + u_2(B) - u_1(A)u_2(B) \neq 0$ . Si  $k = 0$ , l'opérateur est défini par :

$$\left\{ \begin{array}{l} Cr_{1,2}(A, B) = \frac{\gamma Cr_1(A) + Cr_2(B)}{1 + \gamma}, \\ dt_{1,2}(A, B) = \frac{\gamma dt_1(A) + dt_2(B)}{1 + \gamma}, \\ u_{1,2}(A, B) = 0, \\ a_{1,2}(A, B) = \frac{\gamma a_1(A) + a_2(B)}{1 + \gamma}, \end{array} \right. \quad (6.76)$$

où  $\gamma = \frac{u_2(B)}{u_1(A)}$ . Prenons deux exemples pour illustrer cet opérateur. Les tableaux 6.9 et 6.10 présentent les résultats de fusion de deux sources caractérisées par les fonctions de masse  $m_1$  et  $m_2$ , pour la règle orthogonale normalisée ou non et l'opérateur de consensus.

Sur le premier exemple, nous constatons que pour l'opérateur de consensus, le conflit est conservé entre  $d_1$  et  $d_2$ , à la différence des deux autres combinaisons qui soit focalise la

	$d_1$	$d_2$	$d_3$	$D$	$\emptyset$
$m_1$	0.99	0.01	0.00	0.00	0.00
$m_2$	0.00	0.01	0.99	0.00	0.00
$m_1 \oplus m_2$ normalisé	0.00	1.00	0.00	0.00	0.00
$m_1 \oplus m_2$ non normalisé	0.0000	0.0001	0.0000	0.0000	0.9999
opérateur de consensus	0.495	0.010	0.495	0.00	0.00

TAB. 6.9 – Exemple de fusion pour des sources en conflit.

	$d_1$	$d_2$	$d_3$	$D$	$\emptyset$
$m_1$	0.98	0.01	0.00	0.01	0.00
$m_2$	0.00	0.01	0.98	0.01	0.00
$m_1 \oplus m_2$ normalisé	0.490	0.015	0.490	0.005	0.000
$m_1 \oplus m_2$ non normalisé	0.0098	0.0003	0.0098	0.0001	0.9800
opérateur de consensus	0.492	0.010	0.492	0.005	0.00

TAB. 6.10 – Exemple de fusion pour des sources en conflit avec l'introduction de l'ignorance.

masse sur  $d_2$  à cause du faible accord des deux sources soit focalise la plupart de la masse sur l'ensemble vide représentant le fort conflit. Le tableau 6.10 montre que lorsqu'une part de la masse est allouée à l'ignorance, l'opérateur normalisé et l'opérateur de consensus conserve le conflit entre  $d_1$  et  $d_2$ , tandis que l'opérateur non normalisé focalise toujours la masse sur l'ensemble vide (représentant le conflit réel). Ainsi en cas de conflit l'opérateur de consensus permet de prendre une décision sur les singletons, à la différence de la règle orthogonale normalisée ou non. La principale critique de cet opérateur est justement qu'il ne peut donner aucune indication sur le conflit au même titre que la règle normalisée.

Une autre approche présentée dans [104] et [105], et discutée dans [75], propose une combinaison à base de pondérations afin de gérer une nouvelle fois le conflit. Cette combinaison de  $m$  sources  $S_j$  est donnée pour tout  $A \in 2^D$  par :

$$m_L(A) = m(A) + m_c(A), \quad (6.77)$$

où  $m(A)$  représente la masse obtenue par la règle orthogonale non normalisée et  $m_c(A)$  est la masse de conflit associée à  $A$  donnée par :

$$\begin{cases} m_c(A) = w(A, \mathbf{m}) m(\emptyset), \forall A \in 2^D \\ m_c(A) = 0 \text{ sinon,} \end{cases} \quad (6.78)$$

où  $\mathbf{m} = \{m_1, \dots, m_j, \dots, m_m\}$ , et  $\mathcal{P}$  une partition de  $D$  et  $w$  une fonction de poids vérifiant :

$$\sum_{A \in 2^D} w(A, \mathbf{m}) = 1. \quad (6.79)$$

Ainsi, nous retrouvons la combinaison de Smets si  $\mathcal{P} = \{\emptyset\}$  et si  $w(\emptyset, \mathbf{m}) = 1$  (c'est-à-dire que l'ensemble sur lequel la masse du conflit est distribuée est  $\emptyset$ ). Si  $\mathcal{P} = \{D\}$  et si  $w(D, \mathbf{m}) = 1$ , la masse du conflit est attribuée à  $D$  et nous retrouvons la combinaison de Yager. Il est aussi possible de retrouver la combinaison de Dubois et Prade. Dans [104], l'approche retenue consiste à distribuer la masse du conflit à partir d'un apprentissage sur les probabilités pignistiques de Smets (servant dans l'étape de décision) permettant d'optimiser l'erreur quadratique moyenne sur ces probabilités. Un inconvénient majeur est la perte de l'associativité de la combinaison. Haenni [75] relève aussi le problème de la spécification des poids pour chaque opérateur et selon les sous-espaces issus du cadre de discernement.

Nous avons vu le problème cruciale du conflit qui dépend énormément de la combinaison. Peut être que la solution coûteuse proposée par Dezert et Smarandache dans [49] et [50] est une bonne alternative. Comme nous l'avons déjà présentée dans la section 6.1, la modélisation est différente puisque dans ce cas nous considérons l'ensemble des disjonctions et des conjonctions des singletons de  $D$ , noté  $D^D$ . La règle orthogonale non normalisée est alors définie pour tout  $A \in D^D$  par :

$$m(A) = \sum_{\substack{B_1 \cap \dots \cap B_m = A, \\ B_1, \dots, B_m \in D^D}} \prod_{j=1}^m m_j(B_j). \quad (6.80)$$

## 6.4 Décision

Dans le cadre d'applications de classification automatique, la dernière étape concerne la prise de décision, c'est-à-dire le choix de la décision  $d_i$ , ou d'une disjonction. Ce choix va se faire par la maximisation d'un critère. Dans le cadre de cette théorie plusieurs possibilités sont envisageables.

Une fois combinées les fonctions de masses de l'ensemble des sources, il est aisé de calculer les fonctions de croyance et de plausibilité à partir des formules de définitions (*cf.* page 33). En effet ces deux fonctions représentent respectivement la croyance minimale et la croyance maximale. Ainsi, une décision idéale correspond au cas où il existe toujours une décision  $d_i$  dont la croyance est supérieure à la plausibilité de toute autre décision  $d_k$ , ce qui en pratique est rarement vérifié. C'est-à-dire que nous avons :

$$\forall d_k \neq d_i, Pl(d_i) \geq Cr(d_i) > Pl(d_k) \geq Cr(d_k). \quad (6.81)$$

Des situations plus ambiguës se rencontrent souvent, du type :

$$\exists d_k \neq d_i : Pl(d_i) \geq Pl(d_k) \geq Cr(d_i) \geq Cr(d_k), \quad (6.82)$$

ou encore

$$\exists d_k \neq d_i : Pl(d_i) \geq Pl(d_k) \geq Cr(d_k) \geq Cr(d_i). \quad (6.83)$$

Les critères présentés ici dérivés de ces fonctions portent uniquement sur le choix d'un singleton  $d_i$  optimal. Pour des applications plus général d'aide à la décision, il est bon de conserver l'ensemble de l'information obtenue, ou de la résumer de manière synthétique sans cette dernière étape.

### 6.4.1 Maximum de plausibilité

Un premier critère est de choisir le singleton  $d_i$  donnant le maximum de plausibilité. Ainsi, pour l'observation  $x$ , nous déciderons  $d_i$  si :

$$Pl(d_i)(x) = \max_{1 \leq k \leq n} Pl(d_k)(x). \quad (6.84)$$

Ce critère est optimale pour des fonctions de masse dérivées des probabilités [3]. En effet, ce critère est cohérent avec le maximum de probabilité *a posteriori* lorsque la connaissance disponible est bayésienne. L'inconvénient principale de ce critère est qu'il manque parfois de pouvoir discriminant du fait de son caractère *optimiste*. En effet, la plausibilité d'un singleton représente la croyance que nous aurions si toutes les masses des disjonctions incluant ce singleton étaient focalisées sur ce singleton. Prendre le maximum de plausibilité sur les singletons revient donc à déterminer la décision donnant le maximum de chance à chacun des singletons.

De plus, par définition  $Pl(d_i) = Cr(D) - Cr(d_i^c)$ , ainsi décider de  $d_i$  possédant le maximum de plausibilité revient à soutenir la décision qui croit le moins en la décision contraire  $d_i^c$ .

### 6.4.2 Maximum de croyance

Le critère du maximum de croyance consiste à choisir  $d_i$  pour l'observation  $x$  si :

$$Cr(d_i)(x) = \max_{1 \leq k \leq n} Cr(d_k)(x). \quad (6.85)$$

Ce critère est parfaitement identique au précédent dans le cas où la combinaison ne porte que sur les singletons. Les éléments focaux sont cependant rarement réduits au singletons.

Ce critère est bien plus sélectif que le précédent, il constitue un critère plus *pessimiste*. En effet, la fonction de croyance d'une décision fournit la croyance que nous aurions si toutes les masses des disjonctions l'incluant n'étaient pas focalisées sur cette décision (mais sur les autres). Ce critère du maximum de croyance détermine donc la décision qui donne le minimum de chance à chacune des disjonctions.

### 6.4.3 Maximum de croyance sans recouvrement des intervalles de confiance

Dans le cas idéal où les intervalles de confiance ne se recouvrent pas, nous pouvons choisir la décision  $d_i$  pour une observation  $x$  si :

$$Cr(d_i)(x) \geq \max_{1 \leq k \leq n, k \neq i} Pl(d_k)(x). \quad (6.86)$$

Ce critère encore plus stricte que le précédent, n'est intéressant que dans les cas idéaux rarement vérifiés en pratique. Dans ces cas, il n'a pas de risque d'erreur. Du fait qu'il soit bien plus sélectif, employé seul ce critère peut conduire à ne pouvoir prendre de décision sur les seuls singletons.

### 6.4.4 Maximum de croyance avec rejet

Un autre critère proposé dans [99], consiste à choisir  $d_i$  pour l'observation  $x$  si :

$$\begin{cases} Cr(d_i)(x) = \max_{1 \leq k \leq n} Cr(d_k)(x), \\ Cr(d_i)(x) \geq Cr(d_i^c)(x). \end{cases} \quad (6.87)$$

L'ajout de cette seconde condition par rapport au maximum de croyance, permet de ne prendre une décision que si elle est non ambiguë, c'est-à-dire si nous croyons plus en  $d_i$  qu'en son contraire. Plus stricte elle est donc plus discriminante, elle peut donc aussi conduire à aucune décision sur les singletons.

### 6.4.5 Maximum de probabilité pignistique

Un compromis au critère du maximum de plausibilité et du maximum de croyance a été proposé par Smets dans [151], où il introduit la *probabilité pignistique*. Ce nom provient de *pignus* en latin signifiant pari (*bet* en anglais). La probabilité pignistique est une mesure

permettant d'approcher le couple (croyance, plausibilité) en équirépartissant les masses placées sur les éléments différents d'un singleton sur les singletons qui les composent. Ainsi, pour toute décision  $d_i$ , la probabilité pignistique est définie par :

$$bet(d_i) = \sum_{A \in 2^D, d_i \in A} \frac{m(A)}{|A|(1 - m(\emptyset))}, \quad (6.88)$$

où  $|A|$  représente le cardinal de  $A$ . Ainsi le critère du maximum de probabilité pignistique revient à décider  $d_i$  pour l'observation  $x$  si :

$$bet(d_i)(x) = \max_{1 \leq k \leq n} bet(d_k)(x). \quad (6.89)$$

Ce critère est donc un critère *prudent*, bien adapté dans le cas d'ambiguïté. Cependant, comme le fait remarquer [13], la probabilité pignistique nous fait repasser dans un contexte probabiliste peu conforme à la notion de masse.

Dans [51], Dezert propose une extension de cette probabilité pignistique dans le cadre de sa théorie. Ainsi il définit la cardinalité  $\mathcal{C}_{\mathcal{M}}$  de tout élément  $A$  de  $D^D$  par le nombre de parties de  $A$  dans le diagramme de Venn du problème prenant en compte l'ensemble des contraintes d'intégrité de son modèle. La probabilité pignistique généralisée est alors définie pour tout  $A \in D^D$  par :

$$P(A) = \sum_{X \in D^D} \frac{\mathcal{C}_{\mathcal{M}}(X \cap A)}{\mathcal{C}_{\mathcal{M}}(X)} m(X). \quad (6.90)$$

Des règles mixtes ont aussi été envisagées dans lesquelles la plausibilité est employée pour certaines décisions et la croyance pour d'autres. Les décisions considérées par la plausibilité sont ainsi favorisées. Par exemple dans [116] qui traite la détection de mines, notant  $M$  l'évènement "c'est une mine" et  $F$  sinon, la règle de décision est donnée par :

$$\begin{cases} \text{si } Cr(F) > Pl(M) \text{ alors } F, \\ \text{si } Pl(M) \geq Cr(F) \text{ alors } M. \end{cases} \quad (6.91)$$

#### 6.4.6 Règle de décision avec coût

Des règles de décision avec coût ont été proposées dans [42]. Pour toute fonction de coût  $u_{ik}$  les espérances inférieure et supérieure sont définies pour toute décision  $d_i$  par :

$$E^*(d_i) = \sum_{A \in 2^D} m(A) \min_{d_k \in A} u_{ik}, \quad (6.92)$$

et

$$E_*(d_i) = \sum_{A \in 2^D} m(A) \max_{d_k \in A} u_{ik}. \quad (6.93)$$

La fonction  $u_{ik}$  doit exprimer le coût de l'hypothèse  $d_i$  si l'hypothèse  $d_k$  est vraie. Elle peut être établie par une approche probabiliste en utilisant la probabilité pignistique.

Ainsi une décision optimiste peut être obtenue en décidant  $d_i$  si  $d_i = \underset{1 \leq l \leq n}{\operatorname{argmin}} E_*(d_l)$  ; une décision pessimiste en choisissant  $d_i$  si  $d_i = \underset{1 \leq l \leq n}{\operatorname{argmin}} E^*(d_l)$  et une décision intermédiaire en considérant les probabilités pignistiques. Dans [42], la fonction de coût est définie par  $U_{ik} = 1 - \delta_{ik}$ , où  $\delta_{ik}$  représente le symbole de Kronecker. Ainsi, nous avons  $E_*(d_i) = 1 - Pl(d_i)$  et  $E^*(d_i) = 1 - Cr(d_i)$ .

### 6.4.7 Autres règles de décision

Il existe d'autres approches pour l'aide à la décision dans le cadre de la théorie de Dempster-Shafer. Citons par exemple des approches présentées dans [180], à base d'arbres de décision [117] ou encore à partir d'intervalles de probabilités [171]. Les arbres de décision sont également employés dans [66]. Dans [86], une approche à base de hiérarchies est adoptée en s'inspirant des travaux de [147].

## 6.5 Applications

La théorie des croyances peut être appliquée dans un grand nombre de situations ce qui a poussé à l'introduction de fonction de croyance dans toute sorte de domaines. Le principe même de modélisation sur un espace d'hypothèses composées  $2^D$  au lieu de  $D$  est un avantage certain de cette théorie. De plus, les fonctions de masse permettent d'obtenir une modélisation riche des imperfections des données. Nous reprenons ci-dessous les situations proposées par [13] dans lesquelles la théorie des croyances est intéressante :

- Dans les cas limites où toute l'information est connue, cas idéal, la théorie des croyances est alors optimale et se ramène à l'approche bayésienne.
- Lorsqu'une source donne des informations seulement sur certaines décisions, par exemple un sonar peut fournir une information sur le sédiment de surface, alors qu'un sondeur renseigne sur les différentes couches de sédiments.
- Lorsqu'une décision ne peut pas différencier deux hypothèses, il faut alors considérer la disjonction de ces deux éléments, ce qui permet de ne pas introduire d'information supplémentaire forçant leur séparation.
- Lorsqu'il est important de représenter la fiabilité d'une source, elle est modélisée par une masse sur  $m_j(D)$  non nulle.
- Lorsque la fiabilité d'une source dépend d'une décision, l'introduction du coefficient  $\alpha_{ij}$  permet d'affaiblir les masses d'une décision peu fiable pour une source  $S_j$  donnée.
- Lorsqu'il est possible d'introduire des informations *a priori*, pour définir ou modifier les fonctions de masse, ces informations peuvent être introduites dans le calcul des fonctions de masse.
- Lorsque nous souhaitons modéliser des effets de volume partiel (par exemple un pixel qui appartient à plusieurs classes), la théorie des croyances est encore intéressante.

Cette théorie a été employée dans un très grand nombre d'applications, citons en ici quelques unes, sans chercher à être exhaustif.

- En traitement d’images : la théorie de Dempster-Shafer a souvent été abordée en traitement d’images par exemple dans [101] ou dans [66] pour de la reconnaissance manuscrite, mais surtout dans le cas d’images médicales [140, 136, 137, 161, 12, 13, 21] avec pour but de segmenter les images gautier99 pour des images IRM ou [24]) ou de détecter des anomalies (pour des IRM [162, 184]). La fusion d’images issues des rayons X et ultrasons a également été étudiée [52]. D’autres images telles que satellitaires [127, 102], optiques ([8, 164] pour de la segmentation, [64, 65] pour de la fusion de pixels) ont fait l’objet de travaux. Les images sonar haute fréquence ([113] pour de la caractérisation de sédiments) et basse fréquence ([115] pour de la détection de mines enfouies) et davantage les images radar (images LADAR et FLIR [128], pour de la détection de mines [116, 122], des images SAR pour la reconnaissance de navires [86], des images SAR avec des images optiques [99, 100, 68]) ont également fait l’objet de nombreuses études. Dans [112], ce sont directement les données fréquentielles qui sont employées pour de la fusion de classifieurs en vue de la reconnaissance de cibles aériennes.
- En traitement de la parole : la théorie des croyances semble être peu abordée en traitement de la parole [114].
- Données temporelles : dans [84] une étude est proposée pour tenir compte des données temporelles dans le cadre de la théorie de Dempster-Shafer.
- Localisation : la localisation est un problème souvent abordé en fusion d’informations en général. La théorie des croyances a été employée et adaptée pour la localisation de robots ([30, 182], à partir d’images à ultrason [125]), mais aussi pour l’exploitation d’un réseau autoroutier [69].
- Association : la théorie des croyances peut aussi être adaptée pour l’association de données [144, 74].
- Classification : la classification est sans doute l’application la plus étudiée avec tout type de données dans le cadre de cette théorie [103, 44, 186].
- Suivi de cible : le suivi de cible peut aussi être réalisé en partie grâce à cette théorie [149].
- Autres données : un grand nombre de données peuvent être abordées telles que météorologiques [129], des mesures sur les protéines [45], ou des mesures physico-chimiques de la qualité de l’eau de la Seine [43], ...

## 6.6 Réflexions

Si cette théorie semble très séduisante, elle présente néanmoins certaines limites. Le principal défaut de cette approche est sa complexité à croissance exponentielle avec la taille du cadre de discernement. Énoncée pour un cadre de discernement discret, elle a surtout été employée pour des applications de reconnaissance de forme, classification, détection, et donc peu facile d’emploi pour de l’estimation de paramètres. Cependant de récents travaux [159, 134] proposent un élargissement de la théorie au cas de cadres de discernement continu, ouvrant la voie à la résolution de problèmes d’estimation. Le cadre

de discernement s'écrit alors comme un ensemble d'intervalles, ce qui augmente encore la complexité.

Bien sûr l'estimation des fonctions de masse et intrinsèquement le choix de leur modélisation doit être réalisées à partir d'une bonne connaissance des données. Une phase d'apprentissage est souvent indispensable ce qui peut limiter la généralisation, particulièrement si les données évoluent par rapport à celles de l'apprentissage. Des approches adaptatives sont alors à envisager.

La principale hypothèse est l'indépendance cognitive des sources ce qui n'est pas contraignant en soi.

Notons qu'il existe un lien entre la théorie des croyances et celle des probabilités ainsi qu'avec la théorie des possibilités. En effet, nous avons vu que lorsque les éléments focaux sont des singletons, les fonctions de masse sont alors des probabilités. Dans ce cas, les fonctions de croyance et de plausibilité se retrouvent égales aux fonctions de masse. Comme la règle orthogonale de Dempster-Shafer est cohérente avec les probabilités, les probabilités sont un cas particulier de la théorie des croyances.

Lorsque les éléments focaux sont tous inclus les uns dans les autres, les fonctions de croyances sont les fonctions de nécessité dans la théorie des possibilités et les fonctions de plausibilité sont les fonctions de possibilités dans la théorie des possibilités (*cf.* chapitre suivant). Cette situation reste cependant assez théorique et peu probable dans des applications réelles.

Ainsi les notions de probabilité et de possibilité sont incompatibles (sauf s'il n'y a qu'un seul élément focal, cas peu intéressant).



# Chapitre 7

## La théorie des possibilités

Un autre cadre méthodologique important en fusion d'informations est celui de la théorie des possibilités et des sous-ensembles flous dont elle est issue. La théorie des ensembles flous développée par le Professeur Zadeh dans au milieu des années 60 a peu progressé jusqu'au début des années 90. Elle permet une modélisation très souple des informations imprécises. La théorie des possibilités introduite par Zadeh en 1978, puis développée en France par Dubois et Prade [56] est dérivée de la théorie des sous-ensembles flous. Cette approche permet de représenter à la fois les imprécisions et les incertitudes de l'information.

Les applications de la logique floue sont bien plus vastes que son utilisation en fusion d'informations [19, 20]. Dans le cadre de la fusion d'informations la théorie des ensembles flous et des possibilités offrent plusieurs intérêts. En effet, elle permet :

- de représenter les imprécisions et les incertitudes de l'information d'une façon différente de la théorie des croyances.
- d'introduire différentes sémantiques.
- de tenir compte de l'information spatiale en traitement d'images [14].
- de combiner des informations très variées grâce à la richesse et la souplesse des opérateurs proposés.

Dans un premier temps nous rappelons les principes théoriques de la théorie des sous-ensembles flous et des possibilités, puis nous précisons comment l'employer dans le cadre de la fusion d'informations. Pour finir nous présentons quelques applications des théories des sous-ensembles flous et des possibilités à la fusion d'informations.

### 7.1 théorie des sous-ensembles flous et des possibilités

Nous nous contentons ici de rappeler les principes de ces théories, nous nous référons à [55, 19, 20] pour un aperçu plus large de la théorie de sous-ensembles flous (à [18] pour s'exercer à manipuler ces notions) et à [56, 13] pour la théorie des possibilités.

### 7.1.1 Théorie des sous-ensembles flous

#### Sous-ensembles flous

Un ensemble classique ou net est composé d'éléments qui satisfont à des propriétés précises. Ainsi des éléments qui ne satisfont pas ces propriétés n'appartiennent pas à cet ensemble. Notons  $\mathcal{S}$  un ensemble net et  $x, y$  ses éléments. Un sous-ensemble  $A$  de  $\mathcal{S}$  peut s'écrire à partir de sa *fonction caractéristique*  $\chi_A : \mathcal{S} \Rightarrow \{0, 1\}$  par :

$$\chi_A(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{si } x \notin A. \end{cases} \quad (7.1)$$

C'est une *fonction binaire* qui vérifie l'appartenance d'un point  $x$  de  $\mathcal{S}$  à  $A$ . La théorie des sous-ensembles flous propose une fonction qui vérifie l'appartenance graduelle, une *fonction d'appartenance*. Ainsi la fonction caractéristique est une fonction d'appartenance particulière.

**Exemple 7.1.1** Prenons l'exemple de la taille en centimètre d'une personne (ou d'un objet observé). Le sous-ensemble  $A$  des tailles comprises entre 160 et 180 cm a pour fonction caractéristique :

$$\chi_A(x) = \begin{cases} 1 & \text{si } 160 \leq x \leq 180, \\ 0 & \text{sinon.} \end{cases} \quad (7.2)$$

Supposons vouloir caractériser l'ensemble  $B$  des personnes avoisinant les 170 cm, la proposition *avoisinant les 170 cm* n'est pas précise.  $B$  pourrait être représenté par une fonction caractéristique, mais faut-il choisir les tailles comprises entre 160 et 180 cm ou 165 et 175 cm comme étant celles avoisinant 170 ? La fonction d'appartenance permet de généraliser la fonction caractéristique en tenant compte des tailles pas trop éloignées de 170 sans être trop proches (*cf.* figure 7.1).

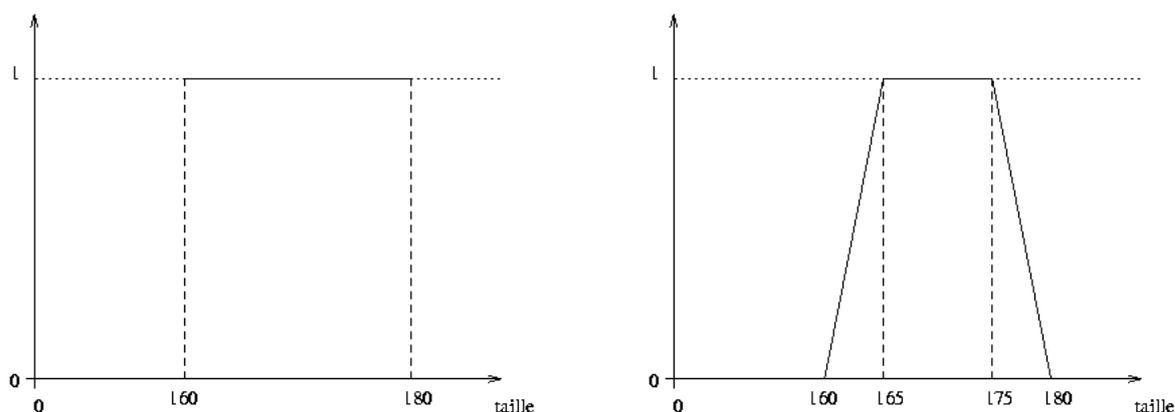


FIG. 7.1 – Fonction caractéristique de l'ensemble  $A$  et fonction d'appartenance de l'ensemble  $B$ .

**Définition 7.1.2** *Un sous-ensemble flou  $F$  de  $\mathcal{S}$  est défini par une fonction d'appartenance  $\mu_F$  qui à tout  $x$  de  $\mathcal{S}$  associe une valeur réelle  $\mu_F(x)$  dans  $[0, 1]$  représentant le degré d'appartenance de  $x$  au sous-ensemble flou  $F$ .*

### Remarques

- Les sous-ensembles flous sont souvent appelés plus simplement ensembles flous. Nous parlons ainsi de la théories des ensembles flous.
- L'intervalle  $[0, 1]$  est ici pris comme définition car il est le plus employé, d'autres intervalles ou ensembles peuvent cependant être utilisés, tels que les treillis.
- Le degré d'appartenance n'est pas une probabilité. Supposons qu'un sonar détecte deux objets  $O_1$  et  $O_2$  sur le fond marin. Ces objets peuvent être une mine plus ou moins dangereuse selon son état. Ainsi  $\mathcal{S}$  est l'ensemble des objets détectables et  $A$  le sous-ensemble flou des objets mortels. Supposons qu'un traitement nous permette de dire que le degré d'appartenance de  $O_1$  à  $A$  est 0.1 et que la probabilité que  $O_2$  appartienne à  $A$  est 0.1. D'où vient le principal danger ? Le degré d'appartenance signifie que l'objet  $O_1$  est proche d'un objet inoffensif, il s'agit peut-être d'un simple rocher. La probabilité que l'objet  $O_2$  appartienne à  $A$  signifie que les objets ressemblant à  $O_2$  sont à 10% des objets mortels et donc à 90% des objets non mortels. Ainsi si vous ne prenez garde à  $O_2$  le risque est d'être mortellement touché, alors qu'avec  $O_1$  vous ne risquez qu'une petite avarie.

### Description de sous-ensembles flous

La fonction d'appartenance définit complètement un sous-ensemble flou. Ainsi différentes notations sont employées pour décrire un sous-ensemble flou. Il peut être noté  $\{(x, \mu_A(x)) : x \in A\}$  ou encore  $\int_{\mathcal{S}} \mu_A(x)/x$ . La fonction d'appartenance d'un sous-ensemble flou permet de définir plusieurs paramètres descriptifs du sous-ensemble.

**Définition 7.1.3** - **Noyau** : *le noyau d'un sous-ensemble flou  $A$  de  $\mathcal{S}$ , noté  $Noy(A)$  est l'ensemble des éléments appartenant totalement à  $A$  :*

$$Noy(A) = \{x \in A : \mu_A(x) = 1\}. \quad (7.3)$$

*Le noyau est un ensemble binaire (i.e. décrit par une fonction binaire). Un sous-ensemble flou  $A$  est dit normalisé si au moins un élément  $x$  appartient complètement à  $A$  (i.e  $Noy(A) \neq \emptyset$ ). Un sous-ensemble flou  $A$  est dit unimodal s'il existe un unique élément  $x$  tel que  $\mu_A(x) = 1$ .*

- **Support** : *le support d'un sous-ensemble flou  $A$  de  $\mathcal{S}$ , noté  $Supp(A)$ , est l'ensemble des éléments appartenant au moins un peu à  $A$  :*

$$Supp(A) = \{x \in A : \mu_A(x) > 0\}. \quad (7.4)$$

*Le support est également un ensemble binaire.*

- **Hauteur** : la hauteur d'un sous-ensemble flou  $A$ , noté  $h(A)$  est la valeur maximale atteinte sur le support de  $A$  :

$$h(A) = \sup_{x \in A} \mu_A(x). \quad (7.5)$$

Ainsi un sous-ensemble flou est normalisé si sa hauteur est égale à 1.

Ses trois définitions sont illustrées sur la figure 7.2.

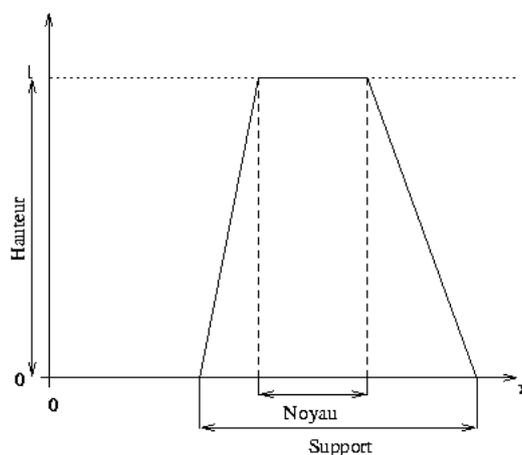


FIG. 7.2 – Noyau, support et hauteur d'un sous-ensemble flou.

**Définition 7.1.4** Le cardinal d'un sous-ensemble flou  $A$  de  $\mathcal{S}$  fini, est le nombre d'éléments qui appartiennent à  $A$  pondéré par leur degré d'appartenance :

$$|A| = \sum_{x \in \mathcal{S}} \mu_A(x). \quad (7.6)$$

Cette définition peut être étendue au cas où  $\mathcal{S}$  n'est pas fini mais mesurable. Si le support de  $A$  est fini, le cardinal de  $A$  est donné par :

$$|A| = \sum_{x \in \text{Supp}(A)} \mu_A(x). \quad (7.7)$$

**Définition 7.1.5** L' $\alpha$ -coupe (ou encore coupe de niveau  $\alpha$ ) d'un sous-ensemble flou est défini par l'ensemble binaire :

$$A_\alpha = \{x \in \mathcal{S} : \mu_A(x) \geq \alpha\}, \quad (7.8)$$

et  $A_\alpha$  est un sous-ensemble binaire de fonction caractéristique :

$$\chi_{A_\alpha}(x) = \begin{cases} 1 & \text{si } \mu_A(x) \geq \alpha, \\ 0 & \text{sinon.} \end{cases} \quad (7.9)$$

Les  $\alpha$ -coupes strictes ou fortes sont données par :

$$A_{\alpha s} = \{x \in \mathcal{S} : \mu_A(x) > \alpha\}. \quad (7.10)$$

Un sous-ensemble flou quelconque peut être décrit à partir de ces  $\alpha$ -coupes. Sa fonction d'appartenance peut ainsi être reconstruite à partir de plusieurs formules. Pour tout  $x$  de  $\mathcal{S}$ , nous avons :

$$\mu_A(x) = \int_0^1 \chi_{A_\alpha}(x) d\alpha, \quad (7.11)$$

$$\mu_A(x) = \sup_{\alpha \in ]0,1]} \min(\alpha, \chi_{A_\alpha}(x)), \quad (7.12)$$

$$\mu_A(x) = \sup_{\alpha \in ]0,1]} (\alpha \chi_{A_\alpha}(x)). \quad (7.13)$$

Ainsi, choisir une  $\alpha$ -coupe revient à seuiller la fonction d'appartenance permettant de sélectionner les éléments ayant un degré d'appartenance au moins égal à  $\alpha$ . Ceci peut être vu comme une étape de *défuzzification* (de l'anglais *fuzzyfication* qui est l'étape de modélisation par le flou (*fuzzy* )) souvent nécessaire pour l'étape de décision lors de la fusion.

**Définition 7.1.6** - Une **quantité floue** est un sous-ensemble flou sur  $\mathbb{R}$ .

- Un **intervalle flou** est une quantité floue convexe (toutes ses  $\alpha$ -coupes sont des intervalles).
- Un **nombre flou** est un intervalle flou dont toutes ses  $\alpha$ -coupes sont des intervalles fermé, de support compact et unimodal. Sur la figure 7.3 nous représentons le nombre flou  $A_{170}$  représentant environ 170 par sa fonction d'appartenance  $\mu_{A_{170}}$ .

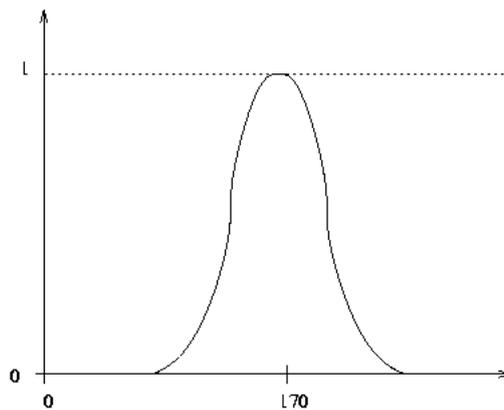


FIG. 7.3 – Exemple de nombre flou représentant *environ 170*.

Un nombre flou peut être vu comme une représentation d'une quantité imprécise. En fusion les nombres flous peuvent ainsi représenter des connaissances imprécises sur des mesures ou des observations.

### Opérations sur les sous-ensembles flous

Les opérations sur les sous-ensembles nets peuvent être généralisées aux sous-ensembles flous.

**Définition 7.1.7** Deux sous-ensembles flous  $A$  et  $B$  de  $\mathcal{S}$  sont dits égaux s'ils ont des fonctions d'appartenance égales en tout élément de  $\mathcal{S}$ . Ainsi :

$$A = B \iff \forall x \in \mathcal{S} \mu_A(x) = \mu_B(x). \quad (7.14)$$

**Définition 7.1.8** Soit deux sous-ensembles flous  $A$  et  $B$  de  $\mathcal{S}$ . Si tout élément  $x$  de  $\mathcal{S}$  appartient toujours moins à  $A$  qu'à  $B$ , alors  $A$  est inclus dans  $B$  ( $A \subseteq B$ ) :

$$A \subseteq B \iff \forall x \in \mathcal{S} \mu_A(x) \leq \mu_B(x). \quad (7.15)$$

L'égalité de  $A$  et de  $B$  est ainsi équivalente à l'inclusion de  $A$  dans  $B$  et de  $B$  dans  $A$ .

**Définition 7.1.9** L'union de deux sous-ensembles flous  $A$  et  $B$  de  $\mathcal{S}$  est le sous-ensemble flou formé des éléments de  $\mathcal{S}$  affectés du plus grand des degrés d'appartenance de  $A$  et de  $B$ . La fonction d'appartenance de  $A \cup B$  est donnée pour tout  $x$  de  $\mathcal{S}$  par :

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)). \quad (7.16)$$

Cela signifie qu'un élément ne peut appartenir à  $A$  ou  $B$  plus fortement qu'il n'appartient à chacun d'eux (cf figure 7.4).

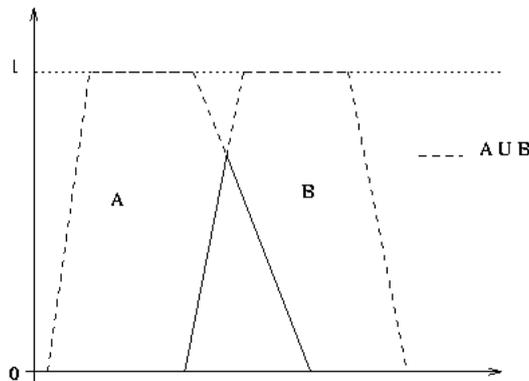
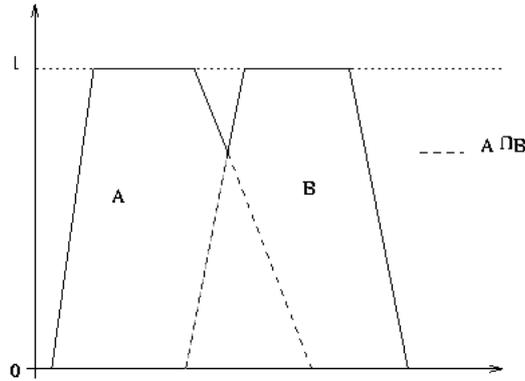


FIG. 7.4 – Union de deux sous-ensembles flous  $A$  et  $B$ .

**Définition 7.1.10** L'intersection de deux sous-ensembles flous  $A$  et  $B$  de  $\mathcal{S}$  est le sous-ensemble flou formé des éléments de  $\mathcal{S}$  affectés du plus petit des degrés d'appartenance de  $A$  et de  $B$ . La fonction d'appartenance de  $A \cap B$  est donnée pour tout  $x$  de  $\mathcal{S}$  par :

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)). \quad (7.17)$$

FIG. 7.5 – Intersection de deux sous-ensembles flous  $A$  et  $B$ .

Cela signifie qu'un élément ne peut appartenir à  $A$  ou  $B$  plus fortement qu'il n'appartient à chacun d'eux séparément (cf figure 7.5).

**Définition 7.1.11** Le complément d'un sous-ensemble flou  $A$  de  $\mathcal{S}$ , noté  $A^c$  est défini par la fonction d'appartenance donnée pour tout  $x$  de  $\mathcal{S}$  par :

$$\mu_{A^c}(x) = 1 - \mu_A(x). \quad (7.18)$$

La plus part des propriétés de ces opérations sur les sous-ensembles nets sont valables pour les opérations sur les sous-ensembles flous. Ainsi si  $A$  et  $B$  sont deux sous-ensembles flous, nous avons :

- L'union et l'intersection sont associatives et commutatives,
- l'union et l'intersection sont idempotentes ( $A \cup A = A$  et  $A \cap A = A$ ),
- si l'univers  $\mathcal{S}$  est considéré comme un ensemble flou de fonction d'appartenance identiquement égale à 1 et l'ensemble vide comme un sous-ensemble flou de fonction d'appartenance identiquement nulle, alors  $A \cap \mathcal{S} = A$ ,  $A \cap \emptyset = \emptyset$ , et  $A \cup \mathcal{S} = \mathcal{S}$ ,  $A \cup \emptyset = A$ ,
- $A \cap B \subseteq A \subseteq A \cup B$ ,
- l'union et l'intersection sont mutuellement distributives,
- l'union et l'intersection sont duales par rapport à la complémentation  $(A \cup B)^c = A^c \cap B^c$  et  $(A \cap B)^c = A^c \cup B^c$  (ce sont les lois de De Morgan),
- $(A^c)^c = A$ ,  $\emptyset^c = \mathcal{S}$  et  $\mathcal{S}^c = \emptyset$ .

Cependant certaines propriétés des opérations binaires ne sont plus vérifiées dans le cas flou, telles que le tiers-exclu ( $A \cup A^c = \mathcal{S}$ ) et la non-contradiction ( $A \cap A^c = \emptyset$ ). En effet, en général :  $A \cup A^c \neq \mathcal{S}$  et  $A \cap A^c \neq \emptyset$ .

### Mesures floues

**Définition 7.1.12** Une **mesure floue** est une fonction  $\nu$  de  $\mathcal{P}(\mathcal{S})$  (i.e. l'ensemble des parties de  $\mathcal{S}$ , ce sont donc des ensembles nets) dans  $[0, 1]$  qui vérifie les axiomes :

- $\nu(\emptyset) = 0$  et  $\nu(\mathcal{S}) = 1$ ,
- *monotonie* :  $\forall A, B \in \mathcal{P}(\mathcal{S}), A \subseteq B \implies \nu(A) \leq \nu(B)$ ,
- *continuité* : Pour toute suite  $(A_n)_n$  d'ensembles emboîtés de  $\mathcal{P}(\mathcal{S})$  ( $A_0 \subseteq A_1 \subseteq \dots \subseteq A_n \subseteq \dots$ , ou  $A_0 \supseteq A_1 \supseteq \dots \supseteq A_n \supseteq \dots$ ), alors  $\lim_{n \rightarrow +\infty} \nu(A_n) = \nu(\lim_{n \rightarrow +\infty} A_n)$ .

Ces mesures introduites par Sugeno pour évaluer l'incertitude sont nommées **mesure de confiance** dans [56]. L'axiome de continuité n'est pas toujours exigé [73]. Aucune contrainte d'additivité n'est exigée (une mesure floue est **additive** si  $\nu(A) = \sum_{i \in A} \nu(\{i\})$  pour tout  $A$  de  $\mathcal{P}(\mathcal{S})$ ). Une conséquence immédiate de l'axiome de monotonie donne les inégalités remarquables :

$$\forall A, B \in \mathcal{P}(\mathcal{S}), \nu(A \cup B) \geq \max(\nu(A), \nu(B)), \quad (7.19)$$

$$\forall A, B \in \mathcal{P}(\mathcal{S}), \nu(A \cap B) \leq \min(\nu(A), \nu(B)). \quad (7.20)$$

Ainsi, les mesures de probabilité sont des mesures floues, mais aussi les fonctions de croyance et de plausibilité de la théorie des croyances et les mesures de possibilité de Zadeh introduites dans la section suivante 7.1.2.

### 7.1.2 Théorie des possibilités

La théorie des possibilités formulée par Zadeh à la fin des années 70 est issue du concept des sous-ensembles flous développé également par Zadeh. Elle a été essentiellement développée en France par Dubois et Prade [56]. Elle repose sur les mesures de possibilité et de nécessité qui permettent de décrire l'incertitude d'un évènement par le degré de possibilité et le degré de possibilité de l'évènement contraire qui s'interprète comme un degré de nécessité (dans le sens de la certitude).

#### Mesures de possibilité et de nécessité

**Définition 7.1.13** Une **mesure de possibilité** est une fonction  $\Pi$  de  $\mathcal{P}(\mathcal{S})$  (i.e. l'ensemble des parties de  $\mathcal{S}$ , ce sont donc des ensembles nets) dans  $[0, 1]$  telle que :

- $\Pi(\emptyset) = 0$  et  $\Pi(\mathcal{S}) = 1$ ,
- $\forall I \subset \mathbb{N}, \forall A_i \subseteq \mathcal{S}, i \in I, \Pi(\cup_{i \in I} A_i) = \sup_{i \in I} \Pi(A_i)$ .

Ainsi dans le cas fini, une mesure de possibilité est une mesure floue. Elle correspond au cas limite de l'équation (7.19). Plus la valeur de la possibilité d'un évènement est proche de 1, plus la réalisation de cet évènement est possible. Si la possibilité d'un évènement est nulle, cela signifie qu'il est impossible. Une mesure de possibilité est bien différente

d'une mesure de probabilité. En effet, considérons un évènement  $A$  de  $\mathcal{S}$  et son contraire  $A^c$ , au moins l'un des deux est tout a fait possible ( $\Pi(A) = 1$  ou  $\Pi(A^c) = 1$ ), alors qu'en probabilité, la probabilité d'un évènement est totalement déterminée par celle de son évènement contraire ( $P(A) + P(A^c) = 1$ ). De plus il peut arriver que  $\Pi(A) = 1$  et  $\Pi(A^c) = 1$ , ce qui traduit l'ignorance totale sur la réalisation de l'évènement  $A$ . Pour lever cette indétermination sur la réalisation ou non de  $A$ , une information supplémentaire sur la possibilité de l'évènement contraire est ajoutée. Les mesures de nécessité, mesures duales des mesures de possibilité, sont introduites dans cet objectif. La nécessité d'un évènement est définie comme la mesure de l'impossibilité de l'évènement contraire.

**Définition 7.1.14** *Une mesure de nécessité est une fonction  $N$  de  $\mathcal{P}(\mathcal{S})$  dans  $[0, 1]$  telle que :*

- $N(\emptyset) = 0$  et  $N(\mathcal{S}) = 1$ ,
- $\forall I \subset \mathbb{N}, \forall A_i \subseteq \mathcal{S}, i \in I, N(\bigcap_{i \in I} A_i) = \inf_{i \in I} N(A_i)$ .

Cette définition est équivalente à la relation reliant nécessité et possibilité :

$$\forall A \subseteq \mathcal{S}, N(A) = 1 - \Pi(A^c). \quad (7.21)$$

La valeur prise par la mesure de nécessité pour un évènement  $A$  indique le degré de certitude avec lequel cet évènement sera réalisé. Plus ce degré est proche de 1, plus il est certain que cet évènement sera réalisé.

Ainsi l'incertitude d'un évènement est décrite par deux valeurs : le degré de possibilité et le degré de nécessité. Notons quelques propriétés sur ces deux mesures :

- $\forall A \subseteq \mathcal{S}, \max(\Pi(A), \Pi(A^c)) = 1$ ,
- $\forall A \subseteq \mathcal{S}, \min(N(A), N(A^c)) = 0$ , c'est-à-dire que deux évènements contraires ne peuvent pas être nécessaires simultanément,
- $\forall A \subseteq \mathcal{S}, N(A) > 0 \implies \Pi(A) = 1$  (en effet  $N(A) > 0 \implies \Pi(A^c) < 1$  et  $\max(\Pi(A), \Pi(A^c)) = 1$ )
- $\forall A \subseteq \mathcal{S}, \Pi(A) < 1 \implies N(A) = 0$ ,
- $\forall A \subseteq \mathcal{S}, N(A) + N(A^c) \leq 1$ ,
- $\forall A \subseteq \mathcal{S}, \Pi(A) + \Pi(A^c) \geq 1$ .

Les deux dernières propriétés montrent la non-additivité de ces mesures, contrairement aux probabilités.

### Distributions de possibilité

Les distributions de possibilité permettent d'attribuer un degré de possibilité à tout élément de  $\mathcal{S}$ , et non plus de  $\mathcal{P}(\mathcal{S})$ .

**Définition 7.1.15** *Une distribution de possibilité est une fonction  $\pi$  de  $\mathcal{S}$  dans  $[0, 1]$  vérifiant la propriété de normalisation :*

$$\sup_{s \in \mathcal{S}} \pi(s) = 1. \quad (7.22)$$

Cette condition correspond à une hypothèse de monde fermé dans lequel au moins un évènement de  $\mathcal{S}$  est complètement possible. Dans le cas d'un monde ouvert, cette condition peut être relâchée.  $\pi(s)$  indique le degré de possibilité pour que l'observation  $x$  soit égale à  $s$ , avec les conventions :

- $\pi(s) = 0$  indique l'impossibilité que  $x$  vaille  $s$ ,
- $\pi(s) = 1$  indique que rien n'empêche que  $x$  vaille  $s$ , ces valeurs sont les plus plausibles.

Les valeurs  $s$  pour lesquelles  $\pi(s) > 0$  sont supposées mutuellement exclusives et l'ensemble de ces valeurs forme l'ensemble flou des valeurs (plus ou moins) possibles de  $x$ . La vraie valeur de  $x$  existe dans  $\mathcal{S}$  (en monde fermé), mais est supposée mal connue. Une distribution de possibilité peut donc représenter l'imprécision attachée à la valeur exacte de  $x$ . Un nombre flou est ainsi une distribution de possibilité qui décrit la vraie valeur de l'observation  $x$ . Il y a une relation d'ordre évidente :  $\pi(s) > \pi(s')$  indique que l'égalité entre  $x$  et  $s$  est plus plausible qu'entre  $x$  et  $s'$ .

Notons une nouvelle fois la différence avec la théorie des probabilités. Une distribution de possibilité n'est pas une distribution de probabilité. En effet, dans le cas fini,  $P(x = s) = 1$  implique  $x = s$ , ce qui n'est pas le cas de  $\pi(s) = 1$ . Si  $\pi(s) = 1$  pour tout  $s$  de  $\mathcal{S}$ , nous ignorons complètement la valeur de  $x$ .

Dans le cas fini, nous pouvons construire une mesure de possibilité à partir d'une distribution de possibilité :

$$\forall A \subseteq \mathcal{S}, \Pi(A) = \sup_{s \in A} \pi(s). \quad (7.23)$$

Inversement, une distribution de possibilité peut être déduit d'une mesure de possibilité :

$$\forall s \in \mathcal{S}, \pi(s) = \Pi(\{s\}). \quad (7.24)$$

Du fait de la dualité, une mesure de nécessité peut être construite à partir d'une distribution de possibilité :

$$\forall A \subseteq \mathcal{S}, N(A) = 1 - \sup_{s \notin A} \pi(s) = \inf_{s \in A^c} (1 - \pi(s)). \quad (7.25)$$

En monde ouvert, il faut prendre garde que  $\Pi(\mathcal{S}) = 1$  n'est alors plus vérifié, ainsi que les propriétés d'implication  $N(A) > 0 \implies \Pi(A) = 1$  et  $\Pi(A) < 1 \implies N(A) = 0$ .

Il est possible de définir des distributions de possibilité conjointes et conditionnelles [19].

### Liens avec l'approche bayésienne et la théorie des croyances

Nous avons déjà soulevé plusieurs différences entre la théorie des sous-ensembles flous et des possibilités avec la théorie des probabilités. Nous reprenons ici l'étude de [58, 13].

Considérons dans le cas fini, une distribution de possibilité comme une famille d'ensembles emboîtés  $\{A_1, \dots, A_n\}$  avec  $A_i \subseteq A_{i+1}$ ,  $\forall i = 1, \dots, n - 1$ , avec des niveaux de confiance associés  $\lambda_i$  donnés par :

$$\lambda_i = N(A_i) = 1 - \Pi(A_i^c). \quad (7.26)$$

Ainsi  $\lambda_i \leq \lambda_{i+1}$ ,  $\forall i = 1, \dots, n-1$  puis que  $N$  est monotone. Soit  $\alpha_1 = 1$ ,  $\alpha_2 \geq \dots \geq \alpha_n$  et  $\alpha_{n+1} = 0$ , les valeurs prises par la distribution de possibilité. Les couples  $(A_i, \lambda_i)$  sont donnés par :

$$A_i = \{s \in \mathcal{S}, \pi(s) \geq \alpha_i\}, \lambda_i = 1 - \alpha_{i+1}. \quad (7.27)$$

A partir des couples  $(A_i, \lambda_i)_{\{1, \dots, n\}}$ , la distribution de possibilité la moins informative (c'est-à-dire la plus grande) qui vérifie  $\lambda_i = N(A_i)$  est donnée pour tout  $s$  de  $\mathcal{S}$  par :

$$\pi(s) = \min_{i=1, \dots, n} \max(1 - \lambda_i, \mu_{A_i}(s)). \quad (7.28)$$

En prenant  $p_i = \alpha_i - \alpha_{i+1}$ , nous avons  $\sum_{i=1}^n p_i = 1$  et  $\pi(s) = \sum_{i, s \in A_i} p_i$ .

Une distribution de possibilité peut donc être vue comme un ensemble aléatoire  $\{(A_i, p_i), i = 1, \dots, n\}$  avec  $\sum_{i=1}^n p_i = 1$ , qui code l'incertitude (les valeurs de probabilité  $p_i$ ) et l'imprécision (la taille des sous-ensembles  $A_i$ ).  $p_i$  est la probabilité pour que la source fournisse exactement  $A_i$  comme représentation de sa connaissance sur l'observation  $x$  et non la probabilité pour que  $x$  appartienne à  $A_i$ . L'ensemble  $\{(A_i, p_i), i = 1, \dots, n\}$  représente des ensembles focaux emboîtés de la théorie des croyances.

$\lambda_i$  peut s'interpréter comme une borne inférieure de la probabilité pour que la vraie valeur de  $x$  soit dans  $A_i$ . En effet, nous pouvons voir une distribution de possibilité comme le codage d'une famille de distributions de probabilité  $\mathbb{P}$  :

$$\mathbb{P} = \{P, P(A_i) \geq \lambda_i, i = 1, \dots, n\}. \quad (7.29)$$

Les mesures de possibilité et de nécessité peuvent alors être vues comme les probabilités supérieures et inférieures de tout évènement  $A$  :

$$\Pi(A) = P^*(A) = \sup_P \{P(A), P \in \mathbb{P}\}, \quad (7.30)$$

$$N(A) = P_*(A) = \inf_P \{P(A), P \in \mathbb{P}\}, \quad (7.31)$$

Ainsi  $\forall A \subseteq \mathcal{S}$  et  $\forall P \in \mathbb{P}$  :

$$N(A) \leq P(A) \leq \Pi(A). \quad (7.32)$$

## 7.2 Fusion d'informations par la théorie des possibilités

Dans cette section nous donnons les différentes étapes de la fusion d'informations dans le cadre de la théorie des sous-ensembles flous et des possibilités. L'étape principale, de combinaison, est ici très riche grâce aux nombreuses études réalisées sur les opérateurs dans le cadre de cette théorie.

### 7.2.1 Modélisation

En reprenant le formalisme de la section 1.2, chaque source  $S_j$  fournit une information représentée par  $M_i^j$  sur la décision  $d_i$  quant à l'observation  $x$ . Une première approche pour définir  $M_i^j$  est de considérer les fonctions d'appartenance de sous-ensembles flous correspondant à l'exactitude d'une décision :

$$M_i^j(x) = \mu_i^j(x), \quad (7.33)$$

où  $\mu_i^j(x)$  représente le degré d'exactitude de la décision  $d_i$  prise pour l'observation  $x$  selon la source  $S_j$ . En terme de classification, c'est le degré d'appartenance de  $x$  à la classe  $C_i$ . Nous avons vu que les fonctions d'appartenance offrent une grande souplesse quant au choix de la modélisation. Cependant cette souplesse peut être vue comme un inconvénient car elle laisse à l'utilisateur le choix pas si simple de définir ces fonctions. De plus les fonctions d'appartenance, si elles modélisent bien le caractère imprécis des informations, l'incertitude n'est considérée que de manière implicite et est peu accessible.

Une seconde approche pour définir  $M_i^j$  est de se placer dans le cadre de la théorie des possibilités, où l'imprécision et l'incertitude sont bien représentées. Dans ce cas l'ensemble de discernement  $D$  peut être choisi comme étant  $\mathcal{S}$  et  $M_i^j$  peut être vu comme un degré de possibilité d'une décision pour une observation :

$$M_i^j(x) = \pi_x^j(d_i), \quad (7.34)$$

qui est donc le degré de possibilité pour que la décision  $d_i$  soit valide pour l'observation  $x$  selon la source  $S_j$ . En terme de classification, c'est le degré de possibilité pour que la vraie classe de l'observation  $x$  soit  $C_i$  selon la source  $S_j$ . La possibilité et la nécessité d'une décision  $d_i$  pour une source  $S_j$  sont alors données par :

$$\Pi_j(\{d_i\}) = \pi_j(d_i), \quad (7.35)$$

$$N_j(\{d_i\}) = \inf\{(1 - \pi_j(d_k)), d_k \neq d_i\}. \quad (7.36)$$

De même la possibilité et la nécessité d'un ensemble quelconque  $A \subseteq D$  sont données par :

$$\Pi(A) = \sup_{s \in A} \pi(s), \quad (7.37)$$

$$N(A) = 1 - \sup_{s \notin A} \pi(s) = \inf_{s \in A^c} (1 - \pi(s)). \quad (7.38)$$

Cette modélisation nécessite de supposer les décisions nettes, tandis que la modélisation par fonction d'appartenance nécessite de les supposer flous.

### 7.2.2 Estimation

La difficulté des deux approches de modélisation réside dans l'estimation des fonctions d'appartenances ou des distributions de possibilités.

Nous pouvons distinguer deux cas [56] celui où l'utilisateur a une idée vague de la répartition de son observation pour une décision donnée de celui où fonctions d'appartenance ou distributions de possibilités doivent être déduites.

Dans le premier cas l'utilisateur peut parfois affecter une sémantique aux variations de l'observation. Ainsi si l'observation  $x$  représente une taille, l'utilisateur peut définir une fonction floue pour les petites tailles, les moyennes et les grandes. Les formes des fonctions sont donc laissées complètement libres à l'utilisateur. Il est souvent fait l'hypothèse que la somme des fonctions d'appartenance en un point est 1, ce qui permet de déterminer de manière unique les fonctions d'appartenance connaissant leur support.

Lorsqu'il est difficile d'employer des variables linguistiques, les fonctions d'appartenance peuvent être définies de manière automatisée. Ainsi il faut déterminer une fonction  $F_i^j$  pour une décision  $d_i$  et une source  $S_j$  telle que :

$$\mu_i^j(x) = F_i^j(x). \quad (7.39)$$

Les fonctions les plus utilisées sont des fonctions de normalisation, monomodales ou multimodales [13]. Mais il est également possible de les déterminer à partir d'algorithmes de classification automatique tels que les C-moyennes floues ou possibilistes [54]. Le problème des C-moyennes floues est que les valeurs d'appartenance à une classe ne sont pas décroissantes en fonction de la distance au centre de la classe, ce qui est évité avec les C-moyenne possibilistes.

Dans le cas des distributions de possibilité un apprentissage probabiliste est souvent réalisé particulièrement à partir d'un histogramme des observations. Ceci nécessite une quantité importante de données souvent présente en traitement du signal. A partir des histogrammes, nous obtenons une estimation des probabilités  $p_i^j$  pour une décision  $d_i$  et une source  $S_j$ . Pour transformer ces probabilités en distributions de possibilité  $\pi_i^j$  supposées discrètes, plusieurs critères ont été proposés [13] comme celui de la relation sur tout les sous-ensembles  $A$  [56] :

$$N(A) \leq P(A) \leq \Pi(A), \quad (7.40)$$

ou de Zadeh :

$$\sum_{i=1}^n p_i^j \pi_i^j = c, \quad (7.41)$$

où  $c$  est une constante de  $[0, 1]$ .

Pour les fonctions d'appartenance il est aussi possible de chercher une ressemblance avec l'histogramme à l'aide de différents critères d'optimisation. Cependant ces approches prennent rarement en compte la spécificité du flou.

### 7.2.3 Combinaison

Un des intérêts majeur de la fusion d'informations par le flou est le grand choix d'opérateurs de fusion permettant la combinaison de fonctions d'appartenance ou des distributions de possibilité. Puisque les fonctions d'appartenance et des distributions de possibilité

sont à valeurs dans  $[0, 1]$ , nous allons noter un opérateur par  $\phi$  une fonction qui à tout  $m$ -uplet  $(a_1, \dots, a_m)$  de  $[0, 1]$  représentant les degrés d'appartenance ou de possibilité de chaque source, associe une valeur dans  $[0, 1]$ .

Le grand nombre d'opérateurs pousse à un classement de ceux-ci afin de permettre le choix de l'opérateur le plus adapté à l'application. Ce choix peut se faire selon plusieurs critères liés aux propriétés des opérateurs. Il est courant de considérer les opérateurs conjonctifs, disjonctifs ou de compromis, c'est-à-dire qui ont des comportements sévères, indulgents ou prudents [73]. Mathématiquement, la combinaison est dite :

- conjonctive si  $\phi(a_1, \dots, a_m) \leq \min_j a_j$  (comportement sévère),
- disjonctive si  $\phi(a_1, \dots, a_m) \geq \max_j a_j$  (comportement indulgent),
- de compromis si  $\min_j a_j \leq \phi(a_1, \dots, a_m) \leq \max_j a_j$  (comportement indulgent).

I. Bloch [11] propose une autre façon de caractériser les opérateurs selon que :

- le résultat ne dépend que des valeurs à combiner (sans autre information) et le comportement est le même quelles que soient ces valeurs,
- le comportement dépend des valeurs à combiner,
- les opérateurs dépendent du contexte, par exemple selon la fiabilité ou le conflit des sources.

Rappelons ici quelques propriétés supplémentaires des opérateurs [73].

**Définition 7.2.1** - **associativité** : Soit  $\phi_2$  un opérateur de  $[0, 1]^2$  à valeurs dans  $[0, 1]$ . Il est associatif si et seulement si pour tout  $a_1, a_2$  et  $a_3$  de  $[0, 1]$  :

$$\phi_2(\phi_2(a_1, a_2), a_3) = \phi_2(a_1, \phi_2(a_2, a_3)), \quad (7.42)$$

ce qui permet d'étendre sa définition à trois arguments. L'associativité n'est pas systématiquement vérifiée, cette propriété permet cependant une certaine adaptativité temporelle de la combinaison.

- **bissymétrie** : Soit  $\phi_2$  un opérateur de  $[0, 1]^2$  à valeurs dans  $[0, 1]$ . Il est bissymétrique si et seulement si pour tout  $a_1, a_2, a_3$  et  $a_4$  de  $[0, 1]$  :

$$\phi_2(\phi_2(a_1, a_2), \phi_2(a_3, a_4)) = \phi_2(\phi_2(a_1, a_3), \phi_2(a_2, a_4)). \quad (7.43)$$

Cette propriété est plus faible que l'associativité et peut être suffisante.

- **commutativité** (ou neutralité ou anonymat) : l'opérateur  $\phi$  de  $[0, 1]^n$  dans  $[0, 1]$  est commutatif si et seulement si :

$$\phi(a_1, \dots, a_n) = \phi(a_{\sigma(1)}, \dots, a_{\sigma(n)}), \quad \forall \sigma, \quad (7.44)$$

où  $\sigma$  est une permutation des indices. Elle est satisfaite par tous les opérateurs utilisés. Ainsi le résultat de la combinaison est indépendant de l'ordre dans lequel l'information est prise en compte.

- **idempotence** (ou unanimité ou consensus) : l'opérateur  $\phi$  de  $[0, 1]^n$  dans  $[0, 1]$  est idempotent si et seulement si :

$$\phi(a, \dots, a) = a, \quad \forall a \in [0, 1]. \quad (7.45)$$

Ainsi une information déjà disponible ne change pas le résultat.

- **unanimité pour les valeurs extrêmes** : l'opérateur  $\phi$  de  $[0, 1]^n$  dans  $[0, 1]$  est unanime pour les valeurs extrêmes si et seulement si :

$$\phi(0, \dots, 0) = 0, \text{ et } \phi(1, \dots, 1) = 1. \quad (7.46)$$

Cette propriété permet de garantir l'intuition sur les cas binaires où toute les valeurs sont soit impossibles soit complètement possibles.

- **monotonie** : l'opérateur  $\phi$  de  $[0, 1]^n$  dans  $[0, 1]$  est monotone par rapport à chaque argument si et seulement si :

$$a'_j > a_j \implies \phi(a_1, \dots, a'_j, \dots, a_m) > \phi(a_1, \dots, a_j, \dots, a_m). \quad (7.47)$$

- **stabilité** pour le changement d'échelle linéaire : l'opérateur  $\phi$  de  $[0, 1]^n$  dans  $[0, 1]$  est stable pour le changement d'échelle linéaire si et seulement si :

$$\phi(ra_1 + t, \dots, ra_m + t) = r\phi(a_1, \dots, a_m) + t, \forall r > 0 \text{ et } t \in \mathbb{R}. \quad (7.48)$$

- **élément neutre** : l'opérateur  $\phi_2$  de  $[0, 1]^2$  dans  $[0, 1]$  possède  $x$  comme élément neutre si et seulement si :

$$\phi_2(a, x) = a, \forall a \in [0, 1]. \quad (7.49)$$

L'existence d'un élément neutre entraîne que les sources donnant cette valeur n'auront pas d'influence sur le résultat de la combinaison. Ceci peut représenter l'indifférence pour ces sources ou l'ignorance totale de ces sources sur l'observation.

Notons aussi que l'importance de la continuité de  $\phi$  qui garantit une certaine robustesse de la combinaison. Il est évident qu'un opérateur de compromis est idempotent, alors qu'un opérateur idempotent et monotone est un opérateur de compromis.

Nous présentons ici quelques opérateurs de manière non exhaustive selon leur comportement sévère, indulgent ou prudent, c'est-à-dire les opérateurs conjonctifs, disjonctifs et de compromis.

### Opérateurs conjonctifs

Les opérateurs conjonctifs combinent l'information à la façon d'un *ET* logique (conjonction). De cette façon le résultat sera proche de 1 si et seulement si toutes les valeurs à combiner sont proches de 1.

Parmi ces opérateurs les *normes triangulaires* ou *t-norme* sont les plus employées et permettent de réduire légèrement cette sévérité.

**Définition 7.2.2** Une *t-norme* est une fonction  $\top : [0, 1]^n \mapsto [0, 1]$  commutative, associative, monotone et dont 1 est l'élément neutre, c'est-à-dire :

$$\top(1, \dots, 1, a) = a, \forall a \in [0, 1]. \quad (7.50)$$

Les t-normes généralisent l'intersection aux sous-ensembles flous. Ainsi si les éléments combinés sont des degrés d'appartenance, le résultat représente le degré d'appartenance au sous-ensemble flou donné par l'intersection des sous-ensembles flous de départ. L'observation ne peut donc pas appartenir plus fortement à l'intersection des sous-ensembles qu'il n'appartient à chacun d'eux.

Notons que :

- $\top(0, a, \dots, a) = 0, \forall a \in [0, 1]$ .
- l'opérateur minimum est une t-norme et toute t-norme  $\top$  est telle que  $\top(a_1, \dots, a_m) \leq \min_j a_j$ , c'est-à-dire que le minimum est la plus grande t-norme.
- les t-normes ne vérifient pas l'idempotence (sauf le min) et la stabilité pour le changement d'échelle linéaire.
- le résultat ne dépend que des valeurs à combiner (sans autre information) et le comportement est le même quelles que soient ces valeurs.

Voici quelques exemples de t-normes :

- t-norme de Zadeh :  $\min_j a_j$ ,
- t-norme probabiliste :  $\prod_{j=1}^m a_j$ ,
- t-norme de Lukasiewicz : pour deux éléments (s'étend par l'associativité à  $m$  éléments)  $\max(0, a_1 + a_2 - 1)$ .
- la t-norme de Weber : pour deux éléments

$$\begin{cases} a_1 & \text{si } a_1 = 1 \\ a_2 & \text{si } a_2 = 1 \\ 0 & \text{sinon} \end{cases} \quad (7.51)$$

C'est la plus petite t-norme.

- t-norme de Yager : pour deux éléments

$$1 - \min(1, [(1 - a_1)^p + (1 - a_2)^p]^{1/p}), \quad (7.52)$$

avec  $p > 0$ . Pour  $p = 1$ , nous retrouvons la t-norme de Lukasiewicz et pour  $p = +\infty$  la t-norme min.

- t-norme archimédienne : une t-norme archimédienne sur  $[0, 1]^2$ , strictement monotone

$$f^{-1}(f(a_1) + f(a_2)), \quad (7.53)$$

où  $f$  est une bijection continue décroissante de  $[0, 1]$  dans  $[0, +\infty]$  telle que  $f(0) = +\infty$  et  $f(1) = 0$ . Une telle fonction est appelée **fonction génératrice**. Une t-norme archimédienne strictement croissante vérifie

$$\top(a_1, a_1) < a_1, \forall a_1 \in [0, 1], \quad (7.54)$$

$$\forall a_1, a_2, a_3 \in [0, 1] \quad a_2 < a_3 \implies \top(a_1, a_2) < \top(a_1, a_3). \quad (7.55)$$

- t-norme de Hamacher : pour deux éléments

$$\frac{a_1 a_2}{\gamma + (1 - \gamma)(a_1 + a_2 - a_1 a_2)}, \quad (7.56)$$

avec  $\gamma > 0$  (si  $\gamma = 1$ , nous retrouvons le produit). Ce type de t-norme peut être intéressant pour ces propriétés d'adaptativité [119]. Une t-norme de Hamacher est une t-norme archimédienne définie à partir d'un générateur multiplicatif.

### Opérateurs disjonctifs

Les opérateurs disjonctifs combinent l'information à la façon du *OU* logique (disjonction). La valeur du résultat de la combinaison sera grande dès lors qu'une des valeurs combinées le sera. Le comportement sera donc indulgent.

Les principaux opérateurs disjonctifs sont les conormes triangulaires ou t-conormes .

**Définition 7.2.3** Une t-conorme est une fonction  $\perp : [0, 1]^n \mapsto [0, 1]$  commutative, associative, monotone et dont 0 est l'élément neutre, c'est-à-dire :

$$\top(0, \dots, 0, a) = a, \forall a \in [0, 1]. \quad (7.57)$$

Les t-conormes généralisent l'union aux sous-ensembles flous. Ainsi si les éléments combinés sont des degrés d'appartenance, le résultat représente le degré d'appartenance au sous-ensemble flou donné par l'union des sous-ensembles flous de départ. L'observation ne peut donc pas appartenir plus fortement à l'union des sous-ensembles qu'il n'appartient à chacun d'eux.

Notons que :

- Si  $\perp(a_1, \dots, a_m)$  est une t-conorme, alors  $\top(a_1, \dots, a_m) = 1 - \perp(1 - a_1, \dots, 1 - a_m)$  est une t-norme (dite t-norme duale). En fait, c'est vrai pour toute fonction de complémentation  $c$  définit pour tout  $x$  de  $[0, 1]$  par :

$$c(x) = f^{-1}(1 - f(x)), \quad (7.58)$$

avec  $f$  fonction de  $[0, 1]$  dans  $[0, 1]$  qui vérifie :

- $f(0) = 0$ ,
- $f(1) = 1$ ,
- $f$  strictement croissante.

Ainsi nous avons les t-conorme  $\perp$  et t-norme duales qui sont données par  $\top(a_1, \dots, a_m) = c(\perp(c(a_1), \dots, c(a_m)))$ .

- $\top(1, a, \dots, a) = 1, \forall a \in [0, 1]$ .
- l'opérateur maximum est une t-conorme et toute t-conorme  $\perp$  est telle que  $\max_j a_j \leq \top(a_1, \dots, a_m)$ , c'est-à-dire que le maximum est la plus petite t-conorme.
- les t-conormes ne vérifient pas l'idempotence (sauf le max) et la stabilité pour le changement d'échelle linéaire.
- le résultat ne dépend que des valeurs à combiner (sans autre information) et le comportement est le même quelles que soient ces valeurs.

Voici quelques exemples de t-conormes :

- t-conorme de Zadeh :  $\max_j a_j$ ,
- t-conorme probabiliste : pour deux éléments (s'étend par l'associativité à  $m$  éléments)  $a_1 + a_2 - a_1 a_2$ ,
- t-conorme de Lukasiewicz : pour deux éléments (s'étend par l'associativité à  $m$  éléments)  $\min(a_1 + a_2 - 1, 1)$ .
- t-conorme de Weber : pour deux éléments

$$\begin{cases} a_1 \text{ si } a_1 = 0, \\ a_2 \text{ si } a_2 = 0, \\ 1 \text{ sinon.} \end{cases} \quad (7.59)$$

C'est la plus petite t-conorme.

- t-conorme de Yager : pour deux éléments

$$\min(1, [a_1^p + a_2^p]^{1/p}), \quad (7.60)$$

avec  $p > 0$ . Pour  $p = 1$ , nous retrouvons la t-conorme de Lukasiewicz et pour  $p = +\infty$  la t-conorme max.

- t-conorme archimédienne : une t-conorme archimédienne sur  $[0, 1]^2$ , strictement monotone

$$\varphi^{-1}(\varphi(a_1) + \varphi(a_2)), \quad (7.61)$$

où la fonction génératrice  $\varphi$  est une bijection continue croissante de  $[0, 1]$  dans  $[0, +\infty]$  telle que  $\varphi(0) = 0$  et  $\varphi(1) = +\infty$ . Une t-conorme archimédienne strictement croissante vérifie

$$\perp(a_1, a_1) > a_1, \quad \forall a_1 \in [0, 1], \quad (7.62)$$

$$\forall a_1, a_2, a_3 \in [0, 1] \quad a_2 < a_3 \implies \perp(a_1, a_2) < \perp(a_1, a_3). \quad (7.63)$$

- t-conorme de Hamacher : pour deux éléments

$$\frac{a_1 + a_2 - (2 - \gamma)a_1 a_2}{\gamma + (1 - \gamma)(1 - a_1 a_2)}, \quad (7.64)$$

avec  $\gamma > 0$ .

### Opérateurs de compromis

Les opérateurs de compromis ont un comportement moins sévère que les opérateurs conjonctifs, et moins indulgent que les opérateurs disjonctifs. Nous en présentons ici quelques uns.

**Minimum ou maximum pondérés** Introduits par Dubois et Prade dans le cadre de la théorie des possibilités [56], ces opérateurs permettent de tenir compte d'une modélisation de la fiabilité des sources. Ainsi, si nous connaissons un coefficient de fiabilité  $w_j \in [0, 1]$  des sources  $S_j$  (qui vaut 1 si la source est complètement fiable et 0 si elle n'est pas du tout fiable), nous pouvons transformer les réponses  $a_j$  des sources par :

$$\max(a_j, 1 - w_j), \quad (7.65)$$

Dans le cas de la théorie des possibilités,  $a_j = \pi_j$  est une distribution de possibilité. Cette transformation revient donc à une disjonction entre une distribution  $\pi_j$  et une distribution constante de valeur  $1 - w_j$ . Ainsi si la source est complètement fiable, la distribution correspondante n'est pas modifiée, tandis que qu'elle devient constante et égale à 1 (ce qui représente l'ignorance totale : tout élément est possible) si la source n'est pas du tout fiable. Une fois cette transformation effectuée, les distributions peuvent être combinées par une conjonction, ce qui donne l'opérateur de minimum pondéré :

$$\min_w(a_1, \dots, a_n) = \min_{j=1}^m \max(a_j, 1 - w_j). \quad (7.66)$$

Si les distributions sont complètement en conflit, la combinaison conjonctive, très sévère, de ces distributions donnera la distribution identiquement nulle. Dans ce cas une combinaison disjonctive est préférable car permet de conserver toutes les décisions possibles par au moins une des deux sources. Ainsi il est important de supposer qu'au moins une des sources est fiable, *i.e.*  $\max_{j=1}^m w_j = 1$ .

De même, il est possible de combiner les sorties  $a_j$  des sources  $S_j$  pondérées par des poids  $w_j$  par un opérateur de maximum pondéré :

$$\max_w(a_1, \dots, a_n) = \max_{j=1}^m \min(a_j, w_j). \quad (7.67)$$

Dans le cas où  $w = (1, 1, \dots, 1)$ , nous retrouvons les opérateurs de conjonction et disjonction min et max. Ces opérateurs vérifient les propriétés d'idempotence, de continuité et de monotonie.

## Opérateurs de moyenne

**Définition 7.2.4** *Un opérateur de moyenne est une fonction moy de  $[0, 1]^n$  dans  $[0, 1]$  telle que*

- $\forall (a_j) \in [0, 1]^n$ ,  $\min_j a_j \leq \text{moy}(a_1, \dots, a_m) \leq \max_j a_j$  et  $\text{moy} \neq \min$  et  $\text{moy} \neq \max$ ,
- *moy est commutative,*
- *moy est monotone par rapport à chaque argument.*

Le premier point montre que *moy* est idempotente. Les moyennes ne sont cependant pas en général associatives. Les médianes sont les seules à être associatives.

$\alpha$	$moy(a_1, a_2)$	commentaire
$-\infty$	$\min(a_1, a_2)$	valeur limite
-1	$\frac{2a_1a_2}{a_1 + a_2}$	moyenne harmonique avec $m(0, 0) = 0$
0	$(a_1a_2)^{-1/2}$	moyenne géométrique
+1	$\frac{a_1 + a_2}{2}$	moyenne arithmétique
+2	$\sqrt{\frac{a_1^2 + a_2^2}{2}}$	moyenne quadratique
$+\infty$	$\max(a_1, a_2)$	valeur limite

TAB. 7.1 – Exemple d'opérateurs de moyenne bissymétriques, continus et strictement croissants pour une fonction  $k$  de type  $x^\alpha$ .

**Définition 7.2.5** Une médiane est une fonction  $med$  de  $[0, 1]^2$  dans  $[0, 1]$  telle que

$$med(a_1, a_2) = \begin{cases} a_1 & \text{si } a_2 \leq a_1 \leq \alpha \text{ ou } \alpha \leq a_1 \leq a_2 \\ a_2 & \text{si } a_1 \leq a_2 \leq \alpha \text{ ou } \alpha \leq a_2 \leq a_1 \\ \alpha & \text{si } a_2 \leq \alpha \leq a_1 \text{ ou } a_1 \leq \alpha \leq a_2 \end{cases} \quad (7.68)$$

avec  $\alpha$  un paramètre de  $[0, 1]$ .

Les opérateurs de moyenne bissymétrique, continus et strictement croissants définis sur  $[0, 1]^m$  s'écrivent :

$$moy(a_1, \dots, a_m) = k^{-1} \left( \sum_{j=1}^m w_j k(a_j) \right), \quad (7.69)$$

où  $k$  est une fonction de  $[0, 1]$  dans  $[0, 1]$  continue et strictement croissante.  $k$  s'interprète comme un changement d'échelle ou de dynamique des valeurs à combiner [13]. Une fois la transformation  $k$  effectuée, les données sont combinées par une moyenne. Les moyennes les plus employées sont obtenues pour les fonctions de type  $k(x) = x^\alpha$ ,  $\forall x \in [0, 1]$  et  $\alpha \in \mathbb{R}$ . Le tableau 7.1 donnent différentes moyennes sur  $[0, 1]^2$  obtenues pour différentes valeurs de  $\alpha$  et pour  $w_j = \frac{1}{2}$  pour tout  $j$ . Le min et le max sont obtenus comme valeurs limites pour  $\alpha$ .

**Moyennes pondérées ordonnées (OWA)** Introduit par Yager [175], l'opérateur OWA (*Ordered Weighted Average*) est toujours étudié pour ses propriétés intéressantes [176, 67, 173].

**Définition 7.2.6** L'opérateur OWA est défini sur  $[0, 1]^n$  à valeurs dans  $[0, 1]$  par :

$$OWA_w(a_1, \dots, a_m) = \sum_{j=1}^m w_j a_{(j)}, \quad (7.70)$$

où  $w = (w_1, \dots, w_m)$  est un vecteur de poids,  $w_j \in [0, 1]$  tel que  $\sum_{j=1}^m w_j = 1$ , et la notation  $(.)$  indique une permutation des indices telle que  $a_{(1)} \leq \dots \leq a_{(m)}$ .

Les poids  $w_j$  portent ainsi sur les rangs des quantités  $a_j$  et non sur les sources  $S_j$ . Nous pouvons remarquer les cas particuliers :

- si  $w_1 = 1$  (et donc  $w_j = 0, \forall j > 1$ ), l'opérateur  $OWA_w$  est le min ;
- si  $w_m = 1$ , l'opérateur  $OWA_w$  est le max ;
- si  $w_j = 1$  pour un  $j$  donné, l'opérateur  $OWA_w$  est une statistique d'ordre  $j$  ;
- si  $m$  est impair et  $w_{(m+1)/2} = 1$ , l'opérateur  $OWA_w$  est la médiane ;
- si  $m$  est pair et  $w_{\frac{m}{2}} = w_{\frac{m}{2}+1} = \frac{1}{2}$ , l'opérateur  $OWA_w$  est la médiane.

Les opérateurs OWA sont idempotents, continus, monotones et stables pour le changement d'échelle linéaire. [73] propose une forme généralisée donnée par :

$$OWA_w^f(a_1, \dots, a_m) = f^{-1} \left( \sum_{j=1}^m w_j f(a_{(j)}) \right), \quad (7.71)$$

où  $f$  est continue et strictement croissante.

**Les intégrales floues** Les intégrales floues sont également des opérateurs de moyenne car elles calculent la valeur moyenne d'une fonction [70, 73]. Elles sont idempotentes, continues, croissantes et comprises entre le minimum et le maximum. Introduites par Sugeno, elles étendent l'intégrale de Lebesgue aux mesures floues. Deux types d'intégrales floues peuvent être considérées : l'intégrale de Sugeno et l'intégrale de Choquet.

**Définition 7.2.7** L'intégrale de Sugeno d'une fonction mesurable  $f$  définie sur  $\mathcal{S}$  dans  $[0, 1]$  relativement à une mesure floue  $\mu$  est donnée par :

$$S_\mu(f) = \int f \circ \mu = \sup_{\alpha \in [0,1]} \min(\alpha, \mu(\{x \in \mathcal{S}, f(x) > \alpha\})). \quad (7.72)$$

Dans le cas fini où  $|\mathcal{S}| = m$ , l'intégrale de Sugeno s'écrit :

$$S_\mu(f) = \max_{j=1}^m \min(f(x_{(j)}), \mu(A_j)), \quad (7.73)$$

où  $(.)$  désigne la permutation telle que  $f(x_{(1)}) \leq \dots \leq f(x_{(m)})$  et  $A_j = \{x_{(j)}, \dots, x_{(m)}\}$ .

Ainsi l'intégrale de Sugeno combine les sorties  $a_j$  des sources  $S_j$  par :

$$S_\mu(a_1, \dots, a_m) = \max_{j=1}^m \min(a_{(j)}, \mu(\{(j), \dots, (m)\})), \quad (7.74)$$

avec  $a_{(1)} \leq \dots \leq a_{(m)}$ .

**Définition 7.2.8** L'intégrale de Choquet d'une fonction mesurable  $f$ , définie de  $\mathcal{S}$  dans  $\mathbb{R}^+$ , relativement à la mesure floue  $\mu$  est donnée par :

$$C_\mu(f) = \int f d\mu = \int_0^{+\infty} \mu(\{x \in \mathcal{S} : f(x) > \alpha\}) d\alpha. \quad (7.75)$$

Dans le cas fini où  $|\mathcal{S}| = m$ , l'intégrale de Choquet s'écrit :

$$C_\mu(f) = \sum_{j=1}^m (f(x_{(j)}) - f(x_{(j-1)})) \mu(A_j), \quad (7.76)$$

avec  $f(x_{(0)}) = 0$ ,  $(.)$  désigne la permutation telle que  $f(x_{(1)}) \leq \dots \leq f(x_{(m)})$  et  $A_j = \{x_{(j)}, \dots, x_{(m)}\}$ .

Ainsi l'intégrale de Choquet combine les sorties  $a_j$  des sources  $S_j$  par :

$$C_\mu(a_1, \dots, a_m) = \sum_{j=1}^m (a_{(j)} - a_{(j-1)}) \mu(\{(j), \dots, (m)\}), \quad (7.77)$$

avec  $a_{(1)} \leq \dots \leq a_{(m)}$  et  $a_{(0)} = 0$ . L'intégrale de Choquet est plus stable pour le changement d'échelle linéaire que l'intégrale de Sugeno.

Voici quelques propriétés liées aux intégrales floues :

- l'intégrale de Choquet par rapport à une mesure additive coïncide avec les opérateurs de moyennes bissymétriques définis par l'équation (7.69) avec pour poids  $w_j$  les mesures  $\mu(\{j\})$ ;
- tout opérateur OWA est une intégrale de Choquet particulière dont la mesure floue est donnée par :

$$\mu(A) = \sum_{i=0}^{j-1} w_{m-i}, \quad \forall A \text{ tel que } |A| = j. \quad (7.78)$$

- toute intégrale de Choquet commutative est un opérateur OWA dont les poids sont donnés par :

$$w_j = \mu(A_{m-j+1}) - \mu(A_{m-j}), \quad \forall j = 2, \dots, m, \quad (7.79)$$

$$w_1 = 1 - \sum_{j=2}^m w_j, \quad (7.80)$$

avec  $A_j$  un sous-espace de  $\mathcal{S}$  tel que  $|A_j| = j$  [70].

- les intégrales de Sugeno et de Choquet contiennent les statistiques d'ordre, en particulier le minimum, le maximum et la médiane ;
- les minimum et maximum pondérés sont des cas particuliers de l'intégrale de Sugeno dont les mesures floues correspondantes respectivement au minimum et maximum pondéré sont :

$$\mu(A) = 1 - \max_{j \notin A} w_j \quad (7.81)$$

$$\mu(A) = \max_{j \in A} w_j. \quad (7.82)$$

L'intégrale de Sugeno est aujourd'hui employée pour de la fusion (fusion de classifieurs [112]) de même que l'intégrale de Choquet (pour de la détection de mines [7]). Ces deux intégrales font toujours l'objet d'étude et de généralisation [71, 183, 96, 72].

### Opérateurs hybrides

Certains opérateurs ne peuvent pas être classés comme conjonctifs, disjonctifs ou hybrides. Nous en donnons ici quelques uns.

**Sommes symétriques** Les sommes symétriques ont été introduits par Silvert en 1979.

**Définition 7.2.9** Les opérateurs de sommes symétriques sont des opérateurs  $\phi$  définis de  $[0, 1]^2$  dans  $[0, 1]$  vérifiant :

- la continuité ;
- la non décroissance ;
- $\phi(0, 0) = 0$  ;
- $\phi(1, 1) = 1$  ;
- $\phi(a_1, a_2) = 1 - \phi(1 - a_1, 1 - a_2)$  (stabilité pour l'inversion) ;
- et telle que :

$$\phi(a_1, a_2) = \left( 1 + \frac{g(1 - a_1, 1 - a_2)}{g(a_1, a_2)} \right)^{-1}, \quad (7.83)$$

où  $g$  est une fonction croissante, continue telle que  $g(0, 0) = 0$ .

**Mélanges de t-normes et t-conormes** Il est possible de réaliser des opérateurs de mélange de t-norme et t-conorme afin d'obtenir des opérateurs de compromis. Ces opérateurs ont cependant des propriétés limitées. Zimmermann et Zysno ont proposé :

$$\phi(a_1, \dots, a_m) = \left( \prod_{j=1}^m a_j \right)^{1-\gamma} \left( 1 - \prod_{j=1}^m (1 - a_j) \right)^{\gamma}, \quad (7.84)$$

où  $\gamma \in [0, 1]$ .

**Minimum pondéré augmenté** Cet opérateur est défini par :

$$\phi(a_1, \dots, a_m) = \min_{j=1}^m w_j a_j + \epsilon \sum_{j=1}^m w_j a_j, \quad (7.85)$$

où les poids  $w_j \in [0, 1]$  sont des coefficients d'importance et  $\epsilon > 0$  définit le poids de la somme pondérée par rapport à l'opérateur de conjonction min. La variation de  $\epsilon$  permet de compenser le pessimisme de l'opérateur de conjonction.

### Opérateurs tenant compte du conflit

Dubois et Prade ont introduit en 1992 des opérateurs adaptatifs en fonction du conflit entre les distributions de possibilité qui se comportent comme un minimum si il y a peu de conflit entre les distributions et comme un maximum si le conflit est fort. Nous considérons ici deux distributions de possibilité  $\pi_1$  et  $\pi_2$  définies sur un espace de discernement  $D$  dont la combinaison donne une distribution de possibilité  $\pi'$ .

La combinaison conjonctive de ces deux distributions est bien adaptée s'il existe un recouvrement même partiel entre ces deux distributions (*i.e.* certaines décisions sont données possibles par les deux sources). En cas contraire, les sources sont en conflit. Une mesure de conflit peut être donnée par :

$$h(\pi_1, \pi_2) = 1 - \max_{d \in D} \min(\pi_1(d), \pi_2(d)), \quad (7.86)$$

qui représente le complément à 1 de la hauteur de l'intersection (min) des deux distributions. Il est possible d'envisager de normaliser cette hauteur, mais le conflit est alors masqué au même titre que pour la combinaison conjonctive de la théorie de l'évidence. Cependant cette mesure de conflit est intuitivement bien adaptée aux distributions aux formes simples telles que trapèze ou triangle, mais l'est moins pour des formes quelconques où un seul point peut entraîner une forte valeur de cette mesure de conflit même si les deux distributions ne diffèrent qu'en ce point [13].

Pour les mêmes raisons que pour les opérateurs de minimum et maximum pondérés, pour un emploi raisonnable d'un opérateur de conjonction, il est souhaitable de supposer qu'au moins une des sources est fiable.

Plusieurs opérateurs adaptatifs, repris dans [13], sont possibles :

$$\pi'(s) = \max \left( \frac{\top(\pi_1(s), \pi_2(s))}{h(\pi_1, \pi_2)}, 1 - h(\pi_1, \pi_2) \right), \quad (7.87)$$

$$\pi'(s) = \min \left( 1, \frac{\top(\pi_1(s), \pi_2(s))}{h(\pi_1, \pi_2)} + 1 - h(\pi_1, \pi_2) \right), \quad (7.88)$$

$$\pi'(s) = \top(\pi_1(s), \pi_2(s)) + 1 - h(\pi_1, \pi_2), \quad (7.89)$$

$$\pi'(s) = \max \left( \frac{\min(\pi_1(s), \pi_2(s))}{h(\pi_1, \pi_2)}, \min(\max(\pi_1(s), \pi_2(s)), 1 - h(\pi_1, \pi_2)) \right), \quad (7.90)$$

$$(7.91)$$

où  $\top$  est une t-norme. Les opérateurs donnés par les équations (7.87) et (7.88) combinent une conjonction normalisée par une distribution de valeur constante égale au conflit. Dans le cas de l'opérateur donné par l'équation (7.90), qui n'est pas associatif, lorsque le conflit est nul la combinaison est strictement conjonctive, alors que lorsque le conflit vaut 1 la combinaison est strictement disjonctive.

Si de plus, une information sur un ordenancement sur la fiabilité des sources est connu, par exemple  $S_1$  est plus fiable que  $S_2$ . Ainsi si les deux sources sont concordantes  $S_2$  peut rendre la fusion plus précise par conjonction, alors qu'au contraire si les deux sources sont en conflit, mieux vaut ne pas tenir compte de  $S_2$ . L'opérateur suivant permet de modéliser cette situation dans le cas de deux distributions à combiner :

$$\pi'(s) = \min(\pi_1(s), \max(\pi_2(s), h(\pi_1, \pi_2))). \quad (7.92)$$

Nous ne supposons ainsi connaître uniquement l'ordre des fiabilités des sources et non une mesure de fiabilité comme dans le cas des opérateurs de minimum et maximum pondérés, ce qui est moins contraignant.

Dans le même ordre d'idée, Yager [177] propose de fusionner uniquement une information sur les préférences des sources sur les décisions. Il suffit donc que chaque source fournisse un ordre de préférence entre les décisions. Ceci nécessite peu de connaissance sur les réponses des sources, mais en contrepartie ne permet pas de modéliser comme précédemment la fiabilité ou le conflit.

Ainsi ce cadre théorique offre la possibilité encore peu développée de modéliser et de tenir compte de la fiabilité et du conflit. Citons par exemple les travaux de Oussalah [121] qui propose une modification de l'opérateur donné par l'équation (7.90) afin de tenir compte progressivement du conflit et de la robustesse. Yager [179] propose de tenir compte de l'*a priori* disponible en particulier de la crédibilité des sources [178], en partant d'un opérateur de conjonction pondéré. Une discussion sur le choix des opérateurs est donnée dans [13, 60].

### 7.2.4 Décision

Une fois les informations issues des différentes sources combinées, la décision se fait à partir du maximum des degrés d'appartenance. Ainsi, nous choisirons  $d_k$  si :

$$\mu_k(x) = \max\{\mu_i(x), i = 1, \dots, n\}, \quad (7.93)$$

où  $\mu_i$  représente la fonction d'appartenance pour la décision  $d_i$  issue de la combinaison.

Il est possible d'évaluer cette décision selon deux critères [13] :

- netteté de la décision : elle est obtenue en comparant  $\mu_k(x)$  à un seuil à définir selon l'application,
- discrimination de la décision : elle est obtenue en comparant  $\mu_k(x) - \mu_{k-1}(x)$  à un seuil à définir selon l'application, où  $\mu_{k-1}(x)$  représente la deuxième plus forte fonction d'appartenance.

Ceci permet de prendre une décision de rejet si ces deux critères ne sont pas vérifiés.

## 7.3 Applications

Nous présentons ici de manière non exhaustive quelques domaines d'applications où la fusion d'informations est réalisée à partir de méthodes issues de la théorie de sous-ensembles flous et de la théorie des possibilités.

- En robotique : [108, 120] utilisent la théorie des possibilités pour la localisation de robots à l'aide de capteurs à infra rouge et odométrique. Dans [108] elle est comparée avec le filtre de Kalman.
- En traitement d'images : [139] se placent dans le cadre de la théorie des sous-ensembles flous et des possibilités pour une reconnaissance de partitions musicales à partir d'images optiques afin d'automatiser l'archivage. [16, 15] étudient la reconnaissance d'images médicales 3D du cerveau pour la détection d'anomalies. Dans [142] deux approches de combinaison (par priorité et adaptative) sont étudiées à des fins de classification d'images satellites. Pour de la vision multi-caméras [48] utilise une combinaison adaptative dans le cadre de la théorie des possibilités. [123] s'intéresse à de la restauration d'images à partir d'intégrales floues, alors que l'approche est étudiée dans [124] pour de la reconnaissance de locuteurs à partir de signal sonore. Pour de la segmentation à partir d'une représentation symbolique de la luminance et des nuances [98] proposent une segmentation floue, la fusion étant réalisée à partir de règles heuristiques. [33, 34, 32] la théorie des possibilités est retenue pour la reconnaissance d'objet dans des images sonar. Dans le cadre de la classification d'images sonar [111] compare une fusion des résultats des classifieurs à partir de méthodes issues de la théorie des possibilités, de la théorie des croyances et des votes.
- En classification : Citons la classification d'objets sous-marins [40], de comportements [118] ou de données radar [112] où l'intégrale de Sugeno et différentes combinaisons (conjonctives et disjonctives) sont dans ce dernier comparées.

# Glossaire

- **Ambiguïté**

L'ambiguïté d'une information est le fait que celle-ci entraîne deux ou plus interprétations. L'ambiguïté peut provenir d'une autre imperfection de l'information, mais pas nécessairement.

- **Argument**

Un argument désigne un témoignage, un signe ou une circonstance qui a un lien avec la chose à prouver. Un argument est pur s'il prouve une proposition dans certains cas et ne dit rien dans les autres cas *i.e.* l'argument ne permet pas de dire quoi que ce soit sur son contraire. Un argument est mixte s'il prouve une proposition dans certains cas et prouve le contraire dans d'autres.

- **Confusion**

La confusion est un type d'incohérence faible, par exemple un radar annonce l'arrivée d'un avion à 16h15, alors qu'il arrive à 16h.

- **Conflit**

Deux informations ou plus sont en conflit si elles sont contradictoires.

- **Complémentarité**

Des sources ou des informations sont complémentaires si elles sont de nature différente et permettent de préciser la réponse du système. Elle peut se mesurer à l'aide de l'entropie conditionnelle.

- **Crédibilité** ou **Confiance**

La crédibilité d'une source ou la confiance en une source quantifie la confiance que l'on peut porter à la *source*. Elle doit tenir compte de toutes les imperfections de la source. La mesure de crédibilité est un degré de confiance d'une source ou de la sortie d'un système de fusion. Attention ce terme peut être employé dans la théorie de Dempster-Shafer pour la croyance.

- **Croyance**

La croyance d'une information est une mesure qui quantifie la confiance qui peut être portée à l'*information*. Au sens de Dempster-Shafer cette mesure est définie précisément. Attention on peut trouver le terme de crédibilité ou d'évidence pour cette mesure.

- **Fiabilité**

La fiabilité est relative à la source. Un degré de fiabilité permet de mesurer l'incertitude de l'information apportée par la source et/ou l'incertitude de la source à donner une information.

– **Incertitude**

L'incertitude caractérise le degré de conformité à la réalité d'une information, ou encore l'assurance d'une source en l'information fournie. Une information incertaine décrit donc une connaissance partielle de la réalité. Pour la mesurer, la réalité doit donc être connue. *Exemple* : Cette lettre arrivera demain.

– **Incohérence**

Deux informations sont incohérentes si elles ne sont pas conforme à une réalité que nous avons. L'incohérence des informations peut ainsi entraîner un conflit. Un type d'incohérence particulier est l'incohérence temporelle, par exemple une source peut annoncer *le train est en gare* alors qu'une autre peut annoncer *le train a 10 min de retard*.

– **Inconsistance**

L'inconsistance représente à la fois le conflit (par exemple une source sonar détecte une mine à un endroit, alors que le sondeur décrit une matière rocheuse), l'incohérence (le conflit entraîne une incohérence dans la conclusion *existence d'une mine en matière rocheuse*), l'incohérence temporelle (par exemple un radar détecte une nappe de pétrole sur la surface de l'eau à 16h, alors qu'un autre radar décrit la surface sans pollution à 16h15) et la confusion qui est une incohérence plus faible (par exemple un radar annonce l'arrivée d'un avion à 16h15, alors qu'il arrive à 16h).

– **Information**

L'information est relative soit à la source soit aux données, caractéristiques ou décisions selon le niveau de fusion. Elle peut être de type numérique ou symbolique. L'information concerne ici tout élément pouvant être codé et fusionné.

– **Imprécision**

L'imprécision d'une information est caractérisée par le contenu de l'information. Elle est relative à l'information ou à la source. Elle se mesure par l'erreur de l'information fournie (défaut quantitatif de l'information). La réalité doit donc être connue ou estimée. Nous distinguons l'imprécision sans erreur (*il pleut beaucoup*) et l'imprécision avec erreur *il pleut 10 mm<sup>3</sup>/h* alors qu'il y en a 15).

– **Incomplétude**

L'incomplétude mesure le manque d'information apporté par la source. Elle peut se mesurer par la différence de la quantité d'information réellement fournie par la source et de la quantité d'information que la source doit fournir (en cas de bon fonctionnement ou pour répondre à un problème posé).

– **Monde ouvert, fermé**

Lorsque l'expert connaît toutes les possibilités de décision du problème, il est dans un monde fermé ou clos. Au contraire, s'il peut envisager d'autres réponses *a priori* non attendues, il doit admettre qu'il est dans un monde ouvert.

– **Nécessité**

La nécessité d'un évènement est la forme dual de la possibilité d'un évènement. La nécessité d'un évènement est l'impossibilité de son contraire. Dans le cadre de la théorie des possibilités, une mesure est associée à la nécessité d'un évènement.

– **Possibilité**

La possibilité concerne la capacité d'un évènement à se produire. Seuls les évènements possibles peuvent être probables. Dans le cadre de la théorie des possibilités, une mesure est associée à la possibilité d'un évènement.

– **Redondance**

La redondance d'informations ou de sources est le fait d'avoir plusieurs fois la même information. Elle peut se mesurer à l'aide de l'entropie. Cette imperfection, n'en est pas systématiquement une ; dans le cadre de la fusion on s'appuie sur la redondance des sources pour confirmer une information.



# Bibliographie

- [1] Data Fusion Lexicon, Data Fusion Subpanel of the Joint Directors of Laboratories Technical Panel for C<sup>3</sup>. Rapport technique, F.E. White, Code 42402, NOSC, San Diego, Californie, 1991.
- [2] P. AARABI : Equivalency of Bayesian decision fusion and neural networks. *In Sensor Fusion : Architectures, Algorithms and Application V*, volume 4385, pages 67–76, Orlando, USA, Avril 2001.
- [3] A. APPRIOU : Probabilités et incertitudes en fusion de données multisenseurs. *Revue Scientifique et Technique de la Défense*, 11:27–40, 1991.
- [4] A. APPRIOU : Formulation et traitement de l'incertain en analyse multisenseurs. *In GRETSI*, pages 951–954, Juan-les-Pins, France, Septembre 1993.
- [5] A. APPRIOU : Situation Assesment Based on Spatially Ambiguous Multisensor Measurements. *International Journal of Intelligent Systems*, 16(10):1135–1166, octobre 2001.
- [6] A. APPRIOU : *Décision et Reconnaissance des formes en signal*, chapitre Discrimination multisignal par la théorie de l'évidence, pages 219–258. Hermes Science Publication, 2002.
- [7] S. AUEPHANWIRIYAKUL, J.M. KELLER et P.D GADER : Generalized Choquet fuzzy integral fusion. *Information Fusion*, 3:69–85, 2002.
- [8] A. BENDJEBBOUR et W. PIECZYNSKI : Segmentation d'images multisenseur par fusion évidentielle dans un contexte Markovien. *Traitement du Signal*, 14(5):453–464, 1997.
- [9] I. BLOCH : Fondements des probabilités et des croyances : une discussion des travaux de Cox et Smets. *In GRETSI*, pages 909–912, Juan-les-Pins, France, Septembre 1995.
- [10] I. BLOCH : Incertitude, imprécision et additivité en fusion de données : point de vue historique. *Traitement du Signal*, 13(4):267–288, 1996.
- [11] I. BLOCH : Information Combination Operators for Data Fusion : A Comparative Review with Classification. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 26(1):52–67, Janvier 1996.
- [12] I. BLOCH : Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters*, 17:905–919, 1996.

- [13] I. BLOCH : *Fusion d'informations en traitement du signal et des images*. Hermes Science Publication, Janvier 2003.
- [14] I. BLOCH : *Traitement de données complexes et commande en logique floue*, chapitre Traitement d'images, pages 219–258. Hermes Science Publication, 2003.
- [15] I. BLOCH, O. COLLIOT, O. CAMARA et T. GÉRAUD : Fusion of spatial relationships for buiding recognition, example of brain structure recognition in 3d mri. *Pattern Recognition Letters*, à paraître 2005.
- [16] I. BLOCH, T. GÉRAUD et H. MAÎTRE : Representation and fusion of heterogeneous fuzzy information in the 3d space for model-based structural recognition - application to 3d brain imaging. *Artificial Intelligence*, 148:141–175, 2003.
- [17] I. BLOCH et H. MAITRE : Fusion de données en traitement d'images : modèles d'information et décisions. *Traitement du Signal*, 11(6):435–446, 1994.
- [18] B. BOUCHON-MEUNIER, L. FOULLOY et M. RAMDANI : *Logique floue, Exercices corrigées et Exemples d'applications*. Janvier 1998.
- [19] B. BOUCHON-MEUNIER et C. MARSALA : *Logique floue, principes, aide à la décision*. Hermes Science Publication, Janvier 2003.
- [20] B. BOUCHON-MEUNIER et C. MARSALA : *Traitement de données complexes et commande en logique floue*. Hermes Science Publication, Janvier 2003.
- [21] A.O. BOUDRAA, A. BENTABET, F. SALZENSTEIN et L. GUILLON : Dempster-Shafer's Basic Probability Assignment Based on Fuzzy Membership Functions. *Electronic Letters On Computer Vision and Image Analysis*, 4(1):1–9, 2004.
- [22] J.J. BRAUN : Sensor data fusion with support vector machine techniques. *In Sensor Fusion : Architectures, Algorithms and Application VI*, volume 4731, pages 98–109, Orlando, USA, Avril 2002.
- [23] A.L. BUCZAC et R.E. UHRIG : Hybrid fuzzy-genetic Technique for Multisensor Fusion. *Intelligent Systems*, 93:265–281, 1996.
- [24] A.S. CAPELLE et C. FERNANDEZ-MALOIGNE : Segmentation of brain tumors by evidence theory : on the use of the conflict information. *In International Conference on Information Fusion*, Stockholm, Sweden, Juin 2004.
- [25] M. CHACHOUA : *Une Théorie Symbolique de l'Entropie pour le Traitement des Informations Incertaines*. Thèse de doctorat, Université d'Angers, 1998.
- [26] S. CHAUVIN : évaluation des performances du modèle Bayésien de fusion appliqué à l'imagerie satellitaire. *In GRETSI*, Juan-les-Pins, France, Septembre 1995.
- [27] J.-G. CHEN et N. ANSARI : Adaptative Fusion of Correlated Local Decision. *IEEE Transactions on Systems, Man, and Cybernetics - Part C : Applications and Reviews*, 28(2):276–281, Mai 1998.
- [28] D.A. CLAUSI : Comparison and Fusion of Co-occurrence, Gabor and MRF Texture Features for Classification of SAR Sea-Ice Imagery. *Atmosphere-Ocean*, 39(3):183–194, 2000.

- [29] D.A. CLAUSI et H. DENG : Fusion of Gabor Filter and Co-occurrence Probability Features for Texture Recognition. *In 16th International conference on Vision Interface*, pages 237–242, Halifax, Canada, Juin 2003.
- [30] A. CLÉRENTIN, L. DELAHOUCHE, E. BRASSART et C. DROCOURT : Une architecture de propagation de l'incertitude pour le problème de la localisation. *Revue des Sciences et Technique de l'Information - Technique et Sciences Informatiques*, 22(7-8):903–933, 2003.
- [31] F. CUZZOLIN : Geometry of Dempster's Rule of Combination. *IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics*, 34(2):961–977, 2004.
- [32] S. DANIEL : *Fusion multisource appliquée à la reconnaissance d'objets dans le milieu sous-marin*. Thèse de doctorat, Université de Rennes 1, Novembre 1998.
- [33] S. DANIEL, B. SOLAIMAN et E.P. MAILLARD : A Global Framework for Object Recognition Through Sonar Data Fusion. *In International Conference on Information Fusion*, volume 2, pages 865–872, Las Vegas, USA, Juin 1998.
- [34] S. DANIEL, B. SOLAIMAN et E.P. MAILLARD : Object Recognition on the Seabottom Using Possibility Theory and Sonar Data Fusion. *In International Conference on Information Fusion*, volume 2, pages 396–402, Las Vegas, USA, Juin 1998.
- [35] B.V. DASARATHY : Sensor Fusion Potential Exploitation - Innovative Architectures and Illustrative Applications. *Proceeding of the IEEE*, 85(1):24–38, Janvier 1997.
- [36] B.V. DASARATHY : Information fusion - what, where, why, when, and how ? *Information Fusion*, 2:75–75, 2001.
- [37] B.V. DASARATHY : Image fusion in the context of Aerospace Applications. *Information Fusion*, 3:1, 2002.
- [38] F. DELMOTTE : Un nouvel opérateur de fusion adaptatif. *Traitement du Signal*, 17(4):299–311, 2000.
- [39] F. DELMOTTE et P. BORNE : Modeling of Reliability with Possibility Theory. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 28(1):78–88, Janvier 1998.
- [40] M. DELPLANQUE, A.M. DESOLT-JOLLY et J. JAMIN : Fusion dissymétrique d'informations incomplètes pour la classification d'objets sous-marins. *Traitement du Signal*, 14(5):511–522, 1997.
- [41] A.P. DEMPSTER : Uper and Lower probabilities induced by a multivalued mapping. *Anal. of Mathematical Statistics*, 38:325–339, 1967.
- [42] T. DENOEU : A  $k$ -Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 25(5):804–813, Mai 1995.
- [43] T. DENOEU : Application du modèle des croyances transférables en reconnaissance de formes. *Traitement du Signal*, 14(5):443–451, 1997.

- [44] T. DENOEU : A Neural Network Classifier Based on Dempster-Shafer Theory. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 30(2):131–150, Mars 2000.
- [45] T. DENOEU et M.H. MASSON : EVCLUS : Evidential Clustering of Proximity Data. *IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics*, 34(1):95–109, Avril 2004.
- [46] J. DESACHY, L. ROUX et L.-H. ZAHZAH : Numeric and symbolic data fusion : A soft computing approach to remote sensing images analysis. *Pattern Recognition Letters*, 17:1361–1378, 1996.
- [47] A.-M. DESOST-LEBRUN : Fusion de données. *Techniques de l'Ingénieur, traité Mesure et Contrôle*, (136):1–9, Décembre 1996.
- [48] S. DEVEUGHELE et B. DUBUISSON : Adaptabilité et combinaison possibiliste : application à la vision multi-caméras. *Traitement du Signal*, 11(6):559–568, 1994.
- [49] J. DEZERT : Foundations for a new theory of plausible and paradoxical reasoning. *Information & Security. An International Journal*, 9, 2002.
- [50] J. DEZERT et F. SMARANDACHE : Fusion of imprecise, uncertain an conflicting beliefs with DSm rules of combination. *arXiv*, pages 1–14, 2004.
- [51] J. DEZERT, F. SMARANDACHE et M. DANIEL : The Generalized Pignistic Transformation. *In Seventh International Conference on Information Fusion*, Stockholm, Sweden, Juin 2004.
- [52] A. DROMIGNY-BADIN, S. ROSSATO et Y.M. ZHU : Fusion de données radioscopiques et ultrasonores via la théorie de l'évidence. *Traitement du Signal*, 14(5):499–510, 1997.
- [53] A. DROMIGNY-BADIN et Y.M. ZHU : Amélioration de la dynamique des systèmes d'imagerie à rayon X par fusion bayésienne. *In Journée thématiques en fusion d'information*, pages 40–44, Paris, France, Octobre 1996. GDR-ISIS.
- [54] D. DUBOIS, M. GRABISCH, O. De MOUZON et H. PRADE : *Logique floue, principes, aide à la décision*, chapitre Classification et diagnostic, pages 219–258. Hermes Science Publication, 2003.
- [55] D. DUBOIS, W. OSTASIEWICZ et H. PRADE : *The Handbooks of Fuzzy Sets Series*, chapitre Fundamentals of Fuzzy Sets, pages 21–124. Kluwer Academic Publishers, Boston, 2000.
- [56] D. DUBOIS et H. PRADE : *Théorie des possibilités*. Masson, Novembre 1987.
- [57] D. DUBOIS et H. PRADE : Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4:244–264, 1988.
- [58] D. DUBOIS et H. PRADE : La fusion d'informations imprécises. *Traitement du Signal*, 11(6):447–458, 1994.
- [59] D. DUBOIS et H. PRADE : Possibility theory and data fusion in poorly informed environments. *IFAC*, 2(5):811–823, 1994.

- [60] D. DUBOIS et H. PRADE : On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142:143–161, 2004.
- [61] M.-P. DUBUISSON-JOLY et A. GUPTA : Color and Texture Fusion : Application to Aerial Image Segmentation and GIS Updating. In *IEEE workshop on Applications of Computer Vision*, pages 2–7, 1996.
- [62] C. DURIEU, M.-J. ALDON et D. MEIZEL : La fusion de données multisensorielles pour la localisation en robotique mobile. *Traitement du Signal*, 13(2):143–166, 1996.
- [63] Z. ELOUEDI, K. MELLOULI et Ph. SMETS : Assessing Sensor Reliability for Multi-sensor Data Fusion Within The Transferable Belief Model. *IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics*, 34(1):782–787, Février 2004.
- [64] S. FABRE : *Apport de l'information contextuelle à la fusion multicapteurs. Application à la fusion pixel*. Thèse de doctorat, École Nationale Supérieure de l'Aéronautique et de l'Espace, Août 1999.
- [65] S. FABRE, A. APPRIOU et X. BRIOTTET : Comparison of two Integration Methods of Contextual Information in Pixel Fusion. In *International Conference on Information Fusion*, pages 831–838, California, USA, Juin 1999.
- [66] N. FERAY et D. DE BRUCQ : Fusion d'information provenant d'une famille d'arbres de décision. application à la reconnaissance de chiffre manuscrits. In *Journée thématiques en fusion d'information*, pages 4–13, Paris, France, Octobre 1996. GDR-ISIS.
- [67] J.M. FERNÁNDEZ SALIDO et S. MURAKAMI : Extending Yager's orness concept for the OWA aggregators to other mean operators. *Fuzzy Sets and Systems*, 139:515–542, 2003.
- [68] L. FOUQUE et A. APPRIOU : An evidential Markovian model for data fusion and unsupervised image classification. In *International Conference on Information Fusion*, Paris, France, Juin 2000.
- [69] A. GATTEIN et P. VANNOORENBERGHE : A Comparative Analysis Of Two Approches Using The Road Network For Tracking Ground Targets. In *International Conference on Information Fusion*, Stockholm, Sweden, Juin 2004.
- [70] M. GRABISCH : Fuzzy intergral in multicriteria decision making. *Fuzzy Sets and Systems*, 69:279–298, 1995.
- [71] M. GRABISCH : The symetric Sugeno integral. *Fuzzy Sets and Systems*, 139:473–490, 2003.
- [72] M. GRABISCH et C. LABREUCHE : Bi-capacities - II : the Choquet integral. *Fuzzy Sets and Systems*, à paraître 2005.
- [73] M. GRABISCH et P. PERNY : *Logique floue, principes, aide à la décision*, chapitre Agrégation multicritère, pages 219–258. Hermes Science Publication, 2003.
- [74] D. GRUYER, C. ROYÈRE et V. CHERFAOUI : Utilisation de la théorie des croyances pour l'association dans le suivi de cibles. *Revue des Sciences et Technique de l'Information - Technique et Sciences Informatiques*, 22(7-8):983–1010, 2003.

- [75] R. HAENNI : Are alternatives to Dempster's rule of combination real alternatives? Comments on "About the belief function combination and conflict management problem" - Lefevre et al. *Information Fusion*, 3:237–239, 2002.
- [76] D.L. HALL : *Mathematical Techniques in Multi-sensor Data Fusion*. ARTECH HOUSE, INC, 1992.
- [77] D.L. HALL et J. LLINAS : An Introduction to Multisensor Data Fusion. *Proceeding of the IEEE*, 85(1):6–23, Janvier 1997.
- [78] C. HARRIS, X. HONG et Q. GAN : *Adaptative Modelling, Estimation and Fusion from Data*. Springer, 2002.
- [79] S.K. HAUTANIEMI et J.P.P. SAARINEN : Multitarget tracking with the IMM and Bayesian networks : empirical studies. In *Sensor Fusion : Architectures, Algorithms and Application V*, volume 4385, pages 47–57, Orlando, USA, Avril 2001.
- [80] F. JANEZ : *Fusion de sources d'information définies sur des référentiels non exhaustifs différents*. Thèse de doctorat, Université d'Angers, Novembre 1996.
- [81] F. JANEZ : Rappels sur la théorie de l'évidence. Rapport technique, Note Technique ONERA, Janvier 1997.
- [82] F. JANEZ et A. APPRIOU : Théorie de l'Evidence et cadres de discernement non exhaustifs. *Traitement du Signal*, 13(3):237–250, 1996.
- [83] F. JANEZ et A. APPRIOU : Theory of evidence and non-exhaustive frames of discernment : Plausibilities correction methods. *International Journal of Approximate Reasoning*, 18:1–19, 1998.
- [84] I. JARKASS et M. ROMBAULT : Reconnaissance crédibiliste de séquences temporelles. *Revue des Sciences et Technique de l'Information - Technique et Sciences Informatiques*, 22(7-8):965–981, 2003.
- [85] A. JOSANG : The Consensus Operator for Combinning Beliefs. *Artificial Intelligence Journal*, 141(1-2):157–170, 2002.
- [86] A.L. JOUSSELME, E. BOSSÉ et A. JOUAN : Analysing an Identity Information Fusion Algorithm Based on Evidence Theory. In *Reasearch and Technology Organisation meeting : Military Data and Information Fusion*, Prague, Czech Republic, Octobre 2003.
- [87] A.L. JOUSSELME, D. GRENIER et E. BOSSÉ : A new distance between two bodies of evidence. *Information Fusion*, 2:91–101, 2001.
- [88] R. KENNES : Computational Aspect of the Möbius Transformation of Graphs. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 22(2):201–223, 1992.
- [89] J. KITTLER et F.M. ALKOOT : Sum versus Vote Fusion in Multiple Classifier Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1): 110–115, 2003.

- [90] G.J. KLIR : Mesures of uncertainty in the Dempster-Shafer theory of evidence. In R.R. YAGER, J. KACPRZYK et M. FEDRIZZI, éditeurs : *Advances in the Dempster-Shafer theory of evidence*, chapitre 2, pages 35–49. John Wiley & Sons, Inc., 1994.
- [91] L.I. KUNCHEVA : Switching Between Selection and Fusion in Combining Classifiers : An Experiment. *IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics*, 32(2):146–156, Avril 2002.
- [92] L.I. KUNCHEVA : A Theoretical Study on Six Classifier Fusion Strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):281–286, Février 2002.
- [93] L.I. KUNCHEVA, J.C. BEZDEK et R.P.W. DUIN : Decision Templates for Multiple Classifier Fusion : An Experimental Comparison. *Pattern Recognition*, 34(2):299–314, 2001.
- [94] L.I. KUNCHEVA, M. SKURICHINA et R.P.W. DUIN : An experimental study on diversity for bagging and boosting with linear classifiers. *Information Fusion*, 3(4):245–258, Décembre 2002.
- [95] L.I. KUNCHEVA, C.J. WHITTAKER et C.A. SHIPP : Limits on the Majority Vote Accuracy in Classifier Fusion. *Pattern Analysis and Applications*, 6:22–31, 2003.
- [96] C. LABREUCHE et M. GRABISCH : The Choquet integral for the aggregation of interval scales in multicriteria decision making. *Fuzzy Sets and Systems*, 137:11–26, 2003.
- [97] L. LAM et C.Y. SUEN : Application of Majority Voting to Pattern Recognition : An Analysis of Its Behavior and Performance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 27:553–568, 1997.
- [98] P. LAMBERT et T. CARRON : Symbolic fusion of luminance-hue-chroma features for region segmentation. *Pattern Recognition*, 32:1857–1872, 1999.
- [99] S. LE HÉGARAT-MASCLE, I. BLOCH et D. VIDAL-MADJAR : Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing ON GEOSCIENCE AND REMOTE SENSING*, 35(4):1018–1031, Juillet 1997.
- [100] S. LE HÉGARAT-MASCLE, I. BLOCH et D. VIDAL-MADJAR : Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover. *Pattern Recognition*, 31(11):1811–1823, 1998.
- [101] S. LE HÉGARAT-MASCLE et R. SELTZ : Automatic change detection by evidential fusion of change indices. *Remote Sensing of Environment*, 91:390–404, 2004.
- [102] F. LEDUC et B. Solaiman F. CAVAYAS : Combinaison of fuzzy sets and Dempster-Shafer theories in forest map updating using multispectral data. In *Sensor Fusion : Architectures, Algorithms and Application V*, volume 4385, pages 323–334, Orlando, USA, Avril 2001.

- [103] E. LEFEVRE, O. COLOT et P. VANNOORENBERGHE : A classification method based on Dempster-Shafer's theory and information criteria. *In International Conference on Information Fusion*, pages 1179–1186, California, USA, Juin 1999.
- [104] E. LEFEVRE, O. COLOT et P. VANNOORENBERGHE : Belief function combination and conflict management. *Information Fusion*, 3:149–162, 2002.
- [105] E. LEFEVRE, O. COLOT et P. VANNOORENBERGHE : Reply to the Comments of R. Haenni on the paper "Belief function combination and conflict management". *Information Fusion*, 4:63–65, 2002.
- [106] E. LEFEVRE, P. VANNOORENBERGHE et O. COLOT : Using Information Criteria in Dempster-Shafer's Basic Belief Assignment. *In International Conference on Fuzzy Systems*, pages 173–178, Séoul, Corée, Août 1999.
- [107] P. LOONIS et M. MÉNARD : Synthèse des approches de classement en reconnaissance des formes. Nouveaux outils pour l'adaptation d'un système à son environnement. *Traitement du Signal*, 14(5):465–483, 1997.
- [108] H. MAAREF, M. OUSSALAH et C. BARRET : Fusion de données capteurs en vue de la localisation absolue d'un robot mobile par une méthode basée sur la théorie des possibilités. Comparaison avec le filtre de Kalman. *Traitement du Signal*, 16(5):345–359, 1999.
- [109] P.F. MARTEAU et G. MÉNIER : Alignement approximatif d'arbres pour la recherche d'information en contexte dans les données XML hétérogènes. fusion d'informations structurées et textuelles. *Technique et sciences informatiques*, 22(7-8):1011–1034, 2003.
- [110] A. MARTIN : L'analyse de données, Septembre 2004.
- [111] A. MARTIN : Fusion d'informations haut niveau - Application à la classification d'images sonar. *In EGC*, Paris, France, Janvier 2005.
- [112] A. MARTIN et E. RADOI : Effective ATR Algorithms Using Information Fusion Models. *In International Conference on Information Fusion*, Stockholm, Sweden, Juin 2004.
- [113] A. MARTIN, G. SÉVELLEC et I. LEBLOND : Characteristics vs decision fusion for sea-bottom characterization. *In Journée d'Acoustique Sous-Marine*, Brest, France, Octobre 2004.
- [114] J. MAUCLAIR et J. PINQUIER : Fusion of descriptors for speech / music classification. *In European Signal Processing Conference, EUSIPCO*, pages 1285–1288, Vienna, Austria, Septembre 2004.
- [115] F. MAUSSANG, M. ROMBAULT, J. CHANUSSOT, A. HÉTET et M. AMATE : Détection en Imagerie SAS par Fusion de Données de Statistiques Locales. *In Journée d'Acoustique Sous-Marine*, Brest, France, Octobre 2004.
- [116] N. MILISAVJEVIC, I. BLOCH, S. VAN DEN BROEK et M. ACHEROY : Improving mine recognition through processing and dempster-shafer fusion of ground-penetrating radar data. *Pattern Recognition*, 36:1233–1250, 2003.

- [117] H.T. NGUYEN et E.A. WALKER : On decision making using belief functions. *In* R.R. YAGER, J. KACPRZYK et M. FEDRIZZI, éditeurs : *Advances in the Dempster-Shafer theory of evidence*, chapitre 14, pages 311–330. John Wiley & Sons, Inc., 1994.
- [118] A. NIFLE et R. REYNAUD : Classification de comportements fondée sur l'occurrence d'évènements en théorie des possibilités. *Traitement du Signal*, 14(5):523–534, 1997.
- [119] M. OUSSALAH : On the use of Hamacher's t-norms family for information aggregation. *Information Sciences*, 153:107–154, 2003.
- [120] M. OUSSALAH, H. MAAREF et C. BARRET : New fusion methodology approach and application to mobile robotics : investigation in framework of possibility theory. *Information Fusion*, 2:31–48, 2001.
- [121] M. OUSSALAH, H. MAAREF et C. BARRET : From adaptative to progressive combination of possibility distributions. *Fuzzy Sets and Systems*, 139:559–582, 2003.
- [122] S. PERRIN, E. DUFLOS, P. VANHEEGHE et A. BIBAUT : Multisensor Fusion in the Frame of Evidence Theory for Landmines Detection. *IEEE Transactions on Systems, Man, and Cybernetics - Part C : Applications and Reviews*, 34(4):485–498, Novembre 2004.
- [123] T.D. PHAM : An image restoration by fusion. *Pattern Recognition*, 34:2403–2411, 2001.
- [124] T.D. PHAM et M. WAGNER : Similarity normalization for speaker verification by fuzzy fusion. *Pattern Recognition*, 33:309–315, 2000.
- [125] E. PIAT et D. MEIZEL : Proposition d'un cadre probabiliste de fusion de croyances. *Traitement du Signal*, 14(5):485–498, 1997.
- [126] L. PIEGON : Genetic algorithms for multiagent fusion system learning. *In* *Sensor Fusion : Architectures, Algorithms and Application V*, volume 4385, pages 87–95, Orlando, USA, Avril 2001.
- [127] L. PIEGON, B. SOLAIMAN, T. TOUTIN et K.P.B. THOMSON : Dempster-Shafer theory for multi-satellite remotely sensed observations. *In* *Sensor Fusion : Architectures, Algorithms and Application IV*, volume 4051, pages 228–236, Orlando, USA, Avril 2000.
- [128] G. POWELL, D. MARSHALL, R. MILIKEN et K. MARKHAM : A Data Fusion System for Object Recognition based on Transferable Belief Models and Kalman Filters. *In* *International Conference on Information Fusion*, Stockholm, Sweden, Juin 2004.
- [129] D. QINGDONG, X. LINGYU et Z. HAI : D-S Evidence Theory Applied to Fault Diagnosis Based on Embedded Sensors. *In* *International Conference on Information Fusion*, Paris, France, Juin 2000.
- [130] I. QUIDU, J.-Ph. MALKASSE et G. BUREL : Classification de mines sous-marines à partir de l'image sonar brute : caractérisation du contour de l'ombre portée par algorithme génétique. *Traitement du Signal*, 20(1):55–76, 2003.

- [131] RESEARCH AND TECHNOLOGY ORGANISATION et NORTH ATLANTIC TREATY ORGANISATION : Massive Military Data Fusion and Visualisation : Users Talk with Developers. *In RTO Meeting proceedings*, Halden, Norway, Septembre 2002.
- [132] RESEARCH AND TECHNOLOGY ORGANISATION et NORTH ATLANTIC TREATY ORGANISATION : Military Data and Information Fusion. *In RTO Meeting proceedings*, Prague, Czech republic, Octobre 2003.
- [133] R. REYNAUD : La fusion de données, du capteur au raisonnement. *Traitement du Signal*, 11(6):431–434, 1994.
- [134] B. RISTIC et Ph. SMETS : Belief theory on the continuous space with an application to model based classification. *In Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1119–1126, Perugia, Italy, Juillet 2004.
- [135] S.A. RIZVI et N.M. NASRABADI : Fusion of FLIR automatic target recognition algorithms. *Information Fusion*, 4:247–258, 2003.
- [136] M. ROMBAUT : Fusion de données images segmentées à l’aide du formalisme de Dempster Shafer. *In GRETSI*, pages 655–658, Vannes, France, Septembre 1999.
- [137] M. ROMBAUT et Y.M. ZHU : Study of Dempster-Shafer theory for image segmentation applications. *Image and Vision Computing*, 20:15–23, 2002.
- [138] K.N. ROSS et R.D. CHANEY : Hidden Markov Models for threat prediction fusion. *In Sensor Fusion : Architectures, Algorithms and Application IV*, volume 4051, pages 300–311, Orlando, USA, Avril 2000.
- [139] F. ROSSANT et I. BLOCH : A fuzzy model for optical recognition of musical scores. *Fuzzy Sets and Systems*, 141:165–201, 2004.
- [140] S. ROSSATO, A. DROMIGNY-BADIN, Y.M. ZHU, V. KAFTANDJIAN, C. ODET, P. GUY, Y. JAYET, G. PEIX et G. GIMENEZ : Amélioration de la fiabilité du contrôle non destructif par fusion de données radioscopiques et ultrasonores. *In Journée thématiques en fusion d’information*, pages 54–58, Paris, France, Octobre 1996. GDR-ISIS.
- [141] C. ROUCHOUZE : Fusion de données : exemple Défense et axes de recherche. *Traitement du Signal*, 11(6):459–464, 1994.
- [142] L. ROUX et J. DESACHY : Fusion d’un grand nombre de sources : un aperçu général des méthodes disponibles en théorie des possibilités. *In Journée thématiques en fusion d’informations*, pages 45–49, Paris, France, Octobre 1996. GDR-ISIS.
- [143] V. ROYER et B. HOELTZENER : Aide à la reconnaissance d’objets à partir d’images. Aspects formel de la fusion numérique symbolique. *In Journée thématiques en fusion d’information*, pages 25–34, Paris, France, Octobre 1996. GDR-ISIS.
- [144] C. ROYÈRE, D. GRUYER et V. CHERFAOUI : Data association with belief theory. *In International Conference on Information Fusion*, Paris, France, Juin 2000.
- [145] G. SAPORTA : Data Fusion and Data Pruning. *In International meeting on Nonlinear Methods and Data Mining*, Rome, Italie, Septembre 2000.

- [146] G. SHAFER : *A mathematical theory of evidence*. Princeton University Press, 1976.
- [147] G. SHAFER et R. LOGAN : Implementing Dempster's Rule for Hierarchical Evidence. *Artificial Intelligence*, 33:271–298, 1987.
- [148] C.E. SHANNON : A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, Octobre 1948.
- [149] T. SILVENNOINEN, P. KORPISAARI et J. SAARINEN : Empirical Investigations on Dempster-Shafer Reasoning System in Attribute Fusion. *In Sensor Fusion : Architectures, Algorithms and Application IV*, volume 4051, pages 247–266, Orlando, USA, Avril 2000.
- [150] Ph. SMETS : The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.
- [151] Ph. SMETS : Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence*, 5:29–39, 1990.
- [152] Ph. SMETS : The variety of ignorance and the need for well-founded theories. *Information Sciences*, 57-58:135–144, 1991.
- [153] Ph. SMETS : Belief Functions : the Disjunctive Rule of Combination and the Generalized Bayesian Theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.
- [154] Ph. SMETS : The  $\alpha$ -junctions : the commutative and associative non interactive combination operators applicable to belief functions. *In D. GABBAY, R. KRUSE, A. NONNENGART et H.J. OHLBACG, éditeurs : Qualitative and quantitative practical reasoning*, pages 131–153. Springer Verlag, Berlin, 1997.
- [155] Ph. SMETS : Imperfect information : Imprecision - Uncertainty. *In A. MOTRO et Ph. SMETS, éditeurs : Uncertainty Management in Information Systems*, pages 225–254. Kluwer Academic Publishers, 1997.
- [156] Ph. SMETS : *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, chapitre Numerical Representation of Uncertainty, pages 265–309. Kluwer, Dordrecht, 1998.
- [157] Ph. SMETS : Data Fusion in the Transferable Belief Model. *In International Conference on Information Fusion*, volume 1, pages 21–33, Paris, France, Juillet 2000.
- [158] Ph. SMETS : Matrix Calculus for Belief Functions. *International Journal of Approximate Reasoning*, 31:1–30, 2002.
- [159] Ph. SMETS : Belief Function on Continuous Spaces. *soumis*, 2004.
- [160] Ph. SMETS et R. KENNES : The Transferable Belief Model. *Artificial Intelligence*, 66:191–234, 1994.
- [161] B. SOLAIMAN, L. LECORNU et C. ROUX : Edge detection through information fusion using fuzzy and evidential reasoning concepts. *In Sensor Fusion : Architectures, Algorithms and Application IV*, volume 4051, pages 267–278, Orlando, USA, Avril 2000.

- [162] A. TALEB-AHMED, L. GAUTIER et M. ROMBAUT : Architecture de fusion de données basée sur la théorie de l'évidence pour la reconstruction d'un vertèbre. *Traitement du Signal*, 19(4):267–283, 2002.
- [163] P.K. VARSHNEY : Special Issue on Data Fusion. *Proceeding of the IEEE*, 85(1):3–5, Janvier 1997.
- [164] P. VASSEUR, C. PEGARD, E. MOUADDIB et L. DELAHOUCHE : Perceptual organization approach based on Dempster-Shafer theory. *Pattern Recognition*, 32:1449–1462, 1999.
- [165] P. VERLINDE : *A Contribution to Multi-Modal Identity Verification using Decision Fusion*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, Paris, France, 1999.
- [166] L. WALD : Definitions and terms of reference in data fusion. *In International Archives of Photogrammetry and Remote Sensing*, volume 32 de part 7-4-3 W6, Valladolid, Espagne, Juin 1999.
- [167] L. WALD : Some Terms of Reference in Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1190–1193, Mai 1999.
- [168] L. WALD : *Data Fusion - Definitions and architectures - Fusion of images of different spacial resolutions*. Les Presses de l'École des Mines, 2002.
- [169] P. WALLEY : Imprecise probabilities. *In* C. B. READ, D. L. BANKS et S. KOTZ, éditeurs : *The Encyclopedia of Statistical Sciences*. Wiley, New York (NY), 1997.
- [170] Z. WANG, K. LEUNG et J. WANG : A genetic algorithm for determining nonadditive set functions in information fusion. *Fuzzy Sets and Systems*, 102:463–469, 1999.
- [171] T. WHALEN : Interval probabilities induced by decision problems. *In* R.R. YAGER, J. KACPRZYK et M. FEDRIZZI, éditeurs : *Advances in the Dempster-Shafer theory of evidence*, chapitre 16, pages 353–374. John Wiley & Sons, Inc., 1994.
- [172] L. XU, A. KRZYZAK et C.Y. SUEN : Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition. *IEEE Transactions on Systems, Man Cybernetics*, 22(3):418–435, Mai 1992.
- [173] Z. XU : Induced uncertain linguistic OWA operators applied to group decision making. *Information Fusion*, "à paraître" 2005.
- [174] R.R. YAGER : On the Dempster-Shafer Framework and New Combination Rules. *Information Sciences*, 41:93–137, 1987.
- [175] R.R. YAGER : On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking. *IEEE Transactions on Systems, Man Cybernetics*, 18(1):183–190, Novembre 1988.
- [176] R.R. YAGER : Constrained OWA aggregation. *Fuzzy Sets and Systems*, 82:89–101, 1996.
- [177] R.R. YAGER : The Power Average Operator. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 31(6):724–731, Novembre 2001.

- [178] R.R. YAGER : A framework for multi-source data fusion. *Information Sciences*, 163:175–200, 2004.
- [179] R.R. YAGER : Structures for prioritized fusion of fuzzy information. *Information Sciences*, 163:175–200, 2004.
- [180] R.R. YAGER et M. Fedrizzi J. KACPRZYK : *Advances in the Dempster-Shafer theory of evidence*. John Wiley & Sons, Inc., 1994.
- [181] Y.P. YEE, E.M. MEASURE, J.L. COGAN et M. BLEIWEIS : Neural networks for sensor fusion of meteorological measurements. *In Sensor Fusion : Architectures, Algorithms and Application V*, volume 4385, pages 77–86, Orlando, USA, Avril 2001.
- [182] Z. YI, H.Y. KHING et Z.X. WEI : Multi-ultrasonic sensor fusion for autonomous mobile robots. *In Sensor Fusion : Architectures, Algorithms and Application IV*, volume 4051, pages 314–231, Orlando, USA, Avril 2000.
- [183] S. YUE, P. LI et Z. YIN : Parameter estimation Choquet fuzzy integral based on Takagi-Sugeno fuzzy model. *Fuzzy Sets and Systems*, à paraitre 2005.
- [184] H. ZHU et O. BASIR : Extended Discounting Scheme for Evidential Reasoning as Applied to MS Lesion Detection. *In International Conference on Information Fusion*, Stockholm, Sweden, Juin 2004.
- [185] Y. ZHU : *Multisensor decision and estimation fusion*. Kluwer Academic Publishers, 2003.
- [186] L.M. ZOUHAL et T. DENŒUX : Une méthode de discrimination non paramétrique basée sur la théorie de Dempster et Shafer. *In GRETSI*, Juan-Les-Pins, France, Septembre 1995.
- [187] L.M. ZOUHAL et T. DENŒUX : An Evidence-Theoric  $k$ -NN Rule with Parameter Optimization. *IEEE Transactions on Systems, Man, and Cybernetics - Part C : Applications and Reviews*, 28(2):263–271, Mai 1998.

# Index

- $\alpha$ -coupe, 68
- ambiguïté, 18, 91
- anonymat, 78
- architecture, 6
- argument, 91
- associativité, 78
  
- bissymétrie, 78
  
- cardinal, 68
- coefficient d'affaiblissement, 39
- combinaison
  - conjonctive, 78
  - de compromis, 78
  - disjonctive, 78
- commutativité, 78
- complémentarité, 17, 91
- confiance, 91
- conflit, 17, 91
- confusion, 17, 91
- conorme triangulaire, 81
- consensus, 78
- critère minimax, 30
- croissance, 91
- crédibilité, 91
  
- degré d'appartenance, 67
- distribution de possibilité, 73
- défuzzification, 69
  
- élément focal, 34
- espérance maximale, 30
  
- fiabilité, 15, 91
- fonction
  - binaire, 66
  - caractéristique, 66
  - d'appartenance, 66
  - d'implicabilité, 37
  - de communalité, 37
  - de croyance, 35
  - de croyance complémentaire, 42
  - de croyances bayésienne, 36
  - de crédibilité, 35
  - de doute, 36
  - de masse, 34
  - de non-croyance, 36
  - de plausibilité, 37
  - génératrice, 80
  - sur les singletons, 42
  - à support simple, 41
  
- grossissement, 40
  
- hauteur, 68
  
- idempotence, 78
- imprécision, 15, 92
- incertitude, 15, 92
- incohérence, 17, 92
- incomplétude, 16, 92
- inconsistance, 17, 48, 92
- indépendance, 14
- information, 92
- intervalle flou, 69
- intégrale
  - de Choquet, 86
  - de Sugeno, 85
  
- masse
  - de croyance, 34
  - élémentaire, 34
- maximum
  - a posteriori, 30

- d'entropie, 30
- de vraisemblance, 30
- mesure
  - de confiance, 72
  - de conflit, 88
  - de nécessité, 73
  - de possibilité, 72
- mesure floue, 72
- monde
  - clos, 19, 92
  - fermé, 19, 92
  - ouvert, 19, 92
- monotonie, 79
- neutralité, 78
- nombre flou, 69
- norme triangulaire, 79
- noyau, 34, 67
- nécessité, 92
- opérateur
  - conjonctif, 79
  - de compromis, 82
  - de moyenne, 83
  - de médiane, 84
  - disjonctif, 81
- possibilité, 92
- quantité floue, 69
- raffinement, 40
- redondance, 16, 93
- sous-ensemble flou, 67
  - normalisé, 67
  - unimodal, 67
- stabilité, 79
- support, 67
- t-conorme, 81
- t-norme, 79
- unanimité, 78, 79
- élément neutre, 79